iScience

CellPress
OPEN ACCESS

Article

Global patterns in culturable soil yeast diversity

Himeshi Samarasinghe, Yi Lu, Renad Aljohani, Ahmad Al-Amad, Heather Yoell, Jianping Xu

jpxu@mcmaster.ca

Highlights

Mean annual rainfall is a positive predictor of global soil yeast diversity

International travel predicts number of shared yeast species between countries

41 novel yeast species were discovered from soils in eight countries

Continued culture-based studies are needed to investigate soil yeast populations

Samarasinghe et al., iScience *24*, 103098
October 22, 2021
https://doi.org/10.1016/j.isci.2021.103098

iScience

CellPress
OPEN ACCESS

Article

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[1,3,*]

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 for identifying 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 correlates with soil yeast community structure and composition worldwide. Evidence for anthropogenic influences on soil yeast communities, directly via travel and indirectly via altered rainfall patterns resulting from climate change, is concerning as we found common infectious yeasts frequently distributed in soil in several countries. Our discovery of 41 putative novel species highlights the continued need for culture-based studies to advance our knowledge of environmental yeast diversity.

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 some *Candida* and *Cryptococcus* species that are opportunistic pathogens of humans (Kurtzman et al., 2011).

It is becoming increasingly apparent that the true extent of global soil yeast diversity is significantly underestimated. Although 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 contributes to the ~1500 recognized yeast species on the planet (Kurtzman et al., 2011; Naranjo-Ortiz and Gabaldón, 2019). The yeast discovery rate has accelerated since the turn of this century, with almost 50% of the currently known yeast species having been described within the last 10 years (Wu et al., 2019; Yurkov, 2018; Naranjo-Ortiz and Gabaldón, 2019). Environmental surveys routinely uncover novel yeast species, accounting for as much as 30% of the total yeast populations, highlighting the need to revise current estimates of global yeast diversity (Groenewald et al., 2018; Yurkov et al, 2016a, 2016b).

Lack of adequate environmental sampling, especially in Asia, Africa, South America, and Central America, limits the discovery of novel yeast species, impacts the characterization of soil yeast communities, and hinders the prediction of global yeast 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 a few shared species between sites, even within the same geographical region. For example, one study found that only eight of the 57 species isolated in soils of

[1]Department of Biology, McMaster University, Hamilton, ON, Canada

[2]Department of Infectious Diseases, South Kensington Campus, Imperial College London, London, UK

[3]Lead contact

*Correspondence: jpxu@mcmaster.ca

https://doi.org/10.1016/j.isci.2021.103098

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 shared among 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 focusing 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 variations in mean annual temperature, mean annual rainfall and soil electrical conductivity explained ~44% of the total variance 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. 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 statistically supported 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 owing to the lack of suitable yeast-specific barcoding primers (Xu, 2016). This is mainly because 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 part of the fungal barcoding DNA, ITS2, to assess global soil fungal diversity and to identify predictors of global diversity patterns (Tedersoo et al., 2014). Owing 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 the 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 to be unique between countries in structure and composition, with 73% of the discovered yeast species found in only one of the sampled countries. Similar to that reported by Tedersoo et al. (2014), mean annual precipitation was the most significant predictor of culturable soil yeast diversity on a global scale. Importantly, we found air traffic volume to be significantly correlated with the number of shared species between countries, suggesting a potential 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.

RESULTS

Yeast isolation and species identification

We isolated a total of 1473 yeasts from 3826 soil samples (Table 1). The yeast isolation rate varied among countries, ranging from 17% in Saudi Arabia to 87% in Canada. The yeast isolation rate 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. No statistically significant correlation was observed between yeast isolation rate and either the length of time between sampling and yeast isolation in the lab, or the geographic distance between sampling locations and our lab where the soil yeasts were isolated ($p > 0.10$).

Table 1. Summary statistics of yeast isolation from global soil samples

Country	Soil samples	Yeast isolates	Known species/ Novel species	Country-specific species	Ascomycete species (isolates)/ Basidiomycete species (isolates)	Shannon diversity index
Cameroon	493	110	10/9	12	18 (106)/1 (4)	2.17
Canada	300	261	34/12	25	37 (179)/9 (82)	3.06
China	340	230	23/5	15	15 (100)/13 (130)	2.54
Costa Rica	388	95	20/2	9	18 (88)/4 (7)	2.21
France	327	175	12/2	3	11 (172)/3 (3)	1.26
Iceland	316	211	11/0	4	5 (25)/6 (186)	1.25
New Zealand	610	155	14/4	5	6 (15)/10 (137)	2.05
Peru	490	139	30/9	20	28 (113)/10 (26)	3.27
Saudi Arabia	562	97	8/1	5	3 (3)/6 (94)	0.91
Total	**3826**	**1473**	**90/44**	**98**		

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 the broad paraphyletic genus "*Candida*" being the most species-rich genus (n = 19 species). Specifically, six "*Candida*" species belonged to clade *Lodderomyces*; two belonged to *Pichia* clade, whereas the clades *Barnettozyma, Candida glaebosa, Cyberlindnera, Kurtzmaniella, Ogataea, Wickerhamomyces/Candida*, and *Yamadazyma* were represented by one species each. The four remaining "*Candida*" species have not been assigned to a clade (Table S1). 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 Operational Taxonomic Units (OTUs) at 98.41% nucleotide similarity. They represent potentially novel yeast species since no existing sequences with >98.41% sequence identity to these OTUs were found in the databases. GenBank accession numbers to the ITS sequences of our 1473 isolates are GenBank: MG817572 to GenBank: MG817630 and GenBank: MW894661 to GenBank: MW896112.

Our rarefaction analyses suggested that sufficient soil sampling was conducted in each country to accurately estimate the true diversity of culturable soil yeasts at the respective sampling sites. 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 (Figure S1). Additional sampling in these locations was not likely to have revealed higher yeast species diversity.

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 yeast species, where 562 soil samples yielded 97 yeast isolates belonging to 9 species, one of which was potentially novel. On the other hand, we obtained 261 yeast isolates from 300 Canadian soil samples, encompassing 46 species, 12 of which were potentially novel. The number of yeast isolates and 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 1).

