Biotic and abiotic factors predict the biogeography of soil microbes in the Serengeti

Bo Maxwell Stevens[1], Derek Lee Sonderegger[2], and Nancy Collins Johnson[1,3]

[1]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA

[2]Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86001, USA

[3]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA

Corresponding Author

Bo Maxwell Stevens (bo_stevens@nau.edu)

Keywords

Soil bacteria, community ecology, grazing, Serengeti National Park, soil texture, phosphorus

Abstract

Field-based observational research is the first step in understanding the factors that predict the biogeography and community structure of soil microbes. The Serengeti National Park in Tanzania is an ideal location for this type of research because active volcanoes generate strong environmental gradients due to ash deposition and a rain shadow. Also, as one of the last remaining naturally grazed ecosystems on Earth, the Serengeti provides insights about the influence of herbivory on microbial communities. We used 16S rRNA amplicons to characterize bacterial and archaeal communities in soils from a 13-year herbivore removal experiment to study the influence of environmental factors and grazing on the natural distribution of soil microbes. We collected soil samples from seven sites, each with three naturally grazed plots and three plots that were fenced to prevent grazing by large mammalian

herbivores. Soil fertility (phosphorus, nitrogen, iron, calcium, organic matter), texture, and pH were measured at each plot. Beta diversity of bacterial and archaeal communities was most strongly correlated with soil texture (R^2 = 32.4%). The abundance of many operational taxonomic units (OTUs) were correlated with soil texture, and the evenness of taxa within samples increased with fine-textured soil. Removal of grazing shifted community structure, with 31 OTUs that were significant indicator taxa of the ungrazed treatment and three OTUs that were significant indicators of the grazed treatment.

Importance

Our results show that in this regional scale study, soil texture was the best environmental predictor, and grazing by large mammals also structures bacterial and archaeal communities. When large mammals are removed, as humans have been doing for millenia, there are cascading effects into the microbial world that can influence ecosystem functions like carbon and nitrogen cycles. These empirical findings from a natural tropical savannah can help inform models of the global distribution and function of soil microbes.

Introduction

Soil bacteria and archaea serve critical functions in natural ecosystems. Many environmental factors can be used to predict the distribution and diversity of communities of bacteria within the soil (Fierer, 2017). Although recent studies have elucidated patterns in the structure of bacterial communities, disentangling their biogeography by determining the underlying mechanisms affecting the distribution of bacteria and archaea (hereafter referred to as microbes) remains a challenge. The Serengeti National Park in Tanzania provides an ideal environmental gradient for investigating the effects of abiotic and biotic factors on microbial communities. Four mechanisms underlying the biogeography of soil microbes have been identified: selection, dispersal, drift, mutation (Hanson *et al.*, 2012). The purpose of this study is

to consider how environmental selection and dispersal by large animals influences the biogeography of soil microbes in the Serengeti. To accomplish this, we examined the diversity and distribution of microbial communities across a natural gradient of abiotic conditions that is superimposed on a replicated experiment that removed large mammalian herbivores for 13 years.

In the Serengeti, local topography and volcanic inputs have created inverse gradients of soil properties and precipitation (Anderson & Talbot, 1965; Ashley *et al.*, 2014). Active volcanoes to the east continue to influence the geologic context with eruptions as recent as 2008 (Sinclair & Arcese, 1995; Vaughan *et al.*, 2008). Ash deposits from the nearby Ngorongoro Volcanic Highlands create gradients of calcium, iron, phosphorus and pH (Anderson & Talbot, 1965; Ruess & Seagle, 1994; Ashley *et al.*, 2014). Many studies have found that pH is a significant driver of microbial community composition globally (Fierer & Jackson, 2006; Lauber *et al.*, 2009; Griffiths *et al.*, 2011; Kaiser *et al.*, 2016). Phosphorus availability is linked to soil pH, calcium, and iron because of chemical interactions on soil particles. Phosphorus availability is highest in neutral pH ranges. Calcium immobilizes phosphorus above neutral pH ranges (Schachtman *et al.*, 1998; Brady & Weil, 2017), and iron can immobilize phosphorus below a neutral pH, indirectly affecting the growth of microbes and plants (Vance *et al.*, 2003; Pigna & Violante, 2003). Increasing phosphorus availability can influence microbial communities by altering assemblages or increasing diversity (Beauregard *et al.*, 2009; Kuramae *et al.*, 2011, 2012; Tan *et al.*, 2012). Abiotic soil characteristics directly affect microbial communities by influencing resource availability and habitat structure, and they indirectly affect bacterial communities by influencing biotic factors such as vegetation and food webs.

Soil microbes exist within a complex web of organisms that includes large grazing animals with many potential mechanisms for affecting microbial communities. The Serengeti hosts one of the largest natural grazing ecosystems in the world with currently over two million large mammalian herbivores (Eby *et al.*, 2014). Moderate levels of grazing can alter soil nutrient

concentrations by transporting nutrients and modifying plant communities (Anderson *et al*. 2007). In ecosystems with high soil fertility where nutrient cycling is dominated by bacteria, grazing may have a positive effect on decomposer bacteria (Bardgett & Wardle, 2003). In addition to the direct effects of urine and dung on soil biota (McNaughton *et al.*, 1997; Bardgett *et al.*, 1998), grazing can indirectly influence microbial diversity by stimulating net primary productivity and aboveground biomass (McNaughton, 1979, 1985), leading to increased photosynthesis and root exudation (Bardgett *et al.*, 1998). Also, grazing may influence the composition of plant communities which may alter the composition of microbial communities (Prober *et al.*, 2015). Comparison of soil microbes between grazed and ungrazed treatments in a long-term field experiment will help elucidate shifts in bacterial communities resulting from the cascade of biotic and abiotic responses to the activities of large mammalian herbivores.

Although no single factor consistently explains the biogeography of soil microbes, certain environmental variables tend to correspond with spatial patterns in the composition and abundance of bacteria. Soil organic matter, pH, redox status, soil moisture, nitrogen and phosphorus availability, and soil texture appear to be important predictors of the structure of soil microbial communities (Fierer, 2017). Our study tests the hypothesis that selection by many of these abiotic factors structures microbial communities. Also, we compared microbial communities inside and outside grazing exclosures to test the hypothesis that seasonal defoliation by large mammalian herbivores structure microbial communities. The goal of this study is to increase our understanding of the biogeography of soil microbes in a naturally grazed, tropical grassland. This knowledge can inform the development of general principles to help predict the structure and function of soil bacterial communities in changing environments.

Methods

Sites

In 1999, a herbivore removal experiment was installed at eight sites within the Serengeti National Park, Tanzania (Fig. 1) (Anderson *et al.*, 2007). Six plots (4 x 4 m) were established at each site with three randomly assigned plots open to grazing and the other three plots had 2 m tall chain-linked fences to exclude grazing by large animals including wildebeest, zebras, Thomson's gazelles, buffalo, and topi (McNaughton, 1985; Anderson *et al.*, 2007). Theft of all the fences at one site in the eastern corridor prevented its inclusion in this study. Soil samples were collected from the remaining seven sites during the rainy season in May, 2012. Precipitation in the Serengeti National Park in 2012 was lowest in the southern (662 mm) and highest in the northern sites (1143 mm; Table 1).

