Metagenomic characterization of soil microbial communities in the Luquillo experimental forest (Puerto Rico) and implications for nitrogen cycling

Smruthi Karthikeyan[1], Luis H. Orellana[1], Eric R. Johnston[1], Janet K. Hatt[1], Frank E. Löffler[3,4], Héctor L. Ayala-del-Río[5], Grizelle González[6], Konstantinos T Konstantinidis[1,2*]

[1] School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[2] School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

[3] Center for Environmental Biotechnology, Department of Microbiology, Department of Civil and Environmental Engineering, Department of Biosystems Engineering and Soil Science, University of Tennessee, Knoxville, Tennessee, USA

[4] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

[5] Department of Biology, University of Puerto Rico at Humacao, Puerto Rico, USA

[6] USDA Forest Service, International Institute of Tropical Forestry, Río Piedras, Puerto Rico, USA

*Address correspondence to Konstantinos Konstantinidis, kostas@ce.gatech.edu

School of Civil & Environmental Engineering, Georgia Institute of Technology. 311 Ferst Drive, ES&T Building, Room 3321, Atlanta, GA, 30332. Telephone: 404-639-4292

The authors declare no conflict of interest.

ABSTRACT

The phylogenetic and functional diversity of microbial communities in tropical rainforests, and how these differ from temperate communities remain poorly described but are directly related to the increased fluxes of greenhouse gases such as nitrous oxide (N_2O) from the tropics. Towards closing these knowledge gaps, we analyzed replicated shotgun metagenomes representing distinct life zones from four locations in the Luquillo Experimental Forest (LEF), Puerto Rico. These soils had a distinct microbial community composition and lower species diversity when compared to temperate grasslands or agricultural soils. Unlike temperate soils, LEF soils showed little stratification with depth in the first 0-30cm, with ~45% of community composition differences explained solely by location. The relative abundances and nucleotide sequences of N_2O reductases (*nosZ*) were highly similar between tropical forest and temperate soils. However, respiratory NO reductase (*norB*) was 2-fold more abundant in the tropical soils, which might be relatable to their greater N_2O emissions. Nitrogen fixation (*nifH*) also showed higher relative abundance in rainforest compared to temperate soils (20% vs. 0.1-0.3% of bacterial genomes in each soil type harbored the gene, respectively). Collectively, these results advance our understanding of spatial diversity and metabolic repertoire of tropical rainforest soil communities, and should facilitate future ecological modeling efforts.

Importance:

Tropical rainforests are the largest terrestrial sinks of atmospheric CO_2 and the largest natural source of N_2O emissions, two critical greenhouse gases for the climate. The microbial communities of rainforest soils that directly or indirectly, through affecting plant growth,

47 contribute to these fluxes remain poorly described by cultured-independent methods. To close
48 this knowledge gap, the present study applied shotgun metagenomics to samples selected from 3
49 distinct life zones within the Puerto Rico rainforest. The results advance our understanding of
50 microbial community diversity in rainforest soils and should facilitate future studies of natural or
51 manipulated perturbations of these critical ecosystems.
52
53 **INTRODUCTION**
54
55 Soil microbiomes are one of the most complex ecosystems owing to microenvironments
56 and steep physicochemical gradients, which can change on a micrometer or millimeter scale (1-
57 3). Temporal and spatial heterogeneity, demographic stochasticity, ecotype mixing, dispersion
58 and biotic interactions are the major drivers of soil microbial diversity in these ecosystems (4, 5).
59 The formation of such "metacommunities" coupled with biogeography and other edaphic factors
60 greatly influence the functional and taxonomic profile of a soil ecosystem at any given location
61 (6).
62 Tropical rainforests ("forests" hereafter) are characterized by humid and wet climate
63 patterns and account for a large portion of the world's total forest cover (7). These forests have
64 high levels of primary productivity (~30% of the total global production) due to large amounts of
65 precipitation coupled with year-long warm temperatures and high levels of light (8).
66 Consequently, high levels of biodiversity are observed in these forest soils with unique microbial
67 genotypic signatures being exclusive to this habitat/location, along with only a few cosmopolitan
68 taxa that are shared with other (non-tropical forest) habitats (9, 10). Although tropical forest soils
69 are critical ecosystems that host a plethora of distinct ecological niches, little is known about the

metabolic potential of tropical soils, especially, across elevation and depth gradients. Describing this metabolic diversity is important for studying and monitoring the microbial activities related to greenhouse gas fluxes, namely, nitrous oxide (N_2O) and carbon dioxide (CO_2) from the tropical soils (11).

Notably, tropical forests represent the largest terrestrial sinks of atmospheric CO_2 and the largest natural source of N_2O emissions (12-15). Natural soils have been reported to contribute over 43% of the total global N_2O emissions, with tropical ecosystems being the highest contributors, having 2 to 4 times higher contributions compared to natural temperate ecosystems (16-19). These soils are also responsible for about 70% of terrestrial nitrogen fixation, which underlies, at least in part, their high rates of net primary productivity (11, 20).

Microbially-mediated nitrification and denitrification are the biotic processes contributing the most to global N_2O soil emissions (60-70%) (19, 21, 22), although chemodenitrification, i.e., ferrous iron generated by ferric iron-reducing bacteria reacting with nitrite to produce N_2O abiotically, is also likely high in iron-rich tropical soils (23). In soils, N_2O is biologically produced as a result of incomplete nitrification, DNRA (dissimilatory nitrite reduction to ammonium) or denitrification respiratory pathways (22, 24, 25). Respiratory nitric oxide reductase (*nor*) is a key contributor to the microbial production of N_2O and is commonly encoded in the genome of denitrifying bacteria as well as some ammonia-oxidizing organisms (22, 26-30).

While both biotic and abiotic processes contribute to N_2O production, consumption of N_2O is exclusively mediated by microbial N_2O reductase (NosZ) activity (31-34). Yet, whether the denitrifying microorganisms in these soils differ from their counterparts in temperate soils and, if their functional genes present in the community reflect the high nitrogen fluxes, remain

93 unanswered questions despite their apparent importance for better management and modeling of
94 tropical soil ecosystems. It has also been demonstrated that tropical forests have significantly
95 higher rates of nitrogen fixation (~70% of total terrestrial nitrogen fixation) compared to other
96 ecosystems, significantly affecting the nitrogen budgets in these ecosystems (3, 35-37). For
97 instance, higher rates of nitrogen fixation in soils have been linked to nitrous oxide emissions (N
98 loss) due to reduced N retention capacities (11, 38, 39). How these ecosystem rates translate to
99 the nitrogen-fixing microbial (sub)community diversity and gene potential remains unclear.
100 The Luquillo Experimental Forest (LEF), also known as the El Yunque National Forest in
101 Puerto Rico (PR), has been a long term ecological research (LTER) site since 1988. The site is
102 dedicated to the assessment of the effects of climate drivers on the biota and biogeochemistry.
103 The forest has been subjected to several disturbance regimes over the last few decades, mostly
104 natural and -to a smaller extent- anthropogenic such as tourism and experimental manipulations
105 (40, 41). This site encompasses distinct “life zones” characterized by sharp environmental
106 gradients even across small spatial scales (40, 42, 43). The broad life zones based on the
107 Holdridge classification system include the rain forest, wet forest, lower montane wet forest, and
108 lower montane rain forest. These life zones are distinguished by elevation, temperature and
109 rainfall patterns in addition to other edaphic factors (44-47). The elevation and rainfall patterns
110 also tend to influence oxygen availability, redox potential, nutrient uptake and organic
111 decomposition rates (44, 47, 48). The dynamic interplay of existing physicochemical gradients
112 and climatic factors gives rise to a complex mosaic of biodiversity patterns observed in this
113 forest (45). Hence, LEF represents an ideal environment to study tropical microbial community
114 diversity patterns and their impacts on carbon and nitrogen cycling. The four sampling sites of
115 this study were chosen to represent the distinct vegetation and life zones within the LEF.

Previous studies in the LEF, and similar forest regions, have mostly focused on the effects of redox dynamics, litter decomposition, nitrogen (N) and other nutrient fertilization on microbial community activity through enzyme assays. Few studies have examined microbial diversity patterns across an elevation gradient and were only based on low-resolution techniques such as terminal restriction fragment length polymorphism analysis (14, 49-53). Furthermore, studies linking marker-gene abundances (related to nitrogen cycling) with *in-situ* flux measurements showed very high N_2O fluxes in the forest soils (54). However, the *nosZ* primers targeted only the typical (Clade I) clades, thereby introducing a primer bias, which can be circumvented by employing metagenomic analyses.

With recent developments in next generation DNA sequencing and associated bioinformatics binning algorithms, near-complete metagenome-assembled genomes (MAGs) can been recovered without cultivation (55, 56), opening new windows into studying soil microbial communities. Here, shotgun metagenomes originating from soils from the four different locations/life zones and three different depths in the LEF were analyzed to describe the microbial community diversity, biogeographical patterns, and metabolic potential differences across samples. Furthermore, the metagenomic data obtained from these soils were also compared to similar data from temperate grasslands in Oklahoma (OK) (57) and agricultural soils from Illinois (IL), USA (56) obtained previously by our team. By analyzing near-complete MAGs, we show that the most abundant microbial population (based on number of reads recruited) at each of the sampling locations represent sequence-discrete populations, similar to those observed in other habitats (58). Using such sequence-discrete populations as the fundamental unit of microbial communities, we subsequently assess the population distribution at high resolution

138 across the sampling sites (biogeography) and the gene content they encoded, with a focus on
139 nitrogen metabolism.
140
141 **RESULTS**
142
143 **Diversity of forest microbial communities**
144 The LEF soil communities were compared to those of intensively studied ecosystems,
145 namely the Oklahoma temperate grassland (OK) (57, 59) and Illinois agricultural soils (IL) (56),
146 which were previously characterized with similar shotgun metagenomics approaches. Shotgun
147 metagenomic sequencing recovered a total of 370 million reads across the 4 sites (Suppl. Table
148 S2). Nonpareil 2.0 (60) was used to estimate sequence coverage, i.e., what fraction of the total
149 extracted community DNA was sequenced. Nonpareil analysis of community diversity (Suppl.
150 Fig. S1) showed that the agricultural Urbana (IL) site had the highest diversity of all the soils
151 compared (NP diversity 24.02; note that NP values are given in log scale), and consequently, the
152 lowest sequence coverage at (only) 37.23%. El Verde and Pico del Este (20-30cm) were the least
153 diverse or most completely sequenced with 87.1% and 73.4% coverage respectively (NP
154 diversity of 19.6 and 20.6 respectively or about 2-3 orders of magnitude less diverse). Overall,
155 OK and IL soils appear to be more diverse than the PR soils by about two orders of magnitude,
156 on average, with an average Nonpareil value of 22.75± 0.37. Nearly complete coverage for El
157 Verde and Pico del Este (20-30cm samples) would require 2.402e+09bp and 8.735e+09bp,
158 respectively, and, for the same level of coverage, the more complex communities in Urbana (IL)
159 would require a substantially higher sequencing effort of 1.282e+12bp. The OK soils had an
160 estimated sequencing depth of 2.063e+11± 1.436e+11 bp.

