scientific reports

OPEN

Assessing the efficacy of eDNA metabarcoding for measuring microbial biodiversity within forest ecosystems

Zachary S. Ladin[1]✉, Barbra Ferrell[2], Jacob T. Dums[3], Ryan M. Moore[2], Delphis F. Levia[4], W. Gregory Shriver[5], Vincent D'Amico[6], Tara L. E. Trammell[1], João Carlos Setubal[7] & K. Eric Wommack[1]

We investigated the nascent application and efficacy of sampling and sequencing environmental DNA (eDNA) in terrestrial environments using rainwater that filters through the forest canopy and understory vegetation (i.e., throughfall). We demonstrate the utility and potential of this method for measuring microbial communities and forest biodiversity. We collected pure rainwater (open sky) and throughfall, successfully extracted DNA, and generated over 5000 unique amplicon sequence variants. We found that several taxa including *Mycoplasma* sp., *Spirosoma* sp., *Roseomonas* sp., and *Lactococcus* sp. were present only in throughfall samples. *Spiroplasma* sp., *Methylobacterium* sp., *Massilia* sp., *Pantoea* sp., and *Sphingomonas* sp. were found in both types of samples, but more abundantly in throughfall than in rainwater. Throughfall samples contained Gammaproteobacteria that have been previously found to be plant-associated, and may contribute to important functional roles. We illustrate how this novel method can be used for measuring microbial biodiversity in forest ecosystems, foreshadowing the utility for quantifying both prokaryotic and eukaryotic lifeforms. Leveraging these methods will enhance our ability to detect extant species, describe new species, and improve our overall understanding of ecological community dynamics in forest ecosystems.

Forest ecosystems are highly complex and dynamic assemblages of biotic and abiotic networks that support diverse biological communities[1]. Forests are also critical to numerous ecological functions from global to local scales[2], which include mediating water cycling, nutrient cycling, carbon storage, and providing habitat for billions of species across the planet[3]. Continued increasing rates of globalization and concomitant anthropogenic effects are leading drivers of forest habitat loss, threatening species diversity[4,5]. Moreover, exacerbating factors such as climate change and biological invasion additionally reduce habitat quality, which may lead to disproportionate losses of biodiversity related to contemporary losses of forest habitat, globally[6]. For these reasons, it is imperative that we gain a better understanding and more accurate means of quantification of the ecological communities that comprise these ever-threatened ecosystems.

The long-standing importance of understanding ecosystems and ecological communities has largely focused on the dynamics of synergistic interactions among organisms and abiotic components[7-9]. Despite the prevalence of theories and differences in approaches to finding underlying generalizable rules[10] for describing and modeling community-level patterns[11,12], understanding ecological communities requires the accurate measurement of biological diversity[13] hereafter 'biodiversity' as pertaining to organismal or species diversity[14]. Leveraging rapid technological advances in recent decades, such as high-throughput sequencing, we are now able to characterize community assemblages within ecosystems with increasing accuracy and detail[15]. Furthermore, unprecedented perspectives of biodiversity and complexity within communities has also provided new frameworks for

[1]Department of Plant and Soil Sciences, University of Delaware, 264 Townsend Hall, Newark, DE 19716, USA. [2]Department of Plant and Soil Sciences, Delaware Biotechnology Institute, University of Delaware, Newark, DE 19716, USA. [3]Biotechnology Program, North Carolina State University, Raleigh, NC 27695, USA. [4]Department of Entomology and Wildlife Ecology, University of Delaware, 250 Townsend Hall, Newark, DE 19716, USA. [5]Departments of Geography and Spatial Sciences and Plant and Soil Sciences, University of Delaware, 216C Pearson Hall, Newark, DE 19716, USA. [6]US Forest Service, Northern Research Station, Newark, DE, USA. [7]Instituto de Química, University of Sao Paulo, São Paulo, SP 05508-000, Brazil. ✉email: zach@udel.edu

approaching the nature of interrelations that can influence community dynamics[16,17]. Within forest ecosystems in particular, the elucidation of interactions among plant-soil-microbiome associations[18] can provide new insight in understanding forest health and function more generally[19,20].

Over the past decade, environmental DNA (eDNA) has become an important method of detecting species occurrence[21–24]. Here, we follow the definition of eDNA as a "complex mixture of genomic DNA from many different organisms found in an environmental sample[25]." Sampling and analyzing eDNA is made possible using high-throughput sequencing and metabarcoding methods that are revolutionizing how we can now approach inventory and monitoring of biological diversity[26–29]. For example, eDNA metabarcoding has been effectively used to detect presence of plants[30], fungi[31], animals[32,33], and microorganisms[34]. Use of eDNA metabarcoding has been especially well-suited for identifying cryptic, rare, or endangered species. For example, eDNA from water samples has been successfully used to identify the presence of a rare and endangered saw-toothed fish[35], and similarly, eDNA has been used to detect the presence of an endangered marine skate species[36].

In general, using eDNA metabarcoding has many practical applications that include detecting biological invasion[37–39], trophic ecology[40,41], reconstructing ancient ecosystems[42], plant-pollinator interactions[43,44], and monitoring air and water quality[45–47]. Despite the impressive use of eDNA metabarcoding as a tool to increase our ability to measure biodiversity, eDNA metabarcoding has been limited to applications which use aquatic samples containing DNA, given the inherently fragile nature of organismal tissue containing DNA and its rapid degradation when exposed to typical environmental forces[48]. For instance, several studies have explored degradation processes of eDNA within aquatic environments[49,50]. Another study exploring persistence and factors influencing eDNA within soils found that eDNA became undetectable after roughly one week, although this may vary as a function of soil moisture, temperature, and habitat characteristics[51]. However, despite these methodological challenges, innovative use of eDNA metabarcoding is increasingly being used to detect species within terrestrial ecosystems[52–55].

It has also been well established that airborne biological particles occur within the atmosphere[56], and several studies have been successful in sampling and sequencing eDNA from airborne samples[46,57–59]. However, very few studies to date have explored the partitioning of microbial communities contained and transported in pure rainwater and throughfall[60,61]. Throughfall is the incident precipitation that passes either directly through the forest canopy, or alternatively is first intercepted and splashed from the canopy or coalesces and drips down to lower strata[62,63]. Along with stemflow, which is the intercepted precipitation that flows along branches and converges on the stem of plants, this canopy-breaching throughfall input can constitute upwards of 90% of precipitation entering a forest[64]. Key to hydrologic and biogeochemical cycling within forests, throughfall and stemflow transport water, solutes, particulates, and nutrients to the soil, where inputs enrich soils thereby influencing growth and dynamics of vegetation[65–67] and crops[68].

Airborne microorganisms that are found in aerosols can also play a role in cloud formation via ice nucleation creating biogeochemical feedback loops with vegetated areas, such as forests, indicated by the bioprecipitation hypothesis[69]. Indeed, studies that have successfully sequenced eDNA within aerosols[59] also lend further support to the bioprecipitation hypothesis[70]. Our study was specifically motivated to further assess the efficacy of not only sequencing eDNA within rainwater, which holds interesting implications unto itself, but to see if we could sequence eDNA from organisms that come into physical contact with rainwater in the forest canopy and understory, enabling rain droplets to acquire and sequester eDNA for sampling. By sampling rainwater that filtered through the forest canopy and understory vegetation we were able to successfully obtain DNA and quantify microbial consortia within terrestrial forest ecosystems. We predicted that throughfall would contain greater taxa diversity and greater bacteria abundances than rainwater that did not contact forest vegetation. We wanted to test the hypothesis that there would be additive effects of bioprecipitation and collection of forest canopy phytobiome communities compared to rainwater alone. This is important given the demonstrated linkages of and need for plant-associated bacteria and key functional roles these interactions can have (e.g., nitrogen fixing within the phylloshphere[61,71].

Materials and methods

Study area. Our study occurred within and near a 16-ha forest located in Newark, Delaware, USA (− 75.74477 W, 39.66249 N) within the mid-Atlantic United States (Fig. 1). This small forest consists of a vegetation community with dominant canopy tree species including *Fagus grandifolia* Ehrh (American beech), *Acer rubrum* L. (red maple), *Quercus* spp. (oak), *Liriodendron tulipifera* L. (tulip poplar), and *Liquidambar styraciflua* L. (sweetgum). Additionally, this forest contains both native and non-native understory woody species including *Lindera benzoin* L. (spicebush), *Viburnum* spp. L. (arrowwood viburnum), *Clethra alnifolia* L. (sweet pepperbush), *Rosa multiflora* Thunb. (multiflora rose), *Eleagnus umbellate* Thunb. (autumn olive), and *Rubus* spp (brambles). We randomly selected three sampling points from a suite of previously-established gridded points[72] each of which were greater than 50 m apart to achieve spatial independence, and 50 m from the forest edge to reduce bias from potential edge effects. We also located two rainwater sampling locations north of the patch on the edge of an adjacent fallow agricultural field that was over 100 m from the forest edge.

