# Global Patterns in Culturable Soil Yeast Diversity

Himeshi Samarasinghe[1], Yi Lu[1], Renad Aljohani[1,2], Ahmad Al-Amad[1], Heather Yoell[1] and Jianping Xu[1]*

[1]Department of Biology, McMaster University, Hamilton, Canada (samaraya@mcmaster.ca; luy129@mcmaster.ca; alamadae@mcmaster.ca; yoellh@mcmaster.ca; jpxu@mcmaster.ca)

[2]Current Address: Department of Infectious Diseases, South Kensington Campus, Imperial College London, UK (r.aljohani20@imperial.ac.uk)

*Corresponding author

## Summary

Yeasts, broadly defined as unicellular fungi, fulfill essential roles in soil ecosystems as decomposers and nutrition sources for fellow soil-dwellers. Broad-scale investigations of soil yeasts pose a methodological challenge as metagenomics are of limited use on this group of fungi. Here we characterize global soil yeast diversity using fungal DNA barcoding on 1473 yeasts cultured from 3826 soil samples obtained from nine countries in six continents. We identify mean annual precipitation and international air travel as two significant predictors of soil yeast community structure and composition worldwide. Anthropogenic influences on soil yeast communities, directly via travel and indirectly via altered rainfall patterns resulting from climate change, are concerning as we found common infectious yeasts frequently distributed in soil in several countries. Our discovery of 41 putative novel species highlights the need to revise the current estimate of ~1500 recognized yeast species. Our findings demonstrate the continued need for culture-based studies to advance our knowledge of environmental yeast diversity.

**Keywords:** fungal diversity; culturable yeasts; *Candida; Cryptococcus*; gene flow; international travel

**Subject area:** Biodiversity and Ecology

## 1. Introduction

Soil is host to an incredible amount of microbial life, with each gram containing over 10 billion cells of bacteria, archaea and fungi (Roesch et al., 2007). Despite their relatively low abundance, soil fungi fulfill essential roles in decomposition of organic material, nutrient cycling and soil fertilization (Frac et al., 2018). This is especially true for yeasts, broadly defined as unicellular fungi, whose numbers rarely exceed thousands of cells per gram of soil. Yet, yeasts in soil ecosystems are essential decomposers and nutrient sources for fellow soil-dwelling protists, bacteria, insects, and nematodes (Botha, 2011; Yurkov, 2018). In fact, yeasts may be the predominant soil fungi in cold biospheres such as continental Antarctica (Connell et al., 2008; Vishniac, 1996). Soil is also a primary reservoir for common pathogenic yeasts of the *Candida* and *Cryptococcus* genera that frequently cause superficial and systemic infections in humans (Kurtzman et al., 2011).

It is becoming increasingly apparent that the true extent of global soil yeast diversity is significantly underestimated. While yeast cells were first observed under the microscope in 1680 by Anton Van Leeuwenhoek, their natural habitats were a topic of contention among mycologists who often associated yeasts with fruit trees and fermentation. It was not until the 1950s that soil was established as a true natural habitat of yeasts where they live and reproduce. The culturing media, incubation temperatures and techniques used in pioneering studies were expanded in later projects to isolate soil yeasts of diverse metabolic and functional profiles (reviewed in Yurkov, 2018). Currently, the diverse array of yeasts recovered and characterized from soils across the globe contribute to the ~1500 recognized yeast species on the planet (Kurtzman et al., 2011). Environmental surveys of yeasts routinely uncover novel species, accounting for as much as 30% of yeast populations, highlighting the need to revise current estimates of global yeast diversity (Groenewald et al., 2018; Yurkov et al., 2016b, 2016a).

Lack of adequate environmental sampling, especially in Asia, Africa, South and Central America, limits the discovery of novel yeast species, characterization of soil yeast communities, and prediction of global diversity patterns. Soil yeast populations often differ in structure and composition between locations. With the exception of a few genera that are widespread in soil such as *Cyberlindnera, Schwanniomyces, Naganishia, Goffeauzyma* and *Solicoccozyma* (Botha, 2011), most yeast species have a fragmented distribution with few shared species between sites, even within the same geographical region. One study found that only eight of the 57 species found in soils of Mediterranean xerophyl forests were shared between the three sampling plots in the same locality (Yurkov et al., 2016a). Another study found only a single species to be present in all three sampled temperate forests in Germany (Yurkov et al., 2012). So far, most environmental surveys reported in the literature have been ecologically/

geographically limited, with sampling often focus on a specific ecological niche within a single locality, region, or country (Into et al., 2020; Li et al., 2020; Monteiro Moreira and Martins do Vale, 2020; Tepeeva et al., 2018).

Factors affecting soil yeast diversity have not been fully elucidated but soil moisture, soil pH, carbon content and nitrogen content have been implicated as contributing variables (reviewed in Botha, 2011). In 2006, Vishniac analyzed prominent yeast species in soil along a latitudinal gradient and found that mean annual temperature, mean annual rainfall and electrical conductivity explained ~44% of the variation in yeast species distributions (Vishniac, 2006). As their sampling locations were limited to the Americas and Antarctica, it is unknown whether the same trends persist on a global scale. One potential contributing factor to yeast species and genotype distributions is anthropogenic influences such as international travel. For example, international travel has increased exponentially in the last few decades with direct implications for the global spread of organisms, most notably infectious disease agents and invasive species. The role of global travel in introducing infectious diseases to new areas and facilitating epidemics is well documented (Findlater and Bogoch, 2018), with the current COVID-19 pandemic being a prime example. International travel is likely affecting soil yeast communities by transferring previously geographically isolated species and genotypes across borders, although a link between the two has not been previously investigated.

Metagenomics is widely applied in the study of environmental microbes to investigate taxonomic diversity, characterize functional groups, and elucidate broad scale patterns (Abbasian et al., 2016; Abia et al., 2018; Egidi et al., 2019; Li and Qin, 2005). However, metagenomics and other culture-independent methods cannot be readily applied to the study of yeasts due to the lack of a suitable yeast-specific barcoding gene (Xu, 2016). Yeasts are phylogenetically diverse and occur among filamentous fungi in two major phyla, Ascomycota and Basidiomycota, within the fungal kingdom. In a 2014 study that has not been surpassed in scale before or since, Tedersoo and colleagues used high throughput sequencing of fungal barcoding DNA to assess global soil fungal diversity and identify predictors of global diversity patterns (Tedersoo et al., 2014). Due to the lack of a sequence-based signature, yeasts were not singled out as a group of interest and thus limited information was presented on soil yeast diversity of the 39 countries included in the study. They identified mean annual precipitation and distance from equator as the two strongest overall predictors of soil fungal diversity on a global scale. It is not clear if and to what extent the same predictors apply to yeast diversity in soil.

Using a global collection of 3826 soil samples, here we assessed the culturable soil yeast diversity in nine countries representing all continents except Antarctica. We found soil yeast populations of each country to be unique in structure and composition

as 73% of the discovered species were not shared between countries. Mean annual precipitation was the most significant predictor of culturable soil yeast diversity on a global scale. We found air traffic volume to be significantly correlated with the number of shared species between countries, implying a link between international travel and transfer of yeast species across borders. Our study overcomes the geographical constraints of many previous studies by identifying soil yeast diversity patterns on a global scale. We also demonstrate that culture-dependent methods provide a more comprehensive framework than metagenomics for studying phylogenetically diverse, but morphologically targeted groups of organisms such as yeasts.

## 2. Materials and Methods

### *2.1 Soil collection*

We collected soil from 53 locations in nine countries encompassing all continents except Antarctica (Figure 1 and Supplementary Table 1). At each location, bags (3cm x 7cm sterile, resealable, polyethylene bags) of topsoil, within 1-3 inches from the surface, were collected following sterile protocols and transported to our lab at McMaster University, Canada. Each bag contained soil from 10 sites in the same area but located at least 2m from each other. Different bags represented soils from at least 100m from each other. The soil in each bag was segregated into ~1g aliquots and stored at 4°C. In total, this study included 3826 soil samples originating from the following countries: Cameroon (493), Canada (300), China (340), Costa Rica (388), France (327), Iceland (316), New Zealand (610), Peru (490) and Saudi Arabia (562).

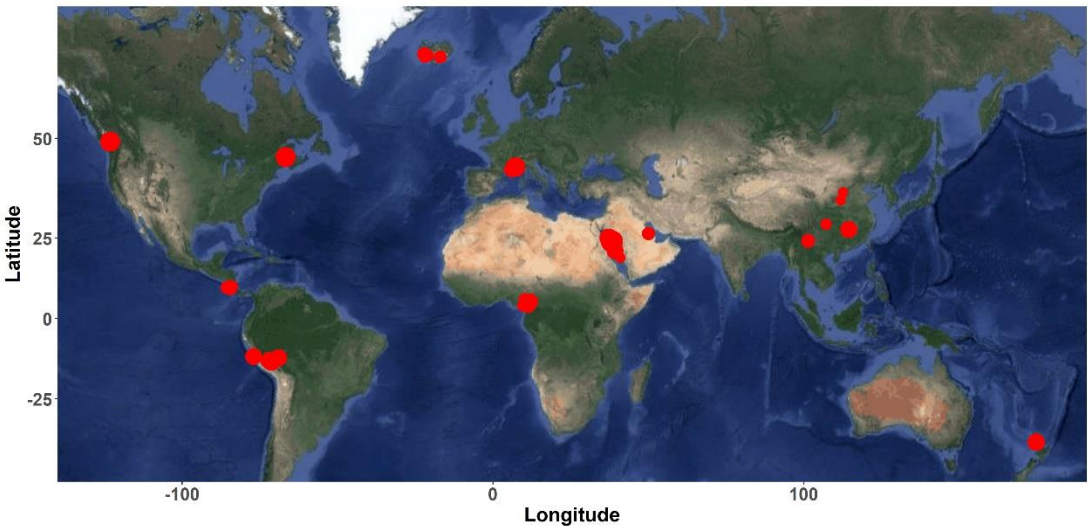


***Figure 1:*** **Soil sampling locations**. Soil was collected from 53 locations, indicated by the red circles, in nine countries. The size of the circle corresponds to the number of samples obtained from that location.