Community composition variation across the forest sites based on 16S rRNA gene sequences.

The number of total 16S-rRNA gene-based OTUs (Operational Taxonomic Unit) observed in each metagenome as well as the Chao1 estimate of total OTUs present reflected the degree of undersampling at each site (Suppl. Fig. S1 and S2), and were also consistent with the Nonpareil coverage estimates (Fig. 1). When Puerto Rico tropical soils (PR) were compared with the agricultural and grassland soils from the United States at the phylum level, *Proteobacteria*, *Acidobacteria* and *Actinobacteria* were the most abundant taxa across all ecosystems. However, in the forest soils, a few highly abundant OTUs dominated the entire soil community whereas in the OK and IL soils, OTUs were more evenly distributed (Suppl. Fig. S2), consistent with the Nonpareil diversity results. Only 1.28% of the total detected OTUs (out of a total 8019, non-singleton OTUs) were shared among all PR samples, while 49.95% of OTUs were exclusive to a particular sampling site in PR, reflecting partly the under-sampling of the extant diversity by sequencing. Only 0.37% of the OTUs (out of a total 13760, non-singleton OTUs) were shared among all the sites across all 3 ecosystems, all of which were assignable to *Alphaproteobacteria*, *Acidobacteria*, *Verrucomicrobia* and *Actinobacteria*.

Further, applying four additional DNA extraction methods on a selected subset of our samples, including two manual phenol chloroform-based methods that are often advantageous for iron rich soils like those in tropical forest, revealed similar levels of diversity, more or less (Suppl.Fig. S3). Hence, the diversity patterns reported here are robust and independent of the DNA method used.

Factors driving community diversity in the forest soils: Multidimensional scaling analysis of beta diversity

The PCoA (Principal Coordinate Analysis) plots, constructed based on the MASH distances among whole metagenomes, showed a clustering pattern that was primarily governed by site/location. Accordingly, site explained 45.22% of the total diversity (Fig. 2B). The non-metric multidimensional scaling (NMDS) analysis of the data revealed only site, pH and soil moisture to be statistically significant physicochemical parameters in explaining the observed community diversity (Fig. 2C, Suppl. Table S3). ANOSIM values also indicated site to be a more important factor than depth, with a P value of 0.001 and 0.94, respectively. Based on the distance-based redundancy analysis (dbRDA), site was the most significant factor, even when the interplay between site and sampling depth was accounted for (Suppl. Table S4). Table 1 shows the partitioning of the variance between the proportion that is explained by constrained axes (i.e., environmental variables measured) and the porportion explained by unconstrained axes (i.e., variance not explained by environmental variables measured). The total variance explained by all (measured) environmental variables was 80.2% (Table. 1), which is remarkably high for a soil ecosystem (61).

Major N cycling pathways

Genes encoding proteins involved in denitrification and nitrogen fixation were the most abundant nitrogen (N) cycling pathway genes detected at different sites. Overall, the forest soils harbored about a 2-3-fold higher abundance of denitrification genes, i.e., *narG*, *nirK*, and *norB* catalyzing the reduction of nitrate, nitrite, and nitric oxide, respectively, compared to the grassland and agricultural soils (Fig. 2A). For instance, the *norB* gene abundance was found to

be at the highest abundance among the denitrification genes, with ~37% (SD 9.5%) of the genomes in the PR soils predicted to contain a *norB* gene, compared to ~17% (SD 4%) and ~14% (SD 1.3%) at IL and OK, respectively. Similarly, *narG* showed a 3-fold higher abundance in the PR soils compared to IL and OK soils (Fig.2B). While denitrification gene abundances appeared higher in the tropical soils, the relative abundance of *nosZ* gene, i.e., 11.6% (SD 3%) of the total genomes across the four locations in the LEF were predicted to encode *nosZ*, similar to *nosZ* relative abundance in IL and OK soils, i.e., 11.75% (SD 5%) and 11.08% (SD 3%), respectively (not statistically significant at p=0.05). Similar to *nosZ*, DNRA gene abundances (namely, *nrfA*) was similar across all sites studied herein (9%, SD 1.9%).

Predominant NosZ clades are shared among soil ecosystems

Placing *nosZ*-encoding reads to a reference *nosZ* phylogenetic tree revealed that atypical clades (clade II *nosZ*), affiliated predominantly with *Opitutus*, *Anaeromyxobacter* and other closely related genera, dominated the *nosZ* gene pool in the tropical forests (Figs. 3, Suppl. Figs.S4-S7). In contrast, a very small fraction of reads (<10% of total *nosZ* reads) were recruited to typical *nosZ* clades (or clade I). Members belonging to the clade II *nosZ* dominated the *nosZ* gene pool in OK and IL soils as well, with IL agricultural soils showing the greatest *nosZ* sequence diversity among the three regions. Notably, *O. terrae*-affiliated sequences represented the most abundant sub-clade (*nosZ* OTUs/sub-clades were defined at the 95% nucleotide sequence identity level) in all regions. Furthermore, most of the *O. terrae*–affiliated reads in the forest soil dataset appeared to be assigned to a single sub-clade, while their counterparts in the OK and IL soils appeared to be more evenly distributed among several closely related *nosZ* sub-clades, i.e., showing higher sequence diversity (Fig. 3, Suppl. Figs. S4-S7). *O. terrae* (strain

DSM 11246/PB90-1) *nosZ* reads at >95% identity made up between 20% and 60% of the total *nosZ* reads recovered from the El Verde site and, together with the second most abundant sub-clade from *Anaeromyxobacter sp*., contributed over 30% of the total *nosZ* reads across all four PR locations (Fig. 5). Despite the significant taxonomic diversity observed in these soils (Suppl Fig. S2), the soils from PR shared several abundant *nosZ* gene sequences/sub-clades at >95 nucleotide identity with soils in OK and IL (Fig. 3). Furthermore, in order to compare the predominant *nosZ* clades across the samples shown here, a new phylogenetic reference tree was constructed based on almost full length sequences obtained from the assemblies/MAGs obtained from the metagenomes studied here (namely PR,OK,IL). The short-reads identified as *nosZ* from the PR soils were placed on this tree and show that the majority of these reads are recruited by the *nosZ* sequences obtained from these assemblies/MAGs, indicating that the *nosZ* sequences across these ecosystems studies here are similar (Suppl. Fig. S8)

Nitrogen fixation potential

The nitrogen fixation genes (mainly *nifH*) were present at a much lower abundance in the lower altitude forest samples. For instance, only ~1-3% of all genomes in the lower altitude samples were predicted to encode *nifH* compared to a ~20% of the genomes in the higher elevation samples (Pico del Este) (Fig. 2A), and almost none of the reads from IL and OK metagenomes appeared to encode *nifH* (<0.1%). Therefore, nitrogen fixation gene abundance patterns indicated a much stronger selection for nitrogen fixation in the tropical forest relative to temperate agricultural or natural prairie soils, especially at higher elevations. Furthermore, no ammonia oxidizing genes (*amoA*) were detected in any of the soils except for Urbana soils (IL), which had a history of fertilizer (N) input.

Recovery of metagenome-assembled genomes (MAGs) representative of each site

In order to test the effect of biogeography (i.e., limits to dispersion) of taxa across the elevation gradient sampled, the distribution of abundant MAGs recovered from each PR sampling site ~~\~~(assembly and MAG statistics provided in Suppl. Table S6) were assessed across the sites using read-recruitment plots (62). Taxonomic assignment using the Microbial Genomes Atlas (63) revealed that the most abundant MAG at site El Verde (lowest elevation), representing 4.39% of the total metagenome, and was affiliated with an unclassified *Verrucomicrobia*. The second most abundant (1.8% of total) was likely a member of the genus *Ca. Koribacter* (*Acidobateria*) followed by an unclassified member of *Acidobacteria* (1.45% of total). The *Verrucomicrobium* MAG was found at an abundance of 1.03% of the total population at Sabana, and at 0.07% and 0.03% in Palm Nido and Pico del Este (highest elevation), respectively. Uneven coverage across the length of the reference sequence and nucleotide sequence identities were observed in the recruitment of short-reads from Palm Nido and Pico del Este as well as with all OK datasets, indicating that the related populations in the latter samples were divergent from the reference MAG (Suppl. Fig. S10). Therefore, at least this abundant low-elevation Verrucomicrobial population did not appear to be widespread in the other samples analyzed here (Suppl. Fig. S10). Similarly, the other abundant MAGs from other sites in the forest soils were unique to the corresponding sites (elevation) from which they were recovered. Almost all MAGs used in the analyses were assignable to a novel family, if not higher taxonomic rank, according to MiGA analysis (when compared to 11,566 classified isolate genomes available in the NCBI prokaryotic genome database), underscoring the large unexplored diversity harbored by the PR tropical rainforest soils. The sequence diversity/complexity as well as sequencing depth limited large-scale recovery of high-quality MAGS.

