



RESEARCH

e-Xtra*

DNA Template Dilution Impacts Amplicon Sequencing-Based Estimates of Soil Fungal Diversity

Sarah C. Castle[†] and **Zewei Song**, Department of Plant Pathology, University of Minnesota, St. Paul; **Daryl M. Gohl**, University of Minnesota Genomics Center, Minneapolis; **Jessica L. M. Gutknecht** and **Carl J. Rosen**, Department of Soil, Water, and Climate, University of Minnesota, St. Paul; **Michael J. Sadowsky**, Department of Soil, Water, and Climate, University of Minnesota, and BioTechnology Institute, University of Minnesota, St. Paul; **Deborah A. Samac**, U.S. Department of Agriculture-Agriculture Research Services, Plant Science Research Unit, St. Paul, MN; and **Linda L. Kinkel**, Department of Plant Pathology, University of Minnesota, St. Paul

Accepted for publication 15 May 2018.

ABSTRACT

Next generation sequencing of taxonomically relevant marker genes has enabled researchers to sample the richness, diversity, and composition of environmental microbiomes at previously unattainable depths. However, molecular methods may have unintended downstream consequences and the inadvertent undersampling of microbial communities may be a significant pitfall in microbiome profiling. One such procedure, dilution of the DNA template prior to polymerase chain reaction (PCR), may improve marker gene amplification by reducing chimeric read formation and decreasing PCR inhibitor concentrations. However, dilution unavoidably reduces target DNA template number per sample. We evaluated the effects of pre-PCR DNA template dilution on estimates of soil fungal microbiome diversity, composition, and species abundance distributions across a collection of 144 agricultural soil samples. Fungal DNA templates were serially diluted at 0-, 10-, 100-, and 1,000-fold and sequence data of diluted templates were compared with those of an identical set of undiluted templates. For three prairie soil samples, in addition to evaluating variation among replicates of individual samples, we serially diluted fungal DNA extracts from soil samples in triplicate and sequenced undiluted and diluted samples. DNA template

dilution significantly reduced estimates of fungal richness and diversity, as compared with undiluted samples. Dilution of DNA template also resulted in reduced relative abundances of rare operational taxonomic units (OTUs) and increased relative abundances of common OTUs. Collectively, changes in OTU abundance distributions following dilution produced substantial shifts in overall fungal community composition. Our results highlight risks associated with sample dilution and point to the potential utility of quantifying pre-PCR template concentration in the estimation of microbiomes. We urge researchers to thoroughly document methods and to reconsider routine dilution of pre-PCR DNA templates particularly for low abundance microbiome samples. As efforts to profile environmental microbiomes using molecular sequencing approaches accelerate, developing an adequate understanding of potential methodological bottlenecks will increase our ability to accurately characterize and compare datasets.

Additional keywords: agriculture, ecology, fungal internal transcribed spacer 2, marker gene survey, microorganism, soil microbiome.

Funding: This study was supported by a grant made to L. Kinkel, J. Gutknecht, C. Rosen, M. Sadowsky, and D. Samac by the University of Minnesota LTARN, a grant made to L. Kinkel and Z. Song from the University of Minnesota MnDRIVE, and by a U.S. Department of Agriculture-National Institute of Food and Agriculture grant 2017-67012-26109 postdoctoral fellowship awarded to S. Castle.

[†]Corresponding author: S. C. Castle; E-mail: seccastle@umn.edu

*The e-Xtra logo stands for “electronic extra” and indicates that two supplementary figures and one supplementary table are published online.

The authors declare no conflicts of interest associated with the presented research. D. M. Gohl is a shareholder and CSO of CoreBiome, Inc. Mention of any trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer. This paper is a joint contribution from the USDA-ARS-Plant Science Research Unit and the University of Minnesota.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2018.

Accurately measuring the diversity and composition of biological communities is a central goal of ecology. It has long been understood that sampling effort—described here as the degree to

which a “sample” of individuals reflects a complete “census” of all individuals—can influence both species abundance distributions and biodiversity estimates for macro- (Coddington et al. 2009; Morisita 1959; Nash 1950; Preston 1962a, b; Wolda 1981) and microorganisms (Kinkel et al. 1995; Zhou et al. 2008). Generally, more intensive sampling efforts, though costly, increase the representation of low abundance species, while shallow sampling efforts emphasize dominant species (May 1988; Nee et al. 1991; Preston 1948). Very diverse communities, such as microbial assemblages, represent a unique challenge in that it is not yet feasible for most researchers to directly measure or completely census entire assemblages in situ (Green et al. 2004). Thus, optimizing the balance between sampling effort and the accuracy of community diversity estimates represents an unresolved challenge in microbiome research, especially with respect to exceedingly complex habitats such as soil.

Microbial ecologists employ marker gene surveys (e.g., 16S rRNA, 18S rRNA, or internal transcribed spacer [ITS]) via high-throughput sequencing of polymerase chain reaction (PCR)-amplified regions of marker gene DNA (Lindahl et al. 2013; Mullis and Faloona 1987) to estimate microbial community composition, richness, and diversity. Though exact methods can vary widely among studies, in short, genomic DNA is extracted from environmental samples, DNA is purified, PCR is used to generate millions of copies of marker genes using region-targeting primers, and target amplicons are sequenced (Lindahl et al. 2013). Lastly, quality filtered sequence data are typically rarefied, subsampled, or scaled prior to analysis (Hughes and Hellmann 2005; Weiss et al. 2017) as an ad hoc way of equalizing sequence library size among samples for comparative statistical analyses. Though the number of quality sequences generated per sample or subsampling/rarefaction depth are often used interchangeably with microbiome “sampling effort,” preceding experimental methodological steps have significant consequences for determining the number, identity, and composition of DNA amplicons in a sample even prior to sequencing.

The number of target DNA marker genes in an environmental sample is determined by several factors including number of target cells (e.g., fungal, prokaryotic, plant, eukaryotic) in the matrix of interest, amount of material extracted/collected, and gene copy number per organism (Kembel et al. 2012). Insufficient sampling (i.e., collecting too little material), large amounts of host or non-microbial DNA, low DNA quality, or poor DNA extraction efficiency are known to alter subsequent estimates of microbial diversity and composition (Gohl et al. 2016a; Gonzalez et al. 2012; Lindahl et al. 2013). Like other methodological variables affecting marker gene makeup of the extracted DNA, adjustments to DNA template concentration including variation in final extraction volume or dilution of DNA template are common. Although primarily used to reduce chimeric read formation or decrease concentrations of PCR inhibitors (Schrader et al. 2012), pre-PCR DNA template is routinely diluted even for nonproblematic samples to ensure sequencing success.