Sample collection. We collected samples of rainwater (*N* = 3) and throughfall (*N* = 10) in Newark, Delaware, USA from a single rain event on 11 August 2018. We collected samples at 3 randomly-selected locations within a 16-ha forest, and 2 open locations north of the forest within an adjacent field (Fig. 1). On 10 August 2018, one hour before an anticipated rain event between 1730 and 1800 h, we deployed two replicate high-density polyethylene 45.72-cm diameter funnels (United States Plastic Corp, Lima, Ohio, USA; see Fig. S1 in Supplementary Materials) at each sampling location spaced roughly 5 m apart. Each funnel was surface sterilized within the laboratory using 100% EtOH, dried, and stored within sterile plastic bags until deployment. Addition-
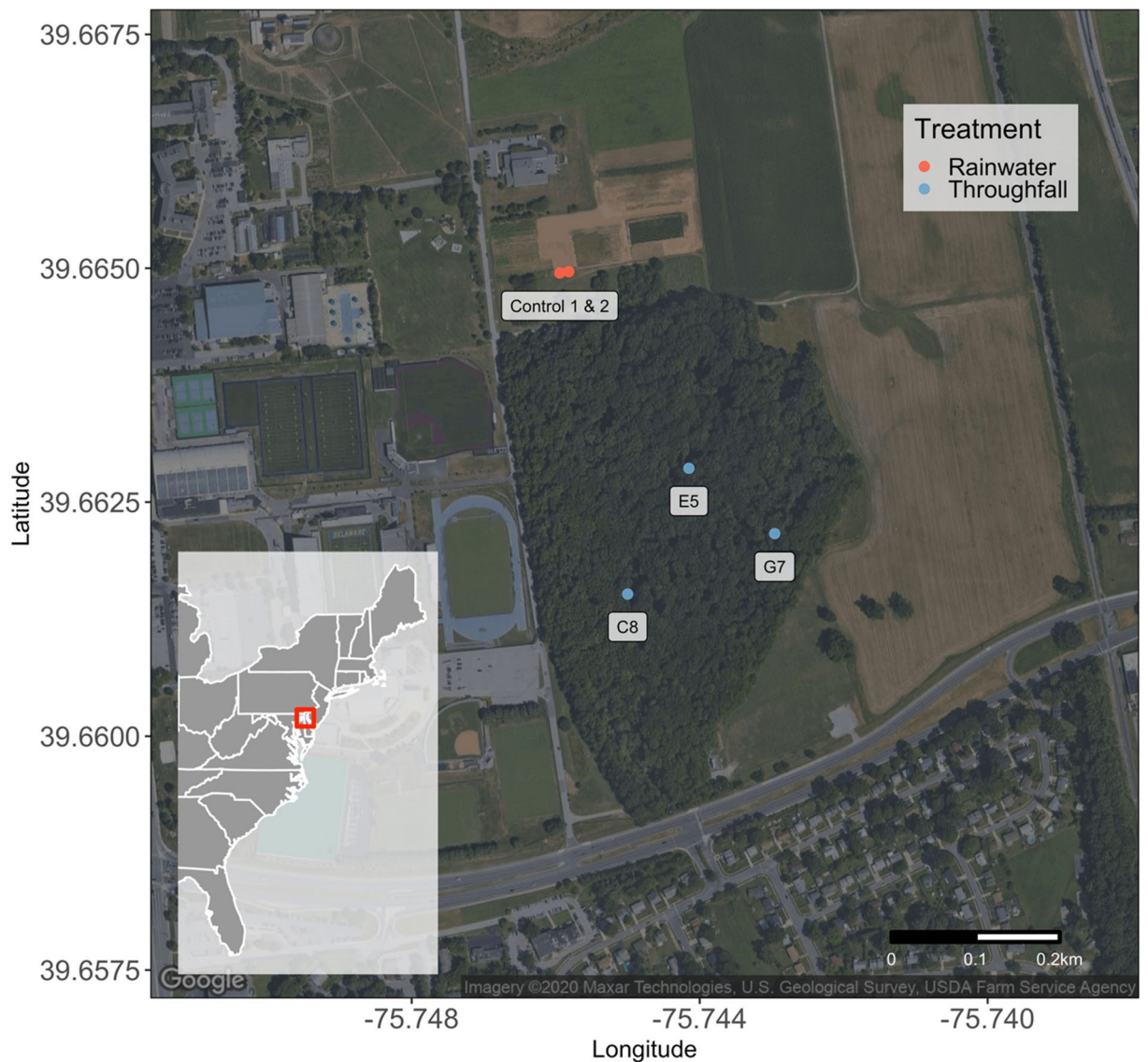
Figure 1. Map of study area showing sampling locations of rainwater (red) and throughfall (blue) treatments within and adjacent to the 16-ha forest (− 75.74477 W, 39.66249 N) in Newark, Delaware, USA. Map produced using program R (ver. 4.0.2; https://www.r-project.org/).

ally, each funnel was fitted with a sterilized 10.16-cm diameter funnel screen (0.5 mm mesh) to keep large debris from falling into rainwater samples. Each funnel was attached to a straight metal post using plastic zip-ties at a ca. 1-m height above the ground. Upon deployment, each funnel was rinsed with 50 mL of sterile de-ionized water, and these samples ($N=8$), were used as an internal control sample to detect any background bacterial DNA contamination in the sampling apparatus. These ddH2O samples were immediately brought to the laboratory, flash frozen in liquid nitrogen, and stored frozen at -80 °C for DNA analysis to detect any potential contamination. Finally, we blocked the bottom of the funnels using sterilized No. 7 rubber stoppers (Fisher Scientific Company, Waltham, Massachusetts, USA). All water samples were collected within 50-mL conical sterile polypropylene centrifuge tubes (Fisher Scientific Company, Waltham, Massachusetts, USA). We then returned the following day (11 August 2018) at 0700 h to collect rainwater samples, which were then immediately transferred to the laboratory, flash frozen in liquid nitrogen, and stored frozen at -80 °C until analysis[73].

Sample processing and 16S sequencing. Frozen samples were removed from -80 °C and immediately thawed in a warm water bath (50 °C), and then passed through sterile 0.22-µm Sterivex filter units without filling bell (Cat# SVGP010, Millipore, Burlington, Massachusetts, USA) using sterile 10-mL B-D Luer-Lok Tip syringes (Fisher Scientific Company, Waltham, Massachusetts, USA). We then extracted DNA from each of the filters using Qiagen DNeasy PowerWater kits (Qiagen, Hilden, Germany). DNA concentration was assessed using a Qubit 4 Fluorometer (Fisher Scientific Company, Waltham, Massachusetts, USA). A total of 21 samples (5 sampling locations each with 2 replicates, 10 ddH2O control samples, 1 DNA extraction control and 1 PCR no template control) were used to generate 16S rRNA amplicon libraries. The V4-V5 region of the 16S rRNA

gene was amplified from 12.5 ng of DNA using the primers (515F-Y (5′-GTGYCAGCMGCCGCGGTAA) and 926R (5′-CCGYCAATTYMTTTRAGTTT)[74] with the Illumina sequencing adapters attached. We chose to use these standard primers which have been previously undergone in silico and in vitro validation[74] to allow comparison of our results to previous research. Moreover, these primers were developed to cover an expanded range of organisms wide enough to capture additional eukaryotic sequences as well[74,75]. Each sample was barcoded following Illumina MiSeq standard procedures for Nextera indices. Samples were sequenced using paired-end 2 × 360 bp on the Illumina MiSeq as part of a 183 total sample equimolar pool at the University of Delaware—Sequencing and Genotyping Center.

Analysis of sequence data. Sequence data were analyzed by first removing forward (515F-Y) and reverse (926R) 16S primers and merging the demultiplexed reads with FLASH (ver. 1.2.11)[76], which were then pooled. We then used the fastp function (ver. 0.20.0)[76] to remove any reads with > 40% of bases having a quality score of < 20, remove any reads having more than one ambiguous base, and remove bases from both the front and back of a given read, when a 4-base sliding window average quality score was < 25. After confirming that primers were successfully removed, mean quality scores for reads ranged between 24 and 38. We then dereplicated our reads using VSEARCH (version 2.13.4)[78] to identify, merge, and sort identical sequences by decreasing abundance. Subsequently, we generated amplicon sequence variants (ASVs) using the unoise3 algorithm[79] implemented via VSEARCH (version 2.13.4)[78]. We then checked for and removed potential chimeras within reads using the '–uchime_denovo' function within VSEARCH (version 2.13.4)[78]. We removed all ASVs that were uniquely detected in ddH2O samples prior to analyses of rainwater and throughfall samples, and report these omitted ASVs (see Table S2 Supplementary Materials). All analyses were implemented with program R (ver. 4.0)[80].

Taxonomic assignment. Due to potential contamination during sequencing, we did not include the ddH2O sample from the Control 2 sampling location within any further analyses. We used the 'vsearch–sintax' function, a non-Bayesian classifier that uses a k-mer matching and bootstrapped confidence interval estimation algorithm for assigning taxonomies to ASVs[81]. However, given the potential for incorrect assignment[82,83], we performed all analyses on ASVs directly, and subsequently for exploratory purposes, presented taxonomic assignment of data up to the rank of genus.

Statistical analysis. After examining the number of reads for each ASV, we discarded any ASVs with reads < 5 (i.e., a 0.1% minimum abundance filter) within our data to avoid inclusion of sequencing errors[84]. After pooling within-sample replicates (i.e., in cases where more than one cryovial of rainwater was collected at a sampling point), we excluded any samples with < 2,000 reads. This list of unique ASVs found within ddH2O samples were then removed from the data, and not included in further analyses. We grouped data categorically into unique ASVs found only within throughfall samples, ASVs found in both throughfall and rainwater samples, and ASVs found only within rainwater samples, and then used a nonparametric Kruskal–Wallis rank sum test to determine significant differences among groups followed by test of multiple comparisons[85] to determine pair-wise differences in counts of ASVs between groups.