Soil analyses

Soil samples were collected from holes (approximately 15 cm deep) created by excavating grasses at all seven sites (n = 42). Within six hours of collection, soils were dried for 48 hours in a solar drier. After 2 weeks, the dry samples were brought to the laboratory and frozen for long-term storage. Frozen soil samples were dried at 103°C and sieved (< 2 mm). Soil organic matter was measured using loss on ignition, 2 g subsamples were weighed, heated to 550°C for 24 hours in a Lindberg HB muffle furnace (Lindberg/MPH, Riverside, MI 49084) then reweighed (Heiri *et al.*, 2001). Soil pH was measured potentiometrically in a 1:2.5 water:soil paste at the Soil Science Laboratory of Sokoine University of Agriculture in Morogoro, Tanzania (Klute, 1986). To measure total phosphorus, calcium, and iron concentrations, 0.3 g subsamples were ground and digested in 7 mL concentrated nitric acid and 3 mL 30% hydrogen peroxide in Milestone 900 Microwave Digestor (Ethos Inc., Bristol, United Kingdom). Samples were digested for 20 minutes and reached a maximum temperature of 425°C. Total soil

phosphorus concentration converted to orthophosphate was quantified via colorimetry (Grimshaw, 1987) on a QuikChem 8000 Series FIA+ (Lachat Instruments, Milwaukee, WI 53218) using QuikChem Method 10-115-01-1-A. Total iron and calcium were measured on an AAnalyst 100 Atomic Absorption Spectrophotometer (Perkin Elmer, Waltham, MA 02451). Samples were compared to in-house standards and external standards produced by Ricca Chemical Company (Arlington, TX 76012) and Hach Company (Loveland, CO 80539).

Soil texture was determined using laser diffraction particle size analysis (Beuselinck *et al.*, 1998). Unsieved soil samples were suspended in water and analyzed on a LS 13 320 Series Laser Diffraction Particle Size Analyzer (Beckman Coulter Brea CA 92821). Particle sizes were grouped according USDA soil texture classification. Our soil textural classification figure was generated from R code available online (Hamilton, 2014). Soil bulk density and total soil nitrogen concentrations were obtained from previous analyses from the same plots (Antoninka *et al.*, 2015). Measurements of soil organic matter, total phosphorus, nitrogen, calcium, and iron were adjusted by bulk density.

Molecular analysis

Amplicons were produced in a two-step protocol (Berry *et al.*, 2011). Samples were amplified in triplicate PCR reactions for the 16S v4 region using the universal bacterial primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (Bates *et al.*, 2011). First round reactions were performed in triplicate in 384 well plates. The 8 µL volumes contained the following: 1 µM each primer, 200 µM each dNTP (Phenix Research, Candler, NC), 0.01 U/µL Phusion HotStart II DNA Polymerase (Life Technologies), 1X HF Phusion Buffer (Life Technologies), 3 mM $MgCl_2$, 6% glycerol, and 1 µL normalized template DNA. Cycling conditions were: 2 minutes at 95°C followed by 20 cycles of 30 seconds at 95°C, 30 seconds at 55°C, 4 minutes at 60°C. Triplicate reactions for each sample were pooled by combining 4 µL from each well, and 2 µL was used to check for results on an agarose gel. The

remainder was diluted 10-fold and used as template in a second PCR reaction in which indexed tails (Caporaso *et al.*, 2012) were added. Second round reaction conditions were identical to the first round except only one reaction was conducted per sample and only 15 total cycles were performed. Indexed PCR products were purified using carboxylated magnetic beads (Rohland & Reich, 2012), quantified by PicoGreen fluorescence, and an equal mass of each sample was combined into a final sample pool. The pool was purified and concentrated, and subsequently quantified by quantitative PCR against Illumina DNA Standards (Kapa Biosystems, Wilmington, MA). Sequencing was carried out on a MiSeq Desktop Sequencer (Illumina Inc, San Diego, CA) running in paired end 2x150 mode. Upon acceptance, sequence will be archived in the NCBI Sequence Read Archive.

Data analysis

The forward reads of the 16S amplicons were imported into QIIME 2 version 2018.11 (Bolyen *et al.*, 2018). Demultiplexing was carried out using minimum quality threshold of q20 and default parameters in QIIME 2. Based on quality threshold, visualized with FastQC version 11.7, reads were trimmed to 139 bases (Andrews, 2010). To determine phylogenetic diversity metrics, a rooted phylogenetic tree was created with MAFFT sequence alignment and FastTree in QIIME 2. QIIME 1.9.1 was used to filter samples below 0.005% abundance (Caporaso *et al.*, 2010; Bokulich *et al.*, 2013). To remove singletons by sample, the `otu_picking_workflow.sh` command in akutils v1.1.1 was performed (Andrews, 2018). Alpha and beta diversity metrics were performed with the q2-diversity plugin for QIIME 2 using the core-metrics-phylogenetic command with a sampling depth of 45000. We used Shannon diversity index to capture bacterial richness and evenness. To estimate alpha diversity with phylogenetic structure, we used Faith's Phylogenetic Diversity index (Faith, 1992). To separate community evenness, we used Pielou's evenness index, where values are constrained to 0 and 1, with higher values representing even abundance of community members. To generate figures, we used the scikit-

bio 0.2.3 (http://scikit-bio.org), matplotlib 3.1.0, and seaborn 0.9.0 python packages. Upon acceptance, our environmental data, correlation matrix, and OTU table will be made publically available on Dryad Digital Repository (https://datadryad.org).

Statistical analyses

In an attempt to account for spatial autocorrelation, beta diversity of bacterial communities were analyzed with single regressions (Anderson *et al.*, 2011) of Bayesian general linear mixed effects models using the "rstanarm" R package version 2.17.4 (Stan Development Team, 2017). Weighted and unweighted UniFrac (Lozupone & Knight, 2005), and Bray-Curtis dissimilarity for all unique non-zero pairs of plots was the response variable (n = 861 for each metric). Standardized values for distance between continuous environmental variables were used to predict beta diversity. Grazing treatment categories were created for all possible combinations of treatments (e.g. beta diversity of grazed vs grazed, grazed vs ungrazed, and ungrazed vs ungrazed samples). We included a random effect for spatial autocorrelation that represented all unique alphabetized combinations of sites. Additionally, we used a random slope for each model.