Effects of dilution on estimates of microbial diversity are well established for mixed microbial communities. For example, studies have shown empirically and computationally that serial dilution of microbial cells reduces bacterial richness and diversity, and alters community composition at dilutions as low as 10^{-5} (Franklin et al. 2001; Yan et al. 2015). Similar to the effects of dilution on whole soil bacterial communities, pre-PCR dilution of mixed-template DNA has been shown to decrease estimates of soil bacterial richness, diversity, and the reproducibility of community composition among replicate samples for 20- or 100-fold dilutions (Kennedy et al. 2014; Wu et al. 2010). Because dilution reduces the number of DNA molecules per sample (Fig. 1), it is expected that the impact of pre-PCR dilution is of greatest consequence for low and medium abundance molecules in mixed-template DNA,

which subsequently have a lower probability of being transferred, and amplified in PCR reactions (Fig. 1). However, the degree to which dilution “bottlenecks” estimates of microbial communities is also likely to depend on the overall sample diversity and initial target molecule number. Thus, we define bottlenecking as a reduction in the number of input DNA template molecules related to a given methodological step in microbiome sample processing.

Here, we explored whether pre-PCR DNA template dilution imposes a potential bottleneck on estimates of fungal community diversity, composition, and species distributions in 144 agricultural and three prairie soil samples. Soil DNA extracts were either PCR amplified and sequenced as “undiluted” DNA templates (0-fold dilution) or were serially “diluted” (10-, 100-, or 1,000-fold) prior to PCR amplification. A qPCR-based quantification of the relative number of fungal DNA molecules was used to select diluted agricultural DNA samples for sequencing, while all dilutions were sequenced for prairie soil DNA samples.

MATERIALS AND METHODS

Soil samples originating from three agricultural fields and an experimental prairie site were collected (0 to 10 cm depth), individual samples were homogenized by hand, materials greater than 2×2 mm were removed, and soil was stored at -80°C prior to DNA

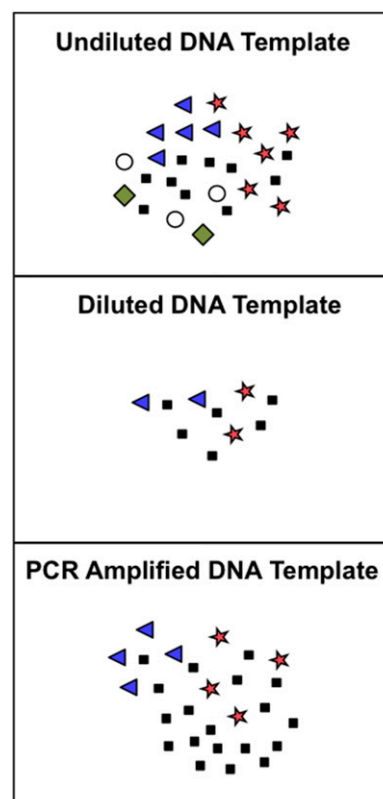


Fig. 1. Pre-PCR dilution reduces the DNA template number per sample with consequences for the types of DNA molecules input to PCR. Low and medium abundance molecules in mixed-template DNA may be most affected due to the lower probability of being transferred to PCR reactions. Composition of PCR-amplified DNA template may differ from the original sample template, with the potential for a reduction in sequenced community richness and diversity. Shapes and colors represent distinct DNA molecules.

extraction. The agricultural field sites, typical of Minnesota farmland, were located at the University of Minnesota Research and Outreach Centers at Waseca, Lamberton, and Grand Rapids, MN, USA. Among all agricultural sites, 144 individual soil samples were collected in 2014 from experimental field plots planted with oat (*Avena sativa*). In 2013, three prairie soil samples were collected from the Cedar Creek Long-Term Ecosystem Science Reserve in central Minnesota, USA within the rhizosphere of the native prairie grass, big blue stem (*Andropogon gerardii*).

Total DNA was extracted from 0.25-g subsamples of each soil sample using the MoBio PowerSoil kit (Mo Bio Laboratories, Inc., Carlsbad, CA), following the manufacturer's protocol for bulk DNA. Among undiluted samples, mean total DNA concentrations were ~ 15.3 ng of DNA μL^{-1} (range: 6.2 to 33.0 ng of DNA μL^{-1}) and were quantified using an absorbance-based measurement on a Biotek Synergy microplate reader (Biotek Inc., Winooski, VT). We amplified and sequenced the rDNA ITS2 amplicon from three sample sets. First for the agricultural soil samples, DNA templates were serially diluted to 0-, 10-, 100-, and 1,000-fold and the ITS2 region was amplified for each dilution. Rather than sequence each dilution, we used a qPCR-based relative quantification method to select dilutions for sequencing. In short, a target final Rn value for the amplification curves of samples was chosen and the dilution of each sample with Rn values closest to that target value was sequenced (Gohl et al. 2016b). In the second sample set, the ITS2 region was PCR amplified from the same 144 agricultural samples using exclusively undiluted DNA (0-fold dilution) and amplicons were sequenced on a separate sequencing run. Observations from agricultural soils were verified using a third set of prairie soil DNA extracts. For three prairie samples, the ITS2 region was amplified in triplicate from undiluted (0-fold dilution) and diluted (10-, 100-, and 1,000-fold) DNA templates, and every dilution of all triplicate amplifications was sequenced (hereafter referred to as *prairie technical replicates*) on a single sequencing run.