277

278 **Functional gene content of the MAGs**

279 The genome sequences of the most abundant MAGs from each location (n=6) were
280 analyzed in more detail to assess the functions they encoded, especially with respect to N cycling
281 pathways (Fig. 4). MAGs from Pico del Este (highest elevation) showed a high abundance of N
282 metabolism related genes compared to MAGs from other sites (Fig. 4). Most notably, genes
283 related to nitrogen fixation were found only in the Pico del Este MAG, which was consistent
284 with the short read analysis datasets showing greater relative abundance of *nifH* at this site.
285 Nitrification (ammonia oxidation related genes) gene clusters were not detected in any of the
286 recovered MAGs. *norB* and *nosZ* genes were found in three out of the six abundant MAGs
287 analyzed. The most abundant El Verde MAG, most closely related to *O. terrae* (AAI = 40 %),
288 possessed a *nosZ* gene, which was congruent with the *nosZ* phylogeny described above (i.e.,
289 ~60% of the *nosZ*-encoded reads from El Verde had a closest match to *O. terrae nosZ*
290 sequences).

291

292 **DISCUSSION**

293

294 The present study reported the taxonomic and gene content diversity of the poorly
295 characterized tropical rainforest soils by using whole-community, shotgun metagenomic
296 sequencing of samples from the El Yunque forest, Puerto Rico. The recovered near-complete
297 MAGs represented several abundant and widespread organisms within this ecosystem that could
298 serve as model organisms for future studies. Furthermore, since the Luquillo Experimental Forest
299 (LEF) within El Yunque is subjected to varying natural as well as experimental (e.g., warming,

phosphorus fertilization) perturbations, our study could also provide a baseline for these perturbations and future soil microbial studies at LEF. Our results revealed that the LEF soils harbor distinct microbial communities at sites with distinct elevation from sea-level. In contrast, and unlike several other soil ecosystems, sampling depth did not have a substantial impact on structuring community diversity, revealing no depth stratification in the LEF soils, at least for the depths sampled here (5-30cm). This could be due to the lack of distinct soil horizons within the first 30cm of the sampling sites, and indicates that the soil formation and/or physicochemical properties in these ecosystems could differ markedly from those in their temperate counterparts (44).

A recent study examining the dominant bacterial phylotypes across the globe found that the predominant phylotypes were widespread across ecosystems. The only exception to this pattern was the forest tropical soils which harbor distinct phylotypes (10). Consistent with these conclusions, the MAGs recovered from each LEF site represented at least novel species and genera, further underlining the under-tapped microbial diversity harbored by tropical forest soils. Currently, the environmental factors driving these diversity patterns remain poorly understood for tropical forest soils (10), but our study provided several new insights into this issue.

In particular, sites El Verde and Sabana (lowest elevation sites) had similar community structure and diversity compared to the two higher-elevation sampling sites with certain MAGs being present at both sites but not in any of the other (higher-elevation) sites examined. This is presumably attributable to both sites having similar climate and vegetation patterns (i.e., Tabonuco forest). On the other hand, Pico del Este was the highest elevation site and experiences almost continuous cloud cover as well as horizontal precipitation. The unique topology of Pico del Este was reflected in distinct and deeply novel MAGs and gene content, which differed

323 markedly from the other three sampling sites within the LEF (PCoA plots, Fig. 2B). The high
324 water content of the Pico del Este soils gives rise to a unique ecosystem dominated by epiphytes
325 (e.g., moss) (64). The epiphytic community has presumably significant impacts on nutrient (e.g.,
326 nitrogen) cycling (65), and influences the water input to the soil, thereby shaping a unique
327 habitat/niche for the soil microbes. Free-living microbes have been shown to be one of the
328 highest contributors to biological N fixation in these forests with high rates of nitrogenase
329 activity associated with the presence of moss/epiphytes (53, 66). Consistent with these previous
330 results and interpretations, the Pico del Este showed an extremely high potential for nitrogen
331 fixation, i.e., it was estimated that 1/5 of the total bacterial genomes sampled possessed genes for
332 N fixation, which is at least 10 times greater than any other site evaluated herein. Accordingly,
333 we found that site (location) alone explained about half (45%) of the beta diversity differences
334 observed among the four sampling sites, which reached ~80% when a few physicochemical
335 parameters namely pH and moisture were also included in the analyses (Fig. 2B, Table 1). This
336 is a remarkably high fraction of beta diversity explained by measured parameters for a soil
337 ecosystem (61) and likely reflected that location and the physical properties that characterized
338 different locations within LEF structured diversity much stronger than in other soil ecosystems.
339 Tropical forests have also been shown to have significantly higher rates of nitrogen fixation
340 compared to other ecosystems, which can exceed the N retention capacity of the soil resulting in
341 large N loss as N_2O (67). The findings reported here on denitrification gene abundances were
342 generally consistent with these previous observations as well.
343 Links between soil community structure and nitrogen cycling can help close the
344 knowledge gaps on how the forest ecosystems impact the release and mitigation of certain highly
345 potent greenhouse gases such as N_2O. The gene abundances observed here, e.g., more than two-

346 fold higher abundance of *norB* (associated with NO reduction to N_2O) and similar *nosZ* (N_2O
347 consumption) abundances in tropical soils relative to temperate soils were consistent with higher
348 N_2O emissions observed previously from the tropics. Further, in acidic soils such as the tropical
349 forest soils evaluated in this study, lack of N limitation can suppress complete denitrification,
350 thereby leading to higher N_2O release compared to other soil ecosystems (35). These
351 interpretations were consistent with our observation that the PR soils harbored a relatively high
352 abundance of respiratory (related to denitrification) *norB* genes as well. Previous studies have
353 also suggested that most denitrifying bacterial genomes possess the genes required to reduce
354 nitrate to nitrous oxide but do not possess the gene responsible for the last step i.e., N_2O
355 reduction to N_2, leading to the release of N_2O gas (Braker and Tiedje, 2003; Richardson et al.,
356 2009; Giles et al., 2012; (22, 26-29), consistent with the findings of our study.
357 It has been established that tropical forest soils are the single highest contributor of
358 natural N_2O emissions. While several abiotic and microbial processes can contribute to soil N_2O,
359 N_2O consumption is an exclusively microbial process, catalyzed by the enzyme product of the
360 *nosZ* genes (34). Based on the assessment of the *nosZ* gene phylogeny, it appears that almost all
361 of the *nosZ* genes from the tropical forest soils studied here belong to a previously overlooked
362 Clade II or atypical *nosZ* genes (32, 34, 68). This clade consists mainly of non-denitrifying, and
363 secondary denitrifying N_2O reducers. Despite the unique phylogenetic diversity harbored by
364 tropical soils in general, the *nosZ* gene sequence diversity appears to be shared between
365 temperate and agricultural soils (Fig 4). These findings imply strong selection pressure for
366 conservation of nitrous oxide reductase sequences across tropical and temperate soil ecosystems
367 that are not apparently applicable to other N-cycling genes and pathways, which warrants further
368 attention in the future.

Integration of functional (e.g., gene expression) data with *in-situ* rate measurements will provide a more complete picture of the composition and functioning in tropical forest soils. The identification of certain biomarker genes such as *nosZ* sequences in our study could facilitate future investigations on biogeochemical N-cycling and greenhouse gas emissions. For instance, the assembled MAGs and gene sequences provided here could be useful for the design of specific PCR assays for assessing transcript levels (activity), allowing potential linking of carbon dioxide, methane, nitrogen, SOM, etc. turnover to the activity of individual populations. It would also be interesting to assess how the findings reported here for the LEF apply (or not) to other tropical forests especially because our study is based on a relative small sample size. While the diversity in the Puerto Rico soils appears to be lower than that in temperate grassland and agricultural soils, and different DNA extraction methods, including phenol-chloroform- and kit-based, provided for similar results (Fig. S3), it is important to note that DNA of the temperate soil samples was extracted using different methods (OK soils were extracted using the PowerSoil kit). Therefore, it would be important to confirm these preliminary findings by using the exact same DNA extraction and sequencing procedures in all soils. Despite the sample size, however, our results showed differences along the elevation gradient sampled at the LEF that are independent of DNA extraction (Suppl.Fig. S3) or sequencing methods, and consistent with our metadata (Fig. 2), and previous process rate measurements. As the gradients at the LEF also provide a natural setting to interpret the potential ramifications of climate change scenarios such as altered participation patterns, the DNA sequences provided here could facilitate future manipulation experiments with an emphasis on understanding and predicting the effects of climate change on microbial community dynamics along the elevation gradient.

MATERIALS AND METHODS

Sampling sites

Soil samples were collected on February 2016 from four locations/sites across the LEF (18.3′ N, 65.80′ W). The four sites namely, Sabana, El Verde field station, Palm Nido and Pico del Este, each located at different elevations from the mean sea level, i.e., 265, 434, 634 and 953 m, respectively, were chosen due to their unique landscape and rainfall patterns, thereby creating distinct ecological niches (Fig. 2A).

The El Yunque forest is categorized into four distinct vegetation zones namely, the Tabonuco, Palo Colorado, Sierra Palm and Dwarf/Elfin forests. Site Sabana and El Verde, which are located at the lowest elevation among the four sites within the LEF, fall under the Tabonuco forest category in terms of vegetation, dominated by the tree species *Dacryodes excelsa* (native to Puerto Rico). They are characterized by canopy cover and low light intensities at the ground level which account for the sparsely vegetated forest floor. However, these sites still harbor the richest flora of all sites (69). Palm Nido is characterized by unstable, wetter soils, steeper slopes and the vegetation is dominated by the Sierra Palm (*Prestoea montana*). The site at the highest elevation, Pico del Este (dwarf forest ecosystem or “elfin woodlands”) is characterized by higher winds, lower temperatures and the vegetation is enveloped by clouds (41, 70) and its main vegetation is comprised of moss and epiphytes. Furthermore, highly acidic soil and continuously water-saturated soils deficient in oxygen are some major characteristics of this ecosystem with most mineral inputs for plants become dissolved in the rain and fog.

