Received: 9 August 2021 | Revised: 9 December 2021 | Accepted: 23 December 2021

DOI: 10.1002/yea.3689

RESEARCH ARTICLE

Yeast WILEY

The search for *Schizosaccharomyces* fission yeasts in environmental metatranscriptomes

Rasha Shraim[1,2,3] | Bart P. S. Nieuwenhuis[3]

[1]The SFI Centre for Research Training in Genomics Data Sciences, National University of Ireland Galway, Galway, Republic of Ireland

[2]Department of Public Health and Primary Care, School of Medicine, Trinity College Dublin, Dublin, Republic of Ireland

[3]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence
Bart P. S. Nieuwenhuis, Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany.
Email: nieuwenhuis@biologie.uni-muenchen.de

Abstract

Fission yeast is an important model organism in evolutionary genetics and cell biology research. Nevertheless, most research is limited to a single laboratory strain and knowledge of its natural occurrence is limited, which reduces our understanding of its life history and hinders isolation of new strains from nature. Understanding the natural diversity of fission yeast can provide insight into its genetic and phenotypic diversity and the evolutionary processes that shaped these. Here, we aimed to identify candidate natural habitats of fission yeasts by searching through a large collection of publicly available environmental metatranscriptomic datasets. Using a custom pipeline, we processed over 13,000 NCBI SRA accessions, from a wide range of 34 different environmental categories. Overall, we found a very low abundance of putative yeast transcripts, with most fission yeast signatures coming from the categories of 'food' and 'terrestrial arthropods'. Additionally, a signal could be found in a variety of marine and fresh aquatic habitats. Our results do not provide a conclusive answer on the natural habitat of fission yeasts, but our analysis further narrows the range of locations where fission yeasts naturally occur.

Take Aways

- We analysed published environmental metatranscriptomes from the NCBI SRA.
- We created a pipeline to select for *Schizosaccharomyces* reads.
- We identified candidate natural environments for *Schizosaccharomyces* spp.

KEYWORDS
biodiversity, environmental DNA, fission yeast, habitat, *Schizosaccharomyces*

1 | INTRODUCTION

While microorganisms make up most of the Earth's biodiversity, the biogeographical patterns of their distribution can often be elusive (Monard et al., 2016). Many produce long-surviving spores that can spread across large distances through air or water or be dispersed by animals (Monard et al., 2016) to germinate and grow where the environment allows them (Becking, 1934). Given such expansive distribution patterns, the environment where a microorganism is isolated from may not necessarily correspond to its growth habitat. To answer the question 'what lives where?', studies analysing the spatial distribution of microorganisms, at first using culturing methods, later using phylogenetically informative genetic markers (Fuhrman, 2009) could thus perform taxonomic profiling (location and abundance of species). This was later extended to functional profiling, which required knowledge of the physiological activity of species (Franzosa et al., 2018). Transcriptomic activity can be used as a proxy for physiological activity and might better reflect the functional niche of a species than its

mere occurrence based on genetic markers (Schneider et al., 2021; Shakya et al., 2019). In the last decade, transcriptomic metacommunity research has given important insights into the occurrence, abundance and interactions of microorganisms in their local communities (Westermann & Vogel, 2021).

Even in highly studied model organisms, the knowledge of the species' natural habitat is not always available, potentially due to their low abundance in nature, or due to under-sampling of specific habitats. Obtaining knowledge of 'where lives what?', what being the organism of interest, is essential to better understand its physiology and molecular genetics. Here, we study the natural occurrence of the model fission yeast species *Schizosaccharomyces pombe* and three related species in the *Schizosaccharomyces* genus.

S. pombe, commonly known as fission yeast, is an important unicellular model organism for eukaryote biology, especially cell biology (Jeffares, 2018; Yanagida, 2002). It has been used in studying cell cycle control, transcription, translation and mating systems among other topics (Hoffman et al., 2015). *S. pombe* was discovered over a century ago and has been a part of biological research since the second half of the 20th century. It was first isolated from East African millet beer (Lindner, 1893). Urs Leupold, considered the main scientist to establish *S. pombe* as a model organism, used a single strain isolated from grape juice in Switzerland (Osterwalder, 1924). Three of its mating type-variant strains have been the basis for most genetic fission yeast studies since (Fantes & Hoffman, 2016; Leupold, 1949). These near-isogenic reference lab strains have made *S. pombe* a convenient and powerful genetically tractable model due to the consistency in the data produced by the different labs studying it (Fantes & Hoffman, 2016).

There are two sides to the motivation of searching for populations of fission yeast in their natural environments. First, knowledge of the natural *diversity* of an organism as well as access to varied strains can further our understanding of its genetics and expressed traits. Second, discovering the natural *environment* of fission yeast may elucidate its evolution in terms of its interaction with this environment.

In fission yeast, the wild type and its many derived mutant strains enable straightforward experimental designs and interpretations (Decottignies et al., 2003); however, this approach is limited by the specificity of each mutation in a single genetic background. While the benefits of working with a single strain are evident, understanding its natural *diversity* can provide valuable insight into genetic and phenotypic diversity, heritability and other evolutionary processes (Jeffares et al., 2015; Parts, 2014). One example comes from baker's yeast where genetic studies of diverse *Saccharomyces cerevisiae* strains helped explain the natural phenotypic variation observed in its oenological or beer-brewing traits (Gallone et al., 2016; Salinas et al., 2012). Another comparative genomics study between natural yeast strains (*S. cerevisiae* and *Saccharomyces paradoxus*) provided insight into the genomic and functional variation underlying their natural diversity and possible evolutionary differences between the two species (Bergström et al., 2014). In fission yeast, natural diversity has expanded our knowledge of evolutionary, genetic and phenotypic aspects of the species (Jeffares, 2018). For example, extending research to include other natural isolates has led to the discovery of meiotic drive mechanisms (Zanders et al., 2014), the identification of gene annotations (Hu et al., 2015) and structural variation in the nuclear (Rhind et al., 2011; Tusso et al., 2019) and mitochondrial (Tao et al., 2019) genome.

Understanding the natural *environment* of fission yeast may provide valuable insight into its evolution and speciation. It is well established that an organism's genotype and its environment interact and that this interaction affects the expressed phenotype. Environmental fluctuations can influence the plasticity of quantitative traits (Via & Lande, 1985). The environment can also dictate the effective population size (Melbinger & Vergassola, 2015). Therefore, identifying the natural non-anthropogenic environments of fission yeast may elucidate their evolutionary history and their current genetic and phenotypic features. One relevant example is the bacteria *Bacillus*, which had been assumed to be a soil bacterium (Hong et al., 2009). The establishment of the gastrointestinal (GI) tract as one of its true habitats clarified features like endospore-formation and protein encasing and resolved observations conflicting with the soil-environment assumption such as its inability to sporulate below 15°C (Hong et al., 2009). Knowledge of the natural environment of fission yeast will inform us on the natural temperature of this environment and thus to which temperatures most physiological processes the species is adapted, a factor of relevance for enzymatic dynamics (Pluskal et al., 2009), but also, for example, meiotic crossover rates (Hyppa et al., 2014). From the growth medium, we can deduce the number of asexual cell cycles that are possible before resources run out, giving information about the number of asexual versus sexual cycles in nature which is mostly unknown (Farlow et al., 2015; Hernández et al., 2021). The density in such a habitat will inform us of intraspecific competition and the need for haploid selfing or the potential for outcrossing (Hernández et al., 2021; Leupold, 1949; Nieuwenhuis et al., 2018).

Therefore, to fully utilize *S. pombe* as a model organism, knowledge of its natural growth environment is needed. Information on its natural diversity and habitat alongside existing biological research will lead to a more holistic appreciation of this organism that may expand its current potential or give rise to new research questions.