Due to the compositional nature of these data[86,87], we first transformed counts of ASVs per sample to the relative proportions, and then to centered log-ratios[88] using the acomp() and clr() functions within the 'compositions' (ver. 1.40–5) R package[89], respectively. Given the sparse nature of our data having a high proportion of zeros within our counts, we used a square-root Bayesian-multiplicative replacement of zeros[90] with the cmultRepl() function from the 'zCompositions' R package (ver. 1.3.4)[91].

To assess differences in community composition and structure between throughfall and rainwater, we independently computed Aitchison distances, which are the pair-wise Euclidean distances of centered log-ratio transformed ASV data[88], among sampling locations, plots, and between treatments using the dist() function within the 'compositions' package (ver 1.40–5)[89]. We subsequently used a nonparametric Kruskal–Wallis rank sum test to test for differences between throughfall and rainwater samples using plot-level Aitchison distances as the response variable.

We also estimated alpha diversity using the package 'DivNet' (ver 0.3.6)[92] to test for differences in microbial community diversity (at the genus rank) between throughfall and rainwater and among sampling plots. We estimated alpha diversity and beta diversity non-directional variation in community structure[93] following suggested methods that use appropriate log-transformed ratios of count data assuming communities are interconnected networks and hence, compositional in nature when estimating alpha and beta diversity[92,94–96]. Alpha diversity (Shannon's diversity index) and beta diversity (Bray–Curtis dissimilarity) indices were estimated for throughfall and rainwater were computed using the 'breakaway' package (ver. 4.7.1)[97]. We finally visually assessed patterns in community composition between throughfall and rainwater treatments by looking at mean and coefficient of variation (CV) of counts among genera and visualizing a hierarchical taxonomic tree. All statistical analyses were performed using program R (ver. 4.0.0)[80].

Results

We successfully extracted DNA from collected ddH2O water, rainwater, and throughfall samples. Sample volumes ranged between 1 and 40 mL, and DNA concentrations within samples ranged between trace amounts < 0.050 to 2.5 ng per μL. All samples with relatively low DNA trace concentrations were either from ddH2O or from rainwater (Table S1 in Supplementary Materials). After first pooling ASV data from sub-samples (i.e., samples from C8_1, E5_1, E5_2, G7_1, G7_2, where we collected and independently analyzed two samples per funnel), and removing all samples with fewer than 2,000 reads, the number of reads per sample ranged between 4922 and 196,015 with a mean of 93,200 (see Fig. S2 in Supplementary Materials). Within each treatment (i.e.,
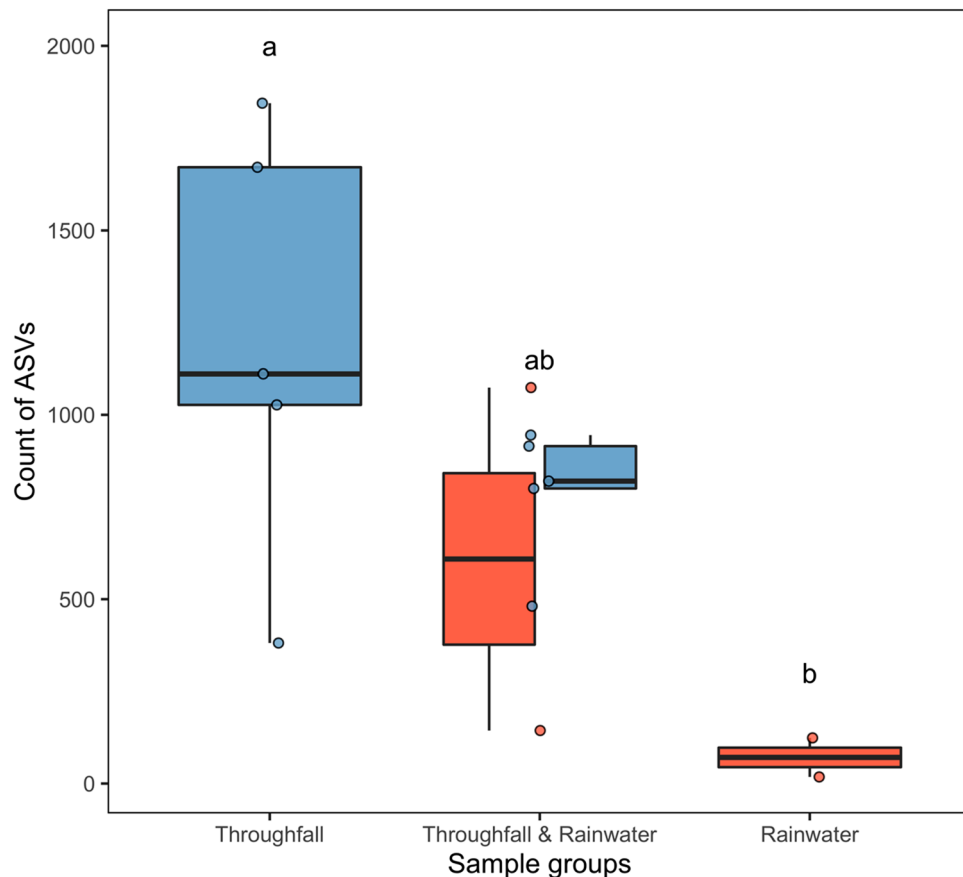
Figure 2. Box-and-whisker plots showing the difference in the count of unique ASVs detected among sampling groups: throughfall, throughfall and rainwater, and rainwater. Rainwater and throughfall treatments are shown in red and blue, respectively. Black lines denote medians, boxes represent 75% quantiles, red and blue dots show sample values, and lowercase letters indicate significant pair-wise differences (α = 0.05).

throughfall and rainwater) the mean number of reads for throughfall and rainwater samples were 68,896 and 32,960, respectively.

On average among samples, there were 4,065 unique ASVs ranging between 352 and 3614 ASVs (Supplementary Materials). Throughfall and rainwater samples had mean counts of 3,066 and 1,443 ASVs, respectively. The number of unique ASVs (mean ± SE) differed among groups (Kruskal–Wallis chi-sq. = 6.82, df = 2, $P < 0.05$), and the count of ASVs found solely within throughfall samples (1170.4 ± 266) was greater than those found in rainwater-only samples (81 ± 58; Z = − 2.51, $P < 0.05$, Fig. 2). However, the mean number of ASVs found in both throughfall and rainwater samples were similar to both ASVs found either in throughfall-only (Z = − 1.67, $P = 0.14$; Fig. 2) or rainwater-only groups (Z = − 1.40, $P = 0.16$; Fig. 2).

The degree of community dissimilarity inferred here based on Aitchison distances among all sampling locations (including both throughfall and rainwater treatments) ranged between 169.4 and 271.4, where 0 indicates having no dissimilarity, and greater values represent increasing dissimilarity (Fig. 3A). Pair-wise Aitchison distances among four plots ranged between 228.0 and 252.1 (Fig. 3B). The Aitchison distance between throughfall and rainwater treatments was 209.9 (Fig. 3C). When comparing the variation of within- and between-treatment community dissimilarity, we found that mean Aitchison distances were similar between throughfall-throughfall versus throughfall-rainwater treatment comparisons (Kruskal–Wallis chi-sq. = 3.628, df = 2, $P = 0.16$; Fig. 3D).