Eleven predictor variables were analyzed for all models; grazing treatment, soil organic matter (SOM), pH, rainfall, percent sand, silt, and clay, and total concentrations of nitrogen, phosphorus, calcium, and iron. Rainfall for 2012 was as determined by satellite measurements from NASA's Global Precipitation Measurement mission (Hou *et al.*, 2013). Means and a 95% credible interval for the posterior distribution are reported in Table 2 and Fig. 7. To estimate variation explained by each model, the 'bayes_R2' function in 'rstanarm' was used to calculate an r-squared (Stan Development Team, 2017). To determine the variation explained by the fixed effect of each model, the full model r-squared (R^2 Full) was partitioned into an r-squared for the fixed (R^2 Fixed) and random effect (R^2 Random) by calculating the sum squared error for the full model and for a null model with random effects only. All models used three chains with

default parameters (family = gaussian, prior = normal, iterations = 2000, adapt = 0.99), and all models converged (Rhat < 1.05).

To summarize overall patterns between bacterial community composition and abiotic factors, we used principal coordinate analysis (PCoA) and the Spearman's rank correlation coefficient in the BIOENV function from scikit-bio. To test the effect of abiotic characteristics on bacterial community composition, we used a distance-based redundancy analysis (db-RDA) (Legendre & Anderson, 1999) performed with forward and backward model selection using the 'capscale' and 'ordistep' functions in the vegan package version 2.5-4 (Oksanen *et al.*, 2018). We performed the db-RDA analysis on Bray-Curtis dissimilarity with default parameters. Percent sand was removed from the selected db-RDA model post hoc because it did not contribute to the overall model. Indicator species analyses of OTUs (operational taxonomic units) were used to determine which bacteria were associated with grazing treatments and sites using the 'multipatt' function in the 'indicspecies' package version 1.7.6 (De Cáceres & Legendre, 2009) package in R (version 3.3.0) with default values. Only OTUs with significant indicator species ($P < 0.05$) are reported. P-values were determined with 999 permutations. The square root of the indicator value (func = "IndVal.g") was used as the test statistic.

Results

Soil nutrient analyses indicated significant edaphic gradients created by the Ngorongoro Volcanic Highlands (Fig. 1; Table 1). As expected, soil throughout the southern Serengeti plains is enriched by volcanic deposits rich in phosphorus, iron, and calcium, with deposits gradually decreasing with increasing latitude. Soil phosphorus concentration was significantly higher in the southern site (6.1 mg cm^{-3}) than the northern site (0.1 mg cm^{-3}; Table 1). Total soil nitrogen concentration was between 1.1 and 2.3 (mg cm^{-3}) throughout the seven sites (Table 1). Soil organic matter ranged from 6.9% (BAL) to 16.2% (BRS; Table 1). Total calcium and iron concentrations were lowest in the northern site (1.8 and 4.5 mg cm^{-3}, respectively) and highest

in the southern site (20.2 and 24.0 mg cm^{-3}, respectively; Table 1). Sites in the north were comprised of sandy loam soil with low phosphorus concentration while sites in the south were silty loam with high phosphorus concentration (Fig. 1; Fig. 2; Table 1). Because of the influence of volcanic ash, many of the environmental variables were highly correlated (Fig. S1). Notably, percent sand, silt, and clay were more than 90% correlated with each other and between 70 and 90% correlated with phosphorus concentration (Fig. S1).

After read quality filtering, sequencing resulted in a total of 5,702,184 reads matching a total of 32,372 OTUs. To remove potentially erroneous OTUs (operational taxonomic units), we used a stringent OTU table filtering threshold that removes singletons by sample and OTUs below 0.005% of total sequence abundance (Bokulich *et al.*, 2013). OTU table quality filtering resulted in 4,431,104 (78.7%) sequences matching 2,782 OTUs (8.6%). For all 42 samples, we had an average of 105,502 sequences per sample, with a minimum of 52,139 and maximum of 162,093 sequences.

Microbial communities in this study were primarily dominated by Actinobacteria (19.5%), Proteobacteria (19.5%), and Acidobacteria (19.2%). Other significant (> 5%) phyla included Verrucomicrobia (8.0%), Firmicutes (6.0%), Chloroflexi (6.0%), and Planctomycetes (5.5%). Phyla with 5% or less relative abundance comprised 16.2% of the overall abundance. Average relative abundances for each site are reported in Fig. 3a. We used four different alpha diversity metrics to determine if environmental variables are related to richness and evenness (Fig. S2). We highlight the strong correlation between percent silt and evenness in Fig. 3c. Average Shannon diversity for each site ranged from 8.8 (± 0.2) in BAL to 9.3 (± 0.1) in BRS and was correlated positively with percent silt (R^2 = 0.17) and total phosphorus concentration (R^2 = 0.14). Overall, evenness was high throughout our study, ranging from 0.86 (± 0.01) to 0.91 (± 0.01). Evenness of communities was positively correlated with silt (R^2 = 0.16), phosphorus (R^2 = 0.25), and pH (R^2 = 0.11), and negatively correlated with rainfall (R^2 = 0.33). Average Faith PD, a phylogenetic diversity estimate, was lowest in BAL (61.2 ± 14.3) and highest in MSB (82.1 ±

6.0) and was not highly correlated with environmental variables. Similarly, richness as measured by the number of observed OTUs, was lowest in BAL (859 ± 242.2) and highest in MSB (1253 ± 156.2) and was not correlated with environmental variables. Relative abundances of many of the 50 most common OTUs were strongly correlated with one or more environmental variables, mainly soil texture (Fig 4; Table S2). Of the 12 strongest correlations, nine had positive and three had negative relationships with percent silt (Fig. S3; Table S2). Indicator species analysis of the grazing treatments revealed that all indicator OTUs were in the bacterial kingdom (Table S3), three OTUs were indicators of grazed plots and 31 were indicators of ungrazed plots (Figure 3b; Table S3).

To estimate the importance of environmental variables on the beta diversity of bacterial communities, we used spearman rank correlation coefficients with distance matrices for three different beta diversity metrics. Phylogenetic (UniFrac) and non-phylogenetic (Bray-Curtis) metrics were similar across four highlighted environmental variables, phosphorus concentration, rainfall, percent silt, and pH (Fig. S4). Despite high correlations between the measurements (Fig. S1) of phosphorus with silt ($r = 0.71$) and pH ($r = 0.75$), spearman correlations were much higher for percent silt (30.5 to 40.6%; Fig. 3d; Fig S4), and low for pH (between -7 and -2.9%; Fig. S4). Phosphorus concentration and rainfall are highly correlated with each other ($r = -0.83$; Fig. S1), and spearman correlations were similar for both variables and all three beta diversity metrics (Fig. S4). Much of the dissimilarity of bacterial communities can be attributed to the presence or absence of OTUs, indicated by the similarities of unweighted UniFrac correlations and Bray-Curtis (Fig. S4). Slightly higher correlations for the first two PCoA axes of weighted UniFrac, when compared to unweighted UniFrac, indicates that some variation in the bacterial communities can be explained by the phylogenetic structure of abundant OTUs (Fig. S4). In addition, we used a dbRDA to understand the effect of environmental variables on Bray Curtis distances of bacterial communities. Model selection of dbRDA indicated that soil texture, rainfall, and phosphorus and iron concentration were important in explaining variation in bacterial

community composition ($F = 2.6$, $p < 0.001$; Fig. S5). Overall, the dbRDA explained 16% of the variation in community composition (adjusted $R^2 = 0.164$).