ITS2 sequence libraries for each sample set were created using dual-index amplification; wherein the first set of PCR reactions was done using KAPA HiFi Hot Start Polymerase and primers 5.8SR_Nextera (5'-TCGTCGGCAGCGTCAGATGTGTATAAG

AGACAGTCGATGAAGAACGCAGCG-3') and ITS4_Nextera (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTCCGCTTATTGATATGC-3'). PCR reactions consisted of 3 μL of template DNA, 0.537 μL of nuclease-free water, 1.2 μL of 5 \times KAPA HiFi buffer, 0.18 μL of KAPA dNTP mix, 0.3 μL of DMSO, 0.3 μL of 5.8SR_Nextera primer, 0.3 μL of ITS4_Nextera primer, 0.12 μL of 25 μM ROX dye, 0.003 μL of 1,000 \times SYBR Green dye, and 0.06 μL of KAPA HiFi Hot Start Polymerase, and were carried out using the following cycling protocol of 95 $^{\circ}\text{C}$ for 5 min, followed by 25 cycles of denaturation at 98 $^{\circ}\text{C}$ for 20s, annealing at 55 $^{\circ}\text{C}$ for 15 min, extension at 72 $^{\circ}\text{C}$ for 1 min, with a final elongation step for 5 min at 72 $^{\circ}\text{C}$. PCR products were diluted 100-fold and 5 μL of products from the first PCR were uniquely indexed with primers in the second PCR step (10 cycles) using the previously described cycle conditions. Amplicons were pooled and a size-selected sample was denatured with NaOH and diluted to 8 pM. Samples were either spiked with 25% PhiX and sequenced on the Illumina HiSeq platform 2 \times 250 bp (Illumina, San Diego, CA), or spiked with 15% PhiX and sequenced on the Illumina MiSeq 2 \times 250 bp using a MiSeq 600 cycle v3 kit (Illumina). Sequencing was carried out at the University of Minnesota Genomics Center, Minneapolis. The average per sample sequence depth varied across the three groups of sequenced samples (serially diluted agricultural, undiluted agricultural, serially diluted prairie). Specifically, the average library size of undiluted agricultural samples was significantly smaller than that of diluted agricultural samples due to variations in the number of samples loaded on each sequencing run and sequencing platform (Supplementary Table S1). We examined the "sequencer effect" by comparing samples sequenced as undiluted templates (0-fold dilution) among sequencing runs (Supplementary Fig. S1).

With 79 randomly selected undiluted agricultural DNA samples, the number of fungal ITS templates per sample was assessed using qPCR-based absolute quantification. A standard curve was constructed ranging from 10^3 to 10^8 ITS2 amplicons from cloned and purified DNA. Briefly, the synthetic ITS1-5.8S-ITS2 template (from *Saccharomyces cerevisiae*) containing a C to G change within the ITS1 primer sequence to correct a primer/template mismatch was inserted into pUCIDT (Integrated DNA Technologies, Coralville, IA). The plasmid was transformed into *E. coli*, purified using a Qiagen Midi Prep kit (Qiagen, Hilden, Germany), linearized with *BsaI* (New England Biolabs, Ipswich, MA), cleaned with 1.8 \times AmPureXP beads (Beckman-Coulter, Brea, CA), and

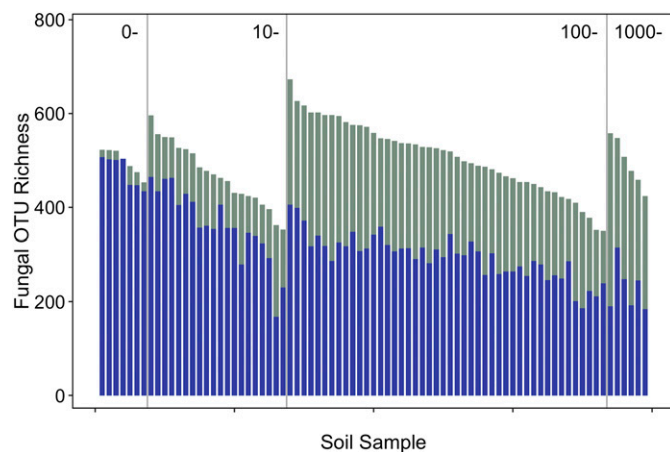


Fig. 2. DNA template dilution reduced fungal operational taxonomic unit (OTU) richness for agricultural soil samples. DNA templates were serially diluted (0-, 10-, 100-, and 1,000-fold; blue bars) prior to PCR amplification for which qPCR-based quantification of the relative number of fungal DNAs was used to select which dilutions were sequenced. Identical undiluted DNA templates were amplified and sequenced (green bars). Bars are ordered by undiluted template fungal OTU richness within each dilution.

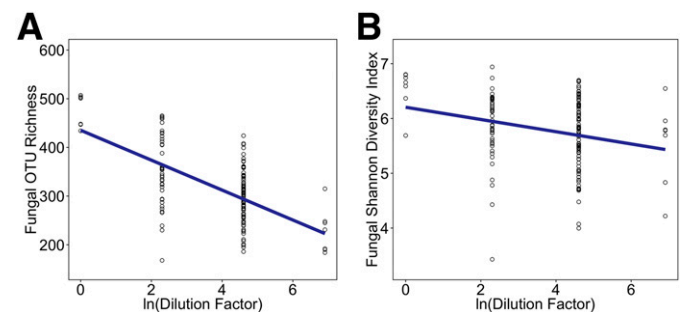


Fig. 3. Fungal operational taxonomic unit (OTU) richness and Shannon H' diversity decrease with pre-PCR DNA template dilution. Fungal richness and diversity declined significantly with pre-PCR dilution of DNA template (richness: $r^2 = 0.38$, $P < 0.0001$, $y = -30.73x + 435.24$; diversity: $r^2 = 0.06$, $P = 0.002$, $y = -0.11 + 6.21$) for 144 agricultural field soil samples from across Minnesota, USA. Samples were serially diluted (0-, 10-, 100-, and 1,000-fold) prior to PCR amplification. qPCR-based quantification of the relative number of fungal DNA molecules was used to select which diluted samples were sequenced.

quantified using PicoGreen (Life Technologies, Carlsbad, CA). DNA samples were amplified for 30 cycles with KAPA HiFi Hot Start Polymerase and the 5.8SR_Nextera and ITS4_Nextera primers. As above, PCR reactions consisted of 3 μ l of template DNA, 0.537 μ l of nuclease-free water, 1.2 μ l of 5 \times KAPA HiFi buffer, 0.18 μ l of KAPA dNTP mix, 0.3 μ l of DMSO, 0.3 μ l of 5.8SR_Nextera primer, 0.3 μ l of ITS4_Nextera primer, 0.12 μ l of 25 μ M ROX dye, 0.003 μ l of 1,000 \times SYBR Green dye, and 0.06 μ l of KAPA HiFi Hot Start Polymerase. The cycle threshold (Ct) values obtained for the agricultural soil samples were compared with the standard curve to determine the concentration of ITS2 sequences in the original samples. Using the measured concentrations of ITS2 molecules in undiluted agricultural samples, the number of ITS2 molecules in paired diluted samples was estimated by dividing by the corresponding dilution factor.