### *2.2 Yeast isolation from soil samples*

Yeasts were isolated at a temperature deemed to be optimal based on the country of origin's mean annual temperature. For each soil sample, approximately 0.1g was added into 5ml of YEPD broth (Yeast Extract-Peptone-Dextrose) in 13ml culture tubes and incubated in a roller drum for 24 hours. The broth contained the antibiotic chloramphenicol (50mg/L) and the selectively toxic fungicide benomyl (5mg/L) to inhibit bacterial and mold growth, respectively. We extended this incubation step to 72 hours for soil samples from Iceland due to slower yeast growth at 14°C. We then plated 100ul of the broth onto solid YEPD containing chloramphenicol and benomyl and incubated at the same temperature for an additional 2-5 days until microbial growth was visible. For each plate that contained morphologically yeast-like colonies, we randomly selected a representative colony and streaked it onto fresh YEPD plates to obtain single colonies. If more than one morphology was present, one representative colony of each type was separately streaked for single colonies. After 2-3 days' incubation, we randomly picked one single colony per yeast isolate and suspended in 50ul nuclease-free water to be used in Polymerase Chain Reaction (PCR).

### *2.3 Yeast identification via ITS sequencing*

We identified the yeasts by sequencing their fungal barcoding gene, the ribosomal Internal Transcribed Spacer (ITS) regions. We performed colony PCR using primers ITS1 (5' TCCGTAGGTGAACCTGCGG 3') and ITS4 (5' TCCTCCGCTTATTGATATGC 3') to amplify the ITS region. The PCR cocktail consisted of 10ul Promega GoTaq Green Mastermix, 5ul nuclease-free water, 2ul each of the primers (2uM) and 1ul cell suspension. The thermocycling conditions were an initial denaturation step at 95°C for 10 minutes followed by 35 cycles of (i) 95°C for 30 seconds, (ii) 55°C for 30 seconds and iii) 72°C for 1 minute. We ran 4ul of the PCR products on a 1.5% agarose gel to check for successful amplification. The remaining PCR products were Sanger sequenced with the ITS1 primer at Eurofins Genomics in Louisville, Kentucky (https://eurofinsgenomics.com/en/home/). We trimmed the low-quality ends of the ABI chromatograms generated from Sanger sequencing and batch converted to FASTA format with DNA Baser's ABI to FASTA converter software (https://www.dnabaser.com/download/Abi-to-Fasta-converter/abi-to-fasta-converter.html). We used BLAST+ applications on the command line to query the multi-FASTA file against the NCBI nucleotide database to detect sequence similarity to existing ITS sequences. The BLAST searches were run remotely (-remote flag) to avoid downloading the entire database onto our servers. The output was compiled into a CSV (Comma Separated Values) file containing the top 10 matches for each query sequence. We inspected the CSV file manually to check for quality and for any inconsistencies in species identity within the top ten matches. The putatively identified species through GenBank searches were further confirmed by manually comparing to

sequences in the curated UNITE database (https://unite.ut.ee/). We assigned species identities to our ITS sequences at a sequence similarity threshold of 98.41% to existing sequences in databases. This threshold was previously determined to be optimum to distinguish yeast species at the ITS locus based on an analysis of 9000 fungal sequences (Vu et al., 2016). Sequences with no matches surpassing this threshold were considered putative novel species.

### *2.4 Statistical analyses of population diversity*

All statistical analyses were conducted in RStudio v.4.0.2 using a combination of base functions and packages including ggmap (Kahle and Wickham, 2013), ggplot2 (Wickham, 2016), and tidyverse (Wickham et al., 2019). We quantified the diversity of yeast populations at our sampling sites by calculating the Shannon diversity index using the package Vegan v.2.5-7 (Oksanen et al., 2020). We conducted rarefaction analyses using the iNEXT package (Hsieh et al., 2016) to determine if sufficient soil sampling was performed in each of the nine countries to accurately estimate their culturable soil yeast diversity.

### *2.5 Relationship between yeast diversity and climate and geographic factors*

Within each country, soil collection sites differed in climatic and environmental conditions, with the exception of New Zealand where sampling was limited to the metropolitan region of Auckland. We assigned the 53 sampling sites to 47 distinct locations based on their geographical coordinates. Using geographical coordinates, we calculated mean annual precipitation and mean annual temperature by averaging monthly data over a 16-year period from 1991-2016, available on Climate Change Knowledge Portal (https://climateknowledgeportal.worldbank.org/). We calculated the elevation and distance from the sampling sites to the equator using Google Maps. We calculated the Shannon diversity index of the yeast populations found at the 47 distinct locations. We constructed mixed models using the package lme4 v.1.1-26 (Bates et al., 2015) where precipitation, temperature, elevation, and distance to equator were set as fixed effects, country was fitted as a random effect and Shannon diversity Index was fitted as the dependent variable.

### *2.6 Air traffic data*

We extracted data on the number of flights occurring between each country-country pair over a 5-year period from 2011-2016 from the Global Transnational Mobility Dataset (Recchi et al., 2019). This dataset is compiled based on a combination of tourism data and distance-adjusted air-traffic data. Next, we calculated the number of yeast species shared between each country-country combination. To assess the correlation between the number of shared species and the volume of air traffic between countries, a linear model was fitted between the two variables.

### *2.7 Comparison to metagenomics study*

We compared our findings to a previous study that used culture-independent methods to investigate global diversity of soil fungi. In 2014, Tedersoo and colleagues extracted DNA directly from soil samples of 39 countries and performed high throughput sequencing of the ITS2 region using primers ITS3 and ITS4 (Tedersoo et al., 2014), covering a portion of the DNA barcoding fragment we sequenced here. Four countries overlapped between the two studies, namely, Cameroon, Canada, China and New Zealand. For each of the four countries, we performed BLAST searches using BLAST+ applications on the command line to identify ITS sequences that appeared in both studies. First, we created custom databases containing our ITS sequences for each country. Next, the ITS sequences for the countries of interest were extracted from Supplementary data table 1 of Tedersoo et al.'s paper and converted to FASTA format. The multi-FASTA file for each country was then queried against its respective custom database using the blastn option. The output was compiled into a CSV file containing the top 5 matches for each query sequence. This CSV file was perused manually to identify significant sequence similarity between query and match sequences.

## Results

### *3.1 Yeast isolation and species identification*

We isolated a total of 1473 yeasts from 3826 soil samples (Table 1). The isolation rate varied among the countries, ranging from 17% in Saudi Arabia to 87% in Canada. The yeast isolation and species distribution data from Cameroon soils have been reported in a previous study (Aljohani et al. 2018). Overall, we observed a slightly negative correlation between the number of soil samples and the yeast isolation rate ($p < 0.05$), as countries with more soil samples did not necessarily yield more yeast isolates.

We successfully assigned species identity to 1367 isolates using the 98.41% sequence identity cut-off to homologous ITS sequences in NCBI and UNITE databases. These strains were categorized into 90 species belonging to 37 genera, with *Candida* being the most species-rich genus (n=19 species). With 60 ascomycetes and 30 basidiomycetes, both major yeast-containing phyla within the fungal kingdom were broadly represented. The 90 species belonged to six Classes, ten Orders and 18 Families. However, two genera, *Nadsonia* and *Holtermanniella,* do not currently have defined Family associations (*incertae sedis*). The remaining 106 yeast strains can be grouped into 44 unique clusters at 98.41% nucleotide similarity. Since no existing sequences with >98.41% ITS sequence identity were found in the databases, these 44 clusters represent potentially novel yeast species. Genbank accession numbers to the ITS sequences of our 1473 isolates are MG817572 to MG817630 and MW894661 to MW896112 (Supplementary dataset 1).

Our rarefaction analyses suggested that sufficient soil sampling was conducted in each country to accurately estimate the true diversity of culturable soil yeasts. Projections for Shannon diversity index beyond the number of soil samples included in the study revealed that the diversity of our yeast populations approached saturation asymptote (Supplementary Figure 1). Additional sampling in these locations was not likely to have revealed higher yeast species diversity.

**Table 1: Summary statistics of yeast isolation from global soil samples**

| Country | Number of soil samples | Number of yeast isolates | Number of known species | Number of novel species | Number of country-specific species | Shannon diversity index |
|---|---|---|---|---|---|---|
| Cameroon | 493 | 110 | 10 | 9 | 12 | 2.17 |
| Canada | 300 | 261 | 34 | 12 | 25 | 3.06 |
| China | 340 | 230 | 23 | 5 | 15 | 2.54 |
| Costa Rica | 388 | 95 | 20 | 2 | 9 | 2.21 |
| France | 327 | 175 | 12 | 2 | 3 | 1.26 |
| Iceland | 316 | 211 | 11 | 0 | 4 | 1.25 |
| New Zealand | 610 | 155 | 14 | 4 | 5 | 2.05 |
| Peru | 490 | 139 | 30 | 9 | 20 | 3.27 |
| Saudi Arabia | 562 | 97 | 8 | 1 | 5 | 0.91 |
| **Total** | **3826** | **1473** | **90** | **44** | **98** | |

### *3.2 Diversity and abundance of culturable soil yeast populations*

The abundance and diversity of soil yeast populations varied significantly between countries. Saudi Arabia ranked lowest among the nine countries in the number of unique species, where 562 soil samples yielded 97 yeast isolates belonging to 9 species, 1 of which was novel. On the other hand, we obtained 261 yeast isolates from 300 Canadian soil samples, encompassing 46 species, 12 of which were novel. The number of yeast isolates and of distinct species found in the seven remaining countries ranged from 95-230 and 11-39 respectively (Table 1). The Shannon diversity index of the soil yeast populations ranged from 0.91 (Saudi Arabia) to 3.27 (Peru). Less diverse populations tended to be dominated by a single yeast species, most notably in France, Iceland, and Saudi Arabia where *Candida subhashii, Goffeauzyma gastrica* and *Cryptococcus deneoformans* predominated the soil yeasts respectively (Figure 2).