To gain a deeper understanding of the relationship of the effects of grazing and abiotic characteristics on beta diversity, we used Bayesian linear mixed effects models. To avoid the multicollinearity between environmental variables, we compared single linear regressions. We used the leave-one-out information criterion to compare models. Similar to Akaike's information criterion, lower values indicate better model fit. Model results are reported in Table S4 and visually represented in Fig. S6. Soil texture and phosphorus provided the best fit for determining the structure of bacterial communities. Results for both weighted and unweighted UniFrac were remarkably similar to those of Bray-Curtis. Percent clay, silt, and sand were three of the top four models. Total phosphorus concentration ranked third on our list of models, the 95% credible interval for phosphorus concentration in relation to beta diversity slightly overlapped zero (CI = [0.00, 0.53]), however, the mean of the posterior distribution (0.27) was the largest and was consistent with an effect, which may indicate that a large portion of the bacterial community structure is determined by levels of this important nutrient. Beta diversity values (y) range between 0.19 and 0.85, and standardized coefficients (x) ranged from 0 to approximately 3. Therefore, a mean posterior distribution for a coefficient with a value of 0.27 (phosphorus) would indicate that for every 1 unit of change in the standardized x-value, beta diversity increases by 0.27, a large difference considering the range of possible values. Linear mixed models with other environmental variables produced fixed effect correlations near zero. Model results for the grazing treatment (CI = [-0.07, -0.02]) are consistent with an effect of grazing on the beta diversity of bacteria (2% of variation explained).

Discussion

Because the Serengeti is a relatively undisturbed grassland, we can assume that the soil microbes observed in this study represent a long-term, stable community, notwithstanding

297 seasonal fluctuations. Furthermore, microbial populations likely reflect interactions with the
298 macrofauna through reciprocal upward and downward trophic cascades (Sinclair *et al.*,
299 2010)(Anderson *et al.*, 2007; Stevens *et al.*, 2018)(Sinclair *et al.*, 2010); Stevens *et al.*, 2018).
300 Results of this study reveals the importance of soil texture and mineral content in structuring
301 microbial communities, and we also show that grazing by large, migratory mammals impacts
302 microbial communities within the topsoil of the Serengeti National Park. The global distribution
303 of bacteria has been linked to a hierarchy of correlated factors, and especially soil pH, organic
304 matter and nitrogen availability (Fierer, 2017). Many studies have found that soil pH is a
305 significant driver of bacterial diversity (Fierer & Jackson, 2006; Lauber *et al.*, 2009; Griffiths *et*
306 *al.*, 2011; Kaiser *et al.*, 2016); but we observed pH to have little relationship with the soil
307 microbes in the Serengeti (Figs. 4, S2-S6; Table S4). A likely reason why our results are not
308 consistent with the literature is that the neutral soil pH range (6.3 to 7.8) observed in this study
309 which is ideal for microbial diversity, as opposed to the larger range of pH (3 - 9) captured
310 globally (Fierer & Jackson, 2006). Furthermore, in contrast to a global analysis showing
311 relationships between soil microbial communities and organic matter and nitrogen availability
312 (Fierer 2017), our study showed no relationship between microbial diversity and these soil
313 variables (Figs. 4, S5, and S6; Table S4). Instead, we found that the richness and evenness of
314 microbes increased in sites with finer textured soil (Figs. 3 and S2) and specific OTUs were
315 highly correlated with texture variables (Fig. 4).

316 For organisms living at the scale of soil particles, the size and manner in which those soil
317 particles coalesce can have a profound influence. Many studies have reported a connection
318 between soil particle size and microbial diversity (Sessitsch *et al.*, 2001; Carson *et al.*, 2010;
319 Chau *et al.*, 2011; Neumann *et al.*, 2013). Sand has larger pores with higher connectivity and
320 therefore retains less water and nutrients than finer textured soil. The strong relationship
321 between microbial community structure and soil texture could be linked to water potential, soil
322 moisture, pore connectivity, and nutrient diffusion, rather than the texture itself (Saxton & Rawls,

2006; Dechesne *et al.*, 2008; Carson *et al.*, 2010; Serna-Chavez *et al.*, 2013; Neumann *et al.*, 2013). A previous study found microbial abundance to be positively correlated with water holding capacity and soil moisture in finer textured soil in the Serengeti (Ruess & Seagle, 1994). Biotic interactions are influenced by water content, pore connectivity, and nutrient diffusion and could also influence microbial abundance and diversity. Small soil pores in fine textured soils may provide microbes a refuge from predation by bacterivorous protozoa and nematodes; and thus, loss through predation is likely to be higher in wetter, coarser soils that enable motility (Hassink *et al.*, 1993; Nielsen *et al.*, 2014). Furthermore, highly connected water-filled pores may favor competitive interactions (Treves *et al.*, 2003). Models of two bacterial species indicate less coexistence in wet soil conditions because connectivity facilitates interactions among microbes such that highly competitive species can more easily exclude poor competitors in saturated soil compared to dry soil (Hardin, 1960; Dechesne *et al.*, 2008; Long & Or, 2009). Overall, the structure of microbial communities in the Serengeti likely reflect both the top-down biotic influences of predation and bottom-up abiotic factors.

Our results indicate that environmental gradients resulting from volcanic inputs of ash influence the biogeography of microbes in the Serengeti, but it is impossible to uncouple the interconnected soil factors that arise from the volcanic deposits. Phosphorus concentration explained 14% of the variation in community composition and was consistent with a strong effect on beta diversity (Table S4; Fig. S6). Soil texture, on the other hand, explained 17% of the variation but was inconsistent with an effect on beta diversity (Table S4; Fig. S6), even with a strong gradient in soil texture (from 25% to almost 60% silt). Therefore, it is likely that both phosphorus concentration and soil texture directly and indirectly (through biotic interactions) influence microbial community structure. Results from our mixed models indicate that further research is necessary to completely disentangle the effects of soil texture and phosphorus concentration. An experimental design created to separate mineralogy from soil texture could help separate these effects.