Bacterial communities found in throughfall had a significantly greater alpha diversity (3.68 ± 0.086), as indicated by Shannon's diversity estimate (mean ± SE), than rainwater (3.14 ± 0.14; Q-statistic = 10,125, $P < 0.0001$; Fig. 4A). Alpha diversity differed among plots (Q-statistic = 3,785, $P < 0.0001$). Each of the forest plots where throughfall samples were collected had greater alpha diversity, i.e., C8 (3.96 ± 0.16), G7 (3.65 ± 0.25), and E5 (3.65 ± 0.25) compared to control (rainwater) plot (3.14 ± 0.302; ß = − 0.299, P < 0.05; Fig. 4B). Alpha diversity was greater at C8 compared to G7 (ß = − 0.2986, $P < 0.05$). However, pair-wise hypothesis tests indicated that alpha diversity was similar between C8 and E5 (ß = − 0.1198, $P = 0.14$), and between E5 and G7 ß = − 0.1718, $P = 0.13$; Fig. 4B). Bray–Curtis dissimilarity estimate (mean ± SE) between throughfall and rainwater treatments was 0.39 ± 0.03. However, we did not detect a difference in Bray–Curtis dissimilarity estimates among any of the independent comparisons between plot pairs ($P > 0.17$ in all cases; Fig. 4C).
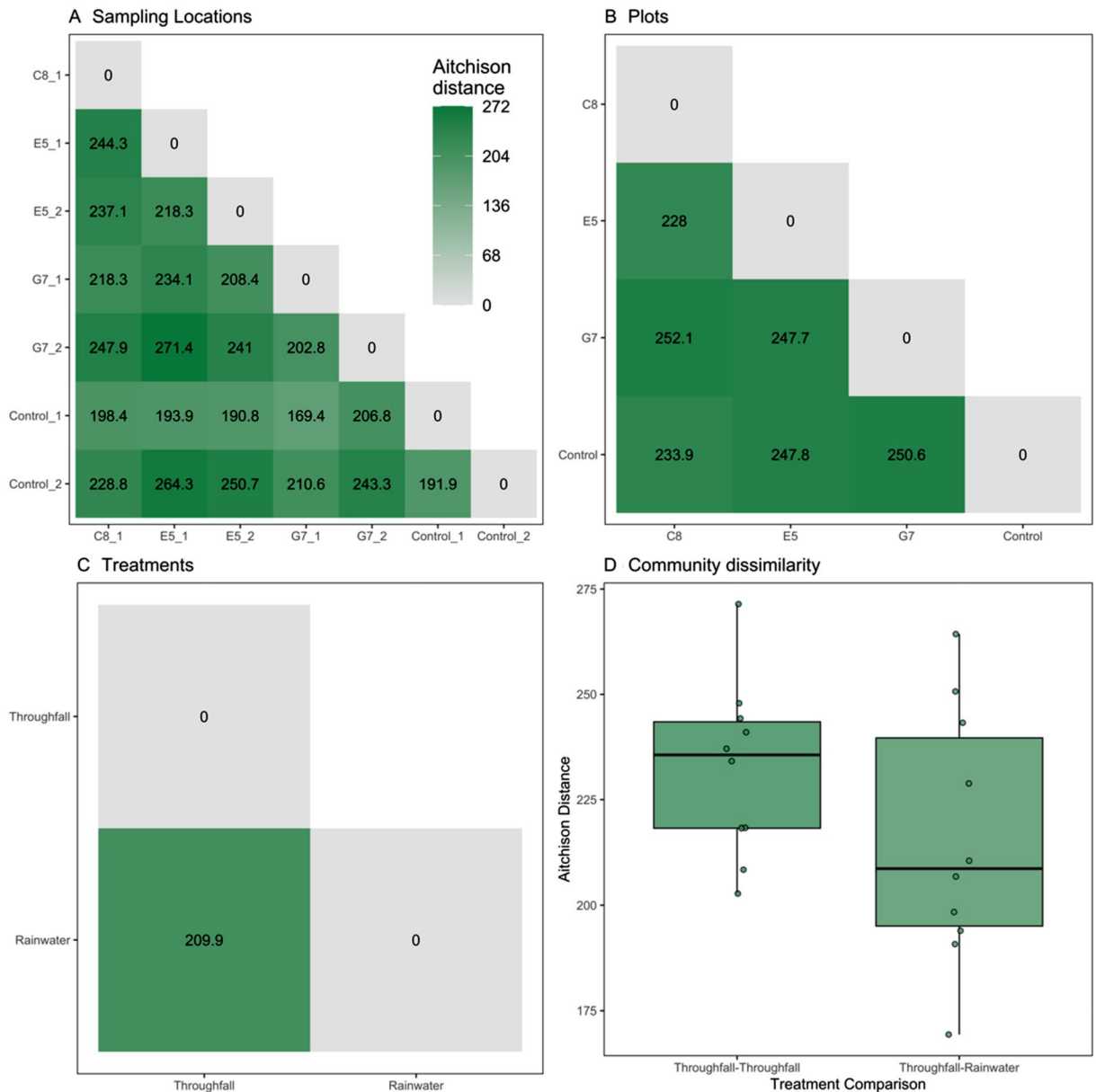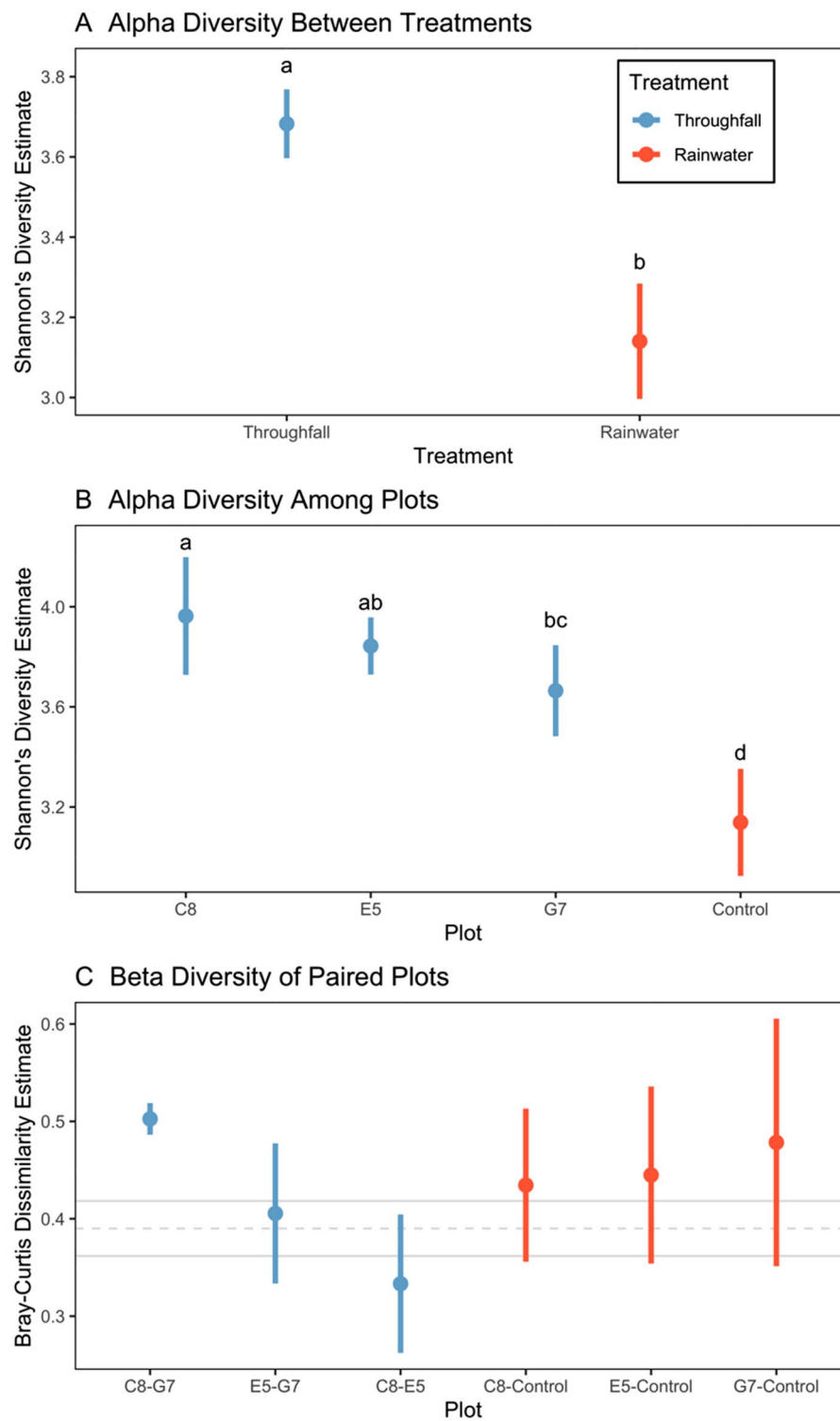
Figure 3. Plots of community dissimilarity based on pair-wise Aitchison distances (i.e., centered log-ratio transformed relative proportions of ASVs) among (**A**) sampling locations, (**B**) plots, and (**C**) throughfall and rainwater treatments. Aitchison distances are shown by a color ramp from light gray to green. A boxplot-and-whisker plot (**D**) compares plot-level mean Aitchison distances within throughfall samples, and between throughfall and rainwater treatments.

In general, the abundance (mean and CV) differed among taxonomically-assigned genera (Kruskal–Wallis chi-sq. = 270.6, df = 149, $P < 0.0001$; Fig. 4). Genera assigned to ASVs that were found in both throughfall and rainwater samples were in general found in greater abundance in throughfall, including *Spiroplasma, Methylobacterium*, *Massilia*, *Pantoea*, and *Sphingomonas* (Fig. 5). Genera found only in throughfall samples include *Mycoplasma*, *Pedobacter*, *Spirosoma*, *Roseomonas*, and *Lactococcus*. (Fig. 5). Among the top 20 in abundance, only ASVs associated with *Parabacteroides* were more abundant in rainwater samples than in throughfall (Fig. 5). Rainwater-only samples were lower in richness (the number of unique taxa), and in general were found in lower abundances, and included genera such as *Parabacteroides* sp., *Kluyvera* sp., and *Xanthomonas* sp. (Fig. 5). Further, visual examination of a hierarchical tree of taxonomically-assigned lineages to ASVs using the 'metacoder' (ver. 0.3.4) R package[98] also showed that throughfall contained greater diversity than rainwater (Fig. 6).