All sequences were analyzed with an in-house wrapper pipeline (<https://github.com/ZeweiSong/FAST/wiki/Fungal-ITS2-Pipeline-Using-Both-Reads>) and de novo sequence clustering was done at 97% sequence similarity using VSEARCH (Rognes et al. 2016). Sequences present only once across the entire dataset were discarded (Edgar 2013). Nonfungal sequences were removed from the operational taxonomic unit (OTU) table using a threshold suggested

by Tedersoo et al. (2015) wherein OTUs with 70% sequence alignment to a UNITE database (Kõljalg et al. 2013) reference at a similarity of 75% or greater were retained. We assigned taxonomy to OTUs at 97% similarity to UNITE database entries, and those not present in the database were classified as unknown fungi. OTUs classified as protists and nonfungal eukaryotes, which comprised approximately 0.67% of all taxonomically assigned sequences, were removed. As recommended for sequence libraries that vary substantially in depth (Weiss et al. 2017), sample libraries were rarefied to uniform depths of 20,000 and 12,000 sequences per sample for agricultural and prairie soil technical replicates, respectively, using a multiple rarefaction method (Song et al. 2015). Subsequent OTU-based analyses were done using QIIME (Caporaso et al. 2010). All sequence data and metadata are available via Figshare (<https://doi.org/10.6084/m9.figshare.6399827>).

Statistical analyses. Data were analyzed in R (R Core Team 2014). The relationship between fungal alpha diversity (number of unique fungal OTUs and Shannon H' diversity index) and dilution factor, fungal molecule number, or library size were examined using linear regression analysis. Differences in beta-diversity were assessed on Hellinger-transformed OTU-based Bray-Curtis dissimilarities using nonmetric multidimensional scaling (NMDS) and the

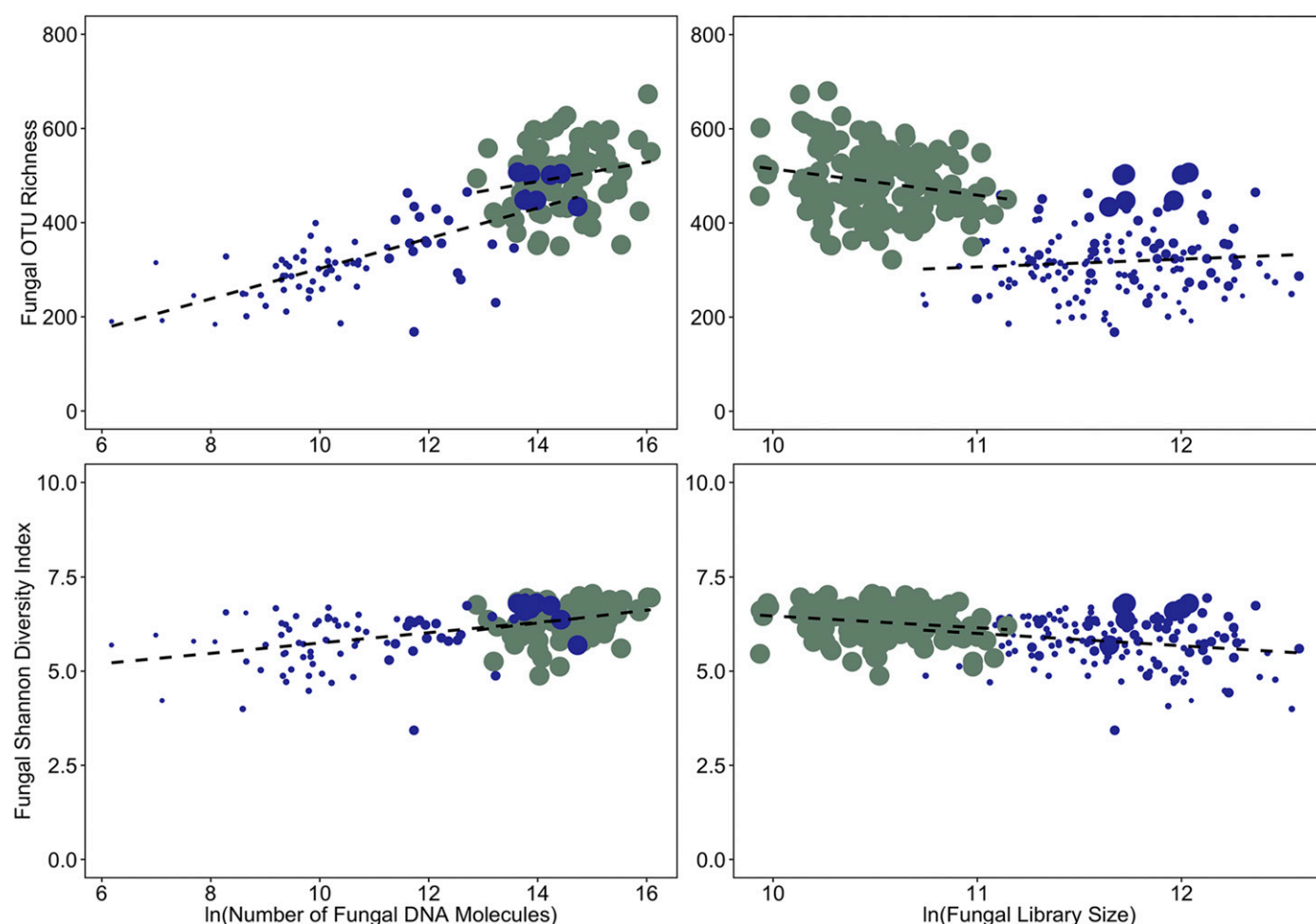


Fig. 4. Fungal richness and diversity increase with input DNA template molecule number, but not sequenced library size. Regression showed significant relationships between sample operational taxonomic unit richness and Shannon H' diversity index by pre-PCR ITS2 template number for diluted (richness: $r^2 = 0.51$, $P < 0.0001$; diversity: $r^2 = 0.12$, $P = 0.0013$) and undiluted (richness: $r^2 = 0.026$, $P = 0.081$; diversity: $r^2 = 0.049$, $P = 0.028$) samples. Relationships between richness and diversity and fungal library size were weak, but generally significant, for diluted (richness: $r^2 = 0.00054$, $P = 0.30$; diversity: $r^2 = 0.031$, $P = 0.020$) and undiluted (richness: $r^2 = 0.039$, $P = 0.0099$; diversity: $r^2 = 0.028$, $P = 0.026$) samples. Template number was measured using qPCR for undiluted samples and predicted for diluted samples by multiplying by dilution factors. Size of marker indicates the dilution (0-, 10-, 100-, 1,000-fold) selected for sequencing as determined by qPCR-based quantification of the relative number of fungal molecules.