Life in Serengeti soil evolved to co-exist beneath one of the largest mammalian migrations on Earth. We would expect the exclusion of large herbivores should have an effect on the microbial communities and our linear mixed model results are consistent with that prediction (Fig. 3b; Table S4; Fig. S6). Further, we found 34 total indicators of grazing, including one OTU within Rhizobiaceae (Table S3), some of which are known to fix nitrogen within plant roots (Spaink *et al.*, 2012). An association between Rhizobiaceae and plants within grazed plots could indicate a cooperative strategy to compensate for aboveground herbivory (Ramula *et al.*, 2019). Specifically, we found three indicators of grazing and 31 indicators of ungrazed plots (Fig. 3b, Table S3). There are many potential mechanisms by which herbivory might influence soil microbial communities (Bardgett *et al.*, 1998). The influx of nutrients from mammalian waste products could cause a shift in bacterial communities, increasing community differences between grazed plots, especially at larger spatial distances, while ungrazed plots maintain higher similarity, as our data implies (Figs. 3 and S6; Tables S3 and S4). Herbivory has altered the composition of plant communities at our sites, which in turn could alter bacterial communities (Anderson *et al.*, 2007; Prober *et al.*, 2015). Arbuscular mycorrhizal fungi can influence the community structure of soil bacteria (Artursson *et al.*, 2006), and the abundance of these fungi has been shown to be higher inside the fences that exclude herbivores and also in the southern sites with fine textured soil and higher phosphorus content (Antoninka *et al.*, 2015; Stevens *et al.*, 2018, 2020). Millions of migratory mammals offer soil microbes dispersal opportunities resulting in more homogenous communities (Vos *et al.*, 2013). Future research is needed to link the mammalian microbiome with the soil microbiome using a time-series of sampling that coordinates with the grazing cycle in the Serengeti.

Much of the biogeography of microbial communities remains a mystery, but our data provides some insights into the distribution of soil microbes in a naturally grazed grassland. This study highlights the importance of soil properties, and especially texture in structuring a

significant portion of microbial evenness and beta diversity. Additionally, we discovered that the removal of mammalian herbivores had a measurable effect on the beta diversity of microbial communities. Together, these results may help inform predictions of the regional biogeography of bacteria in natural, tropical grassland ecosystems. Future studies will benefit from a deeper understanding of microbial functional diversity and the spatial and temporal dynamics of life in the Serengeti soil.

Acknowledgements

The authors would like to thank Michael Anderson, Mark Ritchie, Jeff Propster, Anita Antoninka, Dan Revillini for their advice and input. We also thank Gustavo Zimmer for help with DNA extractions, and Lela Andrews and the NAU Environmental Genetics and Genomics Laboratory (EnGGen), Northern Arizona University, Flagstaff, AZ (nau.edu/enggen) for help with sequencing. Thanks to Emery Cowan for her contribution to the efficacy of the science communication.

Author contributions

NCJ planned, designed the research, and collected samples. BMS performed bioinformatics, statistical analyses, and created figures. DLS and BMS created and analyzed linear models. BMS analyzed and, with the help of NCJ, interpreted data. BMS created figures and wrote the first draft of the manuscript. NCJ edited the final version.

References

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, *et al.*2011. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters* **14**: 19–28.

Anderson TM, Ritchie ME, McNaughton SJ. **2007**. Rainfall and soils modify plant community response to grazing in Serengeti National Park. *Ecology* **88**: 1191–1201.

Anderson GD, Talbot LM. **1965**. Soil factors affecting the distribution of the grassland types and their utilization by wild animals on the serengeti plains, tanganyika. *Journal of Ecology* **53**: 33–56.

Andrews S. **2010**. *FastQC: a quality control tool for high throughput sequence data*.

Andrews L. **2018**. *akutils-v1.2.4: Facilitating analyses of microbial communities through QIIME*. Zenodo.

Antoninka AJ, Ritchie ME, Johnson NC. **2015**. The hidden Serengeti—Mycorrhizal fungi respond to environmental gradients. *Pedobiologia* **58**: 165–176.

Artursson V, Finlay RD, Jansson JK. **2006**. Interactions between arbuscular mycorrhizal fungi and bacteria and their potential for stimulating plant growth. *Environmental Microbiology* **8**: 1–10.

Ashley GM, Beverly EJ, Sikes NE, Driese SG. **2014**. Paleosol diversity in the Olduvai Basin, Tanzania: effects of geomorphology, parent material, depositional environment, and groundwater on soil development. *Quaternary International* **322–323**: 66–77.

Bardgett RD, Wardle DA. **2003**. Herbivore-mediated linkages between aboveground and belowground communities. *Ecology* **84**: 2258–2268.

Bardgett RD, Wardle DA, Yeates GW. **1998**. Linking above-ground and below-ground interactions: how plant responses to foliar herbivory influence soil organisms. *Soil Biology and Biochemistry* **30**: 1867–1878.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. **2011**. Examining the global distribution of dominant archaeal populations in soil. *The ISME Journal* **5**: 908–917.

Beauregard MS, Hamel C, Atul-Nayyar, St-Arnaud M. **2009**. Long-term phosphorus fertilization impacts soil fungal and bacterial diversity but not AM fungal community in Alfalfa. *Microbial Ecology* **59**: 379–389.

Berry D, Mahfoudh KB, Wagner M, Loy A. **2011**. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **77**: 7846–7849.

Beuselinck L, Govers G, Poesen J, Degraer G, Froyen L. **1998**. Grain-size analysis by laser diffractometry: comparison with the sieve-pipette method. *CATENA* **32**: 193–208.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. **2013**. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods* **10**: 57–59.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, *et al.*2018. *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*. PeerJ Inc.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, *et al.*2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, *et al.*2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**: 1621–1624.

Carson JK, Gonzalez-Quiñones V, Murphy DV, Hinz C, Shaw JA, Gleeson DB. **2010**. Low pore connectivity increases bacterial diversity in soil. *Appl. Environ. Microbiol.* **76**: 3936–3942.

Chau JF, Bagtzoglou AC, Willig MR. **2011**. The effect of soil texture on richness and diversity of bacterial communities. *Environmental Forensics* **12**: 333–341.

De Cáceres M, Legendre P. **2009**. Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566–3574.

Dechesne A, Or D, Smets BF. **2008**. Limited diffusive fluxes of substrate facilitate coexistence of two competing bacterial strains. *FEMS Microbiology Ecology* **64**: 1–8.

Eby SL, Anderson TM, Mayemba EP, Ritchie ME. **2014**. The effect of fire on habitat selection of mammalian herbivores: the role of body size and vegetation characteristics. *Journal of Animal Ecology* **83**: 1196–1205.

Faith DP. **1992**. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**: 1–10.

Fierer N. **2017**. Embracing the unknown: disentangling the complexities of the soil microbiome.