Three adjacent soil profiles were taken from each of the four LEF sites (4 sites encompassing 3 lifezones, Palo Colorado was not sampled). For each profile, individual soil

415 cores were taken at each depth (0-5cm, 5-20cm, 20-30cm) using a 3-cm diameter x 15-cm length
416 soil corer (AMS Inc, Idaho) that was decontaminated between samplings by washing with 70%
417 ethanol. Soil samples were stored in sterile Whirl-pak bags and kept on ice during transport and
418 until storage at -80º C. The three cores at each sampling depth were pooled together for
419 community DNA extraction, producing a total of twelve samples across the four sites.
420 Soil pH was determined using an automated LabFit AS-3000 pH Analyzer, and soil
421 extractable P, K, Ca, Mg, Mn, and Zn were extracted using the Mehlich-1 method and measured
422 using an inductively coupled plasma spectrograph at the University of Georgia Agricultural and
423 Environmental Services Laboratories (Athens, GA, USA). Soil extractable P using this method is
424 interpreted as the bioavailable fraction of P. NH_4-N and NO_3-N were measured by first
425 extracting them from soil samples with 0.1 N KCl, followed by the colorimetric phenate method
426 for NH_4^+ and the cadmium reduction method NO_3. The physicochemical conditions at the sites
427 during the time of sampling are provided in Supplementary Table (S1).
428
429 **Community DNA extraction and sequencing**
430 Total DNA from soil was extracted using the FastDNA SPIN KIT (MP Biomedicals, Solon, OH)
431 following manufacturer's procedure with the following modifications (71). Soils were air dried
432 under aseptic conditions followed by grinding employing a mortar and pestle. Cells were lysed
433 by bead beating and DNA was eluted in 50 µl of sterile H_2O. DNA sequencing libraries were
434 prepared using the Illumina Nextera XT DNA library prep kit according to manufacturer's
435 instructions except the protocol was terminated after isolation of cleaned double stranded
436 libraries. Library concentrations were determined by fluorescent quantification using a Qubit HS
437 DNA kit and Qubit 2.0 fluorometer (ThermoFisher Scientific), and samples were run on a High

Sensitivity DNA chip using the Bioanalyzer 2100 instrument (Agilent) to determine library insert sizes. An equimolar pool of the sequencing libraries was sequenced on an Illumina HiSeq 2500 instrument (located in the School of Biological Sciences, Georgia Institute of Technology) using the HiSeq Rapid PE Cluster Kit v2 and HiSeq Rapid SBS Kit v2 (Illumina) for 300 cycles (2 x 150 bp paired end). Adapter trimming and demultiplexing of sequenced samples was carried out by the HiSeq instrument. In total, 12 metagenomic datasets were generated (3 per site for the three depths), and statistic details on each dataset are provided in Supplementary Table S2.

In order to test for any DNA extraction biases of the kit used above, especially for the high iron/clay content that characterizes tropical forest soils and is known to affect the extraction step, four additional DNA extraction methods were performed in parallel on a small subset of samples collected in 2018 from the same sites (6 samples per extraction method for 5 ecxtraction methods covering the 4 sites). The methods included two manual (as opposed to kit-based) phenol-chloroform based methods (72, 73) as well as two other kit-based methods namely; DNeasy PowerSoil and DNeasy PowerSoil Pro (Qiagen Inc.). For this evaluation, the soils were first homogenized and subsequently in five subsamples to use with each method (including the FastDNA SPIN KIT-based method mentioned above). The libraries were constructed and sequenced the same way as described above for the FastDNA SPIN KIT method.

All metagenomic datasets were deposited in the European Nucleotide Archive (ENA) under project PRJEB26500. Additional data is available at http://enve-omics.ce.gatech.edu/data/prsoils.

Bioinformatics analysis of metagenomic reads and MAGs

The paired end reads were trimmed and quality checked using the SolexaQA (74) package with a cutoff of Q>20 ($\geq$99% accuracy per base-position) and a minimum trimmed length of 50 bp.

i) Assembly and population genome binning: Co-assembly of the short reads from the same location was performed using IDBA-UD (75) and only resulting contigs longer than 500 bp in length were used for downstream analysis (e.g. functional annotation and MyTaxa classification). Genes were predicted on the co-assembled contigs using MetaGeneMark (76) and the predicted protein-coding regions were searched against the NCBI All Genome database using Blastp (77). Since the assembly of individual datasets resulted mostly in short contigs (data not shown), the contigs from the co-assembly (combining metagenomes from the three sampling depths, for each site) were used for population genome binning. Contigs longer than 1Kbp were binned using MaxBin (78) to recover individual MAGs (default settings). The resulting bins were quality checked for contamination and completeness using CheckM (79), and were further evaluated for their intra-population diversity and sequence discreteness using fragment recruitment analysis scripts as part of the Enveomics collection (62) essentially as previously described (80).

ii) Functional annotation of MAGs: Genes were predicted for each MAG using MetaGeneMark and the predicted protein-coding regions were searched against the curated Swiss-Prot (81) protein database using Blastp (77). Matches with a bitscore higher than 60 or amino acid identity higher than 40% were used in subsequent analysis. The Swiss-Prot database identifiers were mapped to their corresponding metabolic function based on the hierarchical classification subsystems of the SEED subsystem category (Level 1) (82). The relative abundance of genes mapping to each function was calculated based on the number of predicted genes from each MAG assigned to the function (for read-based assessment, see below). Relative abundance data were plotted in R using the "superheat" package (https://arxiv.org/abs/1512.01524). Individual biomarker genes for each step of the nitrogen cycling pathway were manually verified by

484 visually checking the alignment of the identified sequences by the pipeline outlined above
485 against verified reference sequences.
486 *iii) Functional annotation of short reads*: Protein-coding sequences present in short reads were
487 predicted using FragGeneScan (83) using the 1% Illumina error model. The predicted genes were
488 then searched against the Swiss-Prot database using Blastp (best match). Low quality matches
489 (bitscore < 60) were excluded, and relative abundance of genes mapping to each function was
490 determined as described in the previous section.
491
492 **Community diversity estimation**
493 *i) Nonpareil*: Nonpareil (60) was used to estimate sequence coverage, i.e., what fraction of the
494 total extracted community DNA was sequenced and predict the sequencing effort required to
495 achieve "nearly complete coverage"(≥95%). The default parameters in Nonpareil were used for
496 all datasets. Only one of the two paired reads (forward) for each dataset was used to avoid
497 dependency of the paired reads, which can bias Nonpareil estimates (60).
498 *ii) MASH and multidimensional scaling*: MASH, a tool employing the MinHash dimensionality
499 reduction technique to compare sample-to-sample sequence composition based on k-mers (84),
500 was used to compute pairwise distances between whole metagenomic datasets and construct the
501 distance matrix to be used in multidimensional scaling. Pairwise MASH distances between the
502 metagenomic datasets were computed from the size-reduced sketches (default parameters).
503 PCoA (Principal coordinate analysis) and NMDS (Non-metric multidimensional scaling) were
504 employed to visualize the distance matrix and evaluate the physicochemical parameters driving
505 community diversity, respectively. Furthermore, dbRDA (distance based redundancy analysis),

was used to obtain a finer resolution on the observed compositional variation. All of the above startistical analysis were performed using the vegan package in R (85), with default settings.

iii) 16S rRNA gene fragments recovered from shotgun metagenomes: 16S ribosomal rRNA (16S) gene fragments were extracted from the metagenomic datasets using Parallel-META (86). 16S-carrying reads were classified taxonomically using the GreenGenes database.

Recovered 16S fragments were clustered ('closed-reference OTU picking' strategy using UCLUST (87)) and taxonomically classified based on their best match in the GreenGenes database (88) at an ID $\geq$ 97% in QIIME (89, 90). The relative abundance of the OTUs were calculated based on the number of reads assigned to each OTU. Community composition was assessed based on OTU taxonomic assignments at the genus and the phylum ranks and was compared between the sites based on the relative abundance of OTUs at each site.