In the past few decades, systematical attempts have been made to find and collect natural strains with varying levels of success (Á. Benito et al., 2018, from honey; Gomes et al., 2002, from cachaça; Hellberg, 2013, from environmental substrates). Most strains have been incidentally discovered in microbiology studies of high-sugar products or natural fermentations such as wine, rum, honey and kombucha (see Jeffares, 2018, for detailed list). However, despite being globally distributed, *S. pombe*'s natural origin and dispersal patterns remain largely unknown (Jeffares, 2018; Tusso et al., 2019). So far, these strains have almost exclusively been found as human commensals (Jeffares, 2018). It is therefore tempting to describe this species as a human-associated specialist; however, the *Schizosaccharomyces* clade arose long before primates, approximately 220 million years ago (Rhind et al., 2011). The most closely related

species in the clade, *Schizosaccharomyces cryophilus* and *Schizosaccharomyces octosporus*, split approximately 119 mya from *S. pombe* (see fig. S3 in Rhind et al., 2011). All of the species in this genus have very similar growth forms and nutritional requirements (Hayles & Nurse, 2018), suggesting that their natural niche must have existed equally long, and it is thus probable that fission yeasts have natural growth habitats that are not human associated.

The aim of this research, then, was to search for genetic signatures of the four *Schizosaccharomyces* species and identify their natural biotopes. We explored published metatranscriptomic data in order to identify candidate growth habitats using a bioinformatic pipeline that combines several published tools. Given that where a microorganism is isolated may not correspond to its growth habitat, metatranscriptomes were chosen over metagenomes to select for environments in which *Schizosaccharomyces* spp. are more likely to be transcriptionally active. We utilized a wealth of publicly available biological data that is worth 'reusing' and exploring further (Pasquetto et al., 2017; Stephens et al., 2015). We selected projects across different environments that have been sampled and sequenced to search for signatures of fission yeasts in as many places as possible.

2 | METHODS

The strategy used for the analysis of the available datasets with short read data from environmental samples is outlined in Figure 1 and described in detail below. Briefly, FASTQ files were downloaded from ENA, checked for encoding, filtered and mapped with BBSplit followed by verification using BLAST. The rationale behind this strategy was to use low processing power and rapidly reduce file size and to use BLAST—a slow but precise algorithm—only on likely candidate reads. Scripts are available at https://github.com/rshraim/schizosaccharomyces.

2.1 | Data

Metatranscriptomic metadata was downloaded from NCBI SRA using the query (metatranscriptomic[Source]) AND (illumina[Platform]). The resulting accession list was downloaded in the format of RunInfo, which was then additionally filtered to remove accessions using the criteria: LibrarySelection = 'PCR' or LibraryStrategy = 'AMPLICON' in order to filter out metagenomic (i.e., not metatranscriptomic) data. Additionally, accessions with ScientificName = 'human|gut|Human|sapiens|virus|viral|bacterium|archaeon|bacter|archaea|Bacter' were removed (see below for argumentation). In total, 13,286 accessions from 3473 BioProjects were processed with a median of 2.74×10^9 and mean of 8.86×10^9 for the number of reads. A perl script by Michael Gerth was used to download data from the European Nucleotide Archive (Leinonen et al., 2011), using SRR/ERR/DRR accessions numbers as FASTQ files with ascp from Aspera (aspera/3.7.2; sra_download.pl, Gerth, 2018). A full list of accessions and BioProjects used is given in Table S1. The BioProjects were assigned to 34 different groups representing a variety of environments (air, algae, aquatic, aquatic animal, arthropod, bird, compost, coral, food, fungus, gut, human, hydrocarbon, indoor, industrial, leaf, leaf litter, lichen, mammal, marine, marine animal, marine sediment, micro-organism, mineral, plant, root/rhizosphere, seagrass, sediment, soil, subsurface, synthetic, vertebrates, waste, worm) based on the description of the 'ScientificName' as assigned in SRA. All projects with positive hits for fission yeast (described below) were finally—when possible—manually curated by verification of the original publication—either using the metadata submitted to SRA or from the BioProject or accession numbers if mentioned in the manuscripts.

2.2 | Processing

Before filtering the data, Phred encoding was checked using the script *testformat.sh* from the BBMap package (bbmap/37.28; Bushnell). This is necessary since most downstream tools default to ASCII 33 Phred encoding (also known as Sanger) and must be specified otherwise. FASTQ files with ASCII 64 encoding (also known as Illumina) were reformatted to Sanger encoding with the *reformat.sh* script from the BBMap package. Following this, adapter sequences were trimmed (Illumina TruSeq, Nextera and RNA PCR primer adapters) and reads shorter than 50 bp were discarded using BBDuk (bbmap/37.28; Bushnell). Irrespective of paired-end or single-end sequencing strategy, reads were treated as single-end reads without merging.

2.3 | Mapping

To rapidly find true positive reads and discard all other reads including those from related species, we used the four known fission yeast species from the *Schizosaccharomyces* genus, as well as a variety of other fungal species. Reference genomes from the four available *Schizosaccharomyces* species (*S. pombe*, *S. octosporus*, *S. cryophilus* and *Schizosaccharomyces japonicus*) as well as those from *Protomyces lactucaedebilis*, *Ustilago maydis* and *S. cerevisiae*, representing a taphrinomycete, basidiomycete and saccharomycete (see Table S2 for reference's accession numbers). To reduce false positives in the form of ambiguous reads mapping to the *Schizosaccharomyces* genomes, the *Schizosaccharomyces* genomes were masked for interspersed repeats and low complexity DNA sequences using RepeatMasker, which utilizes RepBase as a repeat reference database (repeatmasker/4.0.7; Bao et al., 2015).

BBSplit (bbmap/37.28; Bushnell, 2014) was then used to map against two artificial references: (1) the masked *Schizosaccharomyces* genomes and (2) the unmasked genomes from the three other fungi. These genomes are similar enough to the *Schizosaccharomyces* species that most ambiguous reads, repeats or ribosomal RNA (rRNA) should map preferentially to this reference, without discarding true *Schizosaccharomyces* reads. The resulting FASTQ files of *Schizosaccharomyces*-mapped reads were converted to FASTA format using *reformat.sh* from the BBMap package (Bushnell, 2014). These