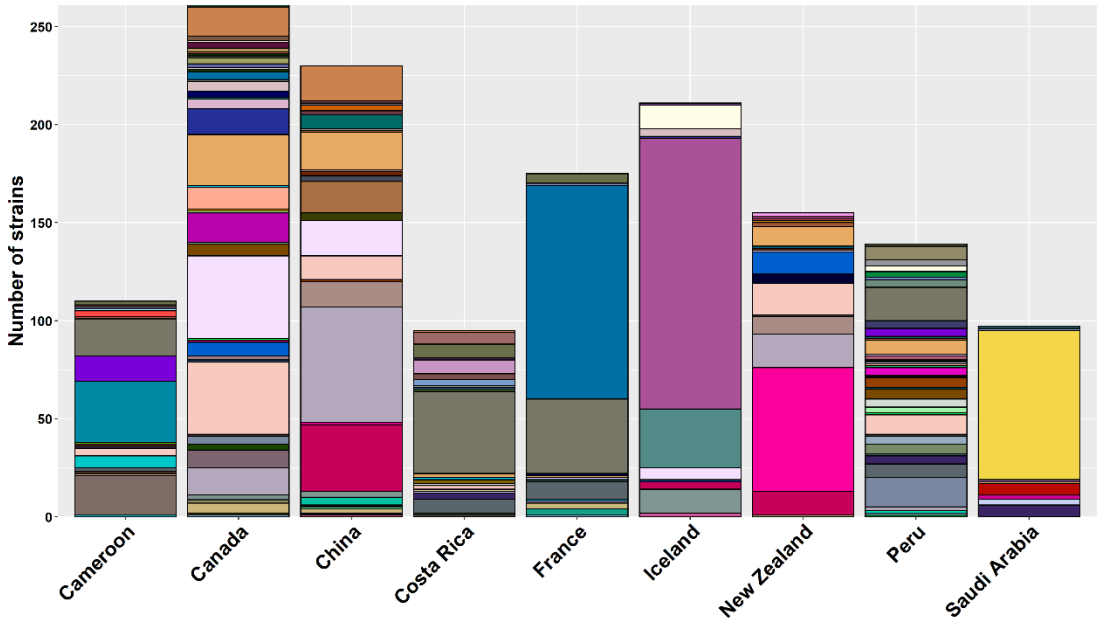


***Figure 2:*** **Culturable soil yeast populations by country.** The X-axis represents the country. Each country is represented by a stacked bar plot. Each colour represents a unique species and the height of the colored sections indicate the abundance of that species.

### *3.2.1 Cameroon*

As shown in the study by Aljohani et al. (2018) for the Cameroonian soil samples, all but one of the 19 yeast species recovered from Cameroonian soil were ascomycetes. The population was dominated by four species: *Cyberlindnera subsufficiens* (28%), *Torulaspora globosa* (18%), *Candida tropicalis* (17%) and *Cyberlindnera saturnus* (12%). Overall, Cameroonian soil contained the second highest number of novel species (n=9) after Canada (n=12).

### *3.2.2 Canada*

The culturable yeast population in Canadian soil was dominated by ascomycetes (37 species): the remaining 9 species were basidiomycetes. *Nadsonia starkeyi-henricii*, a little-known yeast that prefers relatively mild temperatures below 25°C, was the most abundant (16%) followed by the pathogenic *Papiliotrema laurentii* (14%). *Debaryomyces hansenii* (10%), *Barnettozyma californica* (6%) and a novel species (6%) were also present in significant amounts. Canadian soil contained 25 yeast species not found in soil samples of the other eight countries, including cold-adapted yeast *Cystofilobasidium ferigula*, industrial lactose fermenter *Kluyveromyces lactis,* and close relative of Baker's yeast, *Saccharomyces paradoxus.* Canada also had the highest number of novel species (n=12) of the 9 sampled countries.

### *3.2.3 China*

The Chinese culturable soil yeast population consisted of similar numbers of ascomycetes and basidiomycetes (15 and 13 respectively). This population was dominated by strains of *Solicoccozyma aeria* and *Solicoccozyma terrea* that together accounted for 40% of the population. Other yeasts with significant prevalence included *Debaryomyces hansenii* (8%), *Barnettozyma californica* (8%), *Nadsonia starkeyi-henricii* (8%) and a novel species (7%). 15 yeast species were only found in the Chinese soil which also yielded five novel species.

### *3.2.4 Costa Rica*

*Candida tropicalis* was the dominant species in the culturable soil yeast population of Costa Rica with a prevalence of 44%. The frequencies of the remaining 21 species ranged from 1%-7%. Ascomycetous species outnumbered basidiomycetes at 18 to four. Costa Rica was notable for being the only sampled country to contain strains of the common pathogenic yeasts *Candida albicans* (6%) and *Candida orthopsilosis* (3%). Strains of pathogenic *Candida parapsilosis* were also present in Costa Rican soil (3%).

### *3.2.5 France*

The majority of species in the French culturable soil yeast population were ascomycetes (n=11) while three species were basidiomycetes. *Candida subhashii*, a pathogenic *Candida* species first identified in 2009 (Adam et al., 2009), was the dominant yeast in this population with an abundance of 62%. *C. subhashii* was previously determined to have strong antagonistic activity against filamentous fungi and has potential as a biocontrol agent against plant pathogenic fungi (Hilber-Bodmer et al., 2017). The widespread pathogen *Candida tropicalis* was the second most abundant species (22%), followed by *Saccharomyces cerevisiae* (5%). One strain of *Candida parapsilosis* was also detected in French soils.

### *3.2.6 Iceland*

Of the 11 species isolated from Iceland soil, six were basidiomycetes and five were ascomycetes. With an abundance of 65%, *Goffeauzyma gastrica* was the dominant species in the Iceland culturable soil yeast population. *G. gastrica* is a cold-tolerant yeast commonly isolated from environmental sources in Antarctica and is known for its production of antifreeze proteins (Białkowska et al., 2017; Ogaki et al., 2020; Villarreal et al., 2018). *Goffeauzyma gilvescens,* another cold-tolerant yeast commonly found in Antarctica, was the second most abundant (14%), followed by *Candida sake* (6%) and *Solicoccozyma terricola* (6%). Iceland was the only sampled country to not yield any novel yeast species.

### *3.2.7 New Zealand*

Basidiomycete species (n=10) were slightly more prevalent than ascomycete species (n=8) in the New Zealand culturable soil yeast population. *Solicoccozyma phenolica* was the most abundant species with a prevalence of 41%, followed by *Solicoccozyma aeria* (11%), *Papiliotrema laurentii* (10%) and *Solicoccozyma terrea* (8%). We isolated several species with industrial potential from New Zealand soil including *Papiliotrema terrestris,* shown to produce β-galactosidase that was safe for use in food production(Ke et al., 2018), and *Citeromyces matritensis*, an osmotolerant, ethanol-producing yeast shown to be capable of ethanol production from salted algae (Okai et al., 2017).

#### *3.2.8 Peru*

Peru's culturable soil yeast population, consisting of 39 species, ranked the highest among sampled countries in Shannon diversity index. This population was unique in structure and composition as it contained 20 species not found in any other sampled country, including often misidentified pathogen and crude palm oil assimilator *Candida palmioleophila* (Jensen and Arendrup, 2011; NAKASE et al., 1988), rare pathogen *Filobasidium magnum* (Aboutalebian et al., 2020), and halotolerant yeast used in azo dye decolorization *Pichia occidentalis* (Wang et al., 2020). This population contained significantly more ascomycete species (n=29) than basidiomycete species (n=10). Peruvian population was notable for its relative evenness with no single species exceeding 12% in abundance. *Candida tropicalis* was the most prevalent (12%), followed by *Schwanniomyces occidentalis* (11%) and *Papiliotrema laurentii* (7%).

#### *3.2.9 Saudi Arabia*

Saudi Arabian culturable soil yeast population was the least diverse of all sampled countries according to the Shannon diversity index. This population was notable for the overwhelming prevalence of the human pathogenic yeast, *Cryptococcus deneoformans* (78%), the causative agent of fatal fungal meningoencephalitis. The genotypes of C. deneoformans strains from Saudi Arabia have been reported in an earlier study (Samarasinghe et al., 2019). This study was the first to report the environmental presence of *C. deneoformans* in a desert climate. However, overall, the Saudi Arabian soil yeast population consisted of six basidiomycete species and three ascomycete species. One of the species was novel.

### *3.3 Pathogenic yeast species*

According to the information presented in the latest edition of The Yeasts: A Taxonomic Study (Kurtzman et al., 2011), the following 12 species were the most common yeast pathogens of humans worldwide: *Candida albicans, Candida dubliniensis, Candida glabrata, Candida guilliermondii, Candida krusei, Candida lusitaniae, Candida parapsilosis, Candida orthopsilosis, Candida metapsilosis, Candida*

*tropicalis, Cryptococcus neoformans* and *Cryptococcus deneoformans.* We found 220 strains belonging to eight of these species, accounting for 15% of all yeast isolates found in our samples (Figure 3). *C. tropicalis* was both the most abundant and most widespread with 117 isolates originating from Cameroon, Canada, Costa Rica, France and Peru. The 76 *C. deneoformans* isolates were exclusively found in Saudi Arabian soils. Additionally, seven *C. krusei* isolates were found in Costa Rica, six *C. albicans* isolates were found in Costa Rica, five *C. parapsilosis* isolates were found in Costa Rica, France and Saudi arabia, four *C. lusitaniae* isolates were found in Canada and France, four *C. orthopsilosis* isolates were found in Cameroon and Costa Rica and a single *C. glabrata* isolate was found in Costa Rica. Common pathogenic yeasts were not isolated from our natural soils of China, Iceland and New Zealand.