Identification of N cycling genes using ROCker

ROCker (91) was employed for a precise identification and quantification of *nosZ* (encoding nitrous oxide reductase), *norB* (encoding respiratory nitric oxide reductase, cytochrome bc complex associated), *nirK* (encoding nitrite reductase), *narG* (encoding nitrate reductase), *nrfA* (encoding nitrite reductase, DNRA related) *amoA* (encoding ammonia monooxygenase) and *nifH* (encoding nitrogenase) encoding metagenomic reads (http://enve-omics.ce.gatech.edu/rocker/models). Briefly, the short-read nucleotide sequences were searched (using Blastx) against a training set for each abovementioned protein; training sets were manually curated to encompass experimentally verified reference sequences as suggested previously (91). The resulting matching sequences were then filtered using the ROCker compiled model (model for 150bp-long reads for PR and OK soils and 100 bp model for IL soils). Protein

529 abundances (based on the number of reads assigned to the protein) were normalized by
530 calculating genome equivalents. For the latter, the ROCker-filtered read counts were normalized
531 by the median length of the sequences of each protein reference, and the corresponding genome
532 equivalents were calculated as the ratio of NosZ (or another protein of interest) read counts to the
533 RNA polymerase subunit B (*rpoB*), a universal single copy marker, read counts.
534
535 ***NosZ phylogenetic analysis***
536 The NosZ reference protein sequences were aligned were aligned using CLUSTAL Omega (92)
537 and a maximum likelihood reference tree was created using RAxML v 8.0.19 (93) with a general
538 time reversible model option, gamma parameter optimization and '-f a' algorithm. The ROCker
539 identified NosZ-encoding reads were extracted from all datasets, translated into protein
540 sequences using FragGeneScan, and then added to the reference alignment using Mafft (94). The
541 reads were placed in the phylogenetic tree using RAxML EPA algorithm and visualized using
542 iTOL (95).
543
544 **Intra-population diversity assessment based on recovered MAGs**
545 The taxonomic affiliation of individual contig sequences of a MAG was evaluated based on
546 MyTaxa, a homology based classification tool (96). The MiGA (Microbial Genomes Atlas,
547 www.microbial-genomes.org) webserver was used for the taxonomic classification of the whole
548 MAG using the ANI/AAI concept. To assess intra-population diversity and sequence
549 discreteness, each target population MAG was searched against all the reads from each location
550 by Blastn (only contigs longer than 2Kbp were used). Fragment recruitment plots were
551 constructed based on the Blastn matches (threshold values: nucleotide identity ≥75% and

552 alignment length ≥80bp) using the Enveomics collection of scripts (62). The evenness of
553 coverage and sequence diversity of the reads across the length of the reference genome sequence
554 were used to evaluate the presence and discreteness of the population in the chosen dataset.
555
556
557 **Acknowledgments.** This work was supported by the U.S. Department of Energy, Office of
558 Biological and Environmental Research, Genomic Science Program (award DE-SC0006662) and
559 US National Science Foundation (award 1831582). GG was supported by the Luquillo Critical
560 Zone Observatory (National Science Foundation grant EAR-1331841) and the Luquillo Long-
561 Term Ecological Research Site (National Science Foundation grant DEB-1239764). All research
562 at the USDA Forest Service International Institute of Tropical Forestry is done in collaboration
563 with the University of Puerto Rico. We thank María Rivera and Humberto Robles from IITF for
564 their help in soil sampling.
565
566

TABLES:

Table 1: Proportion of total microbial community diversity explained by measured soil environmental factors.

	Inertia	Proportion	Rank
Total	0.1092	1	
Constrained	0.0876	0.8021	6
Unconstrained	0.02161	0.1978	5

Site, sampling depth, pH, total nitrogen, total carbon, moisture data were considered in the analysis

FIGURE LEGENDS

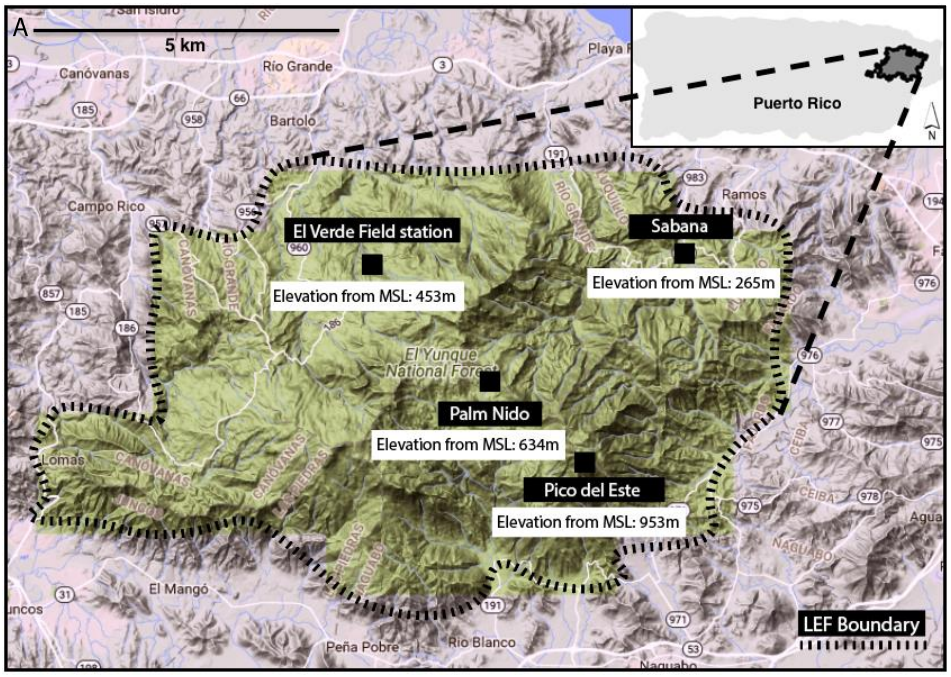
Fig. 1: Sampling location map and microbial community diveristy among the study sites. A. Map of the four sampling sites within the Luquillo Experimental Forest (LEF). B. Principal co-ordinate analysis (PCoA) plots based on MASH distances, colored by sampling site, C. Nonmetric multidimensional scaling (NMDS) plot with the soil physicochemical parameters incorporated. The arrow lengths are proportional to the strength of the correlations obtained between measured soil physicochemical parameters and each ordination axis.

Fig. 2: Abundance of N cycling genes and their distribution across soil ecosystems. A. Abundance of hallmark genes for denitrification, DNRA and nitrogen fixation pathways, represented as genome equivalents (% of total bacterial genomes sampled that carry the gene) in the metagenomes studied (see Figure key). B. Frequency of genomes carrying the respective

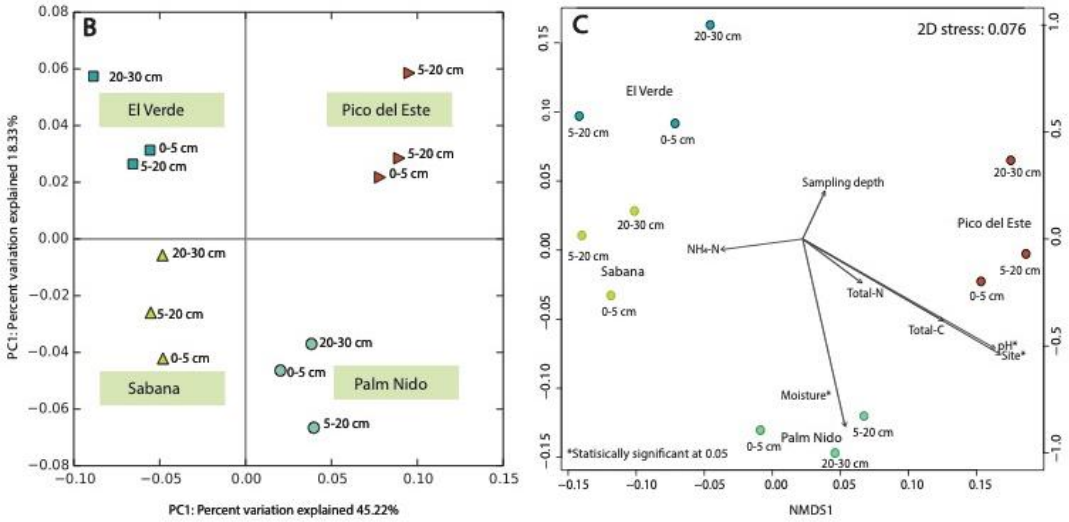
584 denitrifying gene across the three ecosystems studied. Genes denoted by the same letter are not
585 statistically significantly different between ecosystems (ANOVA Tukey test). Statistical
586 significance reported at $p < 0.05$. Note that nitrification genes were not detected in any of the
587 Puerto Rico sites.
588
589 **Fig. 3: Phylogenetic diversity of *nosZ*-encoding sequences recovered in each soil ecosystem.**
590 *nosZ* sequences were identified by the ROCker pipeline and placed in a reference *nosZ*
591 phylogeny as described in the Materials and Methods section. The radii of the pie charts are
592 proportional to the number of reads assigned to each sub-clade and the colors represent the
593 sampling sites from each ecosystem (see Figure key). Sub-clades highlighted in grey indicate the
594 most abundant sub-clades across all three ecosystems whereas the ones highlighted in blue were
595 abundant only in agricultural soils (IL). A. *nosZ* reads from every sampling site recruiting to
596 atypical (Clade II) clades. B. *nosZ* reads recruiting to typical (Clade I) clades. Inset shows the
597 most abundant sub-clade (*Opitutus terrae*) from panel A and its distribution across all sites. Note
598 that in all three ecosystems most of the reads recruit to atypical sub-clades. Suppl. Fig. S7 shows
599 the distribution of the reads among the most abundant sub-clades in detail.
600
601 **Fig. 4: Functions encoded by the recovered population MAGs.** Heatmap showing the relative
602 abundance of genes encoding the major metabolic functions (Level 1 of the SEED subsystem
603 category) for each MAG recovered from the four sites in Puerto Rico. The taxonomic
604 classification of each MAG based on MiGA is shown on the bottom left. The symbols at the
605 bottom of the heatmap denote the presence (or absence) of specific N-cycling genes, namely

606 denitrification and nitrogen fixation. No genes involved in nitrification were detected in any of
607 the bins.

