

 Open access • Posted Content • DOI:10.1101/2020.09.15.299230

Amazonian soil metagenomes indicate different physiological strategies of microbial communities in response to land use change — Source link

Abdul Wadud Khan, Brendan J. M. Bohannan, Kyle M. Meyer, Ann M. Klein ...+4 more authors

Institutions: University of Texas at Austin, University of Oregon, University of Massachusetts Amherst, Lawrence Berkeley National Laboratory ...+1 more institutions

Published on: 17 Sep 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Metagenomics

Related papers:

- [Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe](#)
- [Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities](#)
- [Unravelling the effects of tropical land use conversion on the soil microbiome](#)
- [Soil microbiome responses to the short-term effects of Amazonian deforestation](#)
- [An integrated study to analyze soil microbial community structure and metabolic potential in two forest types.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/amazonian-soil-metagenomes-indicate-different-physiological-41pj9g8j9o>

1 **Amazonian soil metagenomes indicate different physiological strategies of microbial**
2 **communities in response to land use change**

3
4 **Running title:** *Soil metagenomes between forest and pasture*

5
6 Md Abdul Wadud Khan^{a\$}, Brendan J. M. Bohannan², Kyle M. Meyer^{b^}, Ann M. Klein^{b&},
7 Klaus Nüsslein^c, James P. Grover^{a#}, Jorge L. Mazza Rodrigues^{d,e#}

8
9 ^a Department of Biology, The University of Texas, Arlington, TX, USA

10 ^b Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA

11 ^c Department of Microbiology, University of Massachusetts, Amherst, MA, USA

12 ^d Department of Land, Air and Water Resources, University of California, Davis, CA, USA

13 ^e Environmental Genomics and Systems Biology Division, Lawrence Berkeley National
14 Laboratory, Berkeley, CA, USA

15
16 ^{\$}Present Address: Department of Surgical Oncology, The University of Texas MD Anderson
17 Cancer Center, Houston, TX, USA

18 [^]Present Address: Department of Integrative Biology, University of California, Berkeley, CA
19 USA

20
21 [&]Present Address: College of the Siskiyous in Weed, CA, USA

22
23
24 # Corresponding authors: James P. Grover, grover@uta.edu and
25 Jorge L. Mazza Rodrigues, jmrodrigues@ucdavis.edu
26
27
28

1 **ABSTRACT**

2 Despite the global importance in ecological processes, the Amazon rainforest has been subjected
3 to high rates of deforestation, mostly for pasturelands, over the last few decades. In this study,
4 we used a combination of deep shotgun metagenomics and a machine learning approach to
5 compare physiological strategies of microbial communities between contrasting forest and
6 pasture soils. We showed that microbial communities (bacteria, archaea and viruses), and the
7 composition of protein-coding genes are distinct in each ecosystem. The diversities of these
8 metagenomic datasets are strongly correlated, indicating that the protein-coding genes found in
9 any given sample of these soil types are predictable from their taxonomic lineages. Shifts in
10 metagenome profiles reflected potential physiological differences caused by forest-to-pasture
11 conversion with alterations in gene abundances related to carbohydrate and energy metabolisms.
12 These variations in these gene contents are associated with several soil factors including C/N,
13 temperature and $H^+ + Al^{3+}$ (exchangeable acidity). These data underscore that microbial
14 community taxa and protein-coding genes co-vary. Differences in gene abundances for
15 carbohydrate utilization, energy, amino acid, and xenobiotic metabolisms indicate alterations of
16 physiological strategy with forest-to-pasture conversion, with potential consequences to C and N
17 cycles. Our analysis also indicated that soil virome was altered and shifts in the viral community
18 provide insights into increased health risks to human and animal populations.

19

20 **Keywords:** Forest, Shotgun metagenomics, Soil virome, Pasture, KEGG pathways, Nitrogen
21 cycle, Methane cycle

22

23

1 **IMPORTANCE**

2 In the Amazon, forest-to-pasture conversion has a strong negative impact on biological
3 composition that results in alteration of important ecosystem processes. As the most
4 comprehensive survey known to date, this study with shotgun metagenomics approach
5 demonstrates predictive understanding of how an anthropogenic activity shapes microbial
6 taxonomy and functions across ecosystems, and the link between them. We then link the
7 microbiomes with each soil type to explore the adaptation mechanisms and ecological traits of
8 microbial communities in their ecosystems. Among these, carbon metabolism and energy harvest
9 strategies mostly relate to specific environmental challenges of microbial communities,
10 providing insights to understand the mechanisms of ecologically relevant metabolic processes.
11 Despite their critical roles in biogeochemical processes, viruses in terrestrial ecosystem remains
12 understudied. Thus, the inclusion of viral community in this study would shed new lights on their
13 roles in microbial functional traits, and provide novel hypotheses in microbial ecology of
14 terrestrial ecosystem.

15

16 **INTRODUCTION**

17 The Amazon is the largest continuous rainforest on Earth and provides essential
18 ecosystem services at regional and global scales. It harbors the largest collection of plants and
19 animal species in the world (1) and balances the flux of atmospheric gases (2). Despite these
20 known benefits, forest clearing has proceeded at an alarming rate over the last few decades. The
21 process of deforestation and land use change concurrently alters the biological and biochemical
22 compositions of the soil (3–6), which could have important consequences to the associated
23 ecosystem processes.

1 Land use change is predicted to be the most important factor in altering biodiversity in
2 tropical areas for the twenty first century (7). Ecologists have long studied the contribution of
3 plant and animal communities to ecosystem functions in tropical forests (8–10) and the impact of
4 deforestation on these communities (3, 4). In contrast, microbial ecologists have just started to
5 explore similar sets of questions in the Amazon, showing that ecosystem conversion alters
6 bacterial (5, 11–13), archaeal (14–16) and fungal (17) community composition in various
7 Amazonian land use types. While some studies showed that plant diversity increases overall
8 microbial diversity (18, 19) and activity in soil (20), others demonstrated the consequences of
9 deforestation were restricted to specific functional groups of microbes or specific protein-coding
10 genes such as those associated with C and N cycles (14, 21–26). The above studies, however,
11 largely relied on either 16S rRNA or a single functional gene as markers of microbial
12 communities. Although, metagenomics approach has been used in few recent studies to examine
13 the effect at two and four months (27) and of forest-to-agriculture conversion (28) following
14 deforestation. However, these metagenomics studies did not either evaluate the long-term impact
15 of deforestation or functional capacities of the most common land use type following
16 deforestation (i.e., pasture) on ecological processes carried out by soil microorganisms.