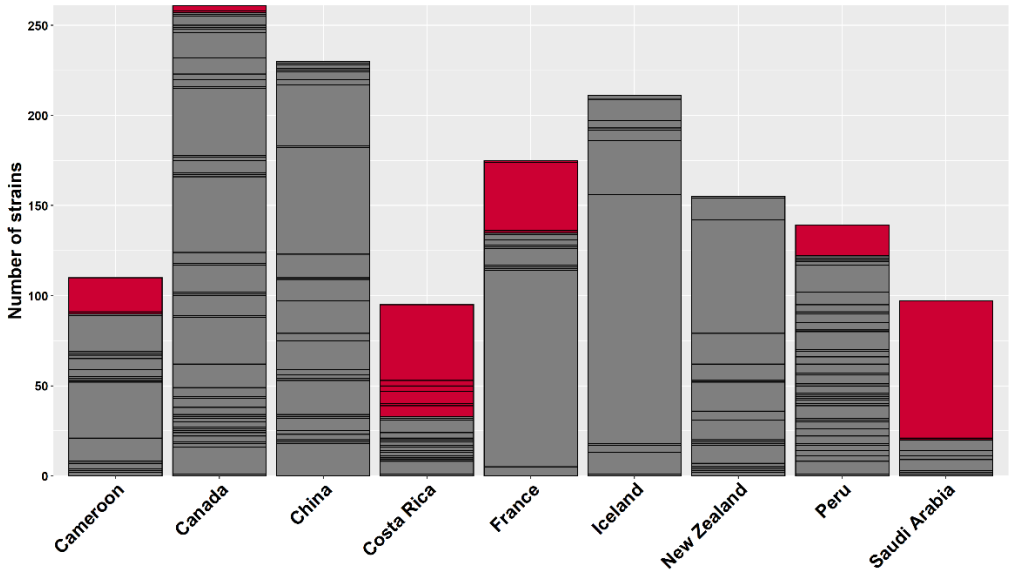


***Figure 3:*** **Pathogenic yeast species and their abundance highlighted in red among countries.** In these stacked bar plots, the pathogenic species are highlighted in red. The height of the red sections indicates their abundance. Soils from China, Iceland and New Zealand did not yield any pathogenic species.

### *3.4 Novel species*

Our yeast population included 44 potentially novel species from eight sampled countries: Our soil samples from Iceland did not yield any novel yeast species. We determined the most closely related genera for 41 species by running BLAST searches in Unite database (remaining three species' ITS sequences were too short for analysis). Our 41 novel species can be categorized into 12 genera (9 ascomycetes, 3 basidiomycetes) with *Wickerhamomyces* containing ten novel species, and *Candida* containing eight. For each genus, we constructed maximum likelihood (ML) trees using RaxML with 1000 bootstraps (Stamatakis, 2014) to determine the taxonomic placement of novel species with respect to all known species of that genus. Our ML trees confirmed the separation of newly discovered species from known species: for an

example, the ML tree of *Wickherhamomyces* species places each novel species at its own distinct node (Figure 4). We observed some geographical clustering where the Cameroonian and Canadian novel species formed their own clusters: the two novel species of Costa Rica and Peru clustered together. The ML trees of the remaining 11 genera can be found in Supplementary Dataset 1.

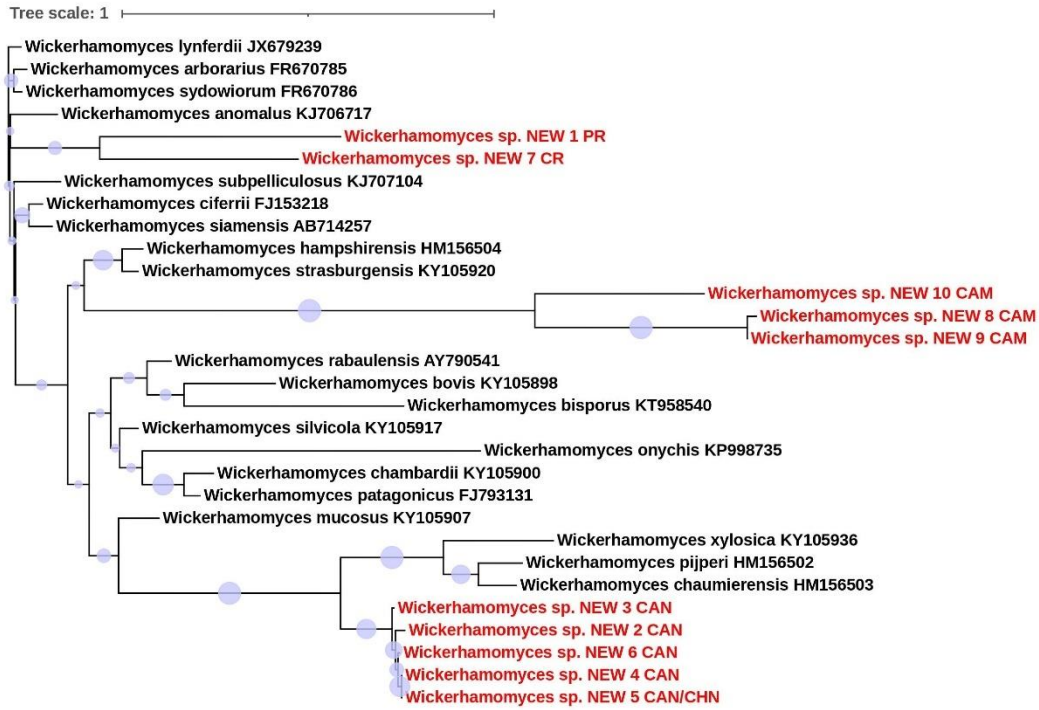


***Figure 4:*** **The maximum likelihood tree of *Wicherhamomyces* species**. The placement of novel species with reference to known *Wickerhamomyces* species is shown. The novel species' country of origin is shown in the OTU labels where CAM = Cameroon, CAN = Canada and CHN = China. The tree was constructed using RaxML with 1000 bootstraps.

### *3.5 Predictors of global culturable soil yeast diversity*

Our 47 distinct sampling locations covered a wide range of global climatic conditions (Table 2) with mean annual precipitation ranging from 0mm (Lima, Peru where there is virtually no rainfall) to 2965mm (Monteverde, Costa Rica), while mean annual temperature ranged from -1.4°C (Svartifoss, Iceland) to 29.6°C (Alqunfudah, Saudi Arabia). Elevation ranged from < 2m (some sites in Svartifoss, Iceland and Auckland, New Zealand) to 4922km above sea level (Rainbow mountain, Peru). Mbandoumou in Cameroon was the closest to the equator (418.5km from equator) while Dimmuborgir in Iceland was the farthest (7280.5km from equator). Four locations, two in Saudi Arabia and two in Cameroon, were removed from further analysis as they did not yield any yeast isolates. The remaining 43 locations varied significantly in culturable yeast diversity as quantified by Shannon diversity index from 0 (only one species was found)

to 2.77 (Fredericton, Canada). According to our mixed model, we found mean annual precipitation to be significantly correlated with the Shannon diversity index ($p = 0.012$, Figure 5). We found no significant correlation between the remaining predictors and Shannon diversity index (Supplementary Dataset 3).

**Table 2: Environmental and geographic characteristics of sampling sites.** Mean annual rainfall, mean annual temperature, distance to equator and elevation of sampling sites are summarized here.

| Country | City | Site code | Mean annual rainfall (mm) | Mean annual temperature (°C) | Distance to equator (km) | Altitude (km) |
|---|---|---|---|---|---|---|
| Cameroon | Babanki | CBB | 1813.312 | 1813.31 | 20.75 | 1173 |
| | Bambui | CBM | 1813.312 | 1813.31 | 20.75 | 1274 |
| | Eloundem | CEL | 1617.423 | 1617.42 | 24.75 | 620 |
| | Makepe | CMK | 2825.192 | 2825.19 | 26.83 | 62 |
| | Mbalgon | CML | 1617.423 | 1617.42 | 25.03 | 556 |
| | Mbandoumou | CMD | 1617.423 | 1617.42 | 24.22 | 719 |
| | Mbingo | CMB | 1813.312 | 1813.31 | 20.75 | 1909 |
| | Njinikejum | CNJ | 1813.312 | 1813.31 | 20.75 | 1573 |
| | Simbock | CSM | 1617.423 | 1617.42 | 24.99 | 643 |
| Canada | Fredericton | CF | 1267.842 | 1267.84 | 5.29 | 10 |
| | Vancouver | CV | 1567.708 | 1567.71 | 9.95 | 31 |
| China | Ailao Mountain | CAC | 1120.492 | 1120.49 | 17.29 | 2782 |
| | Fenyi | CC | 1556.404 | 1556.40 | 17.26 | 307 |
| | Jinfo Mountain | CJ | 1154.85 | 1154.85 | 15.49 | 2085 |
| | Pangquangou Nature Reserve | CT | 467.3346 | 467.33 | 9.66 | 775 |
| | Taihang mountains west of Jincheng | CSX | 515.0269 | 515.03 | 12.11 | 2012 |
| Costa Rica | El Jardin | EJ | 2308.358 | 2308.36 | 27.23 | 152 |
| | La Fotuna | LF | 2964.685 | 2964.68 | 25.51 | 392 |
| | La Paz | LP | 2964.685 | 2964.68 | 25.51 | 1512 |
| | Manuel Antonio | MA | 2308.358 | 2308.36 | 27.23 | 13 |
| | Monteverde | MV | 2964.685 | 2964.68 | 25.51 | 1585 |
| | Playa Hermosa | PH | 2308.358 | 2308.36 | 27.23 | 7 |
| | Playa Samara Beach | SB | 1682.769 | 1682.77 | 26.72 | 7 |
| | Poas Volcano | PV | 2957.59 | 2957.59 | 25.51 | 2350 |
| | Samara Town | ST | 1682.769 | 1682.77 | 26.72 | 13 |
| | Villas Playa Samara Beach Front Hotel | SH | 1682.769 | 1682.77 | 26.72 | 7 |
| France | Hyeres | FHF | 703.8039 | 703.80 | 15.45 | 37 |
| | Uptown/Downtown Nice | FN | 814.7077 | 814.71 | 13.82 | 25 |
| Iceland | Dimmuborgir above Myvatn lake (highlands) | ID | 969.2692 | 969.27 | 1.06 | 283 |
| | Landbrotalaug mini hotspring | IL | 823.4154 | 823.42 | 4.65 | 11 |
| | National park near Svartifoss | ISF | 1904.319 | 1904.32 | -1.40 | 192 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Near Nautholsvik Geothermal Beach near Reykjavik university | IN | 959.4077 | 959.41 | 4.17 | 0 |
| | Skútustaðagígar pseudocraters on Myvatn lake | ISP | 855.5423 | 855.54 | 1.93 | 271 |
| | Thingvellir | IT | 1107.946 | 1107.95 | 3.05 | 89 |
| New Zealand | Auckland | NZA | 1172.973 | 1172.97 | 15.13 | 0 |
| Peru | Amazon | PA | 2139.142 | 2139.14 | 25.51 | 176 |
| | Cusco | PC | 679.0462 | 679.05 | 9.59 | 3322 |
| | Lima | PL | 0 | 0.00 | 19.57 | 138 |
| | Machu Pichu | PM | 563.1885 | 563.19 | 9.50 | 1940 |
| | Rainbow Mountain | PR | 725.7192 | 725.72 | 5.77 | 4922 |
| | Sacred Valley | PS | 636.3077 | 636.31 | 8.52 | 2866 |
| Saudi Arabia | Alqunfudah | SAA | 49.56154 | 49.56 | 29.58 | 1 |
| | Dammam | SAD | 73.69615 | 73.70 | 27.15 | 5 |
| | Jeddah | SAJ | 45.29615 | 45.30 | 29.25 | 15 |
| | Medina | SAM | 65.06154 | 65.06 | 27.18 | 636 |
| | Umluj | SAU | 4.815385 | 4.82 | 27.63 | 14 |
| | Yanbu | SAY | 29.56923 | 29.57 | 27.69 | 9 |