metamds function in R. Differences in estimated community composition for paired undiluted and diluted samples from the same DNA template were evaluated using permutational analysis of variance on distance matrices (PERMANOVA) (Anderson 2001) and the Adonis function in R. Relationships between Bray-Curtis community dissimilarities among undiluted and diluted samples and dilution factor were examined using linear regression analysis.

RESULTS

For the 144 agricultural soil DNA samples, we used a qPCR-based relative quantification method to identify the appropriate dilutions for sequencing and a majority of DNA templates were sequenced as 10- or 100-fold dilutions of the original sample concentration (Fig. 2). For all samples, dilution of the DNA template prior to PCR amplification

markedly reduced estimates of fungal richness and diversity. For diluted agricultural samples, fungal richness and diversity decreased significantly with DNA template dilution (richness: $r^2 = 0.38$, $P < 0.0001$; diversity: $r^2 = 0.06$, $P = 0.002$) (Figs. 2 and 3). Moreover, replicate prairie soil samples—three samples serially diluted and sequenced in triplicate at each dilution—corroborated these observations. As with the agricultural samples, fungal richness and diversity of prairie soil samples were significantly negatively correlated with dilution (richness: $r^2 = 0.92$, $P < 0.0001$; diversity: $r^2 = 0.73$, $P < 0.0001$; Supplementary Fig. S2).

To further elucidate the effects of dilution within agricultural samples, we estimated ITS2 template number in undiluted samples using qPCR-based absolute quantification. Template number in undiluted samples ranged from 393,640 to 9,575,285 molecules and calculated estimates of DNA template molecule number were much

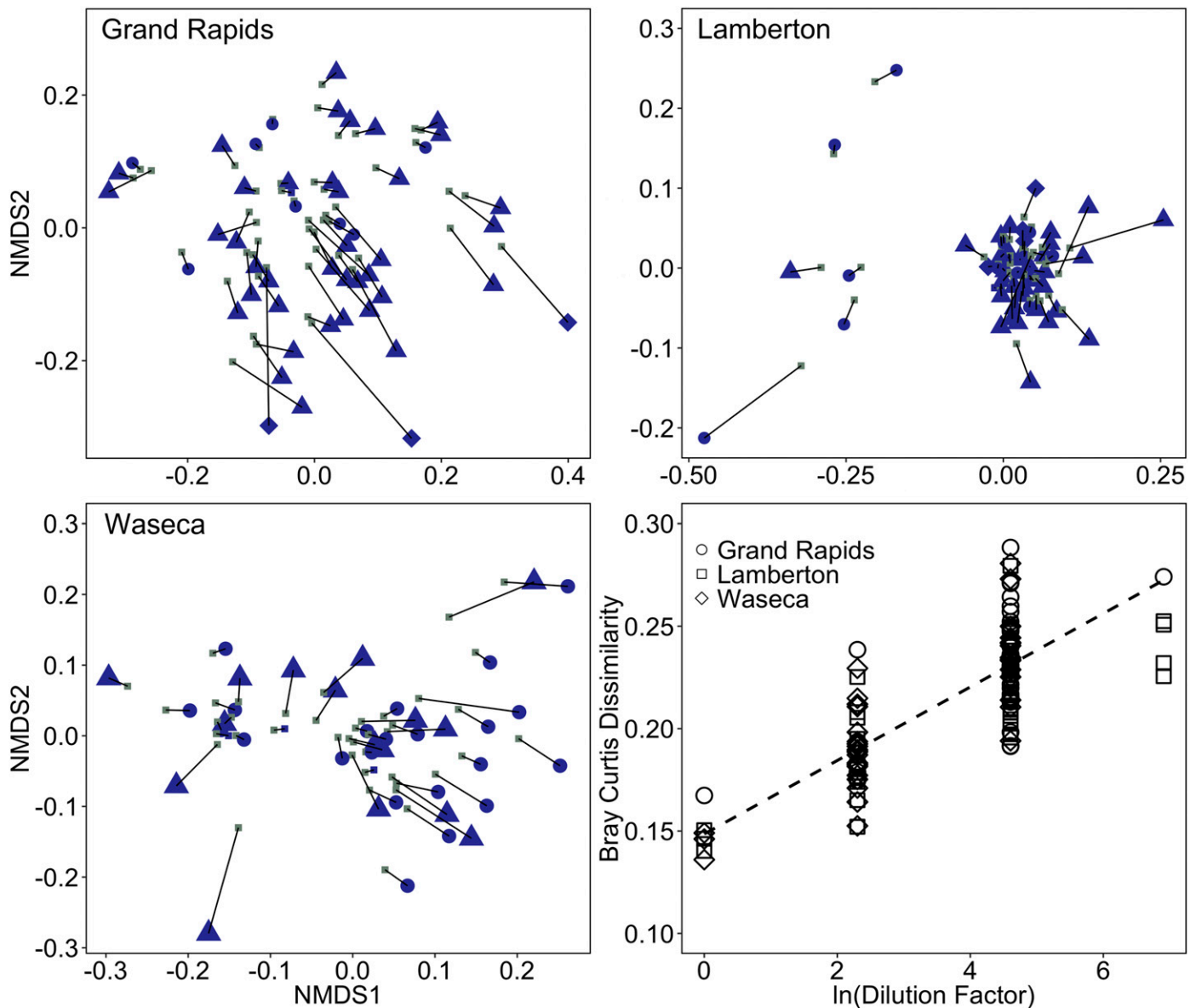


Fig. 5. DNA template dilution alters fungal community composition for agricultural soil samples. Nonmetric multidimensional scaling (NMDS) ordination analysis of fungal community composition based on Hellinger-transformed OTU-based Bray-Curtis dissimilarities for paired undiluted (green) and diluted (blue) soil DNA templates. Shapes indicate sample dilution for undiluted (0-fold dilution; square), 10-fold (circle), 100-fold (triangle), and 1,000-fold dilutions (diamond). Regression analysis showed a significant positive relationship between Bray-Curtis dissimilarities among paired diluted and undiluted samples and dilution factor for all sites together (dashed line; $r^2 = 0.55$, $P < 0.0001$; $y = 0.018x + 0.15$) and for each site individually (Grand Rapids: $r^2 = 0.57$, $P < 0.0001$; Lamberton: $r^2 = 0.38$, $P < 0.0001$; Waseca: $r^2 = 0.70$, $P < 0.0001$).