Discussion

We demonstrated the successful sampling and sequencing of eDNA contained within throughfall and rainwater, and assessed community composition patterns using an appropriate compositionally-aware framework[86,92,95]. This opportunistic method is capable of measuring microbial biodiversity uniquely associated with terrestrial

Figure 4. Community diversity indices (mean ± SE) for (**A**) Shannon's diversity index showing comparison of alpha diversity between throughfall (blue) and rainwater (red) treatments, (**B**) alpha diversity among plots, and (**C**) Bray–Curtis dissimilarity between all unique pairs of plots, with light gray lines indicating mean and SE of community dissimilarity between throughfall and rainwater. Lowercase letters indicate significant differences.
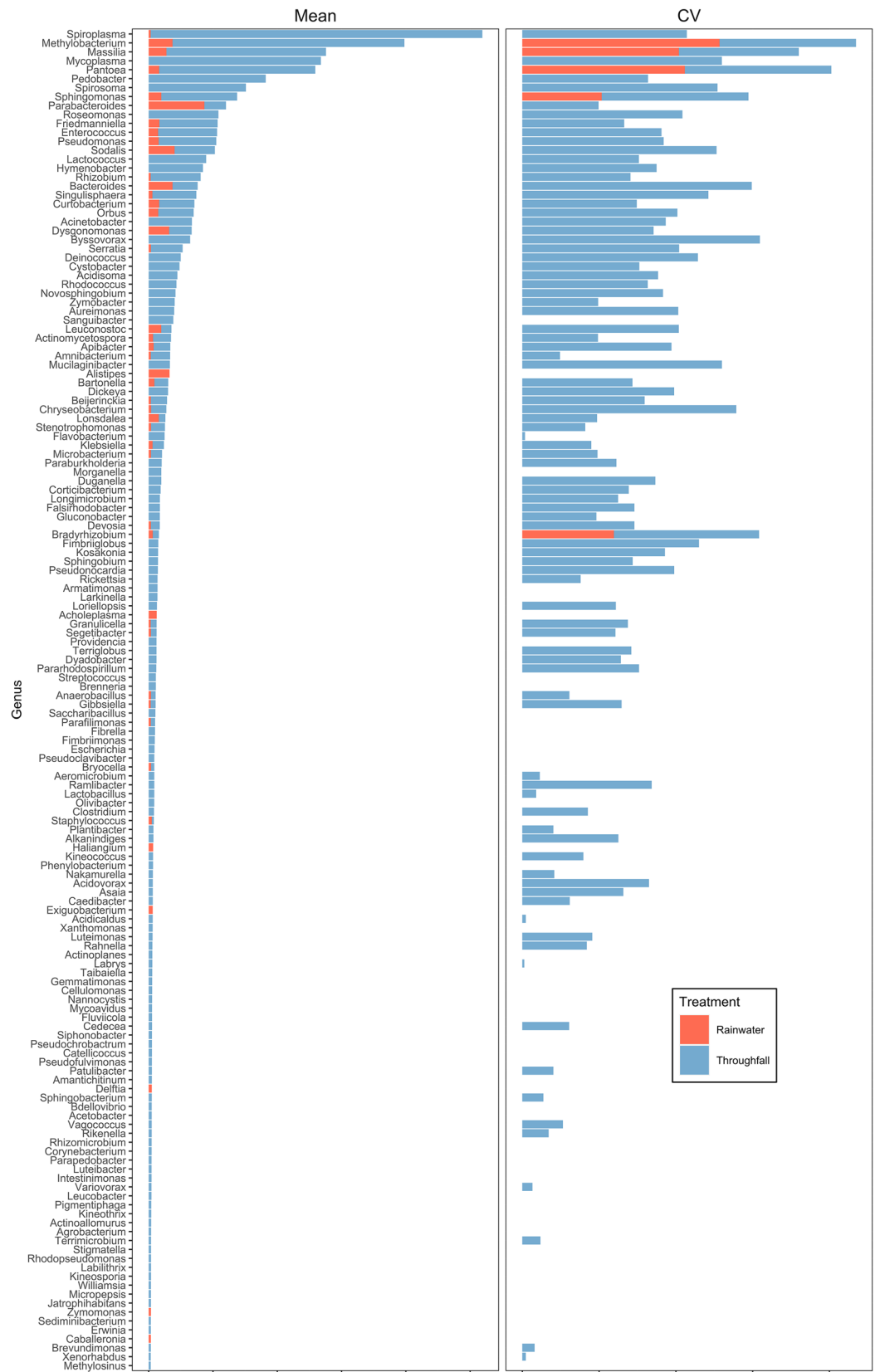
Figure 5. Mean abundance (number of ASV reads) and coefficient of variation (CV) for taxonomically-assigned ASVs for 151 unique bacterial genera detected in rainwater (red) and throughfall (blue) samples.
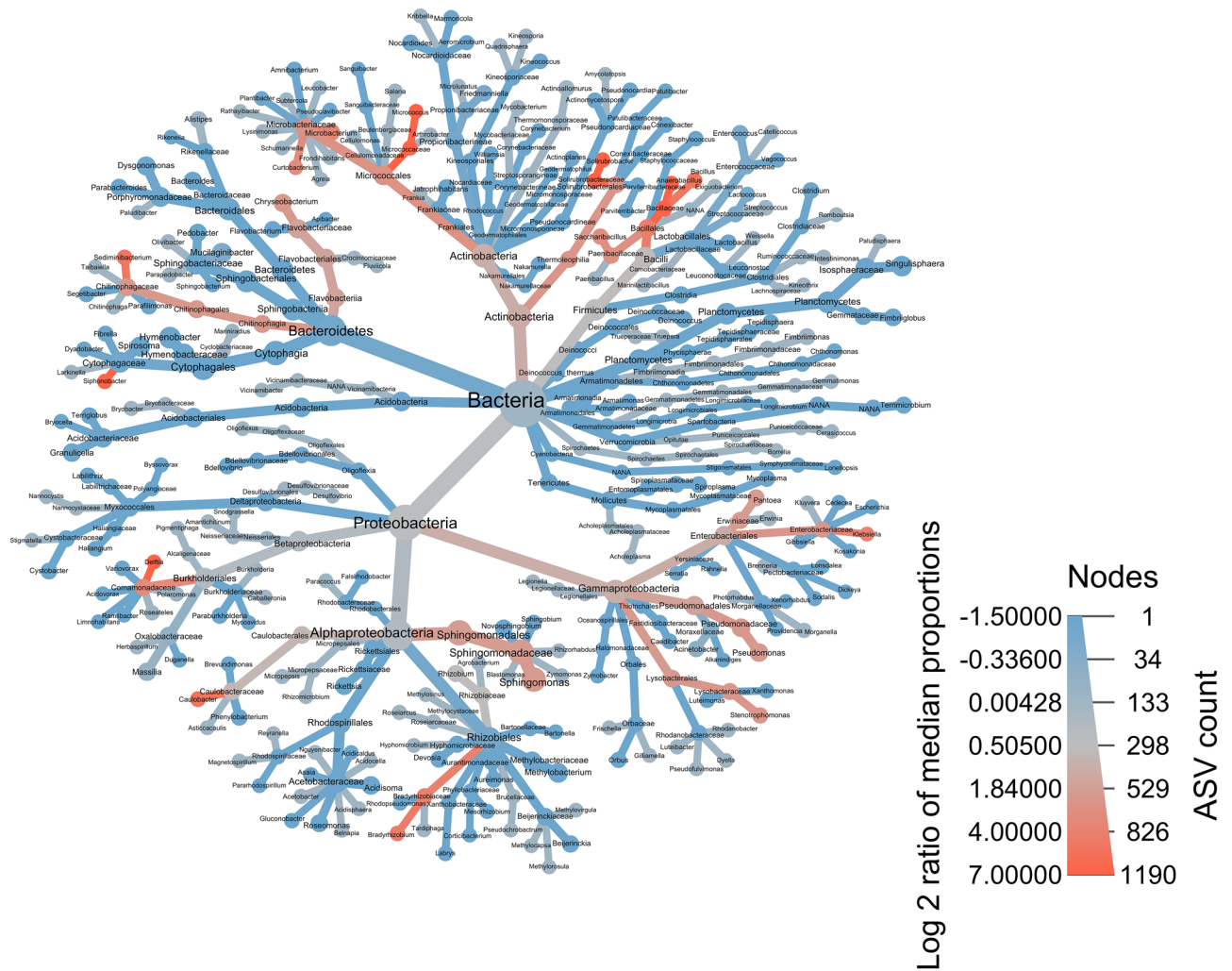
Figure 6. Taxonomic tree showing hierarchical structure of taxonomic ranks (denoted at each graph node) of 221 unique genera, and their membership in community assemblages based on the $\log_2$ of median ratios of genera detected within throughfall (blue), rainwater (red) and taxa found in both throughfall and rainwater (gray) samples. Thickness of graph edges indicate relative abundance of taxa. This figure was produce using the 'metacoder' package (ver. 0.3.4; https://grunwaldlab.github.io/metacoder_documentation/index.html)[77] in program R (ver. 4.0.2).

forest ecosystems. We predicted that throughfall-derived samples would contain greater taxa richness and that these taxa would, in general, would be found in greater abundances, both of which we found to occur (Figs. 4, 5). Our findings complement and lend further support for the promise of using eDNA metabarcoding as a tool for measuring and monitoring biodiversity[99–102].

While our prediction that throughfall samples would contain more diverse taxa than rainwater samples was supported, we also were interested in investigating how diversity might differ between throughfall and rainwater treatments (alpha diversity), among sampling locations (beta diversity), and the amount of within-treatment variation. We discovered that there was a high degree of community dissimilarity (interpreted as ecological distance) both within- and among-sampling locations as measured via Aitchison distances that account for the compositional nature of microbial community data[86]. For example, communities at sampling location E5 within our study area were more dissimilar than throughfall and rainwater sampling plots (Fig. 3B). Yet, we also found greater dissimilarity within-sampling location replicates at E5 (i.e., E5_1 and E5_2; Fig. 3A). However, when comparing the degree of within-treatment community dissimilarity, we found throughfall samples to have more similar microbial communities than rainwater samples. This observed outcome could result from small sampling efforts that may introduce bias.