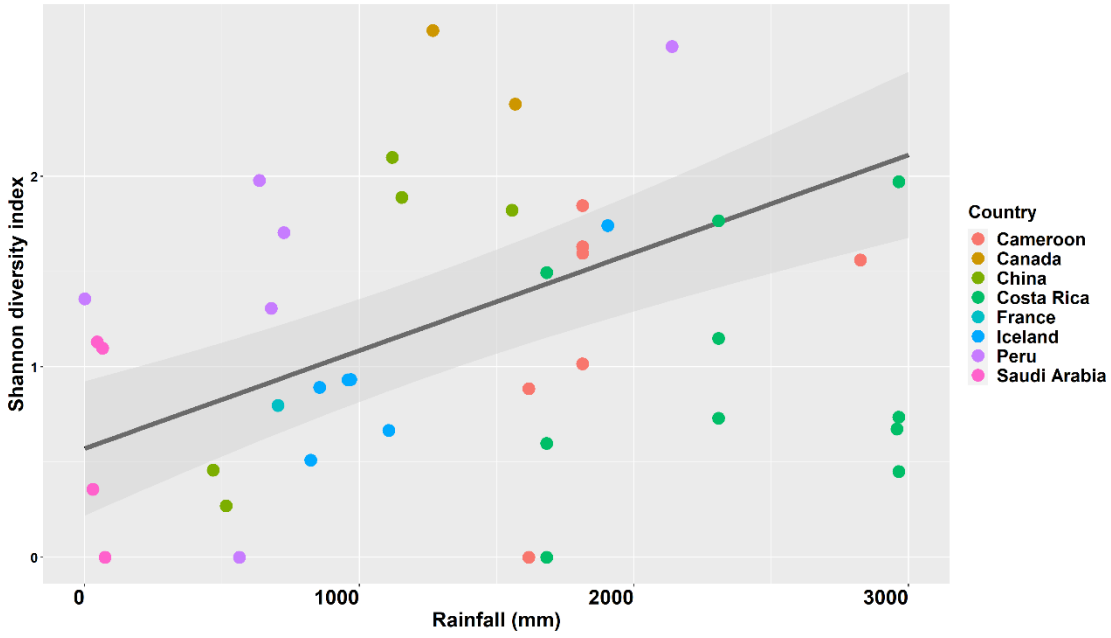


**Figure 5: Mean annual rainfall is significantly correlated with soil yeast diversity.** Here, the Shannon diversity index of our sampling sites is plotted against mean annual precipitation. Sampling sites are colored by country. The line plots the model predictions with associated uncertainty shaded in grey.

### *3.6 Air traffic volume as a predictor of shared species between countries*

36 yeast species were found in more than one country. The following five country pairs had no soil yeast species in common: Iceland-Cameroon, Iceland-Costa Rica, Iceland-France, Iceland-Saudi Arabia, and Cameroon-Saudi Arabia. The number of shared

species between the remaining 31 pairs ranged from one to 11 (Figure 6b). Air traffic volume data extracted from the Global Transnational Mobility Dataset showed that 25 700 496 trips were made between China and France between 2011-2016. During the same period, only 81 trips were made between Iceland and Cameroon (Figure 6a, Supplementary Table 3). We performed a linear regression analysis between air traffic volume, geographic distance, and the number of shared species between countries. While we found no significant correlation between geographic distance and sharedness, air traffic volume was significantly correlated with the number of shared species between countries ($p = 0.003$, Figure 7).

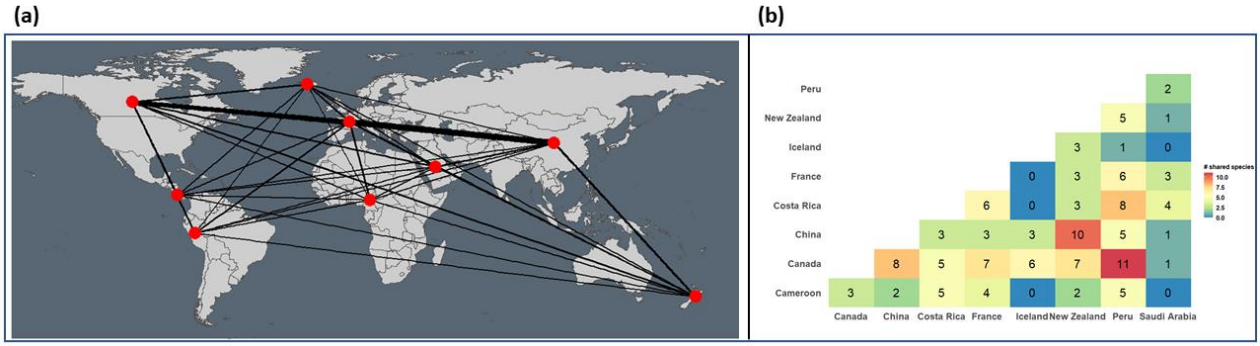


*Figure 6:* **Air traffic volume between countries is correlated with number of shared species.** (a) Volume of air traffic between the nine countries from 2011-2016. Thickness of the line indicates volume. (b) Heat map showing the number of shared species between country pairs.

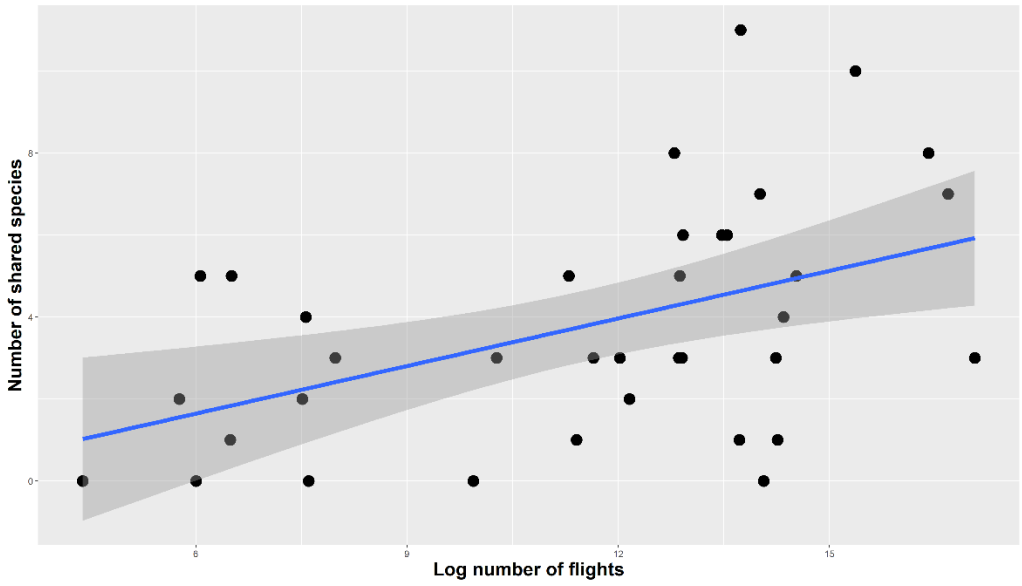


**Figure 7*:* Number of shared species between countries is significantly correlated with traffic volume between them**. We plotted the log number of trips made between country pairs between 2011-2016 against the number of yeast species shared between them.

We constructed neighbour-joining (NJ) trees in MEGA7 (Kumar et al., 2016) based on ITS sequences of the four most shared species in our population: *Debaryomyces*

*hansenii* (7 countries), *Papiliotrema laurentii* (6 countries), *Candida tropicalis* (5 countries) *and Torulaspora delbrueckii* (5 countries). The NJ trees highlighted the lack of strict geographical clustering of isolates by country: for an example, most *P. laurentii* isolates found in New Zealand, Costa Rica, China, Peru, and Cameroon had identical ITS sequences and formed a cluster with most Canadian isolates (Figure 7). The NJ trees for the remaining three species can be found in Supplementary Dataset 2. This result is consistent with the hypothesis of recent long-distance dispersals for many of the shared species.