Cameroon

As shown in the study by Aljohani and colleagues, all but one of the 19 yeast species recovered from Cameroonian soil were ascomycetes (Aljohani et al., 2018). The population was dominated by four species:

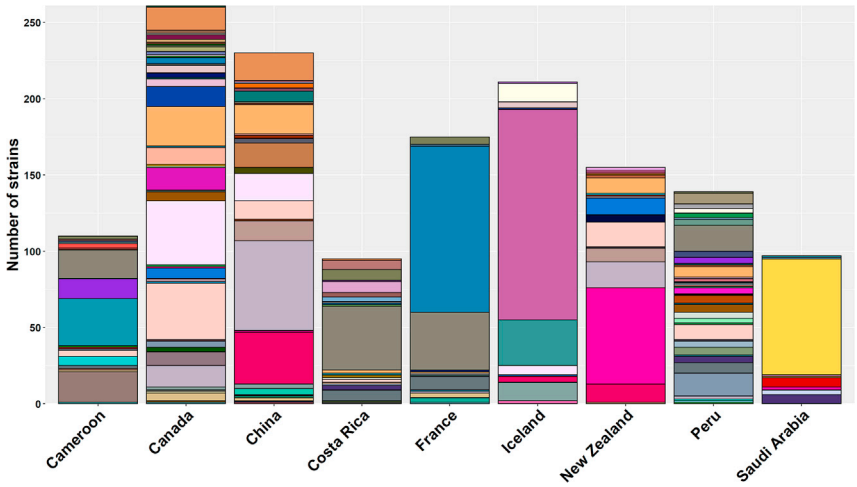

Figure 1. Culturable soil yeast populations by country

The X axis represents the country. Each country is represented by a stacked bar plot. Each color represents a unique species and the height of the colored sections indicate the abundance of that species in the population.
See also Figure S1 and Table S2.

Cyberlindnera subsufficiens (28%), *Torulaspora globosa* (18%), *Candida tropicalis* (17%), and *Cyberlindnera saturnus* (12%). Overall, Cameroonian soil contained the second highest number of potential novel species (n = 9) after Canada (n = 12).

Canada

The culturable yeast population in Canadian soil was dominated by ascomycetes (37 species): the remaining 9 species were basidiomycetes. *Nadsonia starkeyi-henricii*, little-known yeast that prefers relatively mild temperatures below 25°C, was the most abundant (16%) followed by the opportunistic pathogen *Papiliotrema laurentii* (14%). *Debaryomyces hansenii* (10%), *Barnettozyma californica* (6%), and *Candida* sp. (NEW) 8 (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 potential novel species (n = 12) of the nine sampled countries.

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 *Candida* sp. (NEW) 6 (7%). 15 yeast species were only found in the Chinese soil which also yielded five putative novel species.

Costa Rica

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

France

The majority of species in the French culturable soil yeast population were ascomycetes (n = 11) while three species were basidiomycetes. *Candida subhashii*, an opportunistic pathogen 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 opportunistic pathogen *Candida tropicalis* was the second most abundant species (22%), followed by the common Baker's and Brewer's yeast *Saccharomyces cerevisiae* (5%). One strain of *Candida parapsilosis* was also detected in French soils.

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.

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).

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%).

Saudi Arabia

Saudi Arabian culturable soil yeast population was the least diverse of all sampled countries according to the Shannon diversity index. Overall, the Saudi Arabian soil yeast population consisted of six basidiomycete species and three ascomycete species. One of the species was potentially novel. This population was notable for the overwhelming prevalence of the human pathogenic yeast, *Cryptococcus deneoformans* (78%) which is a causative agent of fatal fungal meningoencephalitis. The genotypes of Saudi Arabian *C. deneoformans* strains 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.

Pathogenic yeast species

Based on recent information on yeast trophism (Kurtzman et al., 2011; Opulente et al., 2019), the following 12 species are the most common opportunistic yeast pathogens of humans worldwide: *Candida albicans, Candida dubliniensis, Candida glabrata, Candida guilliermondii* (syn: *Meyerozyma guilliermondii*), *Candida krusei* (*Pichia kudriavzevii*), *Candida lusitaniae* (syn: *Clavispora 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 our global yeast population (Figure 2). *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. In addition, seven *C. krusei* isolates, six *C. albicans* isolates, and one *C. glabrata* isolate 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, and four

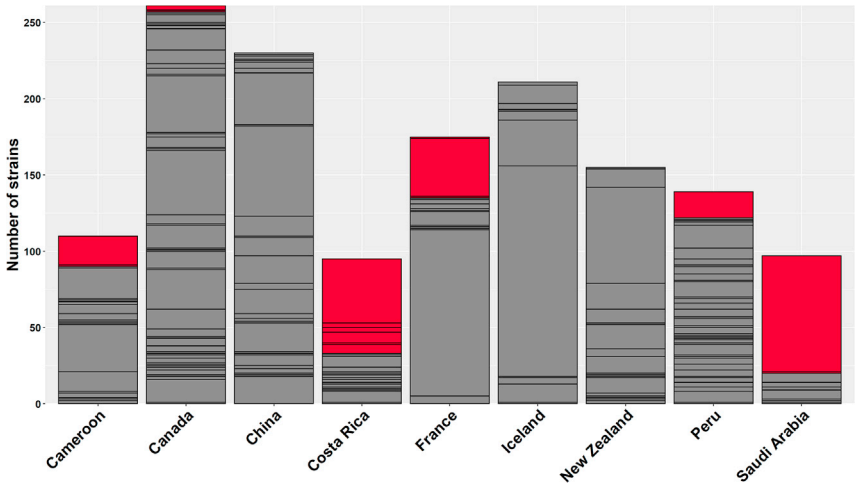

Figure 2. Pathogenic yeast species and their abundance highlighted in red in sampled countries

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.

See also Table S2.

C. orthopsilosis isolates were found in Cameroon and Costa Rica. Soil samples from China, Iceland, and New Zealand did not yield any common yeast pathogens.

Putative novel yeast species

Our yeast population included 44 potentially novel species from eight sampled countries: soil samples from Iceland did not yield any novel yeast species. We determined the most closely related genera for 41 of the 44 putative species by running BLAST searches in the UNITE database. For the remaining three species, we were unable to provide unambiguous identification. Our 41 novel species can be categorized into 12 genera (9 in ascomycetes, 3 in basidiomycetes) with *Wickerhamomyces* containing ten potentially 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 example, the ML tree reveals the distinctness of the potentially novel *Wickherhamomyces* species we isolated in this study (Figure 3). We also observed some geographical clustering where the Cameroonian and Canadian novel species formed their own clusters while the two novel species from Costa Rica and Peru clustered together.