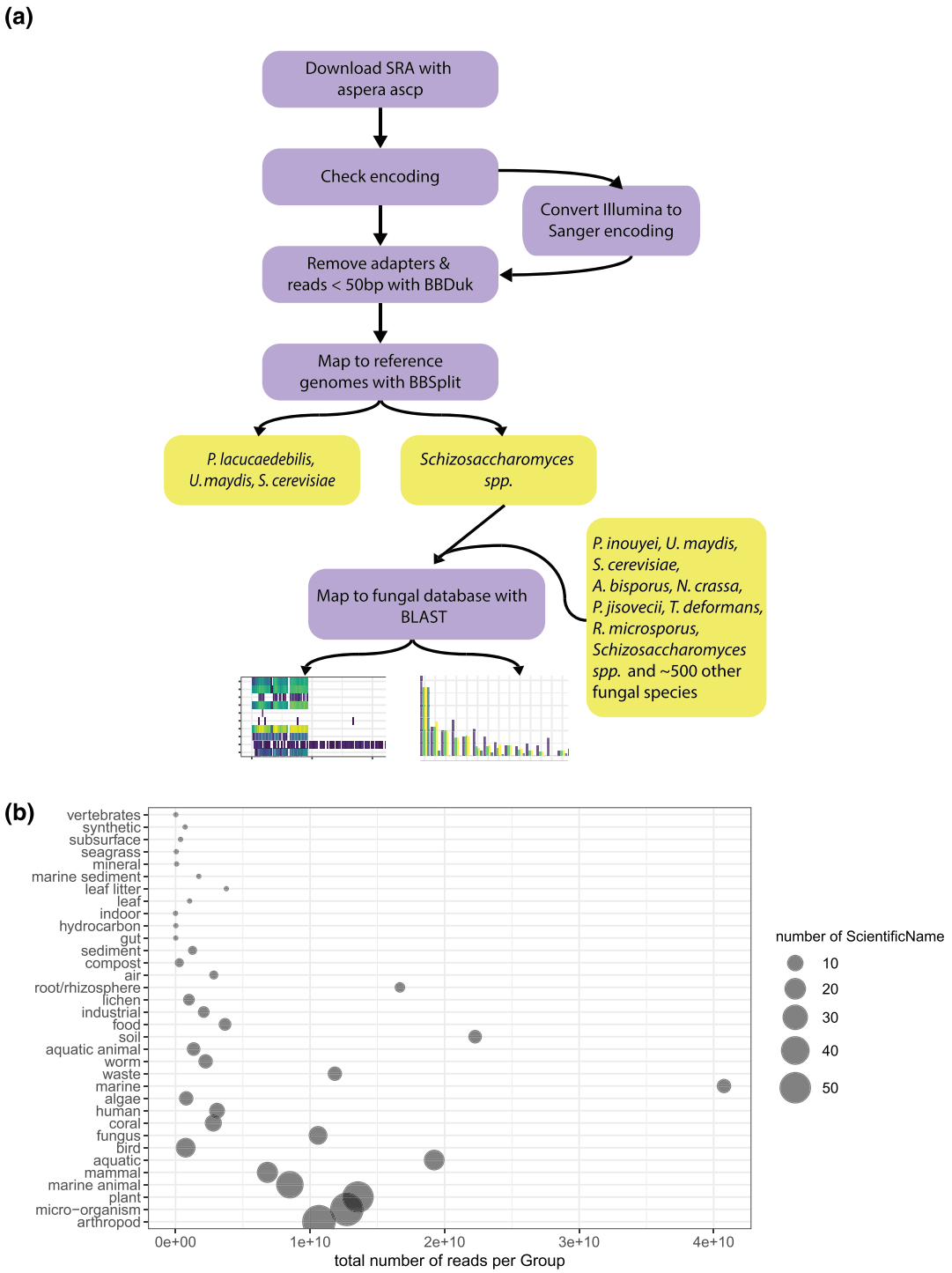

FIGURE 1 (a) Analysis pipeline: Reads were downloaded from the ENA by SRA accessions. Encoding was checked and, if Illumina (ASCII 64), converted to Sanger (ASCII 33). Reads were then trimmed for adapter sequences and any reads shorter than 50 bp were discarded. Trimmed reads were mapped with BBSplit against two references, X and Y (X: the masked *Schizosaccharomyces* species and Y: unmasked *Protomyces lactucaedebilis*, *Ustilago maydis* and *Saccharomyces cerevisiae* genomes). Reads that were mapped to the X reference were then aligned with BLAST against a fungal database and the top BLAST hits of *Schizosaccharomyces* were analysed and visualized in R. (b) Range of processed SRA accessions: Reads from the SRA were grouped into 23 groups according to the environment or organism sequenced (see Section 2). For each organism or environment, the size of the circle indicates the number of different 'ScientificName' identifiers obtained from NCBI SRA metadata (e.g., '*Apis mellifera*' or 'salt marsh'). The total number of reads per group ranged from 5.4×10^5 to 6.06×10^{10} and the number of unique 'ScientificName' identifiers grouped together ranged from 1 to 58. A detailed list of the grouping is available in Table S1 [Colour figure can be viewed at wileyonlinelibrary.com]

were then BLAST aligned against a fungal database containing the following reference genomes chosen as representative species from the major groups in the Fungi kingdom: *P. lactucaedebilis*, *Protomyces inouyei*, *U. maydis*, *S. cerevisiae*, *Rhodotorula toruloides*, *Agaricus bisporus*, *Neurospora crassa*, *Pneumocystis jirovecii*, *Taphrina deformans*, *Rhizopus microsporus* as well as the masked *Schizosaccharomyces* spp. (Table S2) genomes. Reads were aligned using *blastn* with an identity cut-off of 95% (blast/2.6.0+). From the BLAST output, we filtered the top hit based on the proportion of the input read length aligned to the reference, with ties broken successively by highest percent identity and lowest E-value. Due to the high number of hits at rRNA which has high conservation, we performed a second BLAST query using a larger set of over 500 species for which full fungal genomes are available that were downloaded from Ensembl Fungi (full list of fungi accessions in Table S2). For abundance visualization, reads that mapped to multiple positions in the same species genome were counted only once. Reads that mapped to multiple species were counted once for each species. Downstream analysis and visualization were performed in R.

2.4 | Analyses

To validate this pipeline, a positive control approach was used. Artificial FASTQ files were created with 0.01%, 0.1%, 1% and 5% of the reads being *S. pombe* RNA-seq reads and the rest environmental metatranscriptomic reads. Reads were merged from an *S. pombe* transcription study (SRR7291883) and a randomly chosen metatranscriptome sample from the accession list used in this research (SRR3745342; 6,356,979 reads). The different percentage level FASTQs were run through the outlined pipeline, and the results were plotted with the same parameters (identity > 98% and cover > 0.9; cover calculated as length of the read found in reference).

The numbers of BLAST hits per *Schizosaccharomyces* species per sampled environment were summarized in R for all of the batches. Only hits with identity match higher than 98% and cover higher than 0.9 were counted. Based on the SRA information table, the processed accessions belonged to more than 150 unique organisms/environments (listed under 'ScientificName'). Given that many of these were duplicates or closely related environments (e.g., several species of bees or different marine habitats), environments were grouped by their common name for visualization (Table S1). The hits were further assigned a category of position in the well annotated *S. pombe* genome, using a *gff* file downloaded from PomBase (Lock et al., 2018). With the *IRanges v2.20.2* in R (Lawrence et al., 2013), the BLAST hit start and end coordinates were compared against the CDS start and end coordinates from the GFF files that were imported using *ape v5.5* (Paradis & Schliep, 2019). With hits that fell within CDS ranges, the percentage of CDS hits was calculated. We also determined the top five transcripts that hits fell within. Hit counts per transcript were normalized by transcript length. To confirm reads were most likely transcriptomic and not derived from genomic sources, we further analysed if reads were located at transcribed regions (rRNA, ncRNA and exons) or in intronic regions.

3 | RESULTS

We studied the occurrence of fission yeast genetic signatures in large volumes of publicly available transcriptomic data using a relatively low computation intensive method as described in Figure 1. To validate our pipeline, we processed artificial metatranscriptome files with varying percentages (0.01%, 0.1%, 1% and 5%) of *S. pombe* RNA-seq reads to a marine metatranscriptome file (approximately 600, 6300, 64,000 and 334,000 reads, respectively). This dummy dataset was run through the pipeline using the same cut-off of 98% identity and 0.9 cover (i.e., mapped proportion of read length) to classify hits. On average, over 93% of the added *S. pombe* reads were recovered and mapped to *Schizosaccharomyces* spp. (Figure 2). Of those, over 99.5% were correctly classified as *S. pombe* while <0.5% were classified as other *Schizosaccharomyces* species (Figure 2a). No reads from the marine metagenome were classified as *Schizosaccharomyces* reads after filtering. These results suggest a good efficiency and low false positive rate is to be expected with these thresholds using actual data.

In total, 13,286 accessions were processed, for a total of 14,631 FASTQ files (about 6.5 thousand accessions were paired files but were processed as single-end files). Some accessions were discarded either due to a failure in downloading from the ENA or due to an error in evaluating or converting the Phred score encoding. In total, the first BLAST step resulted in approximately 43.5 million *Schizosaccharomyces* spp. hits that mapped to the masked fission yeast genomes of which after filtering by >98% identity and >90% cover, approximately 3.9 million *Schizosaccharomyces* hits remained (in 1.35 million unique reads). We found no association in the variation of raw read counts per accession with the number of positive hits found in different environments (Kendrall's rank correlation tau 0.03057; $z = 1.3954$; $p = 0.163$). Considering that the total number of reads in all accessions and in all accessions with hits is 20.8×10^{10} and 4.4×10^{10}, respectively, suggests a very low presence of fission yeasts in the environments assessed. Due to the low number of hits, we were able to manually curate the results.