lower for diluted samples (482 to 957,528 molecules), indicating that the relative quantification procedure provided a poor estimate of the template abundance values for many of these samples. Among agricultural soil samples, fungal OTU richness and diversity increased with predicted template molecule number when samples were diluted (richness: $r^2 = 0.51$, $P < 0.0001$; diversity: $r^2 = 0.12$, $P = 0.0013$). However, the relationship was weak or nonsignificant between measured fungal ITS2 molecule number and fungal diversity for undiluted samples (Fig. 4) (richness: $r^2 = 0.026$, $P = 0.081$; diversity: $r^2 = 0.049$, $P = 0.028$).

PERMANOVA analysis of OTU-based Bray-Curtis dissimilarities showed that the composition of fungal communities was significantly different among diluted and undiluted agricultural soil samples (Fig. 5) from Grand Rapids ($r^2 = 0.035$; $P = 0.001$), Lamberton ($r^2 = 0.034$; $P = 0.001$), and Waseca ($r^2 = 0.030$; $P = 0.003$). Bray-Curtis dissimilarities among paired diluted and undiluted sample groups increased significantly with dilution for all sites (Fig. 5) ($r^2 = 0.55$, $P < 0.0001$) and for each site separately (Grand Rapids: $r^2 = 0.57$, $P < 0.0001$; Lamberton: $r^2 = 0.38$, $P < 0.0001$; Waseca: $r^2 = 0.70$, $P < 0.0001$). Importantly, agricultural soil samples sequenced as undiluted templates (0-fold dilution) were highly similar in richness, composition, and DNA molecule numbers among sequencing runs (Figs. 4 and 5).

Species abundance distribution (SAD) curves of technical replicate prairie samples were highly dependent on dilution (Fig. 6). Rank-ordering OTUs by abundance revealed that for diluted samples, abundant OTUs were overrepresented, while rare OTUs were underrepresented. By contrast, samples that were amplified from undiluted DNA templates had lower representation of abundant OTUs and greater representation of rare OTUs. Likewise, Preston-type frequency distribution analysis revealed that undiluted samples featured left skewed abundance distributions, while distributions became increasingly right skewed as with increasing sample dilution (Fig. 7).

DISCUSSION

Our objective was to assess effects of DNA template dilution on estimates of fungal microbiome composition, diversity, and

structure. Several broad conclusions emerge from the results of this study. First, DNA template dilutions ranging from 10- to 1,000-fold significantly reduced sequence-based estimates of fungal richness and diversity. Secondly, greater sample dilutions led to greater compositional shifts. Lastly, dilution led to shifts in species abundance distributions, such that dilute samples failed to capture rare OTUs while over representing dominant OTUs. Taken together, our results show that restricting the number of input template DNA molecules to PCR reactions can impose a strong bottleneck on estimates of soil fungal composition and diversity.

We examined template dilution effects on soil fungal communities, where the ITS2 template number in undiluted samples ranged from ~0.4 to 9.5 million per sample—counts that were well above the final rarefaction depth. However, ~65% of the 100-fold diluted samples and all 1,000-fold diluted samples had input molecule numbers lower than 20,000 molecules per sample. With technological advances in next generation sequencing, it is increasingly common for studies to achieve deep sequencing for environmental samples. However, results presented here demonstrate that such sequencing depths for fungi may not improve the accuracy of estimates of microbiome diversity and composition if the read depth or rarefaction depth greatly exceeds the number of input DNA templates.

It is generally assumed that species abundance distributions in microbial communities are skewed such that few taxa are common and many taxa are rare (Curtis et al. 2002). The data presented here show that the shape of the SAD curve is highly dependent on the number of template DNA molecules input to PCR reactions. Greater sampling effort, represented by undiluted samples and larger input molecule numbers, yielded left-skewed SAD curves. By contrast, low sampling efforts that were represented by dilute samples, led to right-skewed SADs and low estimates of species richness. Moreover, our results support the idea that the Preston “veil line” (Preston 1948)—the point at which rare species of the community are under sampled or not detected—shifts toward inclusion of higher frequencies of rare taxa as sampling effort or template concentration increases (Fig. 8). This lack of rare sequences could also result from rare ITS2 sequences not amplified in the first few rounds of PCR, thus leading to their underrepresentation or absence in sequencing results. Regardless, we posit that the shift in species abundance distribution results from sample dilution restricting the representation of marker genes from low and medium abundance organisms while highly abundant marker genes are over-amplified. As next generation sequencing further develops, a potential way to avert dilution-associated issues may be through the use of PCR-free library preparation. However, this area requires further investigation.

Although the 1,000-fold, and even the 100-fold dilutions of DNA templates are somewhat extreme dilutions relative to what is typically reported in the microbiome literature, we observed that even 10-fold dilution of template DNA can result in significant reductions in soil fungal richness and diversity. Even so, soil tends to have greater fungal biomass and diversity than other fungal habitats (Cuadros-Orellana et al. 2013) and we would expect that for fungal communities with relatively lower biomass and diversity, such as those on or within plant leaves or roots, that template dilution is likely to impose a more serious bottleneck to measurements of diversity and composition. In contrast, among bacterial communities, which generally exhibit higher sequence numbers and diversity in environmental samples, bottlenecks as a consequence of sample template dilution may pose less of a limitation at low dilutions (i.e., 10-fold) and may only affect diversity and composition at dilutions greater than 100-fold (Franklin et al. 2001; Wu et al. 2010; Yan et al. 2015). We encourage researchers to target molecule quantification