17 In addition, viruses remain neglected in Amazon soils even though they are the most
18 abundant biological entity, with an estimated 10^{31} on the planet (29). Viruses not only serve as
19 the largest reservoir of genetic material (30–32), but also play an important role in mediating
20 lateral gene transfer (33) and controlling microbial populations (34). Despite this, very few
21 empirical studies have been conducted on soil viromes (Graham et al. 2019; Williamson et al.
22 2017), and none in the context of deforestation in the Amazon rainforest.

23 To examine the effect of forest-to-pasture conversion, the most common land use type

1 following deforestation in the Amazon rainforest, we collected soils samples from pristine forest
2 and 38 years old pastureland, and performed deep sequencing with whole genome shotgun
3 approach, aiming to: (i) estimate the composition of microorganisms and protein-coding genes,
4 and their relationships in primary forest and established pasture; (ii) identify differential
5 abundances of key protein-coding genes involved in ecologically relevant biochemical processes;
6 and (iii) examine the Amazon soil virome and gain a comprehensive understanding of the
7 alterations happening with deforestation.

8

9 **RESULTS**

10 **Metagenome profiles.** We obtained a mean of about 636 million ($\pm 12\%$) reads per sample,
11 totaling 6.4 billion paired-end reads with an average length of 171 bp. Following quality filtering
12 of metagenomes through MG-RAST (version 3.3.6), the sequence reads were mapped to the
13 KEGG (Vol. 36) and RefSeq (Vol. 37) databases, where each read was annotated according to
14 the KEGG Orthology groups (KOs) and species (bacteria, archaea and viruses) associated with
15 the identified alignments. While the vast majority of the sequence reads (83.7%) passed the
16 quality control by MG-RAST, only 13.4% had protein-coding (KOs) and taxonomic annotations.
17 From KO sequencing reads ranging 60,373,709-76,838,373; taxonomic reads of bacteria ranging
18 783,394-1,143,853, archaea ranging 20,661-51,138; and viral reads ranging 18,307-32,903, we
19 obtained a total number of unique 5,607 KO genes; 23,949 bacterial, 706 archaeal, and 1,774
20 viral taxa at species level. Before downstream analyses, all samples were rarefied to their lowest
21 sequencing depth to bring them to equal values in order to remove sequencing effort bias. The
22 rarefaction plots showed that diversity levels of these metagenomic datasets tend to saturate,
23 suitable for a comprehensive analysis of the microbial communities and their physiological

1 potentials in both land use types (**Fig. SF1**).

2 **Impact of land use change on bacterial and archaeal communities.** We observed substantial
3 variation of major bacterial and archaeal taxa between forest and pasture soils. With over 26%
4 [$\pm 0.82\%$; standard error of mean (S.E.M.)] in forest and 19% ($\pm 0.22\%$; S.E.M.) in pasture ($P <$
5 0.001), *Alphaproteobacteria* was the dominant bacterial class in both ecosystems and showed the
6 largest variation in response to forest-to-pasture conversion (**Fig. SF2A**). Other major bacterial
7 groups that differed significantly between forest and pasture ($P < 0.001$) include: *Actinobacteria*
8 and *Firmicutes*, which increased from 13.1% ($\pm 0.21\%$; S.E.M.) to 16.65% ($\pm 0.41\%$; S.E.M.) and
9 11.39% ($\pm 0.52\%$; S.E.M.) to 14.4% ($\pm 0.17\%$; S.E.M.), respectively. *Euryarchaeota* and
10 *Thaumarchaeota*, were the two most differentially abundant archaeal phyla ($P < 0.001$), where
11 the former increased from 39.3% ($\pm 0.59\%$; S.E.M.) to 55.48% ($\pm 0.52\%$; S.E.M.), and latter
12 decreased from 27.41% ($\pm 0.47\%$; S.E.M.) to 7.73% ($\pm 0.72\%$; S.E.M.) following deforestation
13 (**Fig. SF2B**). We also identified differentially abundant bacterial (**Table ST1A**) and archaeal
14 (**Table ST1B**) species in each ecosystem. Using criteria set by DESeq2 and Random Forest
15 methods in this study, we identified 131 bacterial species that were enriched in forest soils, and
16 189 species that were enriched in pasture soils. *Bradyrhizobium* was the most dominant bacterial
17 genus in the forest soils, whereas *Bacillus* and *Arthrobacter* dominated pasture soils. With the
18 same criteria, 101 archaeal species were estimated to be differentially abundant in forest, where
19 an unclassified crenarchaeotal genus was the most dominant. On the other hand, pasture soils had
20 51 species that were differentially abundant with an unclassified euryarchaeal genus being the
21 most dominant.

22

1 **Diversity patterns of taxonomic lineages and protein-coding genes.** Forest-to-pasture
2 conversion increased alpha [$P < 0.01$, (**Fig. 1A**)] but decreased beta [$P < 0.05$ (**Fig. 1B**)]
3 diversities of bacterial species. While archaeal diversity did not follow a similar alpha diversity
4 pattern (**Fig. 1C**), beta diversity was statistically different between forest and pasture [$P < 0.05$,
5 (**Fig. 1D**)]. On the other hand, diversities of KO genes resembled those of bacteria [alpha: $P <$
6 0.05 , (**Fig. 1E**), and beta: $P < 0.01$, (**Fig. 1F**)]. We then asked whether the above bacteria and
7 archaea diversity results co-varied with protein-coding genes diversity values. In this analysis,
8 we used the principal coordinate loadings of the first axis (PCo1), which explained the highest
9 variation in the metagenome compositions. Our data showed strong compositional similarities of
10 KO gene profiles with bacterial [$R^2 = 0.95$, $P < 0.05$ (**Fig. 2A**)] and archaeal [$R^2 = 0.93$, $P < 0.05$
11 (**Fig. 2B**)] species profiles.

12
13 **Impact of land use change on major functional characteristics.** The ordination plots based on
14 principal coordinate analysis (PCoA) showed that the taxonomic composition in the forest soil
15 clustered apart from that in the pasture [ANOSIM: $R = 0.74$, $P = 0.007$ and $R = 0.99$, $P = 0.006$,
16 respectively (**Fig. 2C and 2D**)]. Next, we fit KEGG level 2 functions as vectors onto the
17 ordination and several functional attributes were significantly associated with the bacterial PCo1
18 (28.1%) across different soil samples (**Fig. 2C**). Positive correlations of PCo1 with energy
19 metabolism ($r^2 = 0.92$, $P < 0.001$), cell growth and death ($r^2 = 0.98$, $P < 0.001$), xenobiotics
20 biodegradation and metabolism ($r^2 = 0.96$, $P < 0.001$), and metabolism of other amino acids ($r^2 =$
21 0.96 , $P < 0.01$), as indicated by the direction of the arrows, demonstrated that increasing
22 gradients of these functions were strongly associated with forest samples, whereas pasture
23 samples were located at opposite ends of these gradients. Similarly, pasture sites had increasing

1 gradients of carbohydrate metabolism ($r^2 = 0.86$, $P < 0.01$), cell motility ($r^2 = 0.89$, $P < 0.001$),
2 metabolism of terpenoids and polyketides ($r^2 = 0.58$, $P < 0.05$), nucleotide ($r^2 = 0.67$, $P < 0.05$),
3 cell communications ($r^2 = 0.71$, $P < 0.05$) and biosynthesis of other secondary metabolites ($r^2 =$
4 0.91 , $P < 0.01$), whereas forest sites were at decreasing gradient of them. Most of these
5 functional traits were also estimated to be associated with the archaeal species (**Fig. 2D**), where
6 strength and direction of associations were alike. Similar trends were also observed when
7 performing a comparison for relative abundances of functional categories at KEGG level 2
8 between land use types (**Fig. SF3**).