Nature Reviews Microbiology **15**: 579–590.
Fierer N, Jackson RB. **2006**. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 626–631.
Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. **2011**. The bacterial biogeography of British soils. *Environmental Microbiology* **13**: 1642–1654.
Grimshaw HM. **1987**. The determination of total phosphorus in soils by acid digestion. In: Rowland AP, ed. Chemical analysis in environmental research. NERC/ITE, 92–65.
Hamilton N. **2014**. USDA Textural Soil Classification. *R-bloggers*.
Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. **2012**. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **10**: 497–506.
Heiri O, Lotter AF, Lemcke G. **2001**. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *Journal of Paleolimnology* **25**: 101–110.
Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, Oki R, Nakamura K, Iguchi T. **2013**. The global precipitation measurement mission. *Bulletin of the American Meteorological Society* **95**: 701–722.
Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Küsel K, Rillig MC, Rivett DW, Salles JF, *et al.*2017. Where less may be more: how the rare biosphere pulls ecosystems strings. *The ISME Journal* **11**: 853–862.
Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrumpf M, Daniel R. **2016**. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Scientific Reports* **6**: 33696.
Klute A. **1986**. Water retention: laboratory methods. *Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods* **SSSA book series**: 635–662.
Kuramae EE, Gamper H, van Veen J, Kowalchuk G. **2011**. Soil and plant factors driving the community of soil-borne microorganisms across chronosequences of secondary succession of chalk grasslands with a neutral pH. *FEMS Microbiology Ecology* **77**: 285–294.
Kuramae EE, Yergeau E, Wong LC, Pijl AS, van Veen JA, Kowalchuk GA. **2012**. Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiology Ecology* **79**: 12–24.
Lauber CL, Hamady M, Knight R, Fierer N. **2009**. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**: 5111–5120.
Legendre P, Anderson MJ. **1999**. Distance-Based Redundancy Analysis: Testing Multispecies Responses in Multifactorial Ecological Experiments. *Ecological Monographs* **69**: 1–24.
Lozupone C, Knight R. **2005**. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**: 8228–8235.
McNaughton SJ. **1979**. Grazing as an optimization process: grass-ungulate relationships in the Serengeti. *The American Naturalist* **113**: 691–703.
McNaughton SJ. **1985**. Ecology of a grazing ecosystem: The Serengeti. *Ecological Monographs* **55**: 259–294.
McNaughton SJ, Zuniga G, McNaughton MM, Banyikwa FF. **1997**. Ecosystem catalysis: soil urease activity and grazing in the serengeti ecosystem. *Oikos* **80**: 467–469.
Neumann D, Heuer A, Hemkemeyer M, Martens R, Tebbe CC. **2013**. Response of microbial communities to long-term fertilization depends on their microhabitat. *FEMS Microbiology Ecology* **86**: 71–84.
Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P, *et al.*2018. Package 'vegan'. Community ecology package, R package version 2.5-4. https://CRAN.R-project.org/package=vegan.

Pigna M, Violante A. **2003**. Adsorption of sulfate and phosphate on andisols. *Communications in Soil Science and Plant Analysis* **34**: 2099–2113.

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, Adler PB, Bakker JD, *et al.*2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters* **18**: 85–95.

Ramula S, Paige KN, Lennartsson T, Tuomi J. **2019**. Overcompensation: a 30-year perspective. *Ecology* **100**: e02667.

Rohland N, Reich D. **2012**. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**: 939–946.

Ruess RW, Seagle SW. **1994**. Landscape patterns in soil microbial processes in the serengeti national park, tanzania. *Ecology* **75**: 892–904.

Saxton KE, Rawls WJ. **2006**. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal* **70**: 1569–1578.

Serna-Chavez HM, Fierer N, Bodegom PM van. **2013**. Global drivers and patterns of microbial abundance in soil. *Global Ecology and Biogeography* **22**: 1162–1172.

Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E. **2001**. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl. Environ. Microbiol.* **67**: 4215–4224.

Sinclair ARE, Arcese P. **1995**. *Serengeti II: dynamics, management, and conservation of an ecosystem*. University of Chicago Press.

Sinclair ARE, Metzger K, Brashares JS, Nkwabi A, Sharam G, Fryxell JM. **2010**. Trophic cascades in African savanna: Serengeti as a case study. In: Trophic cascades: predators, prey and the changing dynamics of nature. 255–274.

Spaink HP, Kondorosi A, Hooykaas PJJ. **2012**. *The Rhizobiaceae: molecular biology of model plant-associated bacteria*. Springer Science & Business Media.

Stan Development Team. **2017**. *RStanArm: Bayesian applied regression modelling via Stan*.

Stevens BM, Propster JR, Öpik M, Wilson GWT, Alloway SL, Mayemba E, Johnson NC. **2020**. Arbuscular mycorrhizal fungi in roots and soil respond differently to biotic and abiotic factors in the Serengeti. *Mycorrhiza*.

Stevens BM, Propster J, Wilson GWT, Abraham A, Ridenour C, Doughty C, Johnson NC. **2018**. Mycorrhizal symbioses influence the trophic structure of the Serengeti. *Journal of Ecology* **106**: 536–546.

Tan H, Barret M, Mooij MJ, Rice O, Morrissey JP, Dobson A, Griffiths B, O'Gara F. **2012**. Long-term phosphorus fertilisation increased the diversity of the total bacterial community and the phoD phosphorus mineraliser group in pasture soils. *Biology and Fertility of Soils* **49**: 661–672.

Vance CP, Uhde-Stone C, Allan DL. **2003**. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytologist* **157**: 423–447.

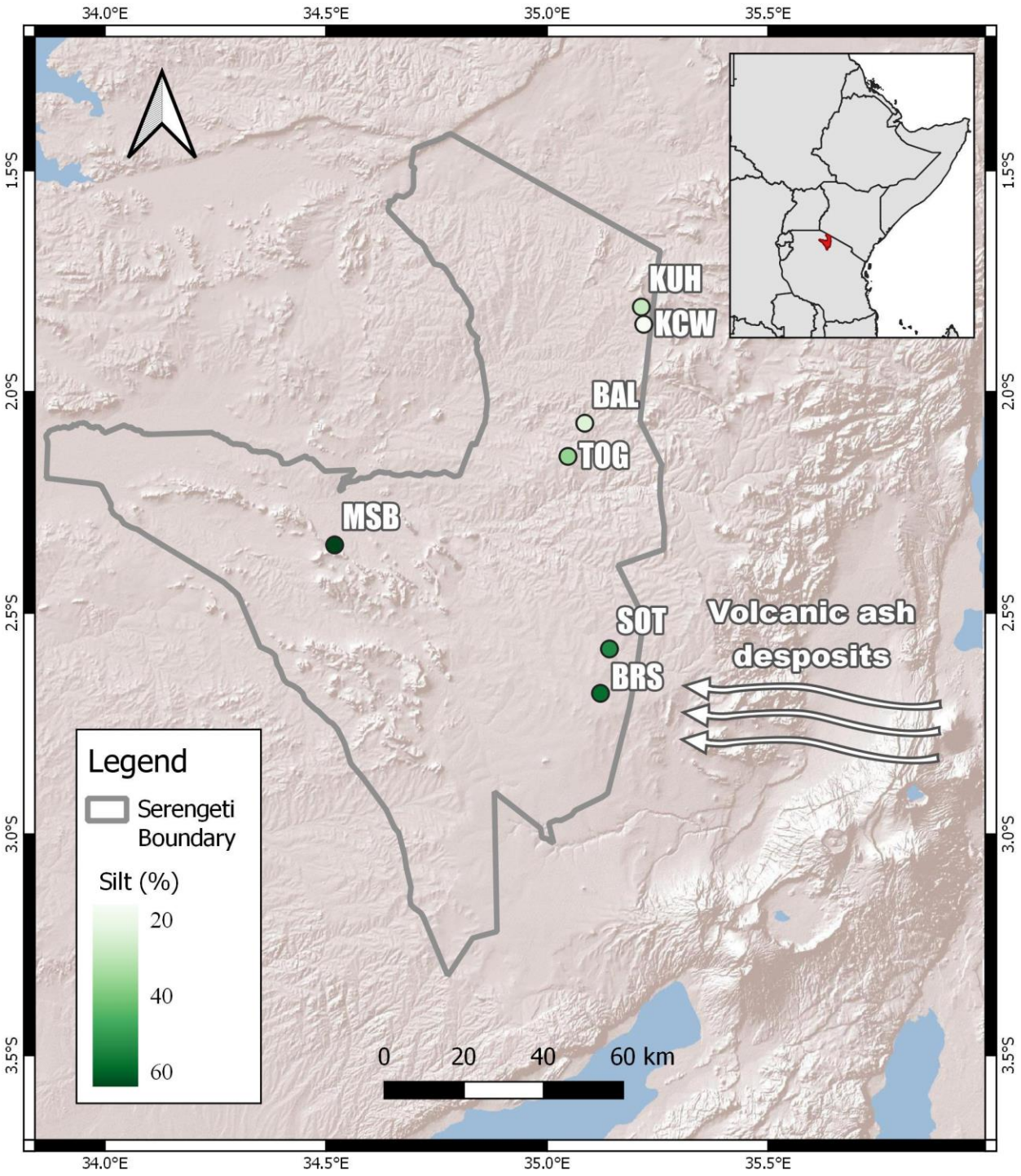
Vaughan RG, Kervyn M, Realmuto V, Abrams M, Hook SJ. **2008**. Satellite measurements of recent volcanic activity at Oldoinyo Lengai, Tanzania. *Journal of Volcanology and Geothermal Research* **173**: 196–206.

Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. **2013**. Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiology Reviews* **37**: 936–954.

Figures and Tables

Table 1. Annual precipitation (AP), and edaphic properties of the seven study sites in the Serengeti National Park, Tanzania.

See attached .xlsx spreadsheet



552

553 Figure 1. Locations of the seven study sites within the Serengeti National Park,

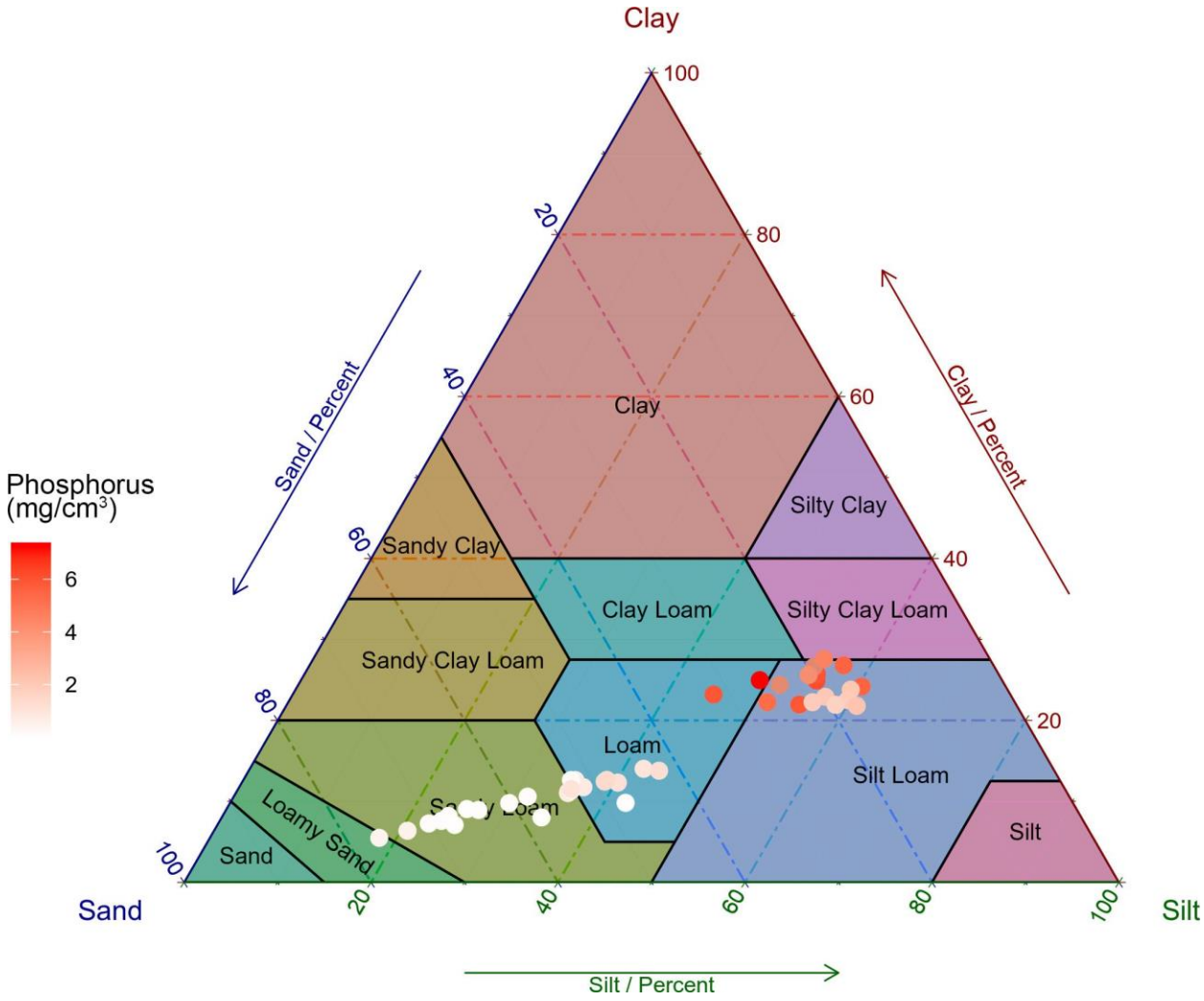
554 Tanzania: Balatines (BAL), Barafu (BRS), Klein's Camp West (KCW), Kuku Hills

555 (KUH), Soit le Motonyi (SOT), Togora (TOG), and Musabi Plains (MSB). The gradient in soil

556 texture is illustrated in green, darker color indicates higher percent silt.