The reads that were obtained from the initial BBMap step should have a higher similarity with the *Schizosaccharomyces* species than our group of 10 'other' fungi (see Section 2). BLAST of these hits against a set of >500 full unmasked fungal genomes confirmed the higher similarity to *Schizosaccharomyces* for the majority (67%) of the reads. Nevertheless, approximately 32% of the reads had a higher similarity to other fungi, and 0.8% of the reads had equal values for coverage, e-value and identity between *Schizosaccharomyces* and the other fungal species (Figure 2b). Ambiguous reads were found in a variety of environments at low frequencies and in few accessions (Figure 2c). These reads mostly mapped to *rRNA* regions in *S. pombe* (see also below). All ambiguous reads as well as those with higher similarity to other fungi were discarded for further analyses.

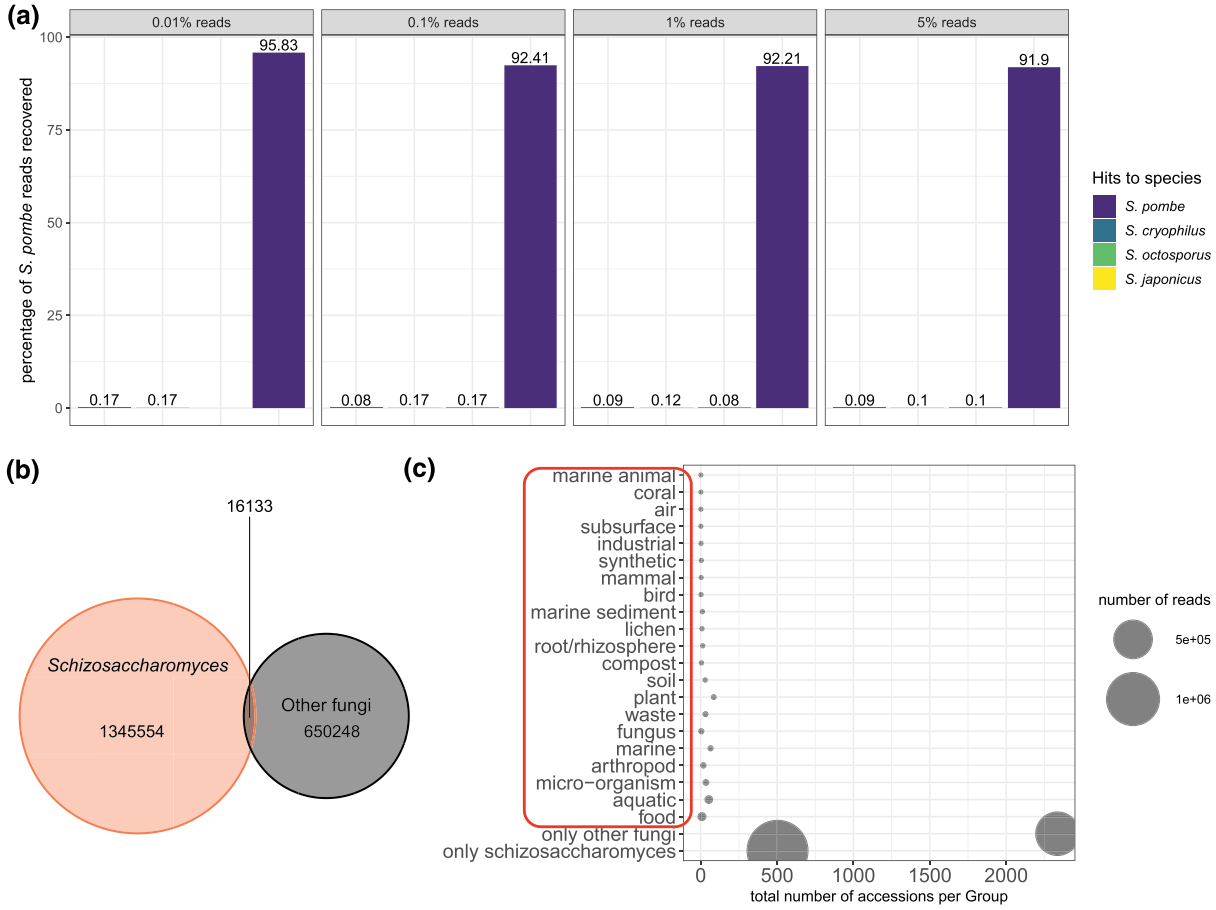

FIGURE 2 (a) Artificial metatranscriptome files with varying percentages of *Schizosaccharomyces pombe* RNA-seq reads were run through the pipeline. A random sample of reads was added to a marine metagenome file (approximately 334,000, 64,000, 6300 and 600 reads of *S. pombe* to 6.4 million metagenome reads). Respectively, 92%, 93%, 93% and 96% of the added *S. pombe* reads were recovered. Of those, over 99.5% were classified as *S. pombe* while <0.5% were classified as other *Schizosaccharomyces* species. (b) Venn diagram of number of reads that map only to *Schizosaccharomyces* ($n = 1{,}345{,}554$), only to other fungi in the BLAST query database ($n = 650{,}248$) or to both categories ($n = 16{,}133$). (c) The number of NCBI accessions (*x*-axis) and the number of reads (given by point size) only mapping to *Schizosaccharomyces* or only mapping to other fungi categories. The groups in the box show the distribution per environmental groups (see Section 2 for grouping) in the overlapping reads from (b) [Colour figure can be viewed at wileyonlinelibrary.com]

Hits of *Schizosaccharomyces* spp. were found in 20 of the 34 different environmental groups (Figure 3). We will discuss the potential for the different environments to harbour fission yeast and what are likely to be true hits or contaminates. We will (i) describe the projects with the highest number of hits, (ii) how the reads are distributed over the genome (Figure 4) and (iii) the environments with the most consistent results.

When normalized by total number of reads in each group, the 'fungi' group had the highest number of positive hits. The majority of the reads (approximately 75%) in the 'fungi' group were from the BioProject PRJNA666900, co-cultures of Fibrobacter sp. UWB7 with *Anaeromyces robustus* or *Caecomyces churrovis*, followed by PRJNA654076 (~12%) and PRJNA654077 (~10%), both co-cultures of *Ophiostoma piceae* and *Pseudomonas putida*. The description on SRA suggests clean cultivated samples, where no fission yeasts would be expected; unfortunately, publications associated with these BioProjects could not be found. In the next group, 'soil', 99% of the reads came from the BioProject PRJNA621679, which surveyed soil microbial communities from a watershed in Colorado, USA. Similarly, over 99% of the reads from the 'trees/plants' group were from the BioProjects PRJNA572120, PRJNA571995 and PRJNA572130, which all surveyed switchgrass phyllosphere microbial communities in Michigan, USA (Howe et al., 2021).