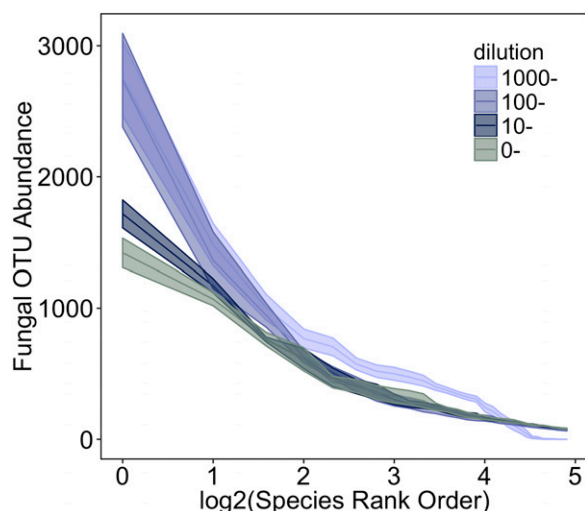


Fig. 6. DNA template dilution increased representation of abundant taxa while decreasing abundance of rare taxa. Ninety-five percent confidence intervals (shaded areas) and means (lines) of operational taxonomic unit (OTU) abundances by \log_2 transformed OTU rank order for three prairie soil samples that were serially diluted, amplified for ITS2, and sequenced in triplicate at each dilution level. The horizontal scale is truncated to the 30 most abundant OTUs for visualization purposes.

to assess whether sample dilution is likely to impose bottlenecks on estimation of microbiome richness, diversity, and composition for both bacterial and fungal communities.

Conclusions. We conducted qPCR-based absolute quantification of fungal DNA template abundance prior to sequence library generation to ensure that template number used in PCR exceeded the intended sequencing depth on a per sample basis. Our results demonstrate the validity of such a practice, as routine dilution of microbiome samples can have negative consequences for estimates of diversity and composition and may do more harm than good when inhibitors are not a hindrance to amplification. The present study underscores the importance of consistent documentation of methodologies for microbiome research including the quantity and concentration of target DNA template into PCR reactions, as opposed to total DNA concentration in environmental samples which includes nontarget DNA (i.e., fungal, bacterial, host, soil animal, and plant), and whether samples underwent dilution or concentration adjustment prior to PCR. As the field of microbiome research moves toward synthetic analyses involving many datasets and intra-laboratory investigations, we encourage researchers to give consideration to the consequences of implementing varying methodologies.

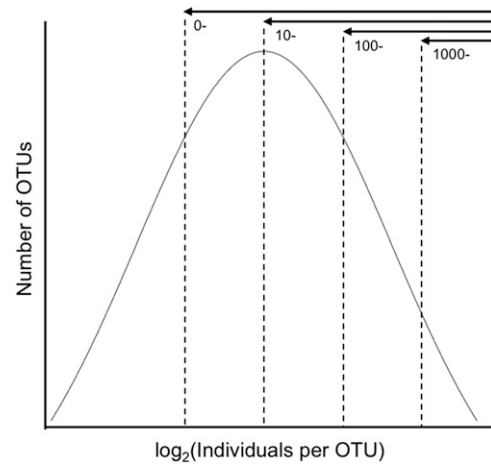


Fig. 8. Conceptualized shifts in lognormal abundance distributions of operational taxonomic units (OTUs) with DNA template dilution. Dilution moves the Preston “sampling veil” to the right, such that mid- and low-abundance species are underrepresented in diluted samples.