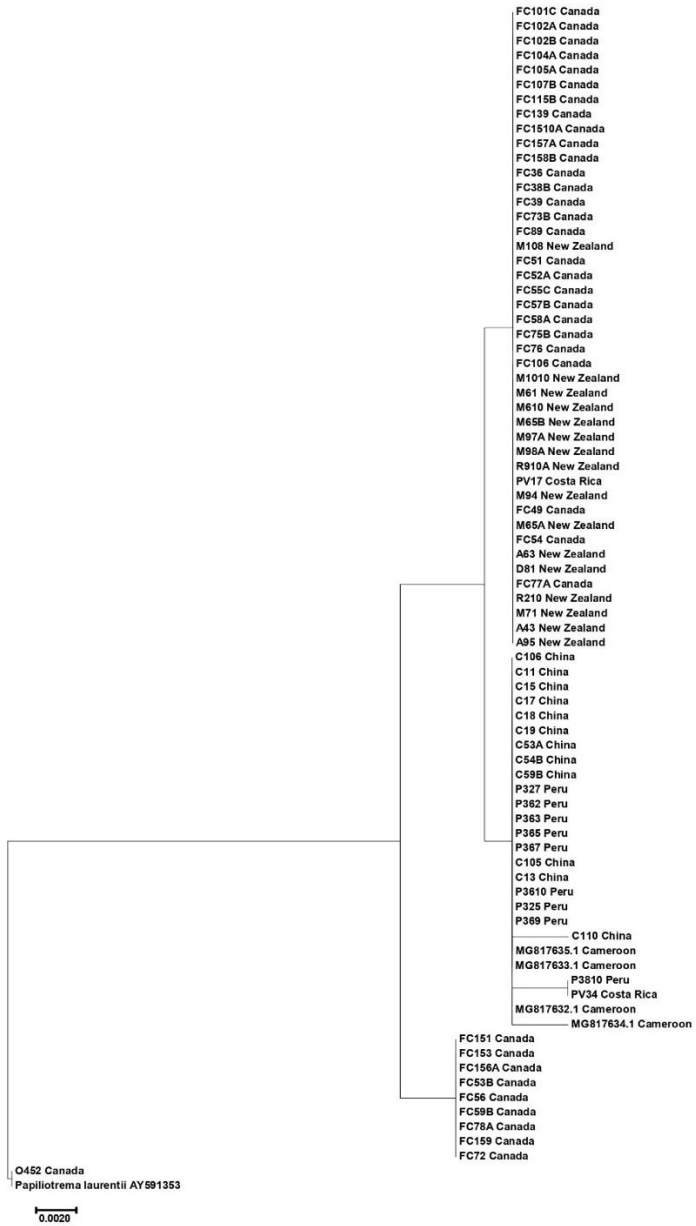


***Figure 8:*** **The neighbour-joining tree of *Papiliotreme laurentii* global isolates found in our study**. No geographical clustering is observed, suggesting frequent gene flow between populations.

### *3.7 Comparison to culture-independent, metagenomics approach*

Our study exclusively used culture-dependent methods to investigate global diversity of culturable soil yeasts. We compared our findings to an excellent study conducted by Tedersoo and colleagues where a metagenomics approach was taken to study global soil fungal diversity (Tedersoo et al., 2014). For the four countries that overlapped between our study and that of Tedersoo et al. (2014), we conducted a detailed side-by-side comparison of results, shown in Table 2. Below we highlight results from two countries.

From Cameroonian soils, Tedersoo et al. obtained 278 fungal OTUs (Operational Taxonomic Units) with species identity determined for 72. Limited information is available for 103 OTUs that were only annotated with a variation of either "fungal_sp" or "uncultured_fungus". The remaining 103 OTUs were annotated to higher fungal phylogenetic ranks such as Kingdom, Phylum, Class, Order or Family. Our study obtained 110 ITS sequences from Cameroonian soil yeasts. Species identity was determined for 94, genus was established for another nine and the remaining seven were annotated as novel species of the fungal kingdom. Our BLAST analyses revealed no overlapping sequences between the two datasets.

For China, Tedersoo et al. found 572 fungal OTUs, 125 of which were annotated to the species level. The remaining OTUs were annotated to the genus or a higher taxonomic rank. We obtained 230 ITS sequences from Chinese soil yeasts: we established species identity for 209 sequences and genus identity for another three. The remaining 18 sequences, clustered into three groups, were identified as putative novel fungal species. Four sequence overlapped between the two studies and their species annotations were consistent between the two studies.

***Table 3:*** **Comparison of findings between Tedersoo et al. and current study for four countries (Tedersoo et al., 2014).** Tedersoo et al. used metagenomics to determine the diversity of all soil fungi. The current study used culture-dependent methods and fungal DNA barcoding

| | | *Tedersoo et al.* | | *Current study* | |
|---|---|---|---|---|---|
| | | Number | Proportion | Number | Proportion |
| *Cameroon* | Total OTUs found | 272 | 100% | 110 | 100% |
| | OTUs with species determined | 73 | 27% | 94 | 85% |
| | Higher taxonomy determined | 52 | 19% | 16 | 15% |
| | Unannotated | 147 | 54% | 0 | 0% |
| *Canada* | Total OTUs found | 413 | 100% | 261 | 100% |
| | OTUs with species determined | 104 | 25% | 230 | 88% |

| | | | | | |
|---|---|---|---|---|---|
| | Higher taxonomy determined | 198 | 48% | 31 | 12% |
| | Unannotated | 111 | 27% | 0 | 0% |
| *China* | Total OTUs found | 572 | 100% | 230 | 100% |
| | OTUs with species determined | 145 | 25% | 209 | 91% |
| | Higher taxonomy determined | 235 | 41% | 21 | 9% |
| | Unannotated | 192 | 34% | 0 | 0% |
| *New Zealand* | Total OTUs found | 269 | 100% | 155 | 100% |
| | OTUs with species determined | 73 | 27% | 146 | 94% |
| | Higher taxonomy determined | 87 | 32% | 9 | 6% |
| | Unannotated | 109 | 41% | 0 | 0% |

## 4. Discussion

Despite being one of the most accessible ecological niches, soil remains an enigmatic source of yeast diversity and ecology. Given that most yeast species are not geographically widely distributed, extensive environmental sampling across diverse regions, habitats and climates is required to uncover new species and diversity patterns. Elucidating global trends and dynamics would also allow us to predict the structure and diversity of soil yeast populations in unsampled locations. Using a set of global soil samples from nine countries in six continents, we address this knowledge gap by characterizing global patterns and predictors of culturable soil yeast diversity. Our study uncovered 134 soil yeast species among 1473 isolates, including 41 previously undescribed species. We identified mean annual precipitation and air traffic volume as significant predictors of soil yeast communities on a global scale. Our findings highlight the influence of both climatic factors and anthropogenic activity on soil yeast populations across the globe.

We found mean annual precipitation to be the strongest predictor of culturable soil yeast diversity across both local and global scales. Previous metagenomic studies have established mean annual precipitation as one of the climatic variables associated with soil fungal diversity (Egidi et al., 2019; Tedersoo et al., 2014). Our results confirm that this trend persists for culturable yeast communities in global soils as well. Vegetation is not a likely mediating factor in the observed positive correlation between precipitation and soil yeast diversity as Tedersoo et al. found plant diversity to be uncoupled from soil fungal richness (Tedersoo et al., 2014). Fungal communities in dry, semi-arid soils contain significantly more Ascomycota fungi than Basidiomycota (Abed

et al., 2013; Murgia et al., 2019; Suleiman et al., 2019). We found a reversal of this trend in global soil yeast communities where basidiomycetous yeasts were found to be more prevalent in sampling sites receiving less rainfall ($p < 0.05$). Some soil-dwelling, basidiomycetous yeasts are known to produce biofilms that allow them to persist in low moisture, oligotrophic conditions (Spencer and Spencer, 1997). Low moisture, and resulting lack of nutrients, could favor cellular structures and metabolic activities of yeasts in one Phylum over the other, creating rainfall-associated global diversity patterns observed in our study. Given our findings, we hypothesize that extreme rainfall and drought events brought on by global warming are likely shifting the established landscape of soil yeast communities. This is especially alarming given the significant presence of pathogenic yeasts we detected in the soils. 15% of all yeast isolates found in our study belong to common pathogenic yeast species capable of causing deadly systemic infections. Altered rainfall patterns, and resulting changes in soil microclimates, could cause outgrowths of pathogenic species and lead to emergence of new fungal infections. With soil ecosystems being a primary source of bacterial and fungal infections, any changes and shifts in soil microbiomes could pose a significant threat to human health.

Each of the nine countries investigated in our study was unique in the composition and structure of its culturable soil yeast population. 73% of the yeast species found in our study (98 out of 134) were specific to a single country. The fragmented nature of soil yeast distributions has been noted in previous studies where only a few species were found to be shared between sampling sites, even within the same region or country (Yurkov, 2018). The nine countries included in our study are separated by thousands of kilometers, with the two closest countries being France and Iceland (2235 km). Geographic isolation was most likely a key factor in limiting the spread and exchange of yeasts between populations, at least until recently when anthropogenic activity has strongly improved the connectivity between countries and continents.

Our findings suggest that human activities can contribute to the changing yeast distribution in soil environments across the globe. International travel has increased exponentially in the past few decades with international tourist arrivals increasing from 25 million in 1950 to a record-high 1.4 billion in 2018 (UNWTO, 2018). The global air transportation network has the small-world property where most countries can be reached from each other via a few flight hops (Wandelt and Sun, 2015). While we found unique yeast species in most localities, 36 yeast species (~25% of all species found) were shared between at least two countries. Countries with more flights occurring between them had more yeast species in common, which implicates human travel as a likely facilitator in the spread of endemic yeasts across geographical borders. The lack of geographical clustering of the most shared species in our population supports gene flow between populations in different countries. The covid-19 pandemic has

555 greatly shifted the political and economical landscape of our planet. Tourism both
556 within and between countries has seen a drastic drop with accompanying tightening
557 of borders between countries. The potential impact of the COVID-19 pandemic on
558 culturable yeast populations remains to be determined.