608 **FIGURES**
609



610



611
612
613 **Figure 1: Sampling location map and microbial community diveristy among the study sites.**
614
615

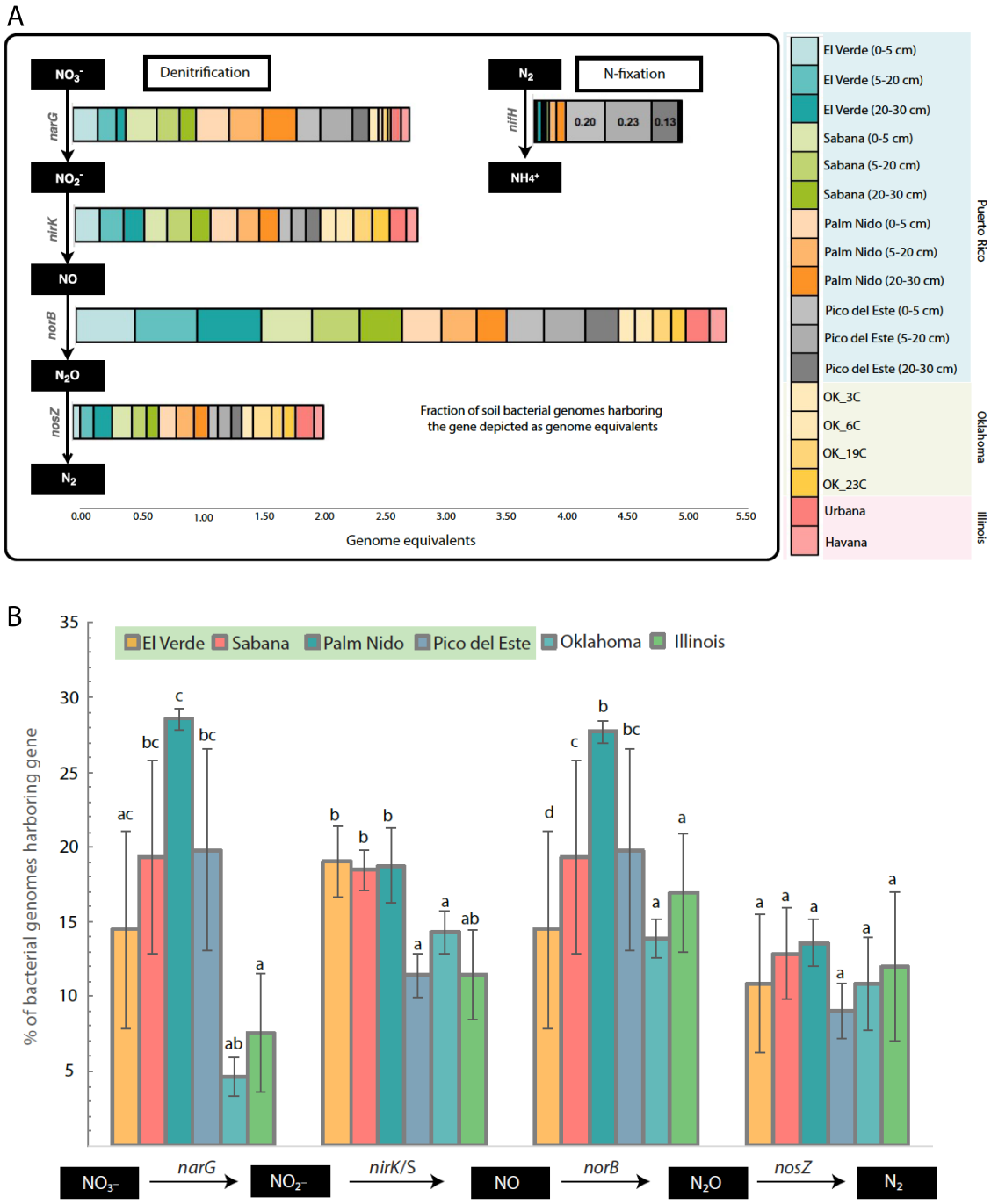
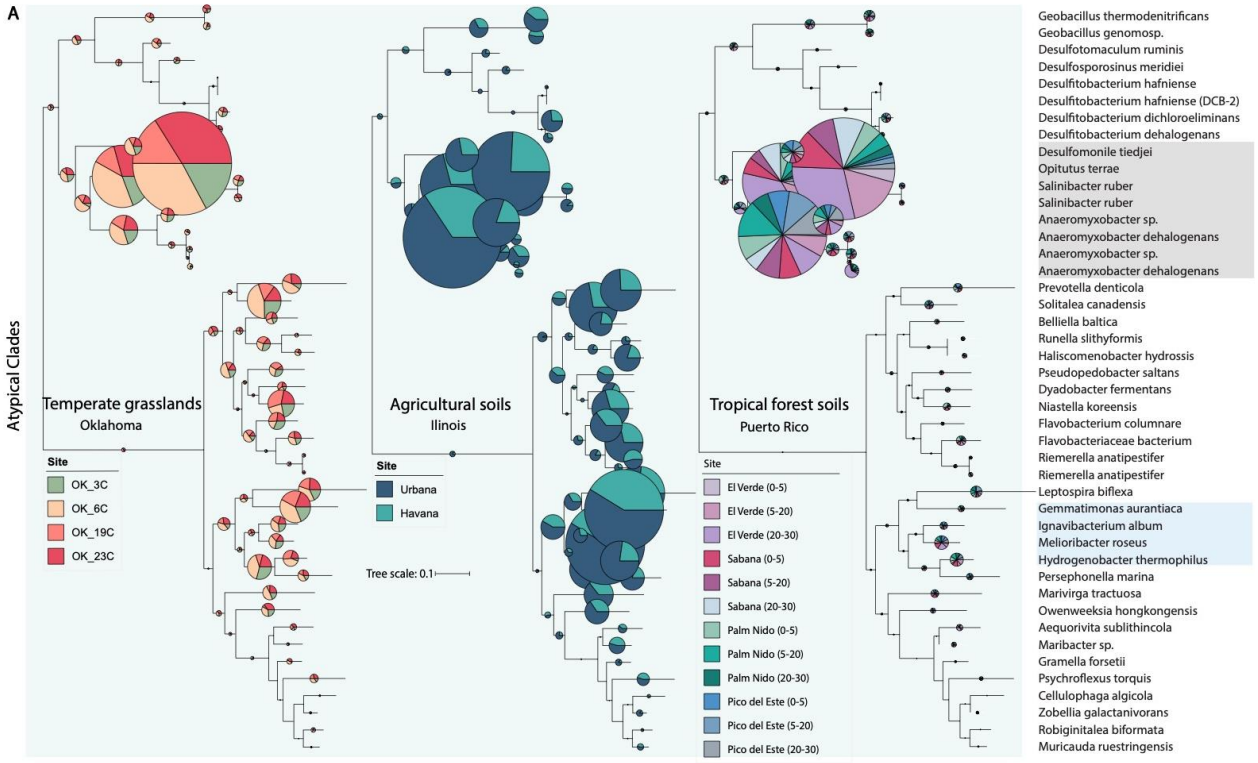


Figure 2: Abundance of N cycling genes and their distribution across soil ecosystems.



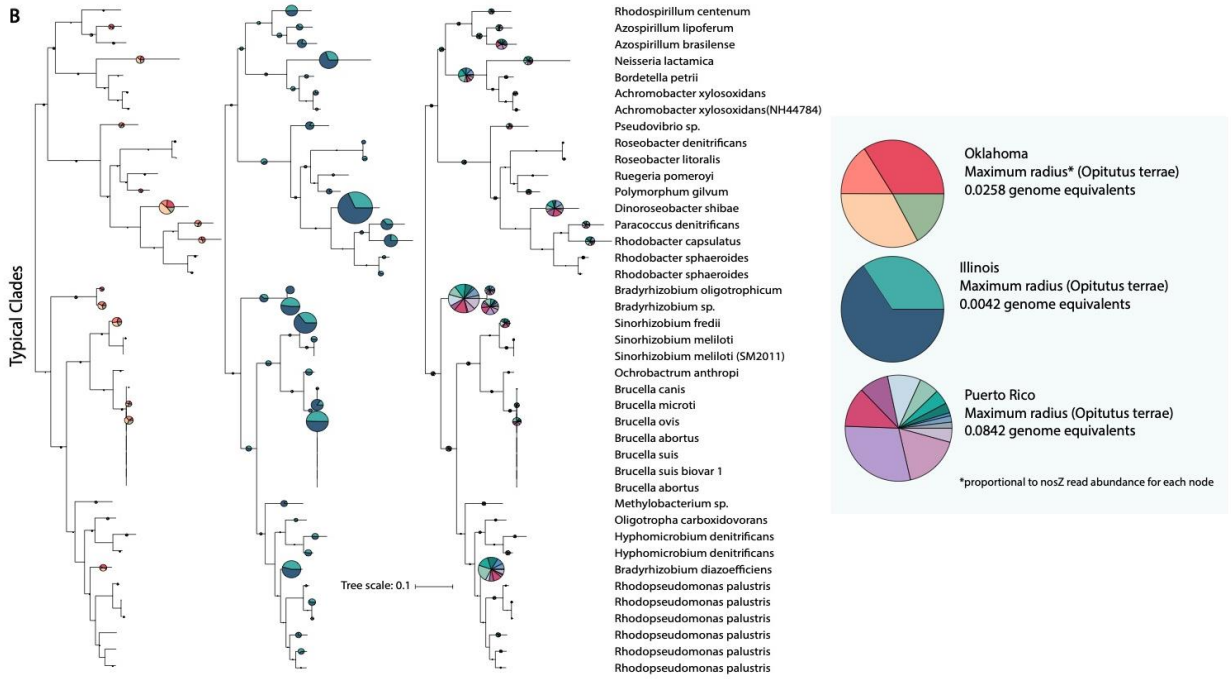
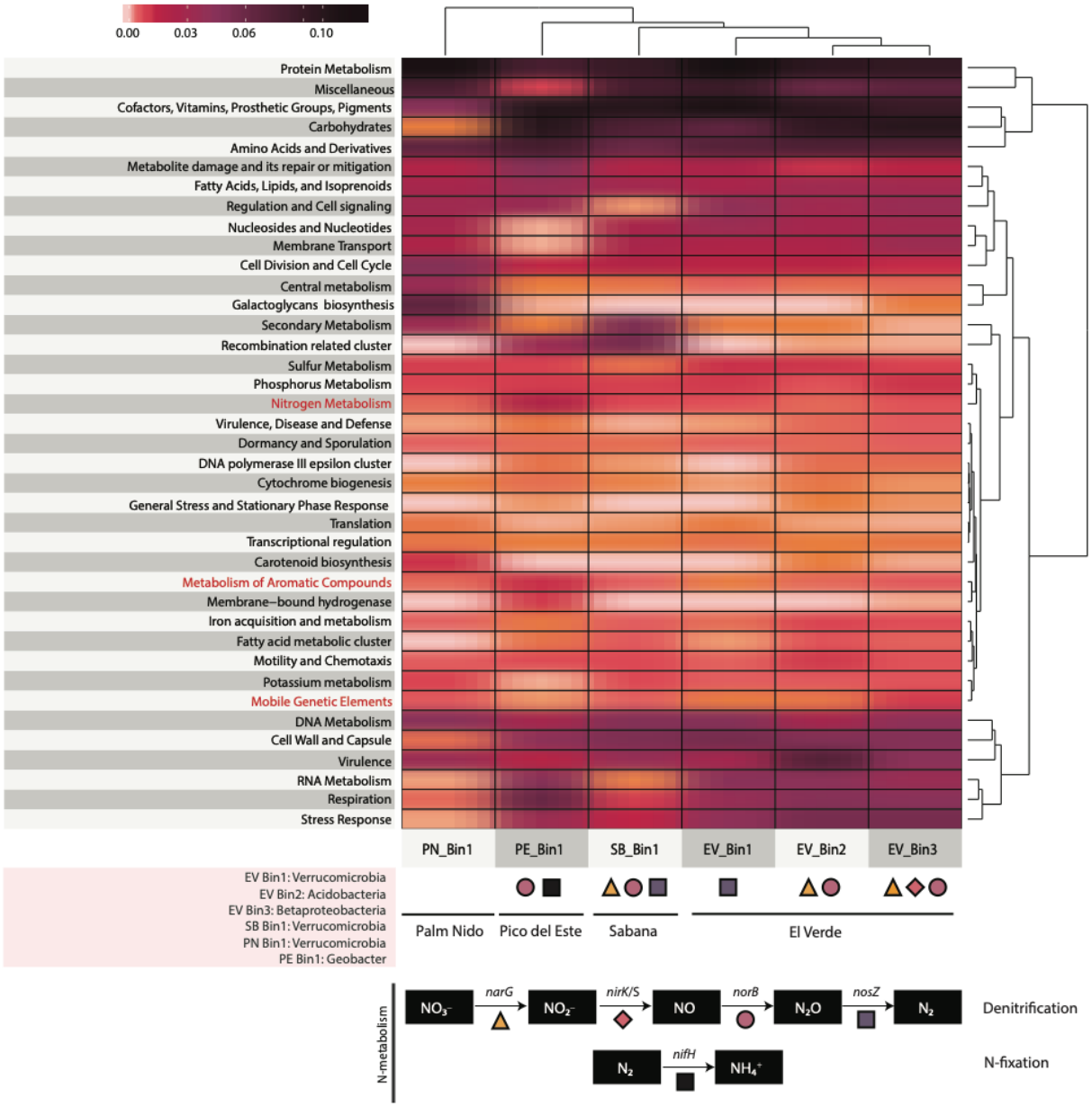