All hits in the aforementioned projects, as well as the three 'co-culture' projects, exclusively mapped to the *rRNA* region of the mitochondrial DNA (Figure S1) which suggests that these might be derived from an alternative, possibly bacterial, source. *S. pombe*'s genome has many signatures of horizontal gene transfer from bacteria, of which many occurred before the radiation of the *Schizosaccharomyces* clade (Rhind et al., 2011). Investigations of sequences of these regions of the full NCBI blast database (accessed July 2021) did not yield further insights. Even though the vast majority of hits (96%) across accessions fell in *rRNA* genes, of which 98% specifically within the mitochondrial *rRNA* region, there were additionally hits at other locations, which did

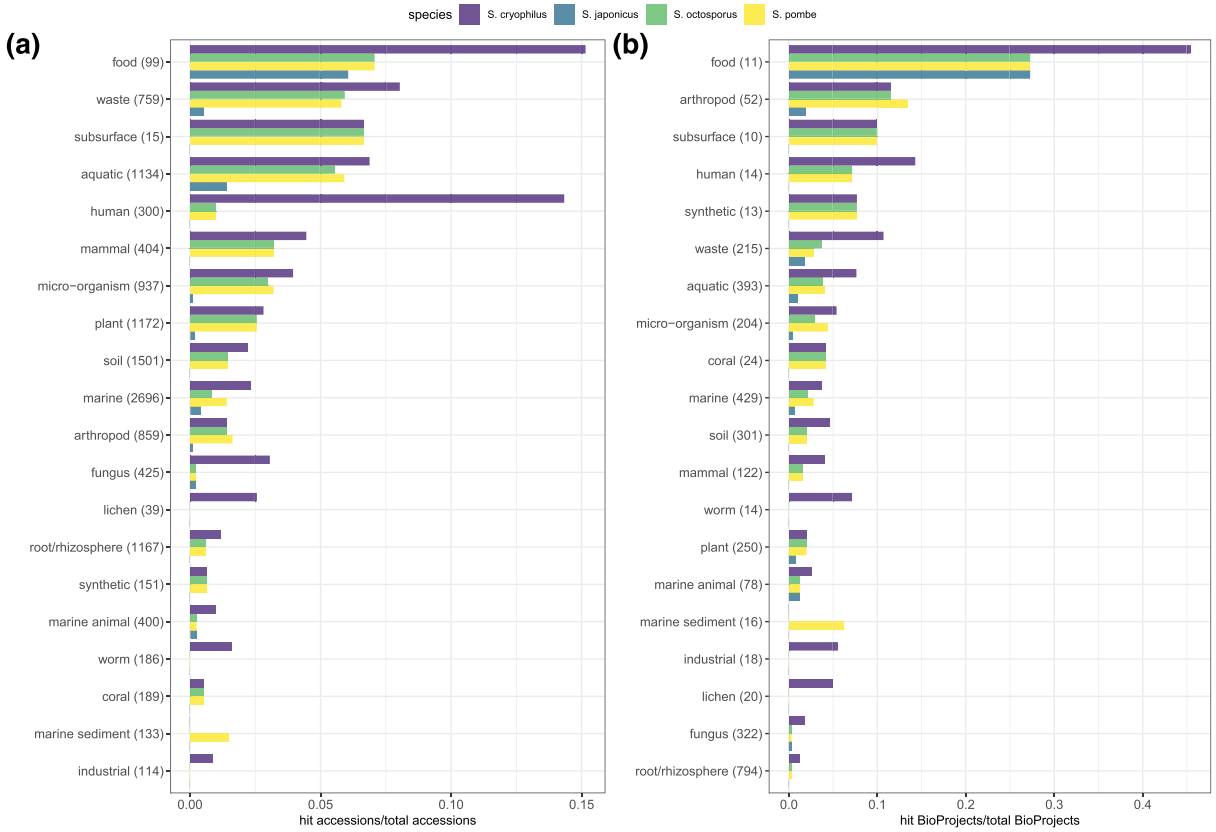

FIGURE 3 The proportion of accessions (a) or BioProjects (b) with positive *Schizosaccharomyces* spp. hits per environmental group. The total number of NCBI accessions (in a) or BioProjects these accessions are derived from (in b) is shown in parentheses next to each group. Note that different axes for the figures [Colour figure can be viewed at wileyonlinelibrary.com]

not greatly affect the BioProjects with mapping reads (Table 1). We assume that accessions with BLAST hits exclusively in the mitochondria are likely not derived from fission yeast. The remaining 4% of hits are localized to loci all over the genome. The assignment of >99.7% of the reads to transcribed regions with only 104 reads in introns from a total of 45,256 reads (conservative estimate that excludes the *rRNA* hits; Table 2) further suggests that the hits observed deviate from expected proportion of intronic reads ($p < 0.001$; χ^2 goodness-of-fit test assuming 3% intronic regions; Wood et al., 2002) and are thus unlikely to be due to genomic contamination during sample handling. Additionally, the total number of reads per gene for all genes with hits is correlated with experimental abundance of expression data for *S. pombe* (both in vegetative growth and under G1 arrest) as would be expected for these genes ($p < 0.001$; linear regression assuming Poisson distribution with log–log transformation of the gene-length-corrected counts and 'MM.mRNA.cpc' or 'MN.mRNA.cpc' from table S4 from Marguerat et al., 2012; Figure S2). This correlation is maintained when removing hits to *rRNA*, which are extreme outliers.

A few accessions show consistent coverage over the entirety of the *S. pombe* genome (Figures 4 and S1), specifically one accession from BioProjects PRJNA510232 (yak *Bos taurus* testes), PRJNA678138 (marine sponge sample) and PRJEB29369 (bacterial marine sample). Based on the descriptions, the first of these samples are likely a contaminant either during library preparation or during sequencing. We discuss the marine samples further in the discussion.

When considering the proportion of accessions and BioProjects with fission yeast hits, few signals can be observed. The strongest signals of presence based on BioProject number and environment were in *food*, *human*, *arthropods* and a variety of harsh environments (Figure 3b,c). *Food* shows the strongest signal in the proportions of accessions and BioProjects. *Human* accessions were all almost exclusively from a single BioProject (vaginal metatranscriptomes). The *arthropods* (a category containing terrestrial animals such as mites, insects and spiders) had a rather high number of hits per BioProject, but not by proportion of accessions.

4 | DISCUSSION

Schizosaccharomyces yeasts are scientifically and economically important microorganisms with little understood patterns of biogeographical origins and distribution (Hoffman et al., 2015; Jeffares, 2018; Tao et al., 2019; Tusso et al., 2019; Yanagida, 2002) and the location specimen have been isolated from are human-associated environments, such as beer and wine (Á. Benito et al., 2018; Jeffares, 2018). However, these environments may not correspond to their natural growth

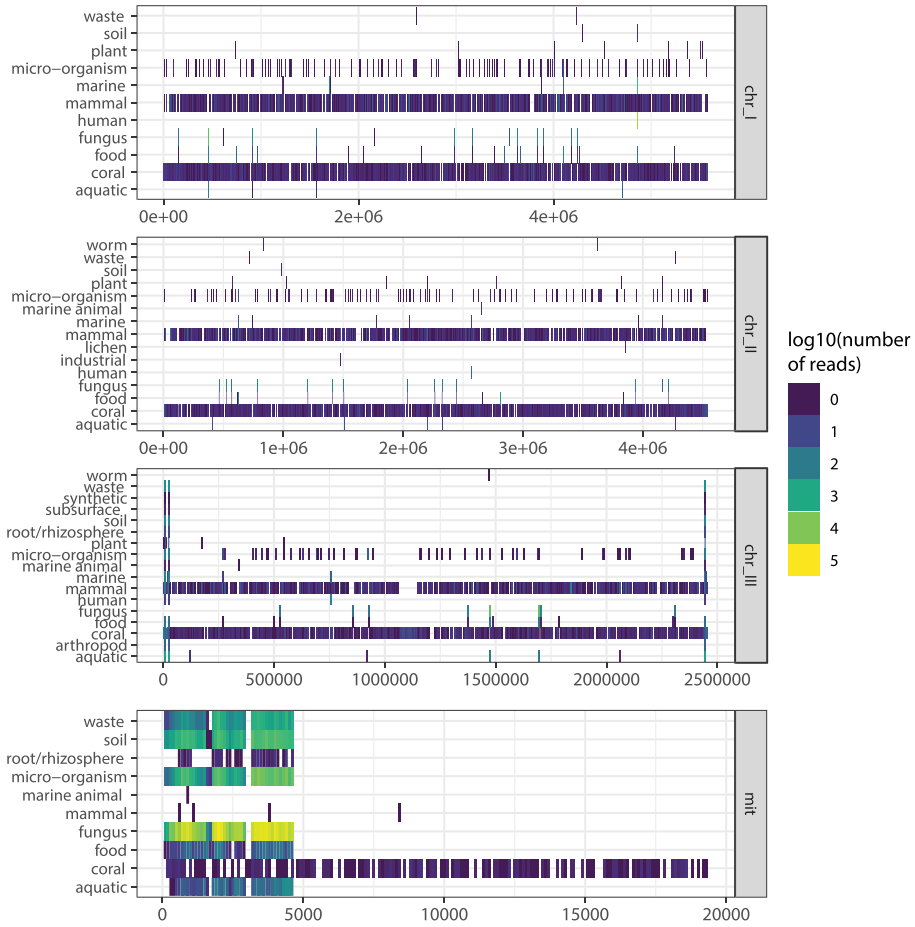

FIGURE 4 The distribution of reads along the genome of *Schizosaccharomyces pombe* (values in basepairs) given for each environmental group with hits for the specific chromosome (chr_I, chr_II or chr_III; bin size of 4 kb) or the mitochondrion (mit; bin size of 1 kb). Each coloured region represents the log-transformed number of reads in the specific bin [Colour figure can be viewed at wileyonlinelibrary.com]