559 The current estimate of ~1500 yeast species in existence is almost certainly a
560 significant underestimate (Kurtzman et al., 2011). Both culture-dependent and
561 independent studies routinely isolate novel yeast species from the environment. Our
562 study is one of many recent surveys to find previously undescribed species accounting
563 for as much as 30% of natural yeast populations (Yurkov, 2018), implying that every
564 one of three yeast species recovered from the environment is likely to be a new one.
565 Investigators often turn to natural soils in search of novel yeast strains with commercial
566 and biotechnological potential. A novel strain of *Pichia kudriavzevii* (syn. *Candida*
567 *krusei*) isolated from soil in a sugarcane filed in Thailand was shown to be more
568 thermotolerant and produce more ethanol than the Thai industrial strain
569 *Saccharomyces cerevisiae* TISTR 5606 (Pongcharoen et al., 2018). Presence of species
570 of the genus *Kazachstania* in mixed cultures of *Saccharomyces cerevisiae* gives rise to
571 fermented wines with diverse aroma profiles: however, *Kazachstania* species are unable
572 to complete fermentation in monocultures (Jood et al., 2017). Discovery of new
573 *Kazachstania* species with more desirable fermentative abilities can aide the full
574 exploitation of this genus in commercial wine fermentation. The thermo and
575 halotolerant yeast *Blastobotrys adeninivorans* is highly useful in a wide range of
576 biotechnological applications including the production of secretory enzymes, as a host
577 for heterologous gene expression and as a biological component in biosensors (Kunze
578 et al., 2017). The metabolic and fermentative capabilities of the novel *Kazachstania*
579 species we found in Peruvian soil and the novel *Blastobotrys* species found in French
580 soil remain to be evaluated.

581 In recent years, researchers have come to view metagenomics as a valuable tool in the
582 investigation of microbial diversity in complex ecological systems. While high
583 throughput sequencing is crucial in unearthing large-scale patterns at higher
584 taxonomic levels, it fails to be adequately informative on targeted groups of organisms
585 such as yeasts. Our findings indicate that culture-dependent methods can succeed
586 where metagenomics may fail in the study of environmental yeast diversity. Limited
587 information on yeast diversity could be extracted from previous metagenomics studies
588 on global soil fungal diversity (Egidi et al., 2019; Tedersoo et al., 2014). Yeasts do not
589 form a monophyletic group that can be easily identified based on sequences alone.
590 Extracting ITS sequences of known yeasts from large metagenomics datasets is a time-
591 consuming task that requires personnel with advanced knowledge of yeast taxonomy.
592 For potentially novel species, the metagenomic approach would completely fail to
593 identify them as yeasts. The metagenomics approach is largely limited by the current

state of knowledge on species taxonomy and annotation. In Tedersoo et al.'s study, species identity was only established for ~25% of the fungal OTUs found in Cameroon, Canada, China and New Zealand: over 30% of the OTUs remained unidentified. Our global collection of soil yeast isolates with identity established and manually validated via ITS sequencing provides a much-needed reference set for future investigators on yeast diversity and taxonomy.

Fungi isolated via culture-dependent methods can be identified as yeasts by morphology and can be further characterized using genomics, metabolomics, and transcriptomics (Xu 2020). Given the relatively low numbers of yeast cells in soil compared to bacteria, mold and other fungi, their DNA can easily escape detection in metagenomics studies, which could explain the lack of overlapping yeast sequences between our study and that of Tedersoo et al. (2014). Our use of culture-dependent methods ensured that the yeasts we found were alive and more likely to be true soil-dwellers as opposed to dead cells temporarily transferred to soil from elsewhere. Selective enrichment and culturing from soil samples in the lab remains the most effective ways of identifying and studying yeasts.

## 5. Conclusions

Our investigation into global patterns in culturable soil yeast diversity reaffirms soil as an important reservoir of environmental yeast species, both known and yet undiscovered. Precipitation emerges as the main predictor of soil yeast diversity across local and global scales. Ongoing global warming crisis and accompanying changes in rainfall could lead to expansion of pathogenic yeasts that already account for a sizable proportion of soil yeast communities. Our findings implicate international travel as a facilitator in the movement of yeast species across borders, with phylogenetic evidence suggesting long-distance gene flow between yeast populations. More environmental sampling is required to further uncover soil yeast diversity, isolates with commercial and biotechnological value and to monitor species that could pose a threat to human health.

## Acknowledgements

We thank the following people for contributing soil samples: Haoran Jia, Yongjie Zhang, Thomas Harrison, Sebastian Harrison, Arshia Kazerouni, and Greg Korfanty for contributing soil samples for this study. HS is supported by an NSERC CGS-D Scholarship. This research is supported by Natural Sciences and Engineering Research Council (NSERC) of Canada (RGPIN-2020-05732) and by the Global Science Initiative Award of McMaster University to JX.

## Author contributions

Study conceived by HS and JX; soil collections were coordinated by JX; lab work performed by HS, YL, RA, AA, and HY; data analyses performed by HS; first manuscript draft written by HS; final draft edited by JX; all authors have read and approved the final version of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest

## References

Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R., Ramadass, K., 2016. Microbial diversity and hydrocarbon degrading gene capacity of a crude oil field soil as determined by metagenomics analysis. Biotechnol. Prog. 32, 638–648. https://doi.org/10.1002/btpr.2249

Abed, R.M.M., Al-Sadi, A.M., Al-Shehi, M., Al-Hinai, S., Robinson, M.D., 2013. Diversity of free-living and lichenized fungal communities in biological soil crusts of the Sultanate of Oman and their role in improving soil properties. Soil Biol. Biochem. 57, 695–705. https://doi.org/10.1016/j.soilbio.2012.07.023

Abia, A.L.K., Alisoltani, A., Keshri, J., Ubomba-Jaswa, E., 2018. Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. Sci. Total Environ. 616–617, 326–334. https://doi.org/10.1016/j.scitotenv.2017.10.322

Aboutalebian, S., Mahmoudi, S., Okhovat, A., Khodavaisy, S., Mirhendi, H., 2020. Otomycosis Due to the Rare Fungi Talaromyces purpurogenus, Naganishia albida and Filobasidium magnum. Mycopathologia 185, 569–575. https://doi.org/10.1007/s11046-020-00439-8

Adam, H., Groenewald, M., Mohan, S., Richardson, S., Bunn, U., Gibas, C.F.C., Poutanen, S., Sigler, L., 2009. Identification of a new species, *Candida subhashii* , as a cause of peritonitis. Med. Mycol. 47, 305–311. https://doi.org/10.1080/13693780802380545

Aljohani, R., Samarasinghe, H., Ashu, T., Xu, J. 2018. Diversity and relationships among strains of culturable yeasts in agricultural soils in Cameroon. Sci. Rep. 8, 15687. http://dx.doi.org/10.1038/s41598-018-34122-2

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Białkowska, A.M., Szulczewska, K.M., Krysiak, J., Florczak, T., Gromek, E., Kassassir, H., Kur, J., Turkiewicz, M., 2017. Genetic and biochemical characterization of yeasts

667 isolated from Antarctic soil samples. Polar Biol. 40, 1787–1803.
668 https://doi.org/10.1007/s00300-017-2102-7

669 Botha, A., 2011. The importance and ecology of yeasts in soil. Soil Biol. Biochem.
670 https://doi.org/10.1016/j.soilbio.2010.10.001

671 Connell, L., Redman, R., Craig, S., Scorzetti, G., Iszard, M., Rodriguez, R., 2008. Diversity
672 of soil yeasts isolated from South Victoria Land, Antarctica. Microb. Ecol. 56,
673 448–459. https://doi.org/10.1007/s00248-008-9363-1

674 Egidi, E., Delgado-Baquerizo, M., Plett, J.M., Wang, J., Eldridge, D.J., Bardgett, R.D.,
675 Maestre, F.T., Singh, B.K., 2019. A few Ascomycota taxa dominate soil fungal
676 communities worldwide. Nat. Commun. 10, 1–9. https://doi.org/10.1038/s41467-
677 019-10373-z

678 Findlater, A., Bogoch, I.I., 2018. Human Mobility and the Global Spread of Infectious
679 Diseases: A Focus on Air Travel. Trends Parasitol.
680 https://doi.org/10.1016/j.pt.2018.07.004

681 Frac, M., Hannula, S.E., Belka, M., Jędryczka, M., 2018. Fungal biodiversity and their
682 role in soil health. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.00707

683 Groenewald, M., Lombard, L., de Vries, M., Lopez, A.G., Smith, M., Crous, P.W., 2018.
684 Diversity of yeast species from Dutch garden soil and the description of six novel
685 Ascomycetes. FEMS Yeast Res. 18, 76. https://doi.org/10.1093/femsyr/foy076

686 Hilber-Bodmer, M., Schmid, M., Ahrens, C.H., Freimoser, F.M., 2017. Competition
687 assays and physiological experiments of soil and phyllosphere yeasts identify
688 Candida subhashii as a novel antagonist of filamentous fungi. BMC Microbiol.
689 17, 1–15. https://doi.org/10.1186/s12866-016-0908-z

690 Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and
691 extrapolation of species diversity ( <scp>H</scp> ill numbers). Methods Ecol.
692 Evol. 7, 1451–1456. https://doi.org/10.1111/2041-210X.12613

693 Into, P., Pontes, A., Sampaio, J.P., Limtong, S., 2020. Yeast Diversity Associated with
694 the Phylloplane of Corn Plants Cultivated in Thailand. Microorganisms 8, 80.
695 https://doi.org/10.3390/microorganisms8010080

696 Jensen, R.H., Arendrup, M.C., 2011. Candida palmioleophila: Characterization of a
697 previously overlooked pathogen and its unique susceptibility profile in
698 comparison with five related species. J. Clin. Microbiol. 49, 549–556.
699 https://doi.org/10.1128/JCM.02071-10

700 Jood, I., Hoff, J.W., Setati, M.E., 2017. Evaluating fermentation characteristics of
701 Kazachstania spp. and their potential influence on wine quality. World J.
702 Microbiol. Biotechnol. 33. https://doi.org/10.1007/s11274-017-2299-1

703 Kahle, D., Wickham, H., 2013. ggmap: Spatial Visualization with ggplot2. R J. 5, 144–

704 161.