Detected differences in alpha diversity (i.e., Shannon's diversity estimates) between throughfall and rainwater treatments, and associated plot-level comparisons additionally supported our predictions that throughfall would contain greater community diversity. Estimated beta diversity from Bray–Curtis dissimilarity of 0.38 also suggests that communities in throughfall and rainwater were distinct relative to the global taxonomic pool among our samples. Our lack in finding significant differences in beta-diversity among paired plot comparisons is likely due to low sample sizes, which future studies may alleviate with increased sampling. We also understand that many

factors can contribute to variation in microbial communities related both to environmental and sampling bias. While our study demonstrates variation within and among treatments, future studies with experimental designs having increased numbers of sampling locations among all treatments, additional within-sampling location replicates, and repeated sampling through time would provide data better-suited to evaluating how local-scale environmental effects (e.g., variation in weather events, vegetation structure, distance-to-edge, wind direction, etc.) may influence microbial diversity in throughfall.

There remain several unknown yet important questions about how microbial communities form and persist within forest ecosystems, how community dynamics over time are influenced by bioprecipitation, and how these might influence forest ecosystem health and function. Quantifying the degree to which bioprecipitation may influence nutrient and mediation of biogeochemical cycling for instance can help better understand how this phenomenon influences forest ecosystems. Supportive results from other studies using flow cytometry to measure bacterial concentrations (cells mL^{-1}) found 5.5 times greater concentrations of bacteria in throughfall, and 10 times greater concentrations in stemflow compared to pure rainwater (0.52×10^5 cells mL^{-1})[103]. In our study, we also found greater abundances of ASVs within throughfall samples compared to pure rainwater, supporting our prediction. Although ASV abundance is not analogous to cell concentration per se, our DNA sample concentrations also reflected this pattern (see Table S1 in Supplementary Materials).

It has long been known that microbial communities associated with the phyllosphere and rhizosphere play important roles in biogeochemical cycling of nutrients, and hence forest ecosystem function and dynamics[18,19]. Given the evidence supporting the bioprecipitation hypothesis[103–105], and the likelihood of microbial agents being transported and input into forests via throughfall and stemflow, more research into this area is warranted. For example, using a combination of measuring $\delta^{17}O$ and $\delta^{18}O$ stable isotopes in forests to estimate nitrogen deposition in pure rainwater and throughfall, along with analyzing DNA obtained from leaf surfaces, rainfall, and throughfall, another study found that nitrate (NO_3^-) was enriched within throughfall compared to pure rainwater via increased nitrifying microbial activity[61]. Despite limitations in interpreting results from previous studies that have not explicitly accounted for the compositional nature of microbial community data[60,103] which can introduce bias, our results of relative patterns of taxa found in throughfall and rainwater support previous findings. In a previous study, microbial communities also differed between rainwater and throughfall samples, and we detected 14 of 15 (93%) of the same bacterial phyla (Fig. 6) within our samples that as those detected within their study[61]. Despite not being able to make direct comparisons about community diversity per se, in general, we detected 82% of the same classes found in another study that explored how variability in storm events can influence microbial community inputs into forests as well as increasing the transfer of bacteria from the canopy to the forest floor via throughfall and stemflow[60].

The implications for future explorations into these phenomena are tantalizing, indeed. Take for example the global-scale transport of nearly 40 million tons of dust particulates per year originating in the Bodel´e depression in the Sahara Desert in Africa and ending up in North and South America[106]. This phenomenon, driven by multiple factors including topography, seasonal patterns in global weather and prevailing winds, all of which are influenced by climate change, are known to transport significant amounts of mineral and trace nutrient inputs to the Amazon rainforest in South America[107,108], and in general, hold important implications for nutrient inputs into forests[109]. Given the overwhelming support of the bioprecipitation hypothesis[70,110–112], it is likely that sampling Saharan dust, for example, using the eDNA metabarcoding approach set forth here, would yield a treasure trove of microbial community data, and would build upon findings from previous studies[101,113] to provide further insight into global-sale macroecological dynamics of interconnected yet geographically disparate biomes. Our study supports the hypothesis directly through our detection of *Pseudomonas* sp. and *Erwinia* sp. that contain specific genes enabling these organisms to facilitate ice-nucleating cloud formation[114].

However, using eDNA metabarcoding as an ecological monitoring tool comes with some inherent challenges and potential pitfalls[102,115]. Depending on the particular application, eDNA metabarcoding may not provide the ideal data to allow for developing and testing certain hypotheses. For instance, after reviewing the current challenges, limitations, and benefits of using eDNA metabarcoding to study population genetics (e.g., to assess allelic diversity within populations), it is predicted that eDNA monitoring will play an important yet, in some cases, complementary role that will lead to broad-scale inference relating to species population genetics and conservation[116].

Next logical steps for the further study of the efficacy in using eDNA within throughfall as a biodiversity monitoring tool for terrestrial ecosystems ought to include the continued examination of addressing aforementioned biases and challenges[57,117], and extending these methods to include the detection eukaryotes as well. Using field methodologies described here, or other standardized methods for sampling throughfall and stemflow[66,67,118], one could simply modify the sample processing methods and sequencing primers. For example, filtering collected samples to physically sort prokaryotic organisms and using taxa-specific primers would increase the taxonomic resolution of taxonomically-assigned reads within samples[119,120]. The use of 18S ribosomal gene (rDNA) primers to sequence ASVs associated with eukaryotes[121] is poised to become a widely-used method for detecting arthropods or vertebrates that occur in difficult-to-sample forest canopy habitat. Indeed, recent studies that have sampled eDNA from surfaces within terrestrial ecosystems provide exciting validation of eDNA metabarcoding for detecting insects[38,122] and vertebrates including amphibians[54,123], reptiles[56], and mammals[32]. We also suggest conducting further controlled experiments within the laboratory that would enable us to 'seed' simplified vegetated microcosms with known microbial consortia, in order to evaluate various methods of sample collection that might include the spraying of vegetation with sterile de-ionized water, while accounting for effects of DNA degradation similar to previous studies. These steps would be helpful in explicitly modeling and addressing the large bias that we acknowledge is associated with variability introduced by the dynamics of rain-providing weather events themselves. Given the support of the bioprecipitation hypothesis, this phenomenon inherently introduces bias potentially during both sample collection and data processing and analysis

(e.g., taxonomic assignment). For these reasons, we envision a field-based experiment that would simulate rain (albeit on a reduced spatial scale), using sterile water, and pumping this simulated rain above the canopy, and collecting samples of throughfall and stemflow to attempt to get a representative and less biased sample of forest-associated communities.

Our demonstration of using a nascent method to measure biodiversity within terrestrial ecosystems additionally holds promise for detecting eukaryotic organisms as well. In a recent controlled study, researchers were able to use eDNA metabarcoding to discriminate between two deer species after sampling recently-browsed twigs containing deer saliva[124]. These results, although from a controlled study, where samples were collected immediately after deer species were in direct contact with the twigs, coupled with our findings warrant conducting future studies aimed at detecting other taxa, including eukaryotes within forest ecosystems. These methods hold much promise for improving our knowledge and ability to measure biodiversity within forest ecosystems, which remains a critical research need given the wide-spread losses in biodiversity due to forest habitat loss and other anthropogenic factors that threaten forest ecosystems globally.

Conclusion

Despite current potential challenges and potential biases that need to be worked out through future study, our results have given us a promising look into how eDNA metabarcoding can be used to characterize terrestrial forest ecosystem microbial communities, and contribute to a more accurate description of community structure and function. Even if we are not quite at the point where using eDNA metabarcoding will be a viable method to assess population genetic diversity, perhaps through building upon recent advances in our understanding of predicting genotype–phenotype relationships[125,126], in conjunction with taking a functional-trait approach to modeling generalizable ecosystem dynamics[11], we will be headed towards a powerful methodology for monitoring global-scale population dynamics among species[39]. We are confident that eDNA metabarcoding will increasingly become a standardized and valuable tool for monitoring global biodiversity, thereby reducing unnecessary costs and impacts from sampling organisms directly, and will find increasing utility for detecting invasive species, detecting cryptic or rare species, and will ultimately help conserve endangered species, and ecological communities associated with imperiled habitats on Earth.