habitats. Here, we rigorously combed through a large number of published metatranscriptomic datasets using a simple bioinformatic pipeline, in search of signatures of the *Schizosaccharomyces* yeasts to identify their natural biotopes. We found a remarkably low number of reads that map to fission yeasts in our datasets. Across the approximately 14,000 accessions analysed, we found signatures of fission yeasts in a variety of environments including soils, food and marine environments and in association with a variety of organisms, including arachnids and insects.

Finding signals of fission yeast in 50% of the *food* BioProjects and 15% of the accessions, an environment where fission yeast is known to live (Jeffares, 2018), suggests that the pipeline functions could be able to pick up signals of fission yeast in environmental metatranscriptomic datasets, also confirming the sensitivity we observed with our dummy datasets. Because *S. pombe* is a model species, we were cautious that hits could be contaminations at laboratories or sequencing facilities. Even though some samples show signs of contamination (e.g., the yak testes samples), the homogeneous mapping pattern helps distinguish these from low abundance true hits. Additionally, the correlation of read abundance with published expression data and the low number of intronic reads suggest minimal contamination from genomic samples. Of the environments tested, those that contain putative fission yeast reads show a very low abundance of hits, and most hits (96%) are in the mitochondrial or nuclear *rRNA* genes which, due to conservation, are not unambiguous (Liu et al., 2012). Our results are therefore not clear-cut and should be considered more as indications for future research on the natural habitats of fission yeast.

Similar to the *food* category, the *arthropod* group of BioProjects also fits in with previous findings. Many reads were found associated with *Apis mellifera*, the most common honeybee, specifically from RNA sequencing of the worker and royal jelly (Maori et al., 2019). This is in line with previous research that identified honey as an environment where fission yeast can be isolated from (S. Benito et al., 2013; Jeffares, 2018). Additionally reads were associated with fruit fly, ant, mosquito and tick. The recent isolation of the fission yeast *S. japonicus* from *Drosophila* further strengthens the association of fission yeasts to insects (Seike et al., 2021), either as vectors between food resources or potentially as host.

Surprisingly, the *Schizosaccharomyces* spp. were also consistently found in a variety of harsh environments. A large proportion being aquatic environments or in association with aquatic animals (sponge, oyster and whales), both in marine, riverine, lake and freshwater (50 BioProjects in total) but also in soil and at deeper 'subsurface' locations. Because these types of environments are highly overrepresented in our dataset, the proportion of accessions and BioProjects is low, but they may be real signals. Signatures of *Schizosaccharomycetes* have been found in metagenomic surveys of marine environments in

TABLE 1 Total number of BioProjects per group with reads mapping to the fission yeast genome for all reads (middle column) and when all mitochondrial and genomic *rRNA* reads are filtered out

	# of BioProjects	
Group name	**All reads**	**Without rRNA**
aquatic	19	30
arthropod	6	6
coral	1	1
food	5	5
fungus	2	6
human	2	2
industrial	1	1
lichen	1	1
mammal	5	5
marine	18	18
marine animal	2	2
micro-organism	8	11
plant	5	5
root/rhizosphere	4	11
soil	6	14
subsurface	1	1
synthetic	1	1
waste	8	23
worm	1	1

TABLE 2 Number of reads mapped to the *Schizosaccharomyces pombe* genome for the top six features

Feature type	Counts
rRNA	1,725,827
mRNA	42,412
ncRNA	2440
3′ UTR	1899
5′ UTR	1119
Intron	104
Other	3213

the past (e.g., *Porites astreoides* coral from the Caribbean Sea in Wegley et al., 2007, and *Avicennia marina* mangroves from the Red Sea in Simões et al., 2015). These observations are remarkable, as fission yeast is not known to be especially salt tolerant (e.g., Yang et al., 2018).

Even though *S. pombe*, *S. japonicus* and *S. octosporus* were present in a variety of trees, forest, leaf, soil and rhizosphere metatranscriptomes, these were a low proportion of the total surveyed accessions. No signs were found in the overrepresented leaf and leaf litter categories, which suggests that the *Schizosaccharomyces* yeasts are probably not forest-associated fungi, contrasting to *S. paradoxus* which is thought to have oak as natural habitat (Kowallik & Greig, 2016). *S. cryophilus* was only notably present in the food, waste and aquatic metatranscriptomes. The species was first discovered as a contaminant of an *S. octosporus* sample and has rarely been used in research (Helston et al., 2010). Interestingly, even though *S. cryophilus* and *S. octosporus* diverged most recently, are most closely related and share 85% orthologue identity on average (Rhind et al., 2011), reads rarely map to both simultaneously. In general, *S. cryophilus* has many fewer hits compared to the other species and is likely less abundant in the environment to start with. It will be interesting to see how the recently described fifth species in the genus, *Schizosaccharomyces osmophilus* (Brysch-Herzberg et al., 2019) for which currently no genome is available, will fit into these findings.

To search for fission yeast and candidate habitats, we analysed publicly available metatranscriptomic data from a variety of studies. Using a large compilation of publicly available datasets presents some logistical challenges. The data presented here were selected based on the available metadata in the NCBI SRA and downloaded from the ENA, but some discrepancy in the metadata and FASTQ file links caused file downloads to fail and required manual formatting of URLs. Additionally, missing or incorrect metadata required further individual research for many accessions. For example, 'metagenomic' and 'metatranscriptomic' were often conflated; the scientific name of the organism was unavailable or ambiguous (such as 'metagenome'). The omission of an associated publication in the metadata furthermore made information on details about sampling, extraction and sequencing methods, or possible contamination identified by the primary researchers inaccessible. Through thorough individual searches for BioProject numbers using online search engines, we were able to obtain a large number of missing manuscript information, but this was not possible for all.

Targeted methods allow researchers to refine and tailor the data collection to the research question of interest and circumvent many of these issues. However, publicly available published biological data are a growing resource worth exploring (Pasquetto et al., 2017; Stephens et al., 2015). Nonetheless, in order to maximize the potential of published data in answering new questions, proper annotation and comprehensive metadata should be the standard. We urge researchers submitting data to public databases to properly annotate their data, mention BioProject numbers in their publications and associate their publications in the online resource. See Bietz and Lee (2009) for a broader discussion on collaboration and databases in metagenomic research.