705 Ke, Q., Fulmer, P., Mizutani, A., 2018. Toxicological evaluation of β-Galactosidase
706 enzyme produced by Papiliotrema terrestris. Regul. Toxicol. Pharmacol. 92, 213–
707 219. https://doi.org/10.1016/j.yrtph.2017.12.002

708 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics
709 Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874.
710 https://doi.org/10.1093/molbev/msw054

711 Kunze, G., Bischoff, F., Chamas, A., Litwinska, K., Matthes, F., Boer, E., 2017.
712 Applications of blastobotrys (arxula) adeninivorans in biotechnology, in: Yeast
713 Diversity in Human Welfare. Springer Singapore, pp. 455–479.
714 https://doi.org/10.1007/978-981-10-2621-8_18

715 Kurtzman, C., Fell, J.W., Boekhout, T. (Eds.), 2011. The Yeasts A Taxonomic Study, 5th
716 edition. ed. Elsevier.

717 Li, A.H., Yuan, F.X., Groenewald, M., Bensch, K., Yurkov, A.M., Li, K., Han, P.J., Guo, L.D.,
718 Aime, M.C., Sampaio, J.P., Jindamorakot, S., Turchetti, B., Inacio, J., Fungsin, B.,
719 Wang, Q.M., Bai, F.Y., 2020. Diversity and phylogeny of basidiomycetous yeasts
720 from plant leaves and soil: Proposal of two new orders, three new families, eight
721 new genera and one hundred and seven new species. Stud. Mycol. 96, 17–140.
722 https://doi.org/10.1016/j.simyco.2020.01.002

723 Li, X., Qin, L., 2005. Metagenomics-based drug discovery and marine microbial
724 diversity. Trends Biotechnol. 23, 539–543.
725 https://doi.org/10.1016/j.tibtech.2005.08.006

726 Monteiro Moreira, G.A., Martins do Vale, H.M., 2020. Soil Yeast Communities in
727 Revegetated Post-Mining and Adjacent Native Areas in Central Brazil.
728 Microorganisms 8, 1116. https://doi.org/10.3390/microorganisms8081116

729 Murgia, M., Fiamma, M., Barac, A., Deligios, M., Mazzarello, V., Paglietti, B.,
730 Cappuccinelli, P., Al-Qahtani, A., Squartini, A., Rubino, S., Al-Ahdal, M.N., 2019.
731 Biodiversity of fungi in hot desert sands. Microbiologyopen 8, e00595.
732 https://doi.org/10.1002/mbo3.595

733 NAKASE, T., ITOH, M., SUZUKI, M., KOMAGATA, K., KODAMA, T., 1988. Candida
734 palmioleophila sp. nov., a yeast capable of assimilating crude palm oil, formerly
735 identified as Torulopsis candida. J. Gen. Appl. Microbiol. 34, 493–498.
736 https://doi.org/10.2323/jgam.34.493

737 Ogaki, M.B., Teixeira, D.R., Vieira, R., Lírio, J.M., Felizardo, J.P.S., Abuchacra, R.C.,
738 Cardoso, R.P., Zani, C.L., Alves, T.M.A., Junior, P.A.S., Murta, S.M.F., Barbosa, E.C.,
739 Oliveira, J.G., Ceravolo, I.P., Pereira, P.O., Rosa, C.A., Rosa, L.H., 2020. Diversity and
740 bioprospecting of cultivable fungal assemblages in sediments of lakes in the
741 Antarctic Peninsula. Fungal Biol. 124, 601–611.

https://doi.org/10.1016/j.funbio.2020.02.015

Okai, M., Betsuno, A., Shirao, A., Obara, N., Suzuki, K., Takei, T., Takashio, M., Ishida, M., Urano, N., 2017. Citeromyces matritensis M37 is a salt-tolerant yeast that produces ethanol from salted algae. Can. J. Microbiol. 63, 20–26. https://doi.org/10.1139/cjm-2016-0259

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Maintainer, H.W., 2020. Package "vegan" Title Community Ecology Package Version 2.5-7.

Pongcharoen, P., Chawneua, J., Tawong, W., 2018. High temperature alcoholic fermentation by new thermotolerant yeast strains Pichia kudriavzevii isolated from sugarcane field soil. Agric. Nat. Resour. 52, 511–518. https://doi.org/10.1016/j.anres.2018.11.017

Recchi, E., Deutschmann, E., Vespe, M., 2019. Estimating Transnational Human Mobility on a Global Scale. SSRN Electron. J. https://doi.org/10.2139/ssrn.3384000

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. 1, 283–290. https://doi.org/10.1038/ismej.2007.53

Samarasinghe, H., Aljohani, R., Jimenez, C., Xu, J., 2019. Fantastic yeasts and where to find them: The discovery of a predominantly clonal Cryptococcus deneoformans population in Saudi Arabian soils. FEMS Microbiol. Ecol. 95. https://doi.org/10.1093/femsec/fiz122

Spencer, J.F.T., Spencer, D.M., 1997. Ecology: Where Yeasts Live, in: Yeasts in Natural and Artificial Habitats. Springer Berlin Heidelberg, pp. 33–58. https://doi.org/10.1007/978-3-662-03370-8_4

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

Suleiman, M.K., Dixon, K., Commander, L., Nevill, P., Quoreshi, A.M., Bhat, N.R., Manuvel, A.J., Sivadasan, M.T., 2019. Assessment of the diversity of fungal community composition associated with Vachellia pachyceras and its Rhizosphere soil from Kuwait desert. Front. Microbiol. 10, 63. https://doi.org/10.3389/fmicb.2019.00063

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring,

780 M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou,
781 A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J.,
782 Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend,
783 H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C.,
784 Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger, F., Brearley,
785 F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and
786 geography of soil fungi. Science (80-. ). 346.

787 Tepeeva, A.N., Glushakova, A.M., Kachalkin, A. V., 2018. Yeast communities of the
788 Moscow city soils. Microbiol. (Russian Fed. 87, 407–415.
789 https://doi.org/10.1134/S0026261718030128

790 UNWTO World Tourism Barometer and Statistical Annex, November 2018, 2018. .
791 UNWTO World Tour. Barom. 16, 1–36.
792 https://doi.org/10.18111/wtobarometereng.2018.16.1.5

793 Villarreal, P., Carrasco, M., Barahona, S., Alcaíno, J., Cifuentes, V., Baeza, M., 2018.
794 Antarctic yeasts: Analysis of their freeze-thaw tolerance and production of
795 antifreeze proteins, fatty acids and ergosterol. BMC Microbiol. 18, 66.
796 https://doi.org/10.1186/s12866-018-1214-8

797 Vishniac, H.S., 2006. A multivariate analysis of soil yeasts isolated from a latitudinal
798 gradient. Microb. Ecol. https://doi.org/10.1007/s00248-006-9066-4

799 Vishniac, H.S., 1996. Biodiversity of yeasts and filamentous microfungi in terrestrial
800 Antarctic ecosystems. Biodivers. Conserv. 5, 1365–1378.
801 https://doi.org/10.1007/BF00051983

802 Vu, D., Groenewald, M., Szöke, S., Cardinali, G., Eberhardt, U., Stielow, B., de Vries, M.,
803 Verkleij, G.J.M., Crous, P.W., Boekhout, T., Robert, V., 2016. DNA barcoding
804 analysis of more than 9 000 yeast isolates contributes to quantitative thresholds
805 for yeast species and genera delimitation. Stud. Mycol. 85, 91–105.
806 https://doi.org/10.1016/j.simyco.2016.11.007

807 Wandelt, S., Sun, X., 2015. Evolution of the international air transportation country
808 network from 2002 to 2013. Transp. Res. Part E Logist. Transp. Rev. 82, 55–78.
809 https://doi.org/10.1016/j.tre.2015.08.002

810 Wang, X., Wang, Y., Ning, S., Shi, S., Tan, L., 2020. Improving Azo Dye Decolorization
811 Performance and Halotolerance of Pichia occidentalis A2 by Static Magnetic
812 Field and Possible Mechanisms Through Comparative Transcriptome Analysis.
813 Front. Microbiol. 11, 712. https://doi.org/10.3389/fmicb.2020.00712

814 Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
815 York.

816 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund,
817 G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S.,

818 Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D.,
819 Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source
820 Softw. 4, 1686. https://doi.org/10.21105/joss.01686

821 Xu, J., 2016. Fungal DNA barcoding. Genome. 59, 913-932
822 https://doi.org/10.1139/gen-2016-0046

823 Xu, J., 2020. Fungal species concepts in the genomics era. Genome 63, 459-468.
824 https://doi.org/10.1139/gen-2020-0022

825 Yurkov, A.M., 2018. Yeasts of the soil - obscure but precious. Yeast 35, 369–378.
826 https://doi.org/10.1002/yea.3310

827 Yurkov, A.M., Kemler, M., Begerow, D., 2012. Assessment of yeast diversity in soils
828 under different management regimes. Fungal Ecol. 5, 24–35.
829 https://doi.org/10.1016/j.funeco.2011.07.004

830 Yurkov, A.M., Röhl, O., Pontes, A., Carvalho, C., Maldonado, C., Sampaio, J.P., 2016a.
831 Local climatic conditions constrain soil yeast diversity patterns in Mediterranean
832 forests, woodlands and scrub biome. FEMS Yeast Res. 16, fov103.
833 https://doi.org/10.1093/femsyr/fov103

834 Yurkov, A.M., Wehde, T., Federici, J., Schäfer, A.M., Ebinghaus, M., Lotze-Engelhard, S.,
835 Mittelbach, M., Prior, R., Richter, C., Röhl, O., Begerow, D., 2016b. Yeast diversity
836 and species recovery rates from beech forest soils. Mycol. Prog. 15, 845–859.
837 https://doi.org/10.1007/s11557-016-1206-8

838