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), whereas the 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) whereas 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 4). We found no significant correlation between the remaining variables and Shannon diversity index (Data S1).

Relationship between air traffic volume and 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 country the pairs

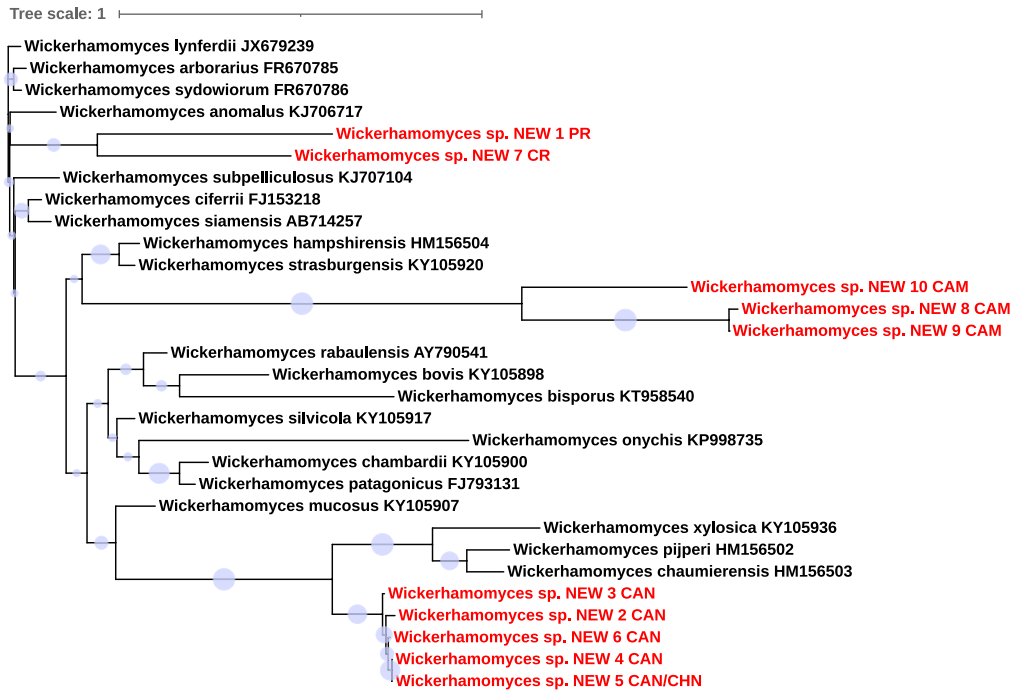

Figure 3. The maximum likelihood tree of *Wicherhamomyces* species based on rDNA ITS sequences

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

ranged from one to 11 (Figure 5B). 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 and 2016. During the same period, only 81 trips were made between Iceland and Cameroon (Figure 5A, Table S3). We performed a linear regression analysis between air traffic volume, geographic distance, and the number of shared species between countries. Although we did not find a significant correlation between geographic distance and the number of shared species, air traffic volume was significantly correlated with the number of shared species between countries ($p = 0.003$, Figures 6 and 7).

We further constructed neighbour-joining (NJ) trees in MEGA7 (Kumar et al., 2016) based on ITS sequences of the four most shared species in our yeast 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 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). This result is consistent with the hypothesis of recent long-distance dispersal for many of the shared species.

Comparison to culture-independent, metagenomics approach

We compared our results to a culture-independent study conducted by Tedersoo and colleagues who used high-throughput sequencing to investigate soil fungal diversity in 39 countries (Tedersoo et al., 2014). We detected 146 OTUs among our global yeast dataset based on the clustering of the ITS2 region at 98.41% sequence identity, whereas the metagenomics study reported a total of 50,589 fungal OTUs, with the number of yeast OTUs not known. Although we were able to annotate all obtained ITS sequences to species level or a higher taxonomic status within the fungal kingdom, ~33% of the fungal OTUs in the metagenomics dataset were merely annotated as environmental_sequence (724, 1.4%), uncultured_soil_fungus (2405, 4.8%), uncultured_ectomycorrhizal_fungus (1407, 2.8%), or uncultured_fungus (11,898, 23.5%). BLAST searches revealed that 26% of our OTUs (38/146) had a significant match (>98.41% identity) in the metagenomics dataset: less than 3% (4/146) matched a fungal OTU from the same country, whereas the remaining 23% (34/146) matched a fungal OTU found in a different country.

Table 2. Environmental and geographic characteristics of sampling sites

Country	City	Site code	Mean annual rainfall (mm)	Mean annual temperature (°C)	Distance to equator (km)	Altitude (meters above sea level)
Cameroon	Babanki	CBB	1813.312	20.75	679.32	1173
	Bambui	CBM	1813.312	20.75	670.44	1274
	Eloundem	CEL	1617.423	24.75	599.4	620
	Makepe	CMK	2825.192	26.83	459.54	62
	Mbalgon	CML	1617.423	25.03	492.84	556
	Mbandoumou	CMD	1617.423	24.22	418.47	719
	Mbingo	CMB	1813.312	20.75	683.76	1909
	Njinikejum	CNJ	1813.312	20.75	693.75	1573
	Simbock	CSM	1617.423	24.99	513.93	643
Canada	Fredericton	CF	1267.842	5.29	5101.56	10
	Vancouver	CV	1567.708	9.95	5470.08	31
China	Ailao Mountain	CAC	1120.492	17.29	2687.31	2782
	Fenyi	CC	1556.404	17.26	3050.28	307
	Jinfo Mountain	CJ	1154.85	15.49	3220.11	2085
	Pangquangou Nature Reserve	CT	467.3346	9.66	4168.05	775
	Taihang mountains west of Jincheng	CSX	515.0269	12.11	3941.61	2012
Costa Rica	El Jardin	EJ	2308.358	27.23	1094.24	152
	La Fotuna	LF	2964.685	25.51	1166.06	392
	La Paz	LP	2964.685	25.51	1132.64	1512
	Manuel Antonio	MA	2308.358	27.23	1042.29	13
	Monteverde	MV	2964.685	25.51	1147.41	1585
	Playa Hermosa	PH	2308.358	27.23	1061.83	7
	Playa Samara Beach	SB	1682.769	26.72	1095.9	7
	Poas Volcano	PV	2957.59	25.51	4786.32	2350
	Samara Town	ST	1682.769	26.72	4851.81	13
	Villas Playa Samara Beach Front Hotel	SH	1682.769	26.72	7280.49	7
France	Hyeres	FHF	703.8039	15.45	7158.39	37
	Uptown/Downtown Nice	FN	814.7077	13.82	7106.22	25
Iceland	Dimmuborgir above Myvatn lake (highlands)	ID	969.2692	1.06	7280.49	283
	Landbrotalaug mini hotspring	IL	823.4154	4.65	7158.39	11
	National park NA Svartifoss	ISF	1904.319	-1.40	7106.22	192
	NA Nautholsvik Geothermal Beach NA Reykjavik University	IN	959.4077	4.17	7117.32	0
	Skútustaðagígar pseudocraters on Myvatn lake	ISP	855.5423	1.93	7277.16	271
	Thingvellir	IT	1107.946	3.05	7131.75	89
New Zealand	Auckland	NA	1172.973	15.13	4089.24	0