9
10 **Differential abundance of protein-coding genes and metabolic pathways.** Next, we focused
11 on determining what metabolic pathways (KEGG level 3) and associated protein-coding genes
12 (KEGG level 4) were responsible for the disparities observed in major functional categories
13 (KEGG level 2) in relation to land use change. A comparative analysis identified 935 KO genes
14 showing large differences between forest and pasture metagenomes, of which 541 belong to
15 KEGG Enzyme Commission (EC) categories. A complete list of KO genes with \log_2FC values
16 and the P -values of Wald tests is provided in **Table ST2**. Our analysis showed that the major
17 differences of forest and pasture metagenomes were associated with genes involved in
18 carbohydrate, energy, and amino acid metabolisms and xenobiotic degradation, which are
19 functionally connected to other pathways (**Fig. 3**).

20 *Carbohydrate metabolism* - Pasture soil samples were estimated to contain elevated
21 numbers of KEGG EC genes associated with the metabolism of starch, sucrose, and galactose
22 (**Table ST2**). In addition, KO genes involved in the transport of mono- and di-saccharides were
23 far more abundant in pasture, of which \log_2FC values ranged from -0.39 to -3.01. Unlike pasture

1 metagenomes, which had elevated counts of glycolysis-specific genes [e.g., ADP-dependent
2 phosphofructokinase/glucokinase (EC:2.7.1.146, EC:2.7.1.147)], forest metagenomes had higher
3 counts of genes specific for gluconeogenesis [e.g., fructose 1,6-bisphosphate phosphatase
4 (EC:3.1.3.11); **Fig. 4**]. In addition, genes associated with the pentose phosphate pathway (PPP)
5 and the conversion of oxaloacetate to phosphoenolpyruvate [PEP; phosphoenolpyruvate
6 carboxykinase (EC4.1.1.49)] had higher relative abundances in the forest metagenome samples.
7 Furthermore, we observed considerable variation between forest and pasture metagenomes in
8 frequencies of genes that are involved in succeeding metabolism of sugars. For example, genes
9 associated with the tricarboxylic acid (TCA) cycle were estimated to be higher in pasture soils
10 (**Fig. 4**). In addition, pasture soils had higher abundances of genes affiliated with fermentation of
11 acetate (e.g., acetyl-CoA hydrolase (EC:3.1.2.1), and butanoate [e.g., butyrate kinase
12 (EC:2.7.2.7)], whereas genes associated with fermentation of propanoate [propionyl-CoA
13 synthetase (EC:6.2.1.17)], and lactate [L-lactate dehydrogenase, cytochrome (EC:1.1.2.3)] were
14 elevated in forest soils.

15 *Energy metabolism* - The most prominent differences observed were related to the genes
16 associated with nitrogen and methane metabolisms (**Fig. 5A**). Forest metagenomes had an
17 overrepresentation of EC genes involved in nitrification [nitronate monooxygenase
18 (EC:1.13.12.16); nitrate reductase 2 alpha subunit (EC1.7.99.4)], denitrification [nitrite
19 reductase, NO-forming (EC:1.7.2.1)] and assimilatory nitrate reduction [nitrate reductase,
20 NADH (EC:1.7.7.1)] (**Fig. 5A**). Conversely, the estimated representation of genes involved in
21 nitrogen fixation was higher in the pasture metagenomes [nitrogenase molybdenum-iron protein
22 alpha, and beta chains (EC:1.18.6.1) with log₂FC values of -5.76 and -5.92, respectively]. In
23 methane metabolism, several genes involved in methanogenesis were enriched in pasture

1 metagenomes [e.g., carbon monoxide dehydrogenase (EC:1.2.7.4); acetyl-CoA
2 decarbonylase/synthase complex subunit delta (EC:2.1.1.245); tetrahydromethanopterin S-
3 methyltransferase (EC:2.1.1.86); and methyl-coenzyme M reductase gamma subunit
4 (EC:2.8.4.1)]. Other genes involved in methanotrophy were enriched in the forest [methylene-
5 tetrahydromethanopterin dehydrogenase (EC:1.5.1.-); formylmethanofuran dehydrogenase
6 subunit A (EC:1.2.7.12)] (**Fig. 5B and Table ST2**).

7 *Amino acid metabolism* - The overall abundance of genes involved in amino acid
8 metabolism was not statistically different between forest and pasture. However, we observed
9 overrepresentation of certain genes in forest metagenomes ($P < 0.05$), especially those encoding
10 for enzymes involved in key steps of oxaloacetate-derivative amino acids such as methionine,
11 threonine (**Fig. SF4A**), PPP-derivative amino acids including histidine (**Fig. SF4B**), and
12 pyruvate derivative branched-chain amino acids (valine, leucine and isoleucine) (**Fig. SF4C**). In
13 addition, degradation genes for branched-chain amino acids were also higher in forest
14 metagenomes (**Table ST2**). As a contrast, pasture metagenomes were associated with the
15 metabolism of alpha-ketoglutarate, another TCA cycle intermediate (**Fig. 4**). In addition, there
16 was increased gene abundance related to the biosynthesis of arginine from glutamate, proline
17 from arginine (**Fig. SF4D**).

18 *Xenobiotics biodegradation and metabolism* - We determined higher gene sequence
19 counts ($P < 0.05$) in forest metagenomes associated with the degradation of diverse aromatic
20 compounds including benzoate, chlorocyclohexane and chlorobenzene, dioxin, nitrotoluene, and
21 polycyclic aromatic hydrocarbon (**Fig. SF5**). These pathways seem to be linked and contribute to
22 the production of the pyruvate and oxaloacetate that might act to support the TCA cycle and

1 replenish its intermediates, a mechanism known as anaplerosis, which are used in the
2 biosynthesis of essential compounds such as amino acids.