Figure 3: Phylogenetic diversity of *nosZ*-encoding sequences recovered in each soil ecosystem.

631



632

633 **Figure 4: Functions encoded by the recovered population MAGs.**

634 **REFERENCES**

635
636
637 1. Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K, Tu Q, Deng Y, He Z,
638 Shi JZ, Yuan MM, Sherry RA, Li D, Luo Y, Schuur EAG, Chain P, Tiedje JM, Zhou J,
639 Konstantinidis KT. 2014. Soil Microbial Community Responses to a Decade of Warming
640 as Revealed by Comparative Metagenomics. Applied and Environmental Microbiology
641 80:1777-1786.

642 2. Fierer N, Bradford Mark A, Jackson Robert B. 2007. Toward an ecological classification
643 of soil bacteria Ecology 88:1354-1364.

644 3. Van Der Heijden Marcel GA, Bardgett Richard D, Van Straalen Nico M. 2007. The
645 unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial
646 ecosystems. Ecology Letters 11:296-310.

647 4. Battin TJ, Sloan WT, Kjelleberg S, Daims H, Head IM, Curtis TP, Eberl L. 2007.
648 Microbial landscapes: new paths to biofilm research. Nat Rev Micro 5:76-81.

649 5. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil
650 microbiome. Nat Rev Micro 15:579-590.

651 6. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial
652 communities. Proc Natl Acad Sci U S A 103:626-631.

653 7. DeAngelis KM, Chivian D, Fortney JL, Arkin AP, Simmons B, Hazen TC, Silver WL.
654 2013. Changes in microbial dynamics during long-term decomposition in tropical forests.
655 Soil Biology and Biochemistry 66:60-68.

656 8. Malhi Y, Phillips OL. 2004. Tropical forests and global atmospheric change: a synthesis.
657 Philosophical Transactions of the Royal Society of London Series B: Biological Sciences
658 359:549.

659 9. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N,
660 Townsend AR, Cleveland CC, Stanish L, Knight R. 2011. Global patterns in the
661 biogeography of bacterial taxa. Environmental Microbiology 13:135-144.

662 10. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ,
663 Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant
664 bacteria found in soil. Science 359:320.

665 11. Pajares S, Bohannan BJM. 2016. Ecology of Nitrogen Fixing, Nitrifying, and
666 Denitrifying Microorganisms in Tropical Forest Soils. Front Microbiol 7:1045.

667 12. Chambers JQ, Silver WL. 2004. Some aspects of ecophysiological and biogeochemical
668 responses of tropical forests to atmospheric change. Philos Trans R Soc Lond B Biol Sci
669 359:463-76.

13. Templer PH, Silver WL, Pett-Ridge J, M. DeAngelis K, Firestone MK. 2008. PLANT AND MICROBIAL CONTROLS ON NITROGEN RETENTION AND LOSS IN A HUMID TROPICAL FOREST. Ecology 89:3030-3040.

14. Cusack DF, Silver WL, Torn MS, Burton SD, Firestone MK. 2011. Changes in microbial community characteristics and soil organic matter with nitrogen additions in two tropical forests. Ecology 92:621-632.

15. K. Firestone M, Davidson E. 1989. Microbiological Basis of NO and N2O Production and Consumption in Soil, vol 47.

16. Frasier R, Ullah S, Moore TR. 2010. Nitrous Oxide Consumption Potentials of Well-drained Forest Soils in Southern Québec, Canada. Geomicrobiol J 27:53-60.

17. Schmidt J, Seiler W, Conrad R. 1988. Emission of nitrous oxide from temperate forest soils into the atmosphere. J Atmos Chem 6:95-115.

18. Díaz-Pinés E, Werner C, Butterbach-Bahl K. 2018. Effects of Climate Change on CH4 and N2O Fluxes from Temperate and Boreal Forest Soils, p 11-27. *In* Perera AH, Peterson U, Pastur GM, Iverson LR (ed), Ecosystem Services from Forest Landscapes: Broadscale Considerations doi:10.1007/978-3-319-74515-2_2. Springer International Publishing, Cham.

19. Butterbach-Bahl K, Baggs EM, Dannenmann M, Kiese R, Zechmeister-Boltenstern S. 2013. Nitrous oxide emissions from soils: how well do we understand the processes and their controls? Philos Trans R Soc Lond B Biol Sci 368.

20. Houlton BZ, Wang Y-P, Vitousek PM, Field CB. 2008. A unifying framework for dinitrogen fixation in the terrestrial biosphere. Nature 454:327.

21. Werner C, Butterbach-Bahl K, Haas E, Hickler T, Kiese R. 2007. A global inventory of N2O emissions from tropical rainforest soils using a detailed biogeochemical model. Global Biogeochemical Cycles 21:GB3010.

22. Giles M, Morley N, Baggs EM, Daniell TJ. 2012. Soil nitrate reducing processes – drivers, mechanisms for spatial variation, and significance for nitrous oxide production. Front Microbiol 3:407.

23. Onley JR, Ahsan S, Sanford RA, Löffler FE. 2018. Denitrification by Anaeromyxobacter dehalogenans, a Common Soil Bacterium Lacking the Nitrite Reductase Genes nirS and nirK. Applied and Environmental Microbiology 84.

24. Braker G, Tiedje JM. 2003. Nitric Oxide Reductase (norB) Genes from Pure Cultures and Environmental Samples. Appl Environ Microbiol 69:3476-3483.

25. Richardson D, Felgate H, Watmough N, Thomson A, Baggs E. 2009. Mitigating release of the potent greenhouse gas N2O from the nitrogen cycle – could enzymic regulation hold the key? Trends Biotechnol 27:388-397.

26. Spiro S. 2012. Nitrous oxide production and consumption: regulation of gene expression by gas-sensitive transcription factors. Philos Trans R Soc Lond B Biol Sci 367:1213-1225.

27. Black A, Hsu PCL, Hamonts KE, Clough TJ, Condron LM. 2016. Influence of copper on expression of nirS, norB and nosZ and the transcription and activity of NIR, NOR and N(2). Microb Biotechnol 9:381-388.

28. Garbeva P, Baggs EM, Prosser JI. 2007. Phylogeny of nitrite reductase (nirK) and nitric oxide reductase (norB) genes from Nitrosospira species isolated from soil. FEMS Microbiol Lett 266:83-9.

29. Zumft WG. 2005. Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme–copper oxidase type. J Inorg Biochem 99:194-215.

30. Higgins SA, Schadt CW, Matheny PB, Löffler FE. 2018. Phylogenomics Reveal the Dynamic Evolution of Fungal Nitric Oxide Reductases and Their Relationship to Secondary Metabolism. Genome Biology and Evolution 10:2474-2489.

31. Braker G, Conrad R. 2011. Diversity, structure, and size of N(2)O-producing microbial communities in soils--what matters for their functioning? Adv Appl Microbiol 75:33-70.

32. Hallin S, Philippot L, Löffler FE, Sanford RA, Jones CM. Genomics and Ecology of Novel N₂O-Reducing Microorganisms. Trends in Microbiology 26:43-55.