(*Continued on next page*)

Table 2. *Continued*

Country	City	Site code	Mean annual rainfall (mm)	Mean annual temperature (°C)	Distance to equator (km)	Altitude (meters above sea level)
Peru	Amazon	PA	2139.142	25.51	1389.72	176
	Cusco	PC	679.0462	9.59	1501.83	3322
	Lima	PL	0	19.57	1336.44	138
	Machu Pichu	PM	563.1885	9.50	1459.65	1940
	Rainbow Mountain	PR	725.7192	5.77	1538.46	4922
	Sacred Valley	PS	636.3077	8.52	1479.63	2866
Saudi Arabia	Alqunfudah	SAA	49.56154	29.58	2122.32	1
	Dammam	SAD	73.69615	27.15	2929.29	5
	Jeddah	SAJ	45.29615	29.25	2362.08	15
	Medina	SAM	65.06154	27.18	2721.72	636
	Umluj	SAU	4.815385	27.63	2780.55	14
	Yanbu	SAY	29.56923	27.69	2666.22	9

Mean annual rainfall, mean annual temperature, distance to equator and elevation of sampling sites are summarized here.

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 patterns and trends 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 potential predictors of culturable soil yeast diversity. Our study uncovered 134 soil yeast species among 1473 isolates, including 41 putative novel 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 drivers of 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. A previous study showed that vegetation was not a strong predictor of soil fungal diversity (Tedersoo et al., 2014). However, factors influencing soil yeast diversity may be different from those affecting overall soil fungal diversity. For example, many yeast species isolated here have been reported as associated with plant materials. Both *P. kudriavzevii* and *M. guilliermondii* are common fruit and leaf inhabitants. *C. tropicalis* and *C. orthopsilosis* are frequently associated with rotting wood and plant materials in different ecosystems (Carvalho et al., 2014; Opulente et al., 2019). Thus, the composition and structure of soil yeast communities may reflect their associations with plants and plant-related substrates present in these soils. In addition, a few yeast species such as *C. tropicalis, C. saturnus, D. hansenii, P. laurentii, S. terrea,* and *S. terricola*, are cosmopolitan and known to be geographically broadly distributed (Kurtzman et al., 2011).

Earlier studies have shown that 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). Here we found a reversal of this pattern 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 can potentially shift the established landscape of soil yeast communities. This is especially alarming given the significant presence of pathogenic yeasts we detected in the soils. Specifically, 15% of all yeast isolates found in our study belong to common opportunistic pathogenic yeast species capable of

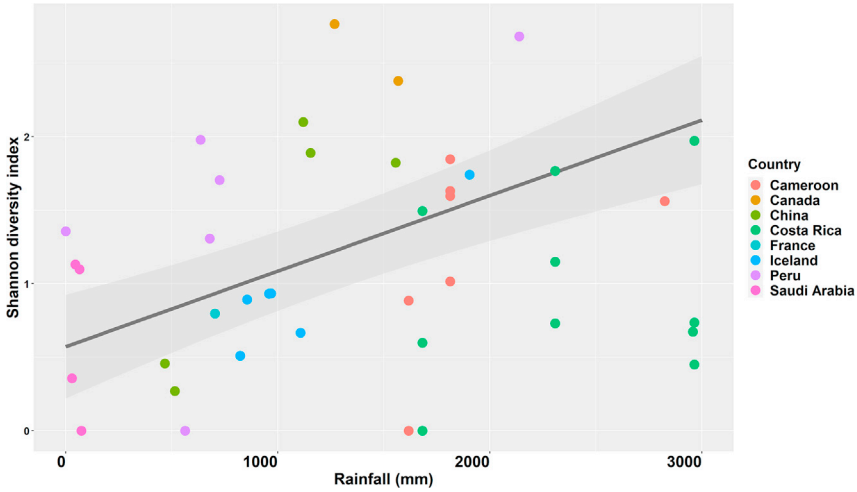

Figure 4. Mean annual rainfall is significantly correlated with soil yeast diversity

Here, the Shannon diversity index of yeast populations is plotted against mean annual precipitation at the sampling sites. Sampling sites are colored by country. The line plots the model predictions with associated uncertainty shaded in gray. See also Data S1.

causing 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 an important source of bacterial and fungal infections, any changes and shifts in soil microbiomes could pose a significant threat to global public 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 and local ecological conditions are likely key factors in limiting the spread and successful colonization of yeasts between populations, at least until recently when anthropogenic activity has strongly improved the connectivity between countries and continents and altered local ecological niches.