References

1. Liang, J. *et al.* Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**, (2016).
2. Myers, N. The world's forests and their ecosystem services. *Nature's Services: societal dependence on natural ecosystems* 215–235 (1997).
3. Miura, S. *et al.* Protective functions and ecosystem services of global forests in the past quarter-century. *For. Ecol. Manag.* **352**, 35–46 (2015).
4. Chaudhary, A. & Brooks, T. M. National consumption and global trade impacts on biodiversity. *World Dev.* **121**, 178–187 (2019).
5. Bowler, D. E. *et al.* Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes. *People and Nature* (2020).
6. Betts, M. G. *et al.* Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* **547**, 441–444 (2017).
7. Hutchinson, G. E., et al. Circular causal systems in ecology. *Ann. NY Acad. Sci.* **50**, 221–246 (1948).
8. MacArthur, R. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**, 533–536 (1955).
9. Odum, H. T. Efficiencies, size of organisms, and community structure. *Ecology* **37**, 592–597 (1956).
10. McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecol. Lett.* **13**, 627–642 (2010).
11. McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21**, 178–185 (2006).
12. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
13. Lovejoy, T. *The global 2000 report to the president (GO Barney, ed.), vol. 2.* (1980).
14. Hawksworth, D. L. *Biodiversity: Measurement and estimation.* vol. 345 (Springer Science & Business Media, 1995).
15. Whitham, T. G. *et al.* A framework for community and ecosystem genetics: From genes to ecosystems. *Nat. Rev. Genet.* **7**, 510–523 (2006).
16. Rosenberg, E. & Zilber-Rosenberg, I. Microbes drive evolution of animals and plants: The hologenome concept. *MBio* **7** (2016).
17. Theis, K. R. *et al.* Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes. *Msystems* **1** (2016).
18. Hawkes, C. V. & Connor, E. W. Translating phytobiomes from theory to practice: ecological and evolutionary considerations. *Phytobiomes* **1**, 57–69 (2017).
19. Uroz, S., Buee, M., Deveau, A., Mieszkin, S. & Martin, F. Ecology of the forest microbiome: highlights of temperate and boreal ecosystems. *Soil Biol. Biochem.* **103**, 471–488 (2016).
20. Baldrian, P. Forest microbiome: diversity, complexity and dynamics. *FEMS Microbiol. Rev.* **41**, 109–130 (2017).
21. Ficetola, G. F., Miaud, C., Pompanon, F. & Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Let.* **4**, 423–425 (2008).
22. Jerde, C. L., Mahon, A. R., Chadderton, W. L. & Lodge, D. M. 'Sight-unseen' detection of rare aquatic species using environmental DNA. *Conserv. Lett.* **4**, 150–157 (2011).
23. Shokralla, S., Spall, J. L., Gibson, J. F. & Hajibabaei, M. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* **21**, 1794–1805 (2012).
24. Stewart, K., Ma, H., Zheng, J. & Zhao, J. Using environmental DNA to assess population-wide spatiotemporal reserve use. *Conserv. Biol.* **31**, 1173–1182 (2017).
25. Taberlet, P., Bonin, A., Coissac, E. & Zinger, L. *Environmental DNA: For biodiversity research and monitoring.* (Oxford University Press, Oxford, 2018).
26. Bohmann, K. *et al.* Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* **29**, 358–367 (2014).
27. Zimmerman, N., Izard, J., Klatt, C., Zhou, J. & Aronson, E. The unseen world: environmental microbial sequencing and identification methods for ecologists. *Front. Ecol. Environ.* **12**, 224–231 (2014).

28. Thomsen, P. F. & Willerslev, E. Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Cons.* **183**, 4–18 (2015).
29. Deiner, K. *et al.* Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).
30. Craine, J. M. *et al.* Molecular analysis of environmental plant DNA in house dust across the United States. *Aerobiologia* **33**, 71–86 (2017).
31. Olson, D. H. *et al.* Mapping the global emergence of *Batrachochytrium dendrobatidis*, the amphibian chytrid fungus. *PLoS ONE* **8**, e56802 (2013).
32. Ishige, T. *et al.* Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biol. Cons.* **210**, 281–285 (2017).
33. Ushio, M. *et al.* Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol. Ecol. Resour.* **17**, e63–e75 (2017).
34. Richards, C. L. *et al.* Detection of pathogenic and non-pathogenic bacteria in drinking water and associated biofilms on the crow reservation, Montana, USA. *Microb. Ecol.* **76**, 52–63 (2018).
35. Simpfendorfer, C. A. *et al.* Environmental DNA detects critically endangered largetooth sawfish in the wild. *Endang. Spec. Res.* **30**, 109–116 (2016).
36. Weltz, K. *et al.* Application of environmental DNA to detect an endangered marine skate species in the wild. *PLoS ONE* **12**, e0178124 (2017).
37. Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K. & Garcia-Vazquez, E. DNA in a bottle—rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS ONE* **12**, e0183347 (2017).
38. Valentin, R. E., Fonseca, D. M., Nielsen, A. L., Leskey, T. C. & Lockwood, J. L. Early detection of invasive exotic insect infestations using eDNA from crop surfaces. *Front. Ecol. Environ.* **16**, 265–270 (2018).
39. Larson, E. R. *et al.* From eDNA to citizen science: emerging tools for the early detection of invasive species. *Front. Ecol. Environ.* **18**, 194–202 (2020).
40. Weiskopf, S. R., Kachel, S. M. & McCarthy, K. P. What are snow leopards really eating? Identifying bias in food-habit studies. *Wildl. Soc. Bull.* **40**, 233–240 (2016).
41. Berry, T. E. *et al.* DNA metabarcoding for diet analysis and biodiversity: a case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecol. Evol.* **7**, 5435–5453 (2017).
42. Sønstebø, J. *et al.* Using next-generation sequencing for molecular reconstruction of past arctic vegetation and climate. *Mol. Ecol. Resour.* **10**, 1009–1018 (2010).
43. Vamosi, J. C., Gong, Y.-B., Adamowicz, S. J. & Packer, L. Forecasting pollination declines through DNA barcoding: the potential contributions of macroecological and macroevolutionary scales of inquiry. *New Phytol.* **214**, 11–18 (2017).
44. Evans, D. M. & Kitson, J. J. Molecular ecology as a tool for understanding pollination and other plant–insect interactions. *Curr. Opin. Insect Sci.* **38**, 26–33 (2020).
45. Radosevich, J., Wilson, W., Shinn, J., DeSantis, T. & Andersen, G. L. Development of a high-volume aerosol collection system for the identification of air-borne micro-organisms. *Lett. Appl. Microbiol.* **34**, 162–167 (2002).
46. Hurley, K. V. *et al.* Car cabin filters as sampling devices to study bioaerosols using eDNA and microbiological methods. *Aerobiologia* **35**, 215–225 (2019).
47. Calderón-Franco, D., Loosdrecht, M. van, Abeel, T. & Weissbrodt, D. G. A novel method to isolate free-floating extracellular DNA from wastewater for quantitation and metagenomic profiling of mobile genetic elements and antibiotic resistance genes. *bioRxiv* (2020).
48. Yang, J. & Zhang, X. eDNA metabarcoding in zooplankton improves the ecological status assessment of aquatic ecosystems. *Environ. Int.* **134**, 105230 (2020).
49. Barnes, M. A. *et al.* Environmental conditions influence eDNA persistence in aquatic systems. *Environ. Sci. Technol.* **48**, 1819–1827 (2014).
50. Harrison, J. B., Sunday, J. M. & Rogers, S. M. Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proc. R. Soc.* **B286**, 20191409 (2019).
51. Sirois, S. H. & Buckley, D. H. Factors governing extracellular DNA degradation dynamics in soil. *Environ. Microbiol. Rep.* **11**, 173–184 (2019).
52. Després, V. *et al.* Primary biological aerosol particles in the atmosphere: a review. *Tellus B Chem. Phys. Meteorol.* **64**, 15598 (2012).
53. Cáliz, J., Triadó-Margarit, X., Camarero, L. & Casamayor, E. O. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc. Natl. Acad. Sci.* **115**, 12229–12234 (2018).
54. Walker, D. M. *et al.* Methodological considerations for detection of terrestrial small-body salamander eDNA and implications for biodiversity conservation. *Mol. Ecol. Resour.* **17**, 1223–1230 (2017).
55. Williams, K. E., Huyvaert, K. P., Vercauteren, K. C., Davis, A. J. & Piaggio, A. J. Detection and persistence of environmental DNA from an invasive, terrestrial mammal. *Ecol. Evol.* **8**, 688–695 (2018).
56. Katz, A. D. *et al.* Environmental DNA is effective in detecting the federally threatened Louisiana pinesnake (*Pituophis ruthveni*). *Environmental DNA*.
57. Valentin, R. E. *et al.* Moving eDNA surveys onto land: Strategies for active eDNA aggregation to detect invasive forest insects. *Mol. Ecol. Resour.* **20**, (2020).
58. Johnson, M. D., Cox, R. D. & Barnes, M. A. Analyzing airborne environmental DNA: a comparison of extraction methods, primer type, and trap type on the ability to detect airborne eDNA from terrestrial plant communities. *Environ. DNA* **1**, 176–185 (2019).
59. Johnson, M. D., Cox, R. D. & Barnes, M. A. The detection of a non-anemophilous plant species using airborne eDNA. *PLoS ONE* **14**, e0225262 (2019).
60. Teachey, M. E., Pound, P., Ottesen, E. A. & Van Stan, J. T. Bacterial community composition of throughfall and stemflow. *Front. For. Glob. Change* **1**, 7 (2018).
61. Guerrieri, R. *et al.* Partitioning between atmospheric deposition and canopy microbial nitrification into throughfall nitrate fluxes in a mediterranean forest. *J. Ecol.* **108**, 626–640 (2020).
62. Levia, D. F., Keim, R. F., Carlyle-Moses, D. E. & Frost, E. E. Throughfall and stemflow in wooded ecosystems. in *Forest hydrology and biogeochemistry* 425–443 (Springer, 2011).
63. Levia, D. F., Hudson, S. A., Llorens, P. & Nanko, K. Throughfall drop size distributions: a review and prospectus for future research. *Wiley Interdiscip. Rev. Water* **4**, e1225 (2017).
64. Levia, D. F. Jr. & Frost, E. E. A review and evaluation of stemflow literature in the hydrologic and biogeochemical cycles of forested and agricultural ecosystems. *J. Hydrol.* **274**, 1–29 (2003).
65. Whitford, W. G., Anderson, J. & Rice, P. M. Stemflow contribution to the 'fertile island' effect in creosotebush, *Larrea tridentata*. *J. Arid Environ.* **35**, 451–457 (1997).
66. Levia, D. F. Jr. & Frost, E. E. Variability of throughfall volume and solute inputs in wooded ecosystems. *Prog. Phys. Geogr.* **30**, 605–632 (2006).