We presented a new approach to the search for the natural habitat of *Schizosaccharomyces* yeasts, based on publicly available metatranscriptomic data. While using metatranscriptomic source data makes it more likely to find actively growing yeast, metagenomic approaches are much more widely used. Substituting metatranscriptomes for metagenomes would allow us to survey many more environments (on NCBI SRA source 'metagenomic' returns 2,909,061 results compared to 46,818 'metatranscriptomic' results [01/08/2021]). Even though our method was able to pick up fission yeast reads with high efficiency, our analysis of over 200 trillion reads did not yield many positive hits, suggesting that fission yeasts occur at

very low densities. Both sequencing methods were specifically designed for low abundance species in metagenomic or transcriptomic data (Castro et al., 2018; Pust & Tümmler, 2021) as well as PCR-based and intense sampling though culturing of isolates (S. Benito et al., 2013; Hellberg, 2013) in the suggested environments might yield a final say on where to find fission yeast. Our results give indications for future research in the search of the fission yeast habitat, focussing on environments associated with insects, and potentially in aquatic environments.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for their comments on this manuscript. We thank Saurabh Pophaly for help with computational questions and for technical support, the members of the evolutionary biology department of LMU for helpful comments and Dirk Metzler for suggestions on the data analyses.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Rasha Shraim https://orcid.org/0000-0002-3351-1179

Bart P. S. Nieuwenhuis https://orcid.org/0000-0001-8159-4784

REFERENCES

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Becking, L. G. M. B. (1934). *Geobiologie of inleiding tot de milieukunde*. W.P. Van Stockum & Zoon.

Benito, Á., Calderón, F., & Benito, S. (2018). *Schizosaccharomyces pombe* isolation protocol. In T. L. Singleton (Ed.), *Schizosaccharomyces pombe* (Vol. 1721, pp. 227–234). Springer. https://doi.org/10.1007/978-1-4939-7546-4_20

Benito, S., Gálvez, L., Palomero, F., Calderón, F., Morata, A., Palmero, D., & Suárez-Lepe, J. A. (2013). *Schizosaccharomyces* selective differential media. *African Journal of Microbiology Research*, *7*(24), 3026–3036. https://doi.org/10.5897/AJMR2013.5684

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Ba, A. N. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, *31*(4), 872–888. https://doi.org/10.1093/molbev/msu037

Bietz, M. J., & Lee, C. P. (2009). Collaboration in metagenomics: Sequence databases and the organization of scientific work. In I. Wagner, H. Tellioğlu, E. Balka, C. Simone, & L. Ciolfi (Eds.), *ECSCW 2009* (pp. 243–262). Springer. https://doi.org/10.1007/978-1-84882-854-4_15

Brysch-Herzberg, M., Tobias, A., Seidel, M., Wittmann, R., Wohlmann, E., Fischer, R., Dlauchy, D., & Peter, G. (2019). *Schizosaccharomyces osmophilus* sp. nov., an osmophilic fission yeast occurring in bee bread of different solitary bee species. *FEMS Yeast Research*, *19*(4), 1–12. https://doi.org/10.1093/femsyr/foz038

Bushnell, B. (2014). *BBMap: A fast, accurate, splice-aware aligner* (37.28) [Computer software]. https://www.osti.gov/servlets/purl/1241166

Castro, J. C., Rodriguez-R, L. M., Harvey, W. T., Weigand, M. R., Hatt, J. K., Carter, M. Q., & Konstantinidis, K. T. (2018). imGLAD: Accurate detection and quantification of target organisms in metagenomes. *PeerJ*, *6*, e5882. https://doi.org/10.7717/peerj.5882

Decottignies, A., Sanchez-Perez, I., & Nurse, P. (2003). *Schizosaccharomyces pombe* essential genes: A pilot study. *Genome Research*, *13*(3), 399–406. https://doi.org/10.1101/gr.636103

Fantes, P. A., & Hoffman, C. S. (2016). A brief history of *Schizosaccharomyces pombe* research: A perspective over the past 70 years. *Genetics*, *203*(2), 621–629. https://doi.org/10.1534/genetics.116.189407

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T. G., Nordborg, M., & Lynch, M. (2015). The spontaneous mutation rate in the fission yeast *Schizosaccharomyces pombe*. *Genetics*, *201*(2), 737–744. https://doi.org/10.1534/genetics.115.177329

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, *15*(11), 962–968. https://doi.org/10.1038/s41592-018-0176-y

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*, *459*(7244), 193–199. https://doi.org/10.1038/nature08058

Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell*, *166*(6), 1397–1410.e16. https://doi.org/10.1016/j.cell.2016.08.020

Gerth, M. (2018). Sra_download.pl. https://github.com/gerthmicha/perlscripts/blob/master/sra_download.pl

Gomes, F. C. O, Pataro, C., Guerra, J. B, Neves, M. J., Corrêa, S. R., Moreira, E. S. A., & Rosa, C. A. (2002). Physiological diversity and trehalose accumulation inSchizosaccharomyces pombestrains isolated from spontaneous fermentations during the production of the artisanal Braziliancachaça. *Canadian Journal of Microbiology*, *48*(5), 399–406. https://doi.org/10.1139/w02-032

Hayles, J., & Nurse, P. (2018). Introduction to fission yeast as a model system. *Cold Spring Harbor Protocols*, *2018*(5), 323–333. https://doi.org/10.1101/pdb.top079749

Hellberg, J. (2013). Finding wild fission yeast: Where are they? *Genetics Society News*, *68*, 59–60.

Helston, R. M., Box, J. A., Tang, W., & Baumann, P. (2010). *Schizosaccharomyces cryophilus* sp. nov., a new species of fission yeast. *FEMS Yeast Research*, *10*(6), 779–786. https://doi.org/10.1111/j.1567-1364.2010.00657.x

Hoffman, C. S., Wood, V., & Fantes, P. A. (2015). An ancient yeast for young geneticists: A primer on the *Schizosaccharomyces pombe* model system. *Genetics*, *201*(2), 403–423. https://doi.org/10.1534/genetics.115.181503

Hong, H. A., To, E., Fakhry, S., Baccigalupi, L., Ricca, E., & Cutting, S. M. (2009). Defining the natural habitat of *Bacillus* spore-formers. *Research in Microbiology*, *160*(6), 375–379. https://doi.org/10.1016/j.resmic.2009.06.006

Howe, A., Stopnisek, N., Dooley, S. K., Yang, F. M., Grady, K. L., & Shade, A. (2021). Genome-centric analyses of seasonal phyllosphere microbiome activities in perennial crops. *BioRxiv*, 2021.04.20.440608. https://doi.org/10.1101/2021.04.20.440608

Hu, W., Suo, F., & Du, L.-L. (2015). Bulk segregant analysis reveals the genetic basis of a natural trait variation in fission yeast. *Genome Biology and Evolution*, *7*(12), 3496–3510. https://doi.org/10.1093/gbe/evv238

Hyppa, R. W., Fowler, K. R., Cipak, L., Gregan, J., & Smith, G. R. (2014). DNA intermediates of meiotic recombination in synchronous *S. pombe* at optimal temperature. *Nucleic Acids Research*, *42*(1), 359–369. https://doi.org/10.1093/nar/gkt861

Jeffares, D. C. (2018). The natural diversity and ecology of fission yeast. *Yeast*, *35*, 253–260. https://doi.org/10.1002/yea.3293

Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., Marsellach, F. X., Iqbal, Z., Lau, W., Cheng, T. M. K., Pracana, R., Mülleder, M., Lawson, J. L. D., Chessel, A., Bala, S., Hellenthal, G., O'Fallon, B., Keane, T., Simpson, J. T., … Bähler, J. (2015). The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nature Genetics*, *47*(3), 235–241. https://doi.org/10.1038/ng.3215

Kowallik, V., & Greig, D. (2016). A systematic forest survey showing an association of *Saccharomyces paradoxus* with oak leaf litter. *Environmental Microbiology Reports*, *8*(5), 833–841. https://doi.org/10.1111/1758-2229.12446

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R., Hoad, G., Jang, M., Pakseresht, N., Plaister, S., Radhakrishnan, R., Reddy, K., Sobhany, S., ten Hoopen, P., Vaughan, R., … Cochrane, G. (2011). The European Nucleotide Archive. *Nucleic Acids Research*, *39*(Database issue), D28–D31. https://doi.org/10.1093/nar/gkq967

Leupold, U. (1949). *Die vererbung von homothallie und heterothallie bei Schizosaccharomyces pombe*. Universität Zürich.

Lindner, P. (1893). *Schizosaccharomyces pombe* n. sp., ein neuer Gärungserreger. *Wochenschrift für Brauerei 10*, 1298–1300.