3 *Others* - Pastures had a higher abundance of several genes affiliated with chemotaxis,
4 flagellar assembly, and regulation of the actin cytoskeleton (**Table ST2**). In addition, we
5 estimated several differentially abundant genes between forest and pasture soils that were
6 associated with plant-microbe and microbe-microbe interactions. For example, genes associated
7 with biosynthesis of other secondary metabolites such as of benzoxazinoid, which are plant
8 secondary metabolites, typically from grass species, and a beta-lactam resistance gene observed
9 to be higher in pasture metagenomes. In contrast, a streptomycin synthesis gene encoding for a
10 streptomycin-6-phosphatase (EC:3.1.3.39) was higher in forest metagenomes. Within the two-
11 component system pathway, the abundance of genes for multidrug transport proteins (K07788,
12 K07789) was greater in forest, while bacitracin resistance response regulator proteins (K11629,
13 K11630) were estimated to be higher in pasture metagenomes. Several genes (K13448, K13429,
14 K13462, K13428, K13447 and K13424) that are known to be associated with plant-pathogen
15 interaction pathways were more abundant in pasture metagenomes, while none were detected in
16 forest metagenomes.

17 **Relationship between the KO gene composition and soil physicochemical properties.** To
18 explore potential drivers of community genomic composition, we performed an environmental
19 fitting of 22 soil variables onto an ordination plot via PCoA. KO genes clustered by land use
20 types [(ANOSIM: $R = 0.98$, $P = 0.007$ (**Fig. SF6**)]. Negative correlations of PCo1 with
21 Carbon/Nitrogen (C/N) ($r^2=0.95$, $P < 0.001$), temperature ($r^2=0.97$, $P < 0.001$), base saturation
22 (V; $r^2=0.78$, $P < 0.01$), and Fe ($r^2=0.72$, $P < 0.01$) demonstrate that pasture samples have
23 increasing gradients of these factors, whilst forest samples are at opposite ends of these

1 gradients. Similarly, forest samples have increasing gradients of exchangeable acidity ($H^+ + Al^{3+}$),
2 Sulfur (S), and Boron (B), whereas pasture samples are opposite to them.

3

4 **Impact of land use change on viral composition and diversity.** While evaluating the impact of
5 land use change to genomic content of microbial communities, we observed a large number of
6 reads associated with viral sequences and asked whether land use change impacted the soil
7 virome. We observed that alpha diversity of viral community is higher in pasture compared to
8 forest (**Fig. 6A**), with significant compositional differences in these two contrasting ecosystems
9 (ANOSIM: $R = 0.67$, $P = 0.008$; **Fig. 6B**). We compared the viral populations at the family level
10 between the two land use types and estimated that *Siphoviridae*, *Podoviridae*, an unclassified
11 family, and *Myoviridae* were most abundant groups, collectively comprised over 80% and 90%
12 of the sequences found in pasture and forest soils, respectively (**Fig. 6C**). While the relative
13 abundance of sequences for the above families did not differ between land uses, we observed
14 significant differences for several viral families, such as *Bicaudaviridae*, *Microviridae*,
15 *Caulimoviridae*, *Tectiviridae* and *Flaviviridae*. We identified 25 viral species that were enriched
16 in forest soils, and 67 species that were enriched in pasture soils (**Table ST3**). In addition, we
17 observed significant compositional associations of viral species with bacteria [$R^2=0.84$, $P <$
18 0.05], archaea [$R^2=0.93$, $P < 0.05$], and KO genes [$R^2=0.86$, $P < 0.05$] profiles (**Figs.S7A, S7B,**
19 **and S7C**).

20

21

22 **DISCUSSION**

23 By applying a multifactorial strategy of deep sequencing (mean 636 million reads per
24 sample) and a machine learning approach, we compared taxonomic differences and generated

1 metabolic maps with differential representation of genes involved in the cycling of nutrients and
2 greenhouse gases in two land use types in the Amazon basin: a pristine forest and a 38 years old
3 pasture. Importantly, we reported diversity and compositional differences of viral communities
4 in forest and pasture, which remained unstudied for these tropical ecosystems. The alteration of
5 bacterial and archaeal compositions and diversities following deforestation, which we observed
6 in this study, mirrors previous observations (5, 14–16). These microbial communities are
7 compositionally in strong concordance with protein-coding (KO) genes, suggesting that soil
8 samples share many of these species, with associated KO genes and metabolic processes. In
9 addition, our study demonstrated that ecosystem conversion results in increased homogenization
10 of protein coding genes (KO), which may be explained by the reduced substrate heterogeneity in
11 the pasture dominated with only two grass species. These shifts in biochemical pathways imply
12 changes in the microbial physiological strategies. Importantly, these differences in metagenomes
13 between forest and pasture are strongly associated with variation of energy and carbohydrate
14 metabolisms; two processes of importance to carbon and nitrogen cycling. While most of the
15 major functional characteristics are maintained in both ecosystems, relative abundances of many
16 specific genes in these two processes substantially differed, indicating the potential for
17 differences in physiological status and nutrient cycling between forest and pasture.

18 Plants are major contributors of organic carbon to soils through rhizodeposition, and
19 microbial communities utilize these carbon compounds (40). The metagenomes of pasture and
20 forest show signatures of resulting resource differences. The pasture soil is rich in cellulosic litter
21 from grasses, and one of the two grass species that dominates in pasture is *Urochloa brizantha*,
22 which stores nutrients including starch materials in their rhizomes (41). Along with animal dung
23 in pastureland, these carbon compounds potentially represent rich sources of fermentable

1 carbohydrates. In the pasture, the increased occurrence of genes involved in starch and sucrose
2 metabolism, and the transport of mono- and di-saccharides suggest that microbes may rely more
3 on exogenous saccharides than other substrates. This may explain why pasture metagenomes
4 have elevated counts of genes for glycolysis, TCA cycle and acetate fermentation. In
5 comparison, the forest metagenome was enriched in genes involved in gluconeogenesis. Higher
6 abundance of pentose phosphate pathway (PPP) genes in forest metagenomes may provide
7 flexibility to glycolysis/gluconeogenesis by funneling glyceraldehyde-3-P, an intermediate of the
8 PPP, into these pathways. Here, we report that forest metagenomes had higher biosynthetic
9 capacity of essential amino acids of oxaloacetate, pyruvate and PPP families, which potentially
10 favors the growth of the bacterivorous and detritivorous animal component of the soil ecosystem.
11 On the other hand, higher gene abundances for amino acid biosynthesis in the α -ketoglutarate
12 family were demonstrated in microbial communities of pastureland. An amino acid of this
13 family, glutamate, is the precursor for chlorophyll synthesis (42), which may support the
14 increased capacity for photosynthesis and carbon fixation of microbial communities in
15 pastureland. This is in agreement with previous findings of increased microbial biomass for the
16 same pasture evaluated in our study (43).