67. Levia, D. F. & Germer, S. A review of stemflow generation dynamics and stemflow-environment interactions in forests and shrublands. *Rev. Geophys.* **53**, 673–714 (2015).
68. Allard, S. M., Ottesen, A. R. & Micallef, S. A. Rain induces temporary shifts in epiphytic bacterial communities of cucumber and tomato fruit. *Sci. Rep.* **10**, 1–10 (2020).
69. Morris, C. E. *et al.* Bioprecipitation: a feedback cycle linking earth history, ecosystem dynamics and land use through biological ice nucleators in the atmosphere. *Glob. Change Biol.* **20**, 341–351 (2014).
70. Sands, D., DC, S., AL, S. & SMET, D. The association between bacteria and rain and possible resultant meteorological implications (1982).
71. Fürnkranz, M. *et al.* Nitrogen fixation by phyllosphere bacteria associated with higher plants and their colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *ISME J.* **2**, 561–570 (2008).
72. Ladin, Z. S., D'Amico, V., Baetens, J. M., Roth, R. R. & Shriver, W. G. Long-term dynamics in local host–parasite interactions linked to regional population trends. *Ecosphere* **7**, e01420 (2016).
73. Lear, G. *et al.* Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *N. Z. J. Ecol.* **42**, 10-50A (2018).
74. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
75. Walters, W. *et al.* Improved bacterial 16S rRNA gene (v4 and v4–5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *Msystems* **1**, e00009-15 (2016).
76. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
77. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one fastq preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
78. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
79. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
80. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, London, 2020).
81. Edgar, R. C. SINTAX: A simple non-bayesian taxonomy classifier for 16S its sequences. *BioRxiv* 074161 (2016).
82. Edgar, R. C. Accuracy of taxonomy prediction for 16S rRNA and fungal its sequences. *PeerJ* **6**, e4652 (2018).
83. Edgar, R. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* **6**, e5030 (2018).
84. Nearing, J. T., Douglas, G. M., Comeau, A. M. & Langille, M. G. Denoising the denoisers: An independent evaluation of microbiome sequence error-correction approaches. *PeerJ* **6**, e5364 (2018).
85. Dunn, O. J. Multiple comparisons using rank sums. *Technometrics* **6**, 241–252 (1964).
86. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
87. Calle, M. L. Statistical analysis of metagenomics data. *Genom. Inform.* **17**, (2019).
88. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A. & Pawlowsky-Glahn, V. Logratio analysis and compositional distance. *Math. Geol.* **32**, 271–275 (2000).
89. van den Boogaart, K. G. & Tolosana-Delgado, R. 'Compositions': a unified R package to analyze compositional data. *Comput. Geosci.* **34**, 320–338 (2008).
90. Martín-Fernández, J.-A., Hron, K., Templ, M., Filzmoser, P. & Palarea-Albaladejo, J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Model.* **15**, 134–158 (2015).
91. Palarea-Albaladejo, J. & Martín-Fernández, J. A. Z, Compositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems* **143**, 85–96 (2015).
92. Willis, A. D. & Martin, B. D. DivNet: Estimating diversity in networked communities. *BioRxiv* 305045 (2018).
93. Anderson, M. J. *et al.* Navigating the multiple meanings of diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
94. Willis, A. D. Rarefaction, alpha diversity, and statistics. *Front. Microbiol.* **10**, 2407 (2019).
95. Quinn, T. P. *et al.* A field guide for the compositional analysis of any-omics data. *GigaScience* **8**, giz107 (2019).
96. Willis, A. & Bunge, J. Estimating diversity via frequency ratios. *Biometrics* **71**, 1042–1049 (2015).
97. Willis, A., Martin, B. D., Trinh, P., Barger, K. & Bunge, J. *Breakaway: Species richness estimation and modeling* (2020).
98. Foster, Z., Sharpton, T. & Grünwald, N. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput. Biol.* **13**, 1–15 (2017).
99. Sales, N. G. *et al.* Assessing the potential of environmental DNA metabarcoding for monitoring neotropical mammals: A case study in the Amazon and Atlantic forest, Brazil. *Mammal Review* (2020).
100. Nguyen, B. N. *et al.* Environmental DNA survey captures patterns of fish and invertebrate diversity across a tropical seascape. *Sci. Rep.* **10**, 1–14 (2020).
101. Waters, S. M., Purdue, S., Armstrong, R. & Detrés, Y. Metagenomic investigation of African dust events in the caribbean. *FEMS Microbiol. Lett.* **367**, fnaa051 (2020).
102. Zinger, L. *et al.* Advances and prospects of environmental DNA in neotropical rainforests. *Adv. Ecol. Res.* **62**, 331–373 (2020).
103. Bittar, T. B., Pound, P., Whitetree, A., Moore, L. D. & Van Stan, J. T. Estimation of throughfall and stemflow bacterial flux in a subtropical oak-cedar forest. *Geophys. Res. Lett.* **45**, 1410–1418 (2018).
104. Redford, A. J., Bowers, R. M., Knight, R., Linhart, Y. & Fierer, N. The ecology of the phyllosphere: Geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ. Microbiol.* **12**, 2885–2893 (2010).
105. Vorholt, J. A. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **10**, 828–840 (2012).
106. Koren, I. *et al.* The bodélé depression: a single spot in the Sahara that provides most of the mineral dust to the Amazon forest. *Environ. Res. Lett.* **1**, 014005 (2006).
107. Artaxo, P., Maenhaut, W., Storms, H. & Van Grieken, R. Aerosol characteristics and sources for the amazon basin during the wet season. *J. Geophys. Res. Atmos.* **95**, 16971–16985 (1990).
108. Swap, R., Garstang, M., Greco, S., Talbot, R. & Kållberg, P. Saharan dust in the Amazon basin. *Tellus B* **44**, 133–149 (1992).
109. Cayuela, C., Levia, D. F., Latron, J. & Llorens, P. Particulate matter fluxes in a mediterranean mountain forest: interspecific differences between throughfall and stemflow in oak and pine stands. *J. Geophys. Res. Atmos.* **124**, 5106–5116 (2019).
110. Cohen, J. M. Are we killing the rain? Meditations on the water cycle and more particularly, on bioprecipitation. *Water Int.* **37**, 670–682 (2012).
111. Fröhlich-Nowoisky, J. *et al.* Bioaerosols in the earth system: climate, health, and ecosystem interactions. *Atmos. Res.* **182**, 346–376 (2016).
112. Hanlon, R. *et al.* Microbial ice nucleators scavenged from the atmosphere during simulated rain events. *Atmos. Environ.* **163**, 182–189 (2017).
113. Kellogg, C. A. *et al.* Characterization of aerosolized bacteria and fungi from desert dust events in Mali, west Africa. *Aerobiologia* **20**, 99–110 (2004).
114. Pandey, R. *et al.* Ice-nucleating bacteria control the order and dynamics of interfacial water. *Sci. Adv.* **2**, e1501630 (2016).

115. Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **17**, e00547 (2019).
116. Adams, C. I. *et al.* Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool?. *Genes* **10**, 192 (2019).
117. Valentin, R., Kyle, K., Allen, M., Welbourne, D. & Lockwood, J. The ecology of aboveground terrestrial eDNA: Its state, transport, and fate on aboveground surfaces. *Authorea Preprints* (2020).
118. Tamaki, M. *et al.* Study on sampling method of rainfall, throughfall, and stemflow to monitor the effect of acid deposition on forest ecosystem. In *Acid rain 2000* 1511–1516 (Springer, 2001).
119. Hospodsky, D., Yamamoto, N. & Peccia, J. Accuracy, precision, and method detection limits of quantitative PCR for airborne bacteria and fungi. *Appl. Environ. Microbiol.* **76**, 7004–7012 (2010).
120. Young, J. M., Weyrich, L. S. & Cooper, A. Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. *Forens. Sci. Int. Genet.* **13**, 176–184 (2014).
121. Hugerth, L. W. *et al.* Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS ONE* **9**, e95567 (2014).
122. Valentin, R. E., Maslo, B., Lockwood, J. L., Pote, J. & Fonseca, D. M. Real-time PCR assay to detect brown marmorated stink bug, *Halyomorpha halys* (stål), in environmental DNA. *Pest Manag. Sci.* **72**, 1854–1861 (2016).
123. Burns, T. J. *et al.* Environmental DNA sampling in a terrestrial environment: Methods to detect a critically endangered frog and a global pathogen. *bioRxiv* (2020).
124. van Beeck Calkoen, S. T. *et al.* The blame game: Using eDNA to identify species-specific tree browsing by red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) in a temperate forest. *For. Ecol. Manag.* **451**, 117483 (2019).
125. Villarreal, C., Padilla-Longoria, P. & Alvarez-Buylla, E. R. General theory of genotype to phenotype mapping: derivation of epigenetic landscapes from n-node complex gene regulatory networks. *Phys. Rev. Lett.* **109**, 118102 (2012).
126. Morris, A., Meyer, K. & Bohannan, B. Linking microbial communities to ecosystem functions: What we can learn from genotype–phenotype mapping in organisms. *Philos. Trans. R. Soc. B* **375**, 20190244 (2020).

Acknowledgements

Author contributions

Z.S.L. led the conceptualization, methodological design, data collection, analysis, and writing of the manuscript. B.F. and J.D. helped conduct laboratory analyses, R.M. contributed to data analysis and interpretation. D.L. helped review and revise the manuscript, and gave input on methodological design. W.G.S., V.D. and T.T. reviewed and revised the manuscript. J.C.S. and E.W. provided critical input into the design of the study, and reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material availlable at https://doi.org/10.1038/s41598-020-80602-9.

Correspondence and requests for materials should be addressed to Z.S.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.