557
558

559



560

561 Figure 2. Soil texture and phosphorus concentration of each plot at seven sites in this study.

562 Plots are colored by soil phosphorus concentration, where darker red indicates fine-textured,

563 high phosphorus soil that result from deposition of volcanic ash from highlands south of the

564 Serengeti National Park.

565

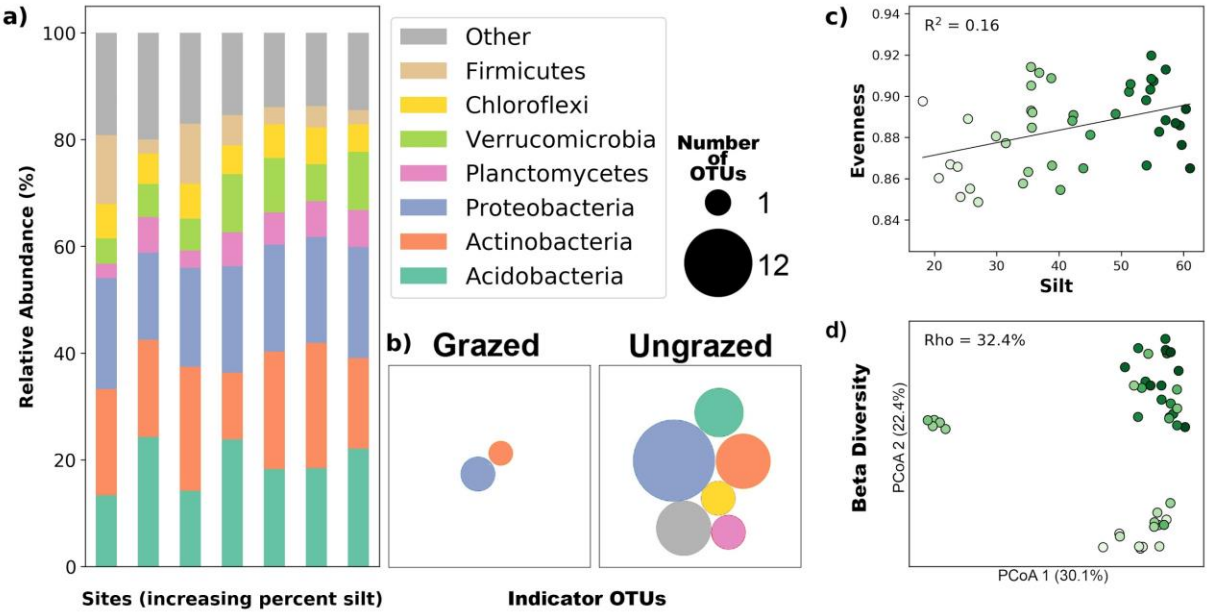


Figure 3. Illustrations of the bacterial community structure across environmental gradients and grazing treatments in the Serengeti. a) Correlation between Pielou's evenness, where 1 is completely even community and 0 in uneven. Points are colored by percent silt, darker colors are higher values, to facilitate interpretation of b) beta diversity as determined by Bray-Curtis dissimilarity. Spearman's rho value is reported to illustrate the correlation between beta diversity and percent silt (green gradient). c) Relative abundance of bacterial phyla. d) Visualization of indicator species grouped by phyla, indicated by colors. Area of each circle represents the number of indicator OTUs within each phylum.

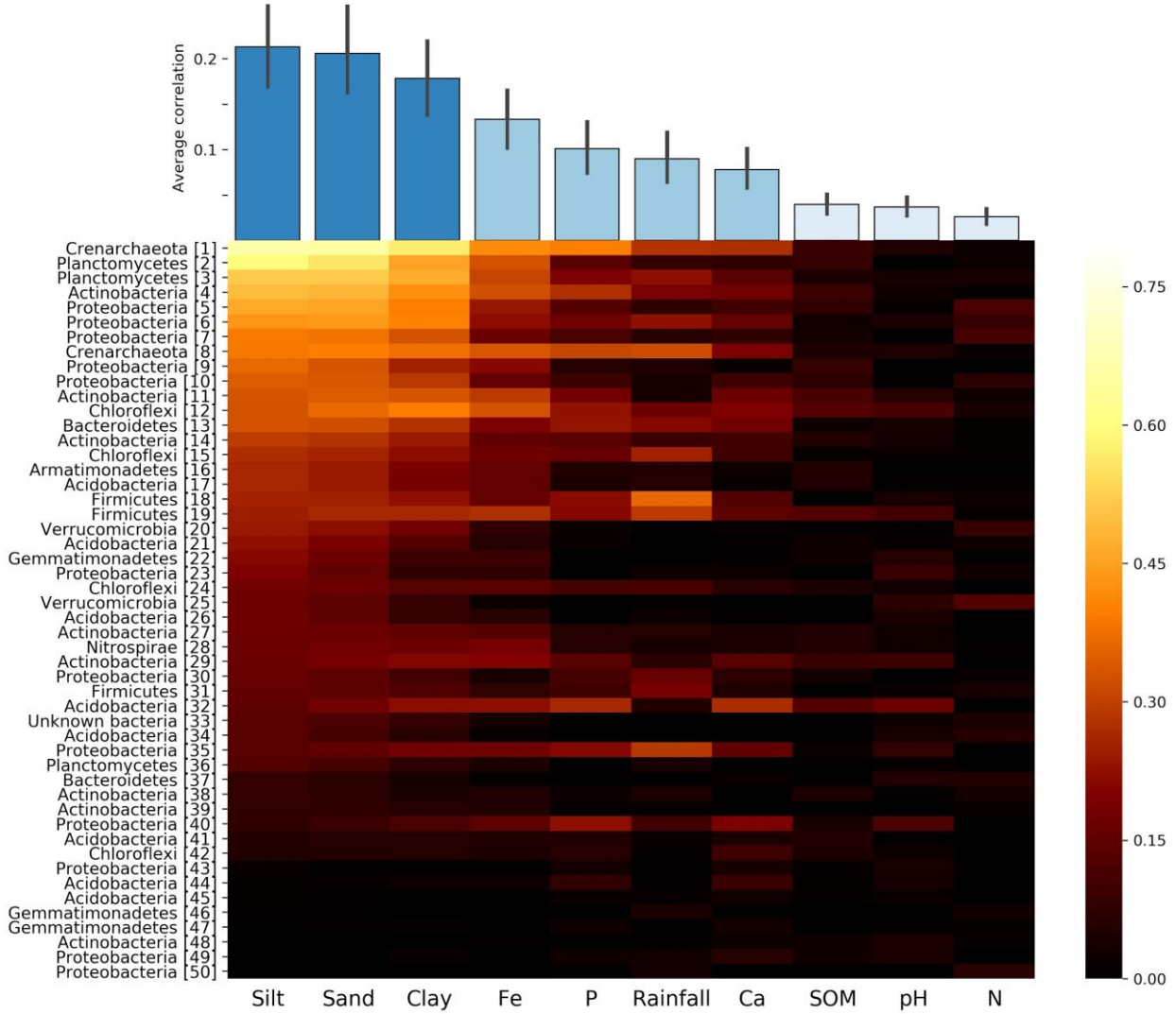


Figure 4. Correlation values for comparisons of relative abundance of 16S rRNA for individual operational taxonomic units (OTUs) and abiotic variables. Phyla for the 50 most abundant OTUs are displayed on the y-axis. The top-left most value is the highest correlation, and values are sorted into descending order for each abiotic variable and OTU. Blue bars above the heatmap represent the average correlation, with a 95% confidence interval, for each environmental variable. Blue shades are used to emphasize visual differences and do not represent a statistical difference.