Our findings suggest that human activities have likely contributed to changing yeast distributions in soil environments across the globe. International travel has increased exponentially in the past few decades

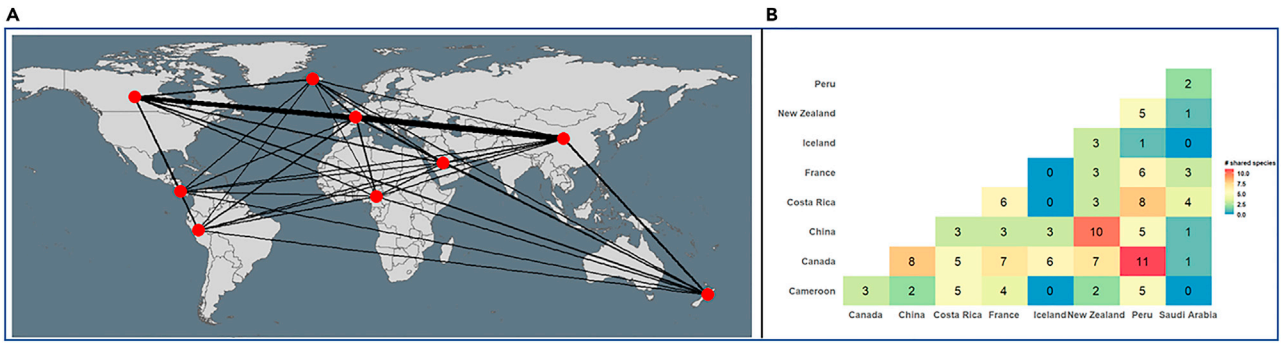

Figure 5. Air traffic volume between countries is correlated with the number of shared species

(A) Volume of air traffic between the nine countries from 2011-2016. Thickness of the line indicates volume. (B) Heatmap showing the number of shared species between country pairs.
See also Table S3.

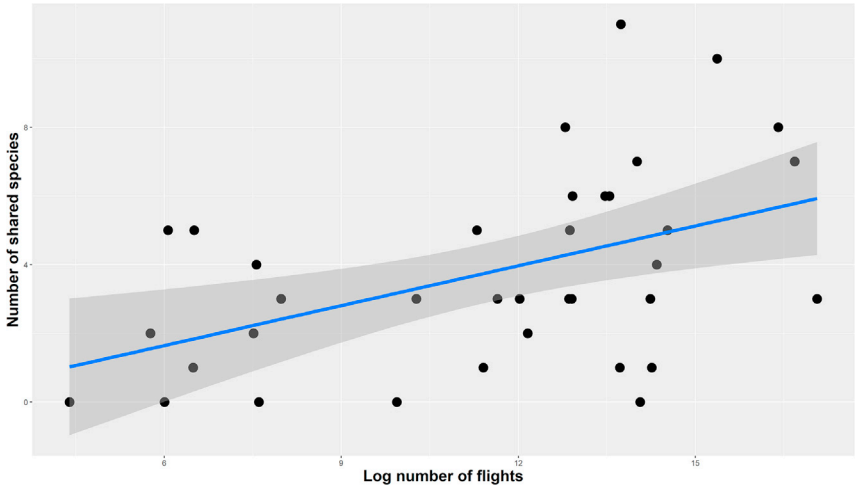

Figure 6. Number of shared species between countries is significantly correlated with traffic volume between them

We plotted the log number of trips that occurred between country pairs from 2011 to 2016 against the number of yeast species shared between them.

with international tourist arrivals increasing from 25 million in 1950 to a record-high 1.4 billion in 2018 (UNWTO World Tourism Barometer and Statistical Annex, 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). Although we found unique yeast species in most localities, 36 yeast species (~25% of all species found) were shared between at least two countries. For some yeast species, their presence in multiple countries could be because of natural dispersal events such as wind throughout that species' history. However, the correlation we found between air traffic volume (but not physical distance) and the number of shared species between countries suggested that human travel has likely facilitated the spread of at least some yeasts across geographical borders. The limited ITS sequence-based geographical clustering of most shared species in our population is also consistent with gene flow between yeast populations in different countries. For *C. tropicalis*, the conclusion of recent gene flow among geographic populations is supported by an analysis of DNA sequences at six gene fragments for 876 global isolates that revealed sharing of multilocus genotypes between countries and continents (Wu et al., 2019). However, the detailed phylogeographic patterns of other shared yeast species remain to be elucidated. The COVID-19 pandemic has greatly shifted the political and economic landscape of our planet. Tourism both within and between countries has seen a drastic drop with accompanying tightening of borders between countries. The potential impact of the COVID-19 pandemic on culturable yeast populations remains to be determined.

Many yeast species have been recently reported and many more remain to be discovered (Kurtzman et al., 2011; Wu et al., 2019; Yurkov, 2018) (Naranjo-Ortiz and Gabaldón, 2019). Both culture-dependent and culture-independent studies routinely discover novel yeast species from the environment. Our study is one of several recent surveys to find previously undescribed species accounting for as much as 30% of natural yeast populations (Yurkov, 2018), implying that every one of three yeast species recovered from the environment could potentially represent a new species. In fact, our results on soil yeast species diversity described here likely represent underestimates of both the total species in these soils as well as the percentage of novel yeast species in the soil. For example, our media and incubation conditions favor fast-growing yeasts under the selected conditions. If both fast-growing and slow-growing yeasts were in the same 0.1 gram of soil, it is likely that we would have recovered the fast-growing ones only due to its growth advantage.

Aside from expanding our understanding of yeast biodiversity, the isolated pure yeast cultures could be of potential applied significance. For example, investigators often turn to natural soils in search of novel yeast strains with commercial and biotechnological potential. In 2018, a novel strain of *Pichia kudriavzevii* isolated from soil in a sugarcane field in Thailand was shown to be more thermotolerant and produce more ethanol than the Thai industrial strain *Saccharomyces cerevisiae* TISTR 5606 (Pongcharoen et al., 2018).