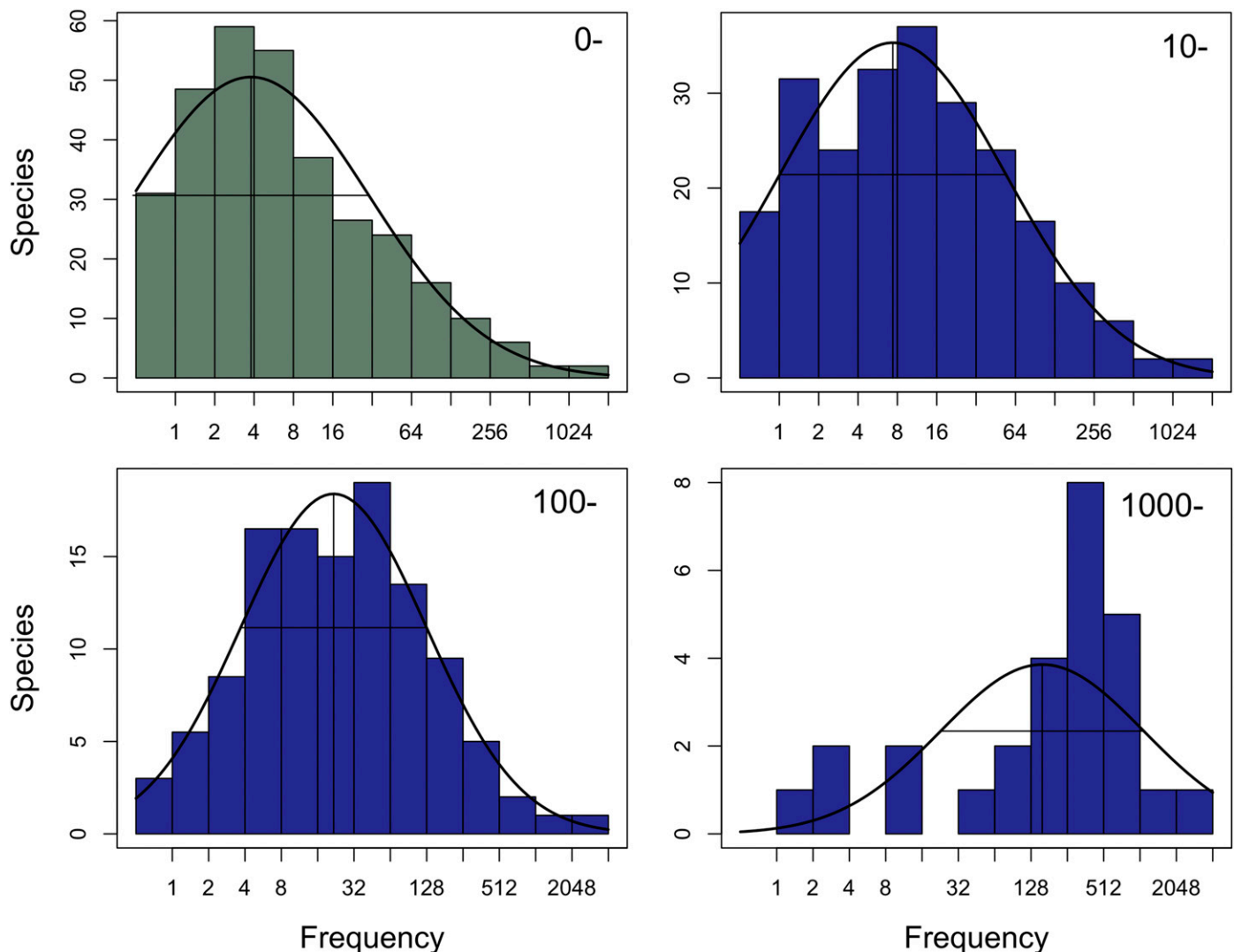


Fig. 7. Preston frequency of fungal operational taxonomic units (OTUs) became increasingly right-skewed with dilution. Prairie soil fungal DNA was PCR-amplified for the ITS2 region as undiluted (0-fold dilution) and serially diluted (10-, 100-, and 1,000-fold) DNA templates prior to sequencing. Fungal OTU frequencies are plotted as Preston’s histograms.

ACKNOWLEDGMENTS

We thank M. Bickell and K. Anderson for LTARN plot support and The University of Minnesota Genomics Center for conducting all molecular sequencing.

LITERATURE CITED

- Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26:32-46.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., and Costello, E. K. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7:335-336.
- Coddington, J. A., Agnarsson, I., Miller, J. A., Kuntner, M., and Hormiga, G. 2009. Undersampling bias: The null hypothesis for singleton species in tropical arthropod surveys. *J. Anim. Ecol.* 78:573-584.
- Cuadros-Orellana, S., Leite, L. R., Smith, A., Medeiros, J. D., Badotti, F., Fonseca, P. L., Vaz, A. B., Oliveira, G., and Góes-Neto, A. 2013. Assessment of fungal diversity in the environment using metagenomics: A decade in review. *Fungal Genom. Biol.* 3:1.
- Curtis, T. P., Sloan, W. T., and Scannell, J. W. 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci.* 99:10494-10499.
- Edgar, R. C. 2013. UPPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10:996-998.
- Franklin, R. B., Garland, J. L., Bolster, C. H., and Mills, A. L. 2001. Impact of dilution on microbial community structure and functional potential: Comparison of numerical simulations and batch culture experiments. *Appl. Environ. Microbiol.* 67:702-712.
- Gohl, D. M., MacLean, A., Hauge, A., Becker, A., Walek, D., and Beckman, K. 2016b. An optimized protocol for high-throughput amplicon-based microbiome profiling. *Nature Protocol Exchange*.
- Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., and Knights, D. 2016a. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* 34:942-949.
- Gonzalez, J. M., Portillo, M. C., Belda-Ferre, P., and Mira, A. 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS One* 7:e29973.
- Green, J. L., Holmes, A. J., Westoby, M., and Oliver, I. 2004. Spatial scaling of microbial eukaryote diversity. *Nature* 432:747-750.
- Hughes, J. B., and Hellmann, J. J. 2005. The application of rarefaction techniques to molecular inventories of microbial diversity. *Methods Enzymol.* 397:292-308.
- Kembel, S. W., Wu, M., Eisen, J. A., and Green, J. L. 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLOS Comput. Biol.* 8:e1002743.
- Kennedy, K., Hall, M. W., Lynch, M. D., Moreno-Hagelsieb, G., and Neufeld, J. D. 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl. Environ. Microbiol.* 80:5717-5722.
- Kinkel, L. L., Wilson, M., and Lindow, S. E. 1995. Effect of sampling scale on the assessment of epiphytic bacterial populations. *Microbiol. Ecol.* 29:283-297.
- Köljal, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., and Douglas, B. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22:5271-5277.
- Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjoller, R., Köljal, U., Pennanen, T., Rosendahl, S., Stenlid, J., and Kauserud, H. 2013. Fungal community analysis by high-throughput sequencing of amplified markers—A user's guide. *New Phytol.* 199:288-299.
- May, R. M. 1988. How many species are there on earth? *Science* 241:1441-1449.
- Morisita, M. 1959. Measuring of interspecific association and similarity between communities. *Mem. Fac. Sci. Kyushu Univ. Ser. E.* 3:65-80.
- Mullis, K. B., and Faloona, F. A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:335-350.
- Nash, C. B. 1950. Associations between fish species in tributaries and shore waters of western Lake Erie. *Ecology* 31:561-566.
- Nee, S., Harvey, P. H., and May, R. M. 1991. Lifting the veil on abundance patterns. *Proc. Biol. Sci.* 243:161-163.
- Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* 29:254-283.
- Preston, F. W. 1962a. The canonical distribution of commonness and rarity: Part I. *Ecology* 43:185-215.
- Preston, F. W. 1962b. The canonical distribution of commonness and rarity: Part II. *Ecology* 43:410-432.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. 2016. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 4:e2584.
- Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. 2012. PCR inhibitors—Occurrence, properties and removal. *J. Appl. Microbiol.* 113:1014-1026.
- Song, Z., Schlatter, D., Kennedy, P., Kinkel, L. L., Kistler, H. C., Nguyen, N., and Bates, S. T. 2015. Effort versus reward: Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS One* 10:e0127234.
- Tedersoo, L., Anslan, S., Bahram, M., Pölme, S., Riit, T., Liiv, I., Köljal, U., Kisan, V., Nilsson, H., Hildebrand, F., and Bork, P. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycKeys* 10:1-43.
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., and Hyde, E. R. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5:27.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia* 50:296-302.
- Wu, J. Y., Jiang, X. T., Jiang, Y. X., Lu, S. Y., Zou, F., and Zhou, H. W. 2010. Effects of polymerase, template dilution and cycle number on PCR based 16S rRNA diversity analysis using the deep sequencing method. *BMC Microbiol.* 10:255.
- Yan, Y., Kuramae, E. E., Klinkhamer, P. G., and van Veen, J. A. 2015. Revisiting the dilution procedure used to manipulate microbial biodiversity in terrestrial systems. *Appl. Environ. Microbiol.* 81:4246-4252.
- Zhou, J., Kang, S., Schadt, C. W., and Garten, C. T. 2008. Spatial scaling of functional gene diversity across various microbial taxa. *Proc. Natl. Acad. Sci.* 105:7768-7773.