33. Richardson D, Felgate H, Watmough N, Thomson A, Baggs E. Mitigating release of the potent greenhouse gas N₂O from the nitrogen cycle – could enzymic regulation hold the key? Trends in Biotechnology 27:388-397.

34. Hallin S, Philippot L, Löffler FE, Sanford RA, Jones CM. 2018. Genomics and Ecology of Novel N2O-Reducing Microorganisms. Trends Microbiol 26:43-55.

35. Zhang J, Cai Z, Cheng Y, Zhu T. 2009. Denitrification and total nitrogen gas production from forest soils of Eastern China. Soil Biol Biochem 41:2551-2557.

36. Orr CH, James A, Leifert C, Cooper JM, Cummings SP. 2011. Diversity and Activity of Free-Living Nitrogen-Fixing Bacteria and Total Bacteria in Organic and Conventionally Managed Soils. Applied and Environmental Microbiology 77:911-919.

37. Townsend Alan R, Cleveland Cory C, Houlton Benjamin Z, Alden Caroline B, White James WC. 2011. Multi-element regulation of the tropical forest carbon cycle. Frontiers in Ecology and the Environment 9:9-17.

38. Hedin LO, Brookshire ENJ, Menge DNL, Barron AR. 2005. The Nitrogen Paradox in Tropical Forest Ecosystems. Annu Rev Ecol Evol Syst 40:613-635.

39. Brookshire EN, Gerber S, Menge DN, Hedin LO. 2012. Large losses of inorganic nitrogen from tropical rainforests suggest a lack of nitrogen limitation. Ecol Lett 15:9-16.

40. S. Brown AEL, S. Silander, L. Liegel. 1983. Research history and opportunities in the Luquillo Experimental Forest. Tech Rep SO-44, US Forest Service p. 128.

41. Gould WA, Gonzalez G, Rivera GC. 2006. Structure and Composition of Vegetation along an Elevational Gradient in Puerto Rico. J Veg Sci 17:653-664.

42. Aide TM, Zimmerman JK, Herrera L, Rosario M, Serrano M. 1995. Forest recovery in abandoned tropical pastures in Puerto Rico. Forest Ecology and Management 77:77-86.

43. Weaver PLG, W. A. 2013. Forest vegetation along environmental gradients in northeastern Puerto Rico. Pages 43-66 in G. González, M. R. Willig, and R. B. Waide, editors. Ecological gradient analyses in a tropical landscape. Ecological Bulletins 54. Wiley-Blackwell, Hoboken, NJ. 2013.

44. Ping C-LM, G. J.; Stiles, C. A.; Gonzalez, G. 2013. Soil characteristics, carbon stores, and nutrient distribution in eight forest types along an elevational gradient, eastern Puerto Rico. Pages 67-86 in G. González, M. R. Willig, and R. B. Waide, editors. Ecological gradient analyses in a tropical landscape. Ecological Bulletins 54. Wiley-Blackwell, Hoboken, NJ. 2013.

45. González G, R. Willig M, Waide R. 2013. Ecological gradient analyses in a tropical landscape: multiple perspectives and emerging themes, vol 54.

46. González G, Lodge DJ. 2017. Soil Biology Research across Latitude, Elevation and Disturbance Gradients: A Review of Forest Studies from Puerto Rico during the Past 25 Years. Forests 8.

47. Van Beusekom AE, González G, Rivera MM. 2014. Short-Term Precipitation and Temperature Trends along an Elevation Gradient in Northeastern Puerto Rico. Earth Interactions 19:1-33.

48. Liptzin D, Silver Whendee L. 2015. Spatial patterns in oxygen and redox sensitive biogeochemistry in tropical forest soils. Ecosphere 6:1-14.

49. Hall SJ, Liptzin D, Buss HL, DeAngelis K, Silver WL. 2016. Drivers and patterns of iron redox cycling from surface to bedrock in a deep tropical forest soil: a new conceptual model. Biogeochemistry 130:177-190.

50. DeAngelis KM, Chivian D, Fortney JL, Arkin AP, Simmons B, Hazen TC, Silver WL. 2013. Changes in microbial dynamics during long-term decomposition in tropical forests. Soil Biol Biochem 66:60-68.

51. Waldrop MP, Balser TC, Firestone MK. 2000. Linking microbial community composition to function in a tropical soil. Soil Biol Biochem 32:1837-1846.

52. Templer Pamela H, Silver Whendee L, Pett-Ridge J, Kristen MD, Firestone Mary K. 2008. Plant and microbial controls on nitrogen retention and loss in a humid tropical forest. Ecology 89:3030-3040.

53. Cantrell SA, Lodge DJ, Cruz CA, Garcia LM, Perez-Jimenez JR, Molina M. 2013. Differential abundance of microbial functional groups along the elevation gradient from the coast to the Luquillo Mountains, p 87-100, Ecological Bulletins **54**, vol **54**.

54. Lammel DR, Feigl BJ, Cerri CC, Nüsslein K. 2015. Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. Front Microbiol 6:1057.

55. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, Küsel K, Rillig MC, Rivett DW, Salles JF, van der Heijden MGA, Youssef NH, Zhang X, Wei Z, Hol WHG. 2017. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J 11:853.

56. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. 2017. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Applied and Environmental Microbiology.

57. Johnston ER, Rodriguez-R LM, Luo C, Yuan MM, Wu L, He Z, Schuur EAG, Luo Y, Tiedje JM, Zhou J, Konstantinidis KT. 2016. Metagenomics Reveals Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem. Front Microbiol 7:579.

58. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist, metagenomics reveal. Environ Microbiol 14:347-355.

59. Luo C, Rodriguez RL, Johnston ER, Wu L, Cheng L, Xue K, Tu Q, Deng Y, He Z, Shi JZ, Yuan MM, Sherry RA, Li D, Luo Y, Schuur EA, Chain P, Tiedje JM, Zhou J, Konstantinidis KT. 2014. Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. Appl Environ Microbiol 80:1777-86.

60. Rodriguez-R LM, Konstantinidis KT. 2014. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics 30:629-635.

61. Zhang X, Johnston ER, Li L, Konstantinidis KT, Han X. 2017. Experimental warming reveals positive feedbacks to climate change in the Eurasian Steppe. ISME J 11:885-895.

62. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1.

63. Rodriguez RL, Gunturu S, Harvey WT, Rossello-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282-w288.

64. Bruijnzeel LA, Proctor J. Hydrology and Biogeochemistry of Tropical Montane Cloud Forests: What Do We Really Know?, p 38-78. *In* Hamilton LS, Juvik JO, Scatena FN (ed), Springer US,

65. Song L, Lu H-Z, Xu X-L, Li S, Shi X-M, Chen X, Wu Y, Huang J-B, Chen Q, Liu S, Wu C-S, Liu W-Y. 2016. Organic nitrogen uptake is a significant contributor to nitrogen economy of subtropical epiphytic bryophytes. Scientific Reports 6:30408.

66. Cusack DF, Silver W, McDowell WH. 2009. Biological Nitrogen Fixation in Two Tropical Forests: Ecosystem-Level Patterns and Effects of Nitrogen Fertilization. Ecosystems 12:1299-1315.

67. Hedin LO, Brookshire ENJ, Menge DNL, Barron AR. 2009. The Nitrogen Paradox in Tropical Forest Ecosystems. Annual Review of Ecology, Evolution, and Systematics 40:613-635.

68. Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C, Rodríguez G, Massol-Deyá A, Krishnani KK, Ritalahti KM, Nissen S, Konstantinidis KT, Löffler FE. 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proceedings of the National Academy of Sciences 109:19709.

69. Johnston MH. 1992. Soil-Vegetation Relationships in a Tabonuco Forest Community in the Luquillo Mountains of Puerto Rico. Journal of Tropical Ecology 8:253-263.

70. Weaver PL. 1995. The Colorado and Dwarf Forests of Puerto Rico's Luquillo Mountains, p 109-141. *In* Lugo AE, Lowe C (ed), Tropical Forests: Management and Ecology doi:10.1007/978-1-4612-2498-3_5. Springer New York, New York, NY.

71. Rodríguez-Minguela CM, Apajalahti JHA, Chai B, Cole JR, Tiedje JM. 2009. Worldwide Prevalence of Class 2 Integrases outside the Clinical Setting Is Associated with Human Impact. Applied and Environmental Microbiology 75:5100-5110.

72. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. Applied and environmental microbiology 66:5488-5491.

73. Tsai YL, Olson BH. 1991. Rapid method for direct extraction of DNA from soil and sediments. Applied and environmental microbiology 57:1070-1074.

74. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485.

75. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420-8.

76. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:e132.

77. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421-421.

78. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26.

79. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25:1043-1055.

80. Konstantinidis KT, DeLong EF. 2008. Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. ISME J doi:http://www.nature.com/ismej/journal/v2/n10/suppinfo/ismej200862s1.html.

81. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45-48.

82. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206-D214.

83. Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38:e191-e191.

84. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17:132.

85. Oksanen J. 2011. Multivariate analysis of ecological communities in R: vegan tutorial v.2.02. http://ccoulufi/~jarioksa/opetus/metodi/vegantutorpdf.

86. Su X, Pan W, Song B, Xu J, Ning K. 2014. Parallel-META 2.0: Enhanced Metagenomic Data Analysis with Functional Annotation, High Performance Computing and Advanced Visualization. PLOS ONE 9:e89323.

87. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.

88. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal 6:610-618.

89. Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG, Knight R. 2011. Using QIIME to analyze 16S rRNA gene sequences from Microbial Communities. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al] CHAPTER:Unit10.7-Unit10.7.

90. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335.

91. Orellana LH, Rodriguez-R LM, Konstantinidis KT. 2017. ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. Nucleic Acids Res 45:e14-e14.

92. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7.

93. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

94. Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066.

95. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242-5.

96. Luo C, Rodriguez-R LM, Konstantinidis KT. 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42:e73-e73.