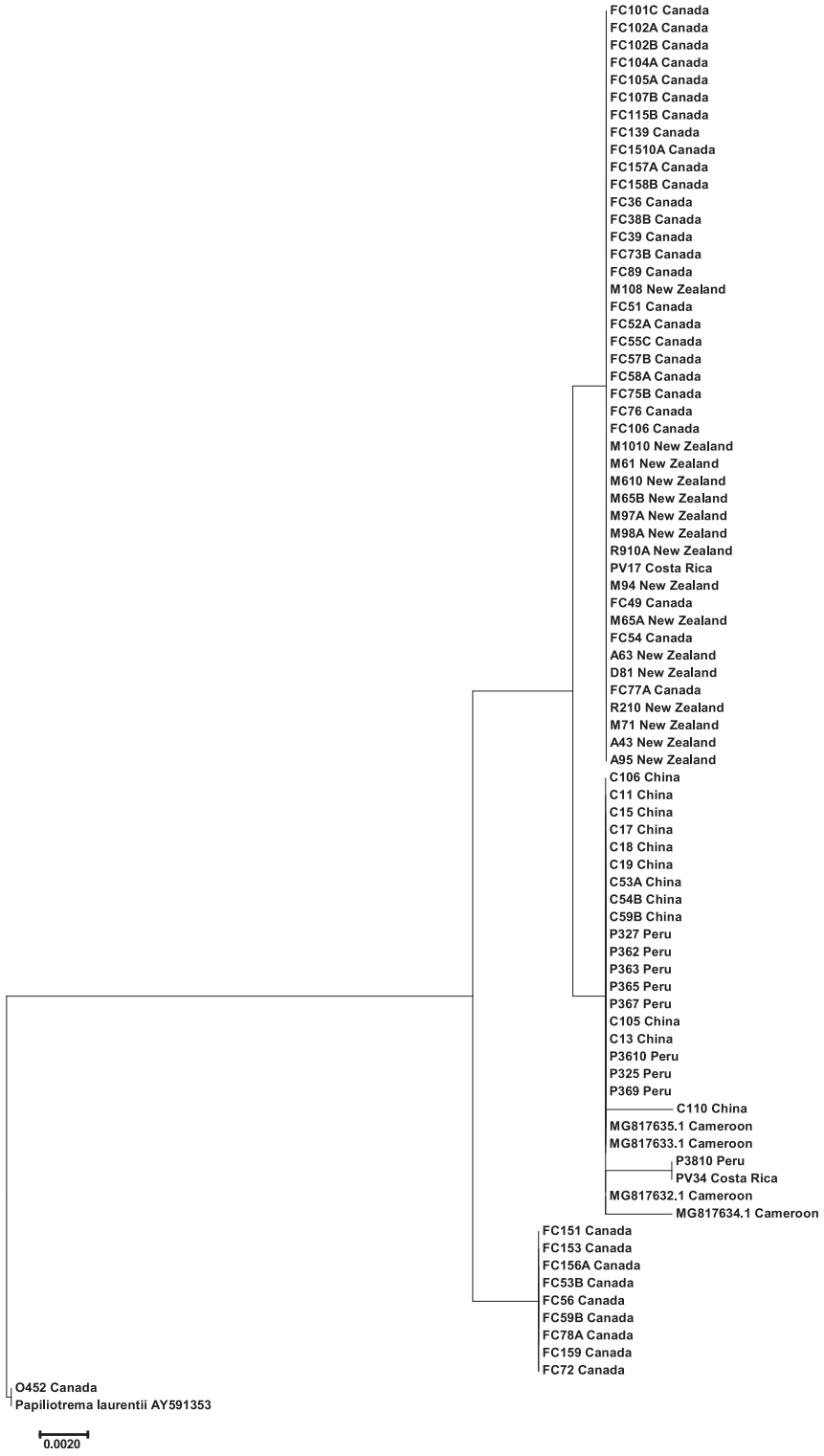

Figure 7. The neighbour-joining tree of Papiliotreme laurentii global isolates found in our study
Limited geographical clustering is observed, suggesting frequent gene flow between populations.

Presence of species of the genus *Kazachstania* in mixed cultures of *Saccharomyces cerevisiae* gives rise to fermented wines with diverse aroma profiles; however, *Kazachstania* species are unable to complete fermentation in monocultures (Jood et al., 2017). The thermo- and halo-tolerant yeast *Blastobotrys adeninivorans* aids in a wide range of biotechnological applications including the production of secretory enzymes, as a host for heterologous gene expression and as a biological component in biosensors (Kunze et al., 2017). At present, the metabolic and fermentative capabilities of the novel yeasts such as those in genera *Kazachstania* and *Blastobotrys*, found in Peruvian and French soils, respectively, and yeasts like *Pichia kudriavzevii* (syn. *Candida krusei*) found in Costa Rican soil are unknown. Discovery of new *Kazachstania* species with more desirable fermentative abilities can aid the full exploitation of this genus in commercial wine fermentation.

In recent years, researchers have come to view metagenomics as a valuable tool in the investigation of microbial diversity in complex ecological systems. The large amount of data generated from high throughput sequencing is crucial for unearthing large-scale patterns at higher taxonomic levels. For example, a 2015 study compared culture-independent vs. culture-dependent characterizations of microbes from hydrocarbon-contaminated soil and found that the two methodologies provided diverging views of microbial communities, with only 8.2% of the fungal OTUs being shared between the two datasets (Stefani et al., 2015). A meta-analysis from 2019 estimated the total number of fungal species to be 7.8-8.8 times that of culturable species (Wu et al., 2019). However, limited information on yeast diversity could be extracted from previous metagenomics studies on global soil fungal diversity (Egidi et al., 2019; Tedersoo et al., 2014). Yeasts cannot be easily identified based on DNA sequences alone: extracting ITS sequences of known yeasts from large metagenomics datasets is a time-consuming task that requires personnel with advanced knowledge of yeast taxonomy. For potentially novel species, the metagenomic approach would completely fail to identify them as yeasts, as they rely on the current state of knowledge on species taxonomy and annotation. In Tedersoo et al.'s study (2014), over 30% of the fungal OTUs remained minimally annotated. In the future, it may be possible to design multiple sets of primers to target known groups of yeasts through the metagenomic approach while excluding their respective, closely related groups of filamentous fungi. However, even with this approach, only those closely related to known yeasts will be amplified whereas novel yeasts may still be missed. Here, using culture-dependent methods, we succeeded in isolating many soil yeasts, which aided in the characterization of global patterns of soil yeast distribution. Our global soil yeast collection 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 approaches (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). In addition, it has been estimated that on average, 40% of fungal DNA in soil are extracellular or comes from cells that were no longer intact, causing estimates of fungal richness to be inflated by as much as 55% (Carini et al., 2016). Selective enrichment and culturing from soil samples in the lab remains an effective approach for studying soil yeasts.

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 point to international travel being a potentially significant contributing factor to 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, obtain isolates with commercial and biotechnological value, and monitor species that could pose a threat to human health.