17 In contrast to the pasture, the forest soil is a more oligotrophic environment, and along
18 with a lower concentration of substrates, there is also a higher diversity of C compounds due to
19 the complex secondary chemistry of a myriad of forest vegetation (44, 45). These complex
20 compounds are degraded by soil microorganisms into a few central aromatic intermediates (*e.g.*,
21 benzoate, catechol, etc.) through peripheral catabolic pathways that are subsequently channeled
22 into the β -keto adipate pathway by *ortho* ring cleavage (46). Among others, nitrate-respiring
23 bacteria, dominated by pseudomonads, cleave the benzene-ring of these aromatic compounds

1 (47–50), which leads to the generation of TCA cycle intermediates. Here, it is noteworthy that
2 *Pseudomonas* spp. and genes related to the degradation of these compounds are mostly enriched
3 in forest soils, suggesting the maintenance of TCA cycle by replacing oxaloacetate and pyruvate
4 that are removed for the biosynthesis of glucose and amino acids. The overrepresentation of
5 genes associated with glyoxylate metabolism in forest may also contribute to carbon anaplerosis
6 by supplying malate and its subsequent conversion to oxaloacetate. Together, the carbohydrate
7 metabolic processes in forest microbial communities not only would maintain a better trade-off
8 between glycolysis and gluconeogenesis but also provide metabolic plasticity for energy harvest
9 and anaplerosis, a strategy that may be critical for the oligotrophic forest soil environment.

10 Prominent differences imposed by ecosystem conversion in the Amazon involve
11 pathways related to energy metabolism and observed differences involve metabolisms of
12 methane and nitrogen, which support previous observations (6, 15, 16, 25, 27, 51). Our data
13 demonstrate that *Euryarchaeota*, where known methanogens belong (25), and several known KO
14 genes associated with methanogenesis are elevated following deforestation. Increased capacity of
15 glycolysis and acetate fermentation may fuel the methanogenesis in pasture, which subsequently
16 may increase the efficiency of fermentation by scavenging acetate and producing carbon dioxide.
17 In contrast, methane-consuming microbial taxa and associated genes involved in this pathway
18 have lower abundances in the pasture, particularly due to the loss of proteobacterial
19 methanotrophs including *Methylomonas*, *Methylobacterium* and *Methylocella*. In addition to
20 differences in methane cycling, our results highlighted differences in pathways of nitrogen
21 cycling. Higher gene content to nitrify nitroalkanes in the forest suggests that nitroalkanes may
22 serve as sole sources of nitrogen, carbon, and energy in its microbial communities (52, 53). Our
23 results also suggest that nitrifying archaeal phylum *Thaumarchaeota* and several species of it are

1 more frequently distributed in forest soils, which was also shown in our previous study using
2 functional marker genes (24). In addition, we observed higher abundance of several species of
3 *Alphaproteobacteria* in forest soils known to be associated with steps of the nitrogen cycle, such
4 as *Nitrobacter* sp. LIP in nitrification, and *Rhodoplanes* sp. HMD01017 in denitrification. Higher
5 capacity for energy-yielding denitrification in forest soils and other oligotrophic habitats leads to
6 the increased production of nitrous oxide and nitrogen gases (Tiedje, 1988), which is evident in
7 the Amazon (55).

8 Soil physicochemical properties not only impact on microbial composition, but also the
9 rate of biogeochemical processes (57, 58). Previous reports suggest that temperature is a key
10 factor that regulates many terrestrial processes including soil respiration (59), nitrification (60),
11 denitrification (61) and methane emission (62). In addition, the C/N ratio has substantial impacts
12 on nitrogen mineralization and nitrification rates (63–65), and therefore on denitrification.
13 Deforestation by slash-and-burn processes causes a variety of changes in chemical factors (i.e.,
14 nutrients) that depend on land use type and depth, time since conversion, soil treatment and
15 amendments, and subsequent land use (66). To the extent that available nitrogen declines, there
16 is likely selective pressure on soil microbes to scavenge atmospheric nitrogen. This probably
17 explains the higher capacity of pasture for nitrogen fixation, consistent with previous works
18 showing increased abundance of the *nifH* gene (22, 67). The higher C/N ratio and the lability of
19 carbon sources, evidenced by overrepresentation of genes associated with starch and sucrose
20 metabolism in the pasture, are suggestive of reduced capacity for nitrification and therefore
21 denitrification. On the other hand, a meta-analysis by Rustad et al. (2001) reported that warming
22 in the range 0.3–6.0°C increased soil respiration rates by 20% and net N mineralization rates by
23 46%. Temperature induced increased respiration would lead to a depletion of oxygen in the

1 terrestrial ecosystem, which subsequently would trigger anaerobic energy-yielding processes
2 such as fermentation, denitrification and methanogenesis. This observation has important
3 implications. The temperature difference between our study sites (24.86°C in forest versus
4 27.14°C in pasture) (15) is similar to what is predicted by 2050 in the Amazon (68). While both
5 soils are sources of greenhouse gases (nitrous oxide, nitric oxide, CO₂, CH₄), increasing
6 temperature have higher potential for the generation of these greenhouse gases from both
7 ecosystems. Therefore, our study in two adjacent ecosystems along temperature gradient
8 provides an opportunity to explore the long-term impact of climate-driven increase in
9 temperature on ecosystem processes in the Amazon soils.

10 Finally, our deep sequencing approach allowed us to obtain a glimpse of the Amazon
11 virome. Viruses are reservoirs of many protein-coding genes of prokaryotic hosts (69), with the
12 potential to influence nutrient cycling, and the adaptation and evolution. For instance,
13 *Pseudomonas* phage PAJU2, a lysogenic phage (70), was estimated to be the most abundant
14 virus in forest soil and its relative abundance is about 8-fold higher compared to the pasture soils,
15 suggesting its important role in maintaining the population size and activity of ecologically
16 important *Pseudomonas* hosts in forest soils. Importantly, the enormous microbial and viral
17 deposits in the forest soil may have novel infectious strains or virulent strains of known
18 environmental microorganisms. Clearing the pristine forest may pose a potential threat to human
19 and animal health. Following the arrival of Europeans in the New World, millions of Native
20 Americans in or near the Amazon basin were killed due to outbreaks of deadly infections,
21 notably smallpox. Our results showed that *Molluscum contagiosum* virus and *Vaccinia* virus, two
22 members of *Poxviridae* family are overrepresented in forest soils. Such soils could preserve
23 cowpox, and even the otherwise extinct smallpox virus. Increased incidences of poxvirus