Liu, K.-L., Porras-Alfaro, A., Kuske, C. R., Eichorst, S. A., & Xie, G. (2012). Accurate, rapid taxonomic classification of fungal large-subunit *rRNA* genes. *Applied and Environmental Microbiology*, *78*(5), 1523–1533. https://doi.org/10.1128/AEM.06826-11

Lock, A., Rutherford, K., Harris, M. A., & Wood, V. (2018). PomBase: The scientific resource for fission yeast. In *Eukaryotic genomic databases* (pp. 49–68). Humana Press. https://doi.org/10.1007/978-1-4939-7737-6_4

López Hernández, J. F., Helston, R. M., Lange, J. J., Billmyre, R. B., Schaffner, S. H., Eickbush, M. T., McCroskey, S., & Zanders, S. E. (2021). Diverse mating phenotypes impact the spread of wtf meiotic drivers in Schizosaccharomyces pombe. *eLife*, *10*, 70812. https://doi.org/10.7554/elife.70812

Maori, E., Garbian, Y., Kunik, V., Mozes-Koch, R., Malka, O., Kalev, H., Sabath, N., Sela, I., & Shafir, S. (2019). A transmissible RNA pathway in honey bees. *Cell Reports*, *27*(7), 1949–1959.e6. https://doi.org/10.1016/j.celrep.2019.04.073

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., & Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, *151*(3), 671–683. https://doi.org/10.1016/j.cell.2012.09.019

Melbinger, A., & Vergassola, M. (2015). The impact of environmental fluctuations on evolutionary fitness functions. *Scientific Reports*, *5*(1), 15211. https://doi.org/10.1038/srep15211

Monard, C., Gantner, S., Bertilsson, S., Hallin, S., & Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Scientific Reports*, *6*(1), 37719. https://doi.org/10.1038/srep37719

Nieuwenhuis, B. P. S., Tusso, S., Bjerling, P., Stångberg, J., Wolf, J. B. W., & Immler, S. (2018). Repeated evolution of self-compatibility for reproductive assurance. *Nature Communications*, *9*(1), 1639. https://doi.org/10.1038/s41467-018-04054-6

Osterwalder, A. (1924). Schizosaccharomyces liquefaciens n. sp., eine gegen freie schweflige Säure widerstandsfähige Gärhefe. *Mitteilungen Aus Dem Gebiete Der Lebensmittel-Untersuchung Und Hygiene. u. Hygiene*, *15*, 5–28.

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parts, L. (2014). Genome-wide mapping of cellular traits using yeast. *Yeast*, *31*(6), 197–205. https://doi.org/10.1002/yea.3010

Pasquetto, I., Randles, B., & Borgman, C. (2017). On the reuse of scientific data. *Data Science Journal*, *16*, 8. https://doi.org/10.5334/dsj-2017-008

Pluskal, T., Nakamura, T., Villar-Briones, A., & Yanagida, M. (2009). Metabolic profiling of the fission yeast *S. pombe*: Quantification of compounds under different temperatures and genetic perturbation. *Molecular BioSystems*, *6*(1), 182–198. https://doi.org/10.1039/B908784B

Pust, M.-M., & Tümmler, B. (2021). Identification of core and rare species in metagenome samples based on shotgun metagenomic sequencing, Fourier transforms and spectral comparisons. *ISME Communications*, *1*(1), 1–4. https://doi.org/10.1038/s43705-021-00010-6

Rhind, N., Chen, Z., Yassour, M., Thompson, D. A., Haas, B. J., Habib, N., Wapinski, I., Roy, S., Lin, M. F., Heiman, D. I., Young, S. K., Furuya, K., Guo, Y., Pidoux, A., Chen, H. M., Robbertse, B., Goldberg, J. M., Aoki, K., Bayne, E. H., … Nusbaum, C. (2011). Comparative functional genomics of the fission yeasts. *Science*, *332*(6032), 930–936. https://doi.org/10.1126/science.1203357

Salinas, F., Cubillos, F. A., Soto, D., Garcia, V., Bergström, A., Warringer, J., Ganga, M. A., Louis, E. J., Liti, G., & Martinez, C. (2012). The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae*. *PLoS ONE*, *7*(11), e49640. https://doi.org/10.1371/journal.pone.0049640

Schneider, A. N., Sundh, J., Sundström, G., Richau, K., Delhomme, N., Grabherr, M., Hurry, V., & Street, N. R. (2021). Comparative fungal community analyses using metatranscriptomics and internal transcribed spacer amplicon sequencing from Norway spruce. *MSystems*, *6*(1), e00884-20. https://doi.org/10.1128/mSystems.00884-20

Seike, T., Sakata, N., Matsuda, F., & Furusawa, C. (2021). Elevated sporulation efficiency in fission yeast *Schizosaccharomyces japonicus* strains isolated from *Drosophila*. *Journal of Fungi*, *7*(5), 350. https://doi.org/10.3390/jof7050350

Shakya, M., Lo, C.-C., & Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics*, *10*, 1–10. https://doi.org/10.3389/fgene.2019.00904

Simões, M. F., Antunes, A., Ottoni, C. A., Amini, M. S., Alam, I., Alzubaidy, H., Mokhtar, N.-A., Archer, J. A. C., & Bajic, V. B. (2015). Soil and rhizosphere associated fungi in gray mangroves (*Avicennia marina*) from the Red Sea—A metagenomic approach. *Genomics, Proteomics & Bioinformatics*, *13*(5), 310–320. https://doi.org/10.1016/j.gpb.2015.07.002

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big Data: Astronomical or genomical? *PLoS Biology*, *13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Tao, Y.-T., Suo, F., Tusso, S., Wang, Y.-K., Huang, S., Wolf, J. B. W., & Du, L.-L. (2019). Intraspecific diversity of fission yeast mitochondrial genomes. *Genome Biology and Evolution*, *11*(8), 2312–2329. https://doi.org/10.1093/gbe/evz165

Tusso, S., Nieuwenhuis, B. P. S., Sedlazeck, F. J., Davey, J. W., Jeffares, D. C., & Wolf, J. B. W. (2019). Ancestral admixture is the main determinant of global biodiversity in fission yeast. *Molecular Biology and Evolution*, *36*, 1975, msz126–1989. https://doi.org/10.1093/molbev/msz126

Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505–522. https://doi.org/10.2307/2408649

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environmental Microbiology*, *9*(11), 2707–2719. https://doi.org/10.1111/j.1462-2920.2007.01383.x

Westermann, A. J., & Vogel, J. (2021). Cross-species RNA-seq for deciphering host–microbe interactions. *Nature Reviews Genetics*, *22*(6), 361–378. https://doi.org/10.1038/s41576-021-00326-y

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., … Nurse, P. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature*, *415*(6874), 871–880. https://doi.org/10.1038/nature724

Yanagida, M. (2002). The model unicellular eukaryote, *Schizosaccharomyces pombe*. *Genome Biology*, *3*(3), 1, comment2003–4. https://doi.org/10.1186/gb-2002-3-3-comment2003

Yang, Y., Liu, Q., Jiang, G., Chen, S., Zhou, L., Sakamoto, N., Kuno, T., Fang, Y., & Yao, F. (2018). Genome-wide screen reveals important roles for ESCRT proteins in drug/ion resistance of fission yeast. *PLoS ONE*, *13*(6), e0198516. https://doi.org/10.1371/journal.pone.0198516

Zanders, S. E., Eickbush, M. T., Yu, J. S., Kang, J.-W., Fowler, K. R., Smith, G. R., & Malik, H. S. (2014). Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *eLife*, *3*, e02630. https://doi.org/10.7554/eLife.02630

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Shraim, R., & Nieuwenhuis, B. P. S. (2022). The search for *Schizosaccharomyces* fission yeasts in environmental metatranscriptomes. *Yeast*, *39*(1), 83–94. https://doi.org/10.1002/yea.3689