Limitations of the study

Our study could be limited by the culturing conditions used which likely favored isolation of fast-growing yeasts over slow-growing species. We did not measure chemical properties such as pH and carbon content of the soil at sampling locations. This additional information could be useful in explaining some of the yeast

diversity patterns observed in our study. Our study could also be limited by the number of sampling locations. If sampling were extended beyond the locations in nine countries surveyed here, more generalizable global diversity patterns with finer detail may have been elucidated.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Soil collection
  - Yeast isolation from soil samples
  - Yeast identification via ITS sequencing
  - Statistical analyses of population diversity
  - Relationship between yeast diversity and climate and geographic factors
  - Air traffic data
  - Comparison to metagenomics study
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103098.

ACKNOWLEDGMENTS

We thank the following people for contributing soil samples: Haoran Jia, Yongjie Zhang, Thomas Harrison, Sebastian Harrison, Arshia Kazerouni, and Gregory Korfanty. HS is supported by an NSERC CGS-D Scholarship. This research is supported by Natural Sciences and Engineering Research Council of Canada (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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: April 13, 2021
Revised: August 9, 2021
Accepted: September 2, 2021
Published: October 22, 2021

REFERENCES

Abbasian, F., Palanisami, T., Megharaj, M., Naidu, R., Lockington, R., and 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., and 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., and 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., and 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., and 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., and Xu, J. (2018). Diversity and relationships among strains of culturable yeasts in agricultural soils in Cameroon. Sci. Reports *81*, 1–11.

Bates, D., Mächler, M., Bolker, B.M., and 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.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., Wit, P. De, Sánchez-García, M., Ebersberger, I., Sousa, F.de, et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. *4*, 914–919. https://doi.org/10.1111/2041-210X.12073.

Białkowska, A.M., Szulczewska, K.M., Krysiak, J., Florczak, T., Gromek, E., Kassassir, H., Kur, J., and Turkiewicz, M. (2017). Genetic and biochemical characterization of yeasts isolated from Antarctic soil samples. Polar Biol. *40*, 1787–1803. https://doi.org/10.1007/s00300-017-2102-7.

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

Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., and Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat. Microbiol. *2*, 16242. https://doi.org/10.1038/nmicrobiol.2016.242.

Carvalho, C., Yang, J., Vogan, A., Maganti, H., Yamamura, D., and Xu, J. (2014). Clinical and tree hollow populations of human pathogenic yeast in Hamilton, Ontario, Canada are different. Mycoses *57*, 271–283. https://doi.org/10.1111/MYC.12156.

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

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

Findlater, A., and Bogoch, I.I. (2018). Human mobility and the global spread of infectious diseases: a focus on air travel. Trends Parasitol. https://doi.org/10.1016/j.pt.2018.07.004.

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

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

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

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

Into, P., Pontes, A., Sampaio, J.P., and Limtong, S. (2020). Yeast diversity associated with the phylloplane of corn plants cultivated in Thailand. Microorganisms *8*, 80. https://doi.org/10.3390/microorganisms8010080.

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

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

Kahle, D., and Wickham, H. (2013). ggmap: spatial Visualization with ggplot2. R. J. *5*, 144–161.

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

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054.

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

C. Kurtzman, J.W. Fell, and T. Boekhout, eds. (2011). The Yeasts A Taxonomic Study, 5th edition (Elsevier).

Li, A.H., Yuan, F.X., Groenewald, M., Bensch, K., Yurkov, A.M., Li, K., Han, P.J., Guo, L.D., Aime, M.C., Sampaio, J.P., et al. (2020). Diversity and phylogeny of basidiomycetous yeasts from plant leaves and soil: proposal of two new orders, three new families, eight new genera and one hundred and seven new species. Stud. Mycol. *96*, 17–140. https://doi.org/10.1016/j.simyco.2020.01.002.

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

Monteiro Moreira, G.A., and Martins do Vale, H.M. (2020). Soil yeast communities in revegetated post-mining and adjacent native areas in Central Brazil. Microorganisms *8*, 1116. https://doi.org/10.3390/microorganisms8081116.

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

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

Naranjo-Ortiz, M.A., and Gabaldón, T. (2019). Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. Biol Rev Camb Philos Soc *94*, 2101–2137. https://doi.org/10.1111/brv.12550.

Ogaki, M.B., Teixeira, D.R., Vieira, R., Lírio, J.M., Felizardo, J.P.S., Abuchacra, R.C., Cardoso, R.P., Zani, C.L., Alves, T.M.A., Junior, P.A.S., et al. (2020). Diversity and bioprospecting of cultivable fungal assemblages in sediments of lakes in the 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., and 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., et al. (2020). Package "Vegan" Title Community Ecology Package Version 2, pp. 5–7.

Opulente, D.A., Langdon, Q.K., BuhV, K., Haase, M.A.B., Sylvester, K., MoriartyV, R., Jarzyna, M., Considine, S.L., Schneider, R.M., and Hittinger, C.T. (2019). Pathogenic budding yeasts isolated outside of clinical settings. FEMS Yeast Res *19*. https://doi.org/10.1093/FEMSYR/FOZ032.

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., Valero, M., et al. (2015). Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol. Ecol. *24*, 525–544. https://doi.org/10.1111/MEC.13048.

Pongcharoen, P., Chawneua, J., and 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., and 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., and Triplett, E.W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. *1*, 283–290. https://doi.org/10.1038/ismej.2007.53.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*. https://doi.org/10.7717/PEERJ.2584.

Samarasinghe, H., Aljohani, R., Jimenez, C., and 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., and 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.

Stefani, F.O.P., Bell, T.H., Marchand, C., Providencia, I.E.de la, Yassimi, A. El, St-Arnaud, M., and Hijri, M. (2015). Culture-dependent and -independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. PLoS One *10*, e0128272. https://doi.org/10.1371/JOURNAL.PONE.0128272.

Suleiman, M.K., Dixon, K., Commander, L., Nevill, P., Quoreshi, A.M., Bhat, N.R., Manuvel, A.J., and 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., et al. (2014). Global diversity and geography of soil fungi. Science *346*, 1256688.

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

UNWTO World Tourism Barometer and Statistical Annex (2018). UNWTO world tour. Barom *16*, 1–36. https://doi.org/10.18111/wtobarometereng.2018.16.1.5.