1 infections in cattle and humans have been reported near the Amazon Region of western
2 Colombia (71) and throughout Brazil (72), coinciding with the increased rate of deforestation in
3 the Amazon. *Wolbachia* phage WO is also overrepresented in forest soils. As one of most
4 widespread reproductive parasites in the biosphere (73), *Wolbachia* influences the reproductive
5 capacity of hosts, including mosquitos, and lytic WO phage is associated with the reduction of
6 *Wolbachia* densities and reproductive parasitism (74). For this and other reasons (e.g., increased
7 puddle formation due to tree stump removal and cattle trampling), forest-to-pasture conversion
8 may drive mosquito-borne disease events, including the recent zika outbreak, and continuous
9 malaria threats in Brazil and Peru near the deforested areas (75). Increased malaria also followed
10 deforestation in Malaysian Borneo (76). Knowledge of the disease ecology of deforestation is
11 still rudimentary, underscoring the need for increased surveillance of potential disease outbreaks
12 near deforested regions.

13
14 **Caveats.** We acknowledge caveats for our study. First, the trade-off between deep sequencing
15 and low sample size prevented a complete exploration of spatial variability in metagenome
16 profiles. However, previous sequencing and empirical studies covering broader spatial ranges in
17 the pristine Amazon and deforested areas are in line with the current observations, suggesting
18 that higher sample sizes will further consolidate the observed differences between forest and
19 pasture. Second, while a vast majority of the sequence reads passed the quality filtering, only
20 13.4% of the sequence reads were annotated, a percentage similar to other soil shotgun
21 metagenomic studies (77, 78). The biology of a vast majority of microbial species remains
22 unknown, limiting full annotation potential genes in metagenomes regardless of database used.
23 Third, metagenomes analyzed in the study provide a snapshot of the functional characteristics of
24 microbial communities at a given time, which may not illuminate important phenomena such as

1 disturbance, recovery, and succession. Fourth, there is a lack of well-curated databases and well-
2 established methods for annotating terrestrial virome at species level. In general, viral sequences
3 are poorly represented at species level in public databases. There are some platforms that are
4 designed and optimized for the detection of viruses of specific kinds such as VirSorter (79) and
5 vConTACT (80) for bacterial and archaeal viruses, and ViromeScan (81) for eukaryotic and
6 human viruses. While there are methods that are recently developed for assembled (VirMap) (82)
7 and unassembled (FastViromeExplorer) (83) metagenomic sequences, none of them has been
8 validated for soil virome. Therefore, in order to be consistent with the algorithm and database
9 used for annotation of prokaryotic taxonomy, we also used RefSeq database to annotate viral
10 taxonomy through the MG-RAST method that provides species level resolution.

11

12

13

14

15 **Materials and Methods**

16 **Site description and sampling.** The study was performed at the Amazon Rainforest Microbial
17 Observatory (ARMO), located at Fazenda Nova Vida in the State of Rondonia, Brazil
18 ($10^{\circ}10'18.71''\text{S}$, $62^{\circ}47'15.67''\text{W}$) (5). This location has one of the highest rates of deforestation
19 in the Brazilian Amazon (84), representing the last four decades of man-made ecosystem
20 alterations occurring in this region. Detailed information about the vegetation and soil properties
21 of our study sites have been described elsewhere (6, 9). The primary forest is a typical wet
22 “Terra Firme” forest with the majority of the trees yet to be identified, while the pasture was
23 established in 1972 after a slash-and-burn procedure followed by aerial seeding of the fast-

1 growing grasses *Urochloa brizantha* and *Panicum maximum*. The sampling strategy and sample
2 processing were described in detail elsewhere (5, 25). Briefly, in each of the forest and pasture
3 ecosystems, five nested quadrats were previously established within a 1-hectare plot. These two
4 sampling plots are 3.66 km apart. For this study, sampling points were located at the corners of
5 each quadrat along a transect, and soil cores were collected with a 10 cm-depth by 5 cm-diameter
6 corer after the removal of the litter layer. The five soil cores for each land use type were
7 collected in 2010, frozen on the spot, transported on dry ice to the laboratory and stored at -80°C
8 until DNA extractions. Soil physicochemical properties were analyzed at the Laboratório de
9 Fertilidade do Solo, Department of Soil Sciences, University of Sao Paulo. At the time of
10 sampling, *in situ* measurements of soil pH averaged 3.96 and 4.24, while soil temperature
11 averaged 24.8°C and 27.2°C for forest and pasture, respectively.

12

13 **Soil DNA extraction and sequencing.** Soil DNA extraction was performed with the
14 PowerLyzer Powersoil DNA isolation kit (Mobio Inc., Carlsbad, CA). Five DNA extractions
15 were performed per soil sample and pooled for metagenomic library preparation with the
16 TrueSeq DNA kit (Illumina Inc., San Diego, CA). Sequencing was performed on the Illumina
17 HiSeq platform to generate 6.4 billion of paired-end reads of 150bp.

18

19 **Sequence processing and annotation of metagenomes.** The Amazon soil metagenomic dataset
20 is publicly available and deposited in the Rapid Annotation using Subsystems Technology for
21 Metagenomes (MG-RAST) database (85). A link for a total of 83 MG-RAST IDs of five forest
22 and five pasture samples is provided in Meyer et al. (2017). Quality filtering of the raw
23 sequences was performed using the default criteria as set by the MG-RAST pipeline (i.e.

1 trimming of low-quality bases, removal of artificial replicate sequences and filtering of
2 sequences with greater than 5 ambiguous bases). As part of the MG-RAST pipeline version 3.2,
3 paired-end reads were joined using the fastq-join algorithm. Single end reads with 150 bp length
4 that could not be joined were retained. Quality filtered sequence reads were searched against the
5 Kyoto Encyclopedia of Genes and Genomes (KEGG) (86), and RefSeq (87) databases for
6 functional and taxonomic (bacterial, archaea and viruses) annotations, respectively. The
7 annotations were conducted at an e-value cutoff of $1e-5$, a minimum identity cutoff of 60%, and
8 a minimum alignment length cutoff of 15 (default parameters) for the annotations of
9 metagenome sequences. Following the annotations, the datasets were exported in biom file
10 format and merged by QIIME 1.8.0 (88) for the downstream analyses.