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

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

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

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

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

Wang, X., Wang, Y., Ning, S., Shi, S., and Tan, L. (2020). Improving azo dye decolorization performance and halotolerance of Pichia occidentalis A2 by static magnetic field and possible mechanisms through comparative transcriptome analysis. Front. Microbiol. *11*, 712. https://doi.org/10.3389/fmicb.2020.00712.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., and Xiang, M. (2019). Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology *10*, 127–140. https://doi.org/10.1080/21501203.2019.1614106.

Xu, J. (2016). Fungal DNA barcoding. Genome. https://doi.org/10.1139/gen-2016-0046.

Xu, J. (2020). Fungal species concepts in the genomics era. Genome *63*, 459–468. https://doi.org/10.1139/GEN-2020-0022.

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

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

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

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Chloramphenicol ≥98% (HPLC)	Sigma-Aldrich	SKU: C0378-5G
Benomyl	Toronto Research Chemicals	Cat No: B161380
Deposited data		
ITS sequences of soil yeasts	This paper	GenBank accession numbers MG817572 to MG817630 and MW894661 to MW896112
Oligonucleotides		
ITS1	Integrated DNA Technologies	5′ TCCGTAGGTGAACCTGCGG 3′
ITS4	Integrated DNA Technologies	5′ TCCTCCGCTTATTGATATGC 3′
Software and algorithms		
DNA Baser DNA Sequence Assembler	Heracle BioSoft S.R.L.	NA
BLAST+ 2.11.0	National Centre for Biotechnology Information	NA
RStudio 4.0.2	R Studio	NA
ITSx 1.1.3	The Bengtsson-Palme Lab	NA
QIIME 2 2021.4.0	Knights and Caporaso labs	NA
Excel 2019	Microsoft Corporation	NA
Other		
GoTaq Green MasterMix	Promega	Cat No: M7123

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jianping Xu (jpxu@mcmaster.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All ITS sequences generated from Sanger sequencing have been deposited in NCBI and accession numbers are listed in the results section. Geographical and climatic information pertaining to soil sampling locations are provided in Tables 1 and S1. Details of soil yeasts isolated in this study are provided in Table S2. Air traffic data of sampled countries are provided in Table S3. Original code used in data analysis and any additional information required to re-analyze the data reported in this paper are available from the lead contact upon request.

METHOD DETAILS

Soil collection

We collected soil from 53 locations in nine countries encompassing all continents except Antarctica (Figure 8 and Table S1). At each location, we set up multiple plots: within each plot, we collected ten samples of topsoil (approximately 1g each) between 1-3 inches from the surface, using sterile protocols. Within each plot, the ten soil samples were at least 2m from each other. The 10 soil samples from the same plot were stored in the same sterile, 3cm x 7cm, resealable polyethylene bag to minimize bag usage and maximize soil representation from each plot. Once transported to our lab at McMaster University, Canada, we segregated the soil in each bag into ten 1g aliquots in 1.5ml Eppendorf tubes and stored at 4°C. Some bags had slightly more than 10g of soil and those were separated into additional tubes. In total, this study

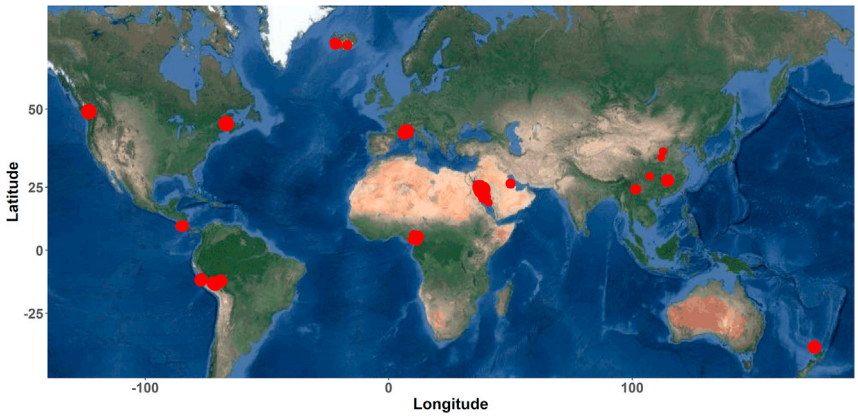

Figure 8. Soil sampling locations. 3826 soil samples were obtained from 53 locations, indicated by the red circles, in nine countries

The size of the circle corresponds to the number of samples obtained from that location.
See also Table S1 and Figure S1.

investigated 3826 aliquot soil samples (from 380 bags) collected from the following countries: Cameroon (493 from 49 bags), Canada (300 from 30 bags), China (340 from 34 bags), Costa Rica (388 from 38 bags), France (327 from 32 bags), Iceland (316 from 31 bags), New Zealand (610 from 61 bags), Peru (490 from 49 bags) and Saudi Arabia (562 from 56 bags).

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. For Iceland soil samples, we extended this incubation step to 72 hours 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).

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 sequenced using the Sanger method 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 (www.DnaBaser.com). 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. 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. In addition, we performed a massBLASTer analysis for the putative novel sequences in the curated UNITE database (https://unite.ut.ee/) to identify the genus they are most closely affiliated with and the associated species hypothesis (SH) codes (Pante et al., 2015).

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.

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 26-year period from 1991-2016, available on Climate Change Knowledge Portal (https://climateknowledgeportal.worldbank.org/). We calculated the elevation of the sampling sites and their distance from 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.

Air traffic data

We extracted data on the number of trips 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.

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, covering a portion of the DNA barcoding fragment we sequenced here (Tedersoo et al., 2014). Four countries overlapped between the two studies, namely, Cameroon, Canada, China, and New Zealand. We conducted BLAST searches of our ITS sequences against the entire metagenomics dataset to identify the potential distributions of our cultured yeast species in the larger global soil samples. First, we extracted the ITS2 region from all of our full ITS sequences using the ITSx software (Bengtsson-Palme et al., 2013). Next, we used the QIIME2 VSEARCH tool to cluster the ITS2 sequences at 98.41% sequence identity into operational taxonomic units (OTU) (Rognes et al., 2016). Finally, we used the blastn functionality from the BLAST+ toolset to query our OTUs against the full OTU dataset from the metagenomics study. The output was compiled into a CSV file containing the top 5 matches for each query OTU sequence. This CSV file was perused manually to identify significant sequence similarity between query and match sequences.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of quantification and statistical analyses conducted in this study can be found in the preceding sections of the STAR Methods. Where appropriate, significance was defined as a p-value of less than 0.05.