11
12 **Data analysis.** Rarefaction plots were generated to check whether sampling effort was sufficient.
13 We then compared the relative abundances of major bacterial, archaeal and viral taxa, and
14 functional categories of genes (KEGG level 2) between forest and pasture ecosystems. Alpha
15 diversities were calculated using species (or gene) richness and the Shannon index (H'). Bray-
16 Curtis dissimilarity was used to calculate the pairwise dissimilarities and perform principal
17 coordinate analysis (PCoA) between metagenomes. To test the hypothesis that forest and pasture
18 have differential metagenome profiles, PCoA plots were generated using the species level
19 abundance dataset of bacteria and archaea, where functional categories at KEGG level 2 were
20 fitted as vectors. Here we used two *R* packages- ‘ape’ (89) and ‘vegan’ (90) to calculate the
21 association of each ecosystem with different functional attributes of the bacterial and archaeal
22 communities.

23

1 **Statistical Analysis.** To determine whether differences in the relative abundances of major
2 taxonomic and functional categories between ecosystems were statistically significant, the
3 Mann-Whitney test with 1000 permutations was used. Performing the same statistical test, we
4 compared the alpha diversities, and pairwise taxonomic distances between samples. Analyses of
5 similarities (ANOSIM) tests were performed to assess whether metagenome compositions were
6 significantly different across the terrestrial ecosystems.

7 We employed the DESeq2 method (91), a negative binomial Wald test, to identify
8 microbial and viral species, and protein-coding genes at KEGG level 4 (KEGG orthology, KO)
9 that were significantly affected by forest-to-pasture conversion. The negative binomial
10 distribution is particularly useful for metagenomic datasets, where observations have high
11 variation among biological replications, and sample variance is typically greater than the sample
12 mean (91). This method also calculates fold-difference in gene abundance, which is represented
13 as the \log_2 -fold-change (\log_2FC). A score of zero indicates that the KO gene or species has
14 statistically the same proportional abundance in metagenomes of both land use types. A positive
15 value for a given gene or species indicates overrepresentation in forest metagenomes compared
16 to pasture, while a negative value indicates the reverse pattern. We also employed a supervised
17 classifier, Random Forest using the *R* package ‘*randomForest*’ (92), to classify a subset of soil
18 species and protein-coding genes that are discriminatory between land uses. To estimate the
19 accuracy of the model, a subset of the samples was randomly partitioned as a training set to build
20 the model, which was then used to classify the remaining samples as test set (cross-validation).
21 With leave-one-out cross validation, suitable for small sample sizes, the average estimated
22 generalization error of the classifier set was calculated and compared with the baseline error for
23 random guessing. We calculated the ratio of baseline error and estimated error, and considered

1 this ratio of two or higher for a good classification, meaning the Random Forest classifier
2 performs at least two-fold better than random guessing. Additionally, this classifier calculates an
3 importance score for each feature (i.e., species or protein-coding gene) by estimating the increase
4 in estimated error when a feature is ignored. In our analysis, we considered 0.0001 as the cutoff
5 for importance score. The use of Random Forest in conjunction with DESeq2 was conservative,
6 in that more genes and species passed only the DESeq2 screen, and fewer passed both. Together,
7 both of the statistical tests increased the robustness in identifying the differentially abundant
8 species and genes between ecosystems with their levels of importance in the ecosystems.

9 QIIME 1.9.0 and GraphPad Prism 7.00 were used for performing statistical analyses and
10 visualizing the results.

11

12 **Availability of data and materials.** Metagenomic sequences for all samples used in this study
13 are available at the Joint Genome Institute IMG/M database under the following accession
14 numbers: 1080879 to 1080888.

15

16 **CONFLICT OF INTEREST**

17 The authors declare that they have no conflict of interests.

18

19 **ACKNOWLEDGMENTS**

20 The authors thank the owners and staff of Agropecuaria Nova Vida for logistical support and
21 permission to work on their property. We are grateful to Siu M. Tsai, Rebecca Mueller, Fabiana
22 Paula, Babur Mirza, and Wagner Piccinini for assistance with fieldwork.

23

1

2

3 REFERENCES

4

- 5 1. Dirzo R, Raven PH. 2003. Global state of biodiversity and loss. *Annu Rev Environ Resour*
- 6 28:137–167.
- 7 2. Betts R, Sanderson M, Woodward S. 2008. Effects of large-scale Amazon forest
- 8 degradation on climate and air quality through fluxes of carbon dioxide, water, energy,
- 9 mineral dust and isoprene, p. 1873–1880. *In Philosophical Transactions of the Royal*
- 10 *Society B: Biological Sciences*. Royal Society.
- 11 3. Fearnside PM. 1999. Biodiversity as an environmental service in Brazil’s Amazonian
- 12 forests: risks, value and conservation. *Environ Conserv* 26:305–321.
- 13 4. Feeley KJ, Silman MR. 2009. Extinction risks of Amazonian plant species PNAS.
- 14 5. Rodrigues JLM, Pellizari VH, Mueller R, Baek K, Jesus EDC, Paula FS, Mirza B,
- 15 Hamaou GS, Tsai SM, Feigl B, Tiedje JM, Bohannan BJM, Nüsslein K. 2013.
- 16 Conversion of the Amazon rainforest to agriculture results in biotic homogenization of
- 17 soil bacterial communities. *Proc Natl Acad Sci U S A* 110:988–993.
- 18 6. Neill C, Melillo JM, Steudler PA, Cerri CC, De Moraes JFL, Piccolo MC, Brito M. 1997.
- 19 Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern
- 20 Brazilian Amazon. *Ecol Appl* 7:1216–1225.
- 21 7. Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E,
- 22 Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld
- 23 M, Poff NLR, Sykes MT, Walker BH, Walker M, Wall DH. 2000. Global biodiversity
- 24 scenarios for the year 2100. *Science* (80-).
- 25 8. Eva HD, Belward AS, De Miranda EE, Di Bella CM, Gond V, Huber O, Jones S,
- 26 Sgrenzaroli M, Fritz S. 2004. A land cover map of South America. *Glob Chang Biol*
- 27 10:731–744.
- 28 9. Feigl B, Cerri C, Piccolo M, Noronha N, Augusti K, Melillo J, Eschenbrenner V, Melo L.
- 29 2006. Biological survey of a low-productivity pasture in Rondônia state, Brazil. *Outlook*
- 30 *Agric* 35:199–208.
- 31 10. Foley JA, Asner GP, Costa MH, Coe MT, DeFries R, Gibbs HK, Howard EA, Olson S,
- 32 Patz J, Ramankutty N, Snyder P. 2007. Amazonia revealed: Forest degradation and loss of
- 33 ecosystem goods and services in the Amazon Basin. *Front Ecol Environ*.
- 34 11. Borneman J, Triplett EW. 1997. Molecular Microbial Diversity in Soils from Eastern
- 35 Amazonia: Evidence for Unusual Microorganisms and Microbial Population Shifts
- 36 Associated with Deforestation APPLIED AND ENVIRONMENTAL MICROBIOLOGY.
- 37 12. Ranjan K, Paula FS, Mueller RC, Jesus E da C, Cenciani K, Bohannan BJM, Nüsslein K,
- 38 Rodrigues JLM. 2015. Forest-to-pasture conversion increases the diversity of the phylum
- 39 Verrucomicrobia in Amazon rainforest soils. *Front Microbiol* 6.
- 40 13. Jesus EDC, Marsh TL, Tiedje JM, Moreira FMDS. 2009. Changes in land use alter the
- 41 structure of bacterial communities in Western Amazon soils. *ISME J* 3:1004–1011.
- 42 14. Navarrete A, Taketani R, ... LM-RB, 2011 U. 2011. Land-use systems affect archaeal
- 43 community structure and functional diversity in western Amazon soils. *Rev Bras Ciência*
- 44 *do Solo* 35:1527–1540.
- 45 15. Khan M, Bohannan B, ... KN-F microbiology, 2019 U. 2019. Deforestation impacts

- 1 network co-occurrence patterns of microbial communities in Amazon soils. *FEMS*
2 *Microbiol Ecol* 95:1–12.
- 3 16. Khan MAW. 2016. RESPONSE OF MICROBIAL NETWORKS AND MICROBIOMES
4 TO THE FOREST-TO-PASTURE CONVERSION IN AMAZON SOILS.
- 5 17. Mueller RC, Rodrigues JLM, Nüsslein K, Bohannan BJM. 2016. Land use change in the
6 Amazon rain forest favours generalist fungi. *Funct Ecol* 30:1845–1853.
- 7 18. Garbeva P, Van Veen JA, Van Elsas JD. 2004. Microbial diversity in soil: Selection of
8 microbial populations by plant and soil type and implications for disease suppressiveness.
9 *Annu Rev Phytopathol*.
- 10 19. Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW,
11 Adler PB, Bakker JD, Cleland EE, Decrappeo NM, Delorenze E, Hagenah N, Hautier Y,
12 Hofmockel KS, Kirkman KP, Knops JMH, La Pierre KJ, Macdougall AS, Mcculley RL,
13 Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N. 2015. Plant
14 diversity predicts beta but not alpha diversity of soil microbes across grasslands
15 worldwide. *Ecol Lett* 18:85–95.
- 16 20. Lange M, Eisenhauer N, Sierra CA, Bessler H, Engels C, Griffiths RI, Mellado-Vázquez
17 PG, Malik AA, Roy J, Scheu S, Steinbeiss S, Thomson BC, Trumbore SE, Gleixner G.
18 2015. Plant diversity increases soil microbial activity and soil carbon storage. *Nat*
19 *Commun* 6.
- 20 21. Mirza BS, Pottap C, Nüsslein K, Bohannan BJM, Rodrigues JLM. 2014. Response of
21 free-living nitrogen-fixing microorganisms to land use change in the amazon rainforest.
22 *Appl Environ Microbiol* 80:281–288.
- 23 22. Paula FS, Rodrigues JLM, Zhou J, Wu L, Mueller RC, Mirza BS, Bohannan BJM,
24 Nüsslein K, Deng Y, Tiedje JM, Pellizari VH. 2014. Land use change alters functional
25 gene diversity, composition and abundance in Amazon forest soil microbial communities.
26 *Mol Ecol* 23:2988–2999.
- 27 23. Gouvêa Taketani R, Mui Tsai S. The Influence of Different Land Uses on the Structure of
28 Archaeal Communities in Amazonian Anthrosols Based on 16S rRNA and amoA Genes.
29 Springer <https://doi.org/10.1007/s00248-010-9638-1>.
- 30 24. Jr GH, Rodrigues J, Ecology BB-A soil, 2016 U. 2016. Land-use change drives abundance
31 and community structure alterations of thaumarchaeal ammonia oxidizers in tropical
32 rainforest soils in Rondônia, Brazil. *Appl Soil Ecol* 107:48–56.
- 33 25. Meyer KM, Klein AM, Rodrigues JLM, Nüsslein K, Tringe SG, Mirza BS, Tiedje JM,
34 Bohannan BJM. 2017. Conversion of Amazon rainforest to agriculture alters community
35 traits of methane-cycling organisms. *Mol Ecol* 26:1547–1556.
- 36 26. Mirza BS, Mcglinn DJ, Bohannan BJM, Nüsslein K, Tiedje JM, Rodrigues JLM, Nojiri H.
37 2020. Diazotrophs Show Signs of Restoration in Amazon Rain Forest Soils with
38 Ecosystem Rehabilitation Downloaded from <https://doi.org/10.1128/AEM>.
- 39 27. Navarrete AA, Venturini AM, Meyer KM, Klein AM, Tiedje JM, Brendan BJ, Nüsslein
40 K, Tsai SM, Rodrigues JLM. 2015. Differential response of Acidobacteria subgroups to
41 forest-to-pasture conversion and their biogeographic patterns in the western Brazilian
42 Amazon. *Front Microbiol* 6.
- 43 28. Goss-Souza D, Mendes LW, Rodrigues JLM, Tsai SM. 2020. Ecological Processes
44 Shaping Bulk Soil and Rhizosphere Microbiome Assembly in a Long-Term Amazon
45 Forest-to-Agriculture Conversion. *Microb Ecol* 79:110–122.
- 46 29. Comeau A, Hatfull G, Krisch H, ... DL-R in, 2008 U. 2008. Exploring the prokaryotic

- 1 virosphere. *Res Microbiol* 159:306–313.
- 2 30. Frost L, Leplae R, ... AS-NR, 2005 U. 2005. Mobile genetic elements: the agents of open
3 source evolution. *Nat Rev Microbiol* 3:722–732.
- 4 31. Paul J, Biotechnology MS-CO in, 2005 U. 2005. Marine phage genomics: what have we
5 learned? *Curr Opin Biotechnol* 16:299–307.
- 6 32. Weinbauer MG, Rassoulzadegan F. 2004. Are viruses driving microbial diversification
7 and diversity? *Environ Microbiol* 6:1–11.
- 8 33. Breitbart M, Miyake J, Letters FR-F microbiology, 2004 U. 2004. Global distribution of
9 nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* 236:249–256.
- 10 34. Mühlhng M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, Post
11 AF, Joint I, Mann NH. 2005. Genetic diversity of marine *Synechococcus* and co-occurring
12 cyanophage communities: Evidence for viral control of phytoplankton. *Environ Microbiol*
13 7:499–508.
- 14 35. Pratama A, Microbiology J van E-T in, 2018 U. 2018. The 'neglected' soil virome–
15 potential role and impact. *Trends Microbiol* 26:649–662.
- 16 36. Paez-Espino D, Eloee-Fadrosch E, Nature GP-, 2016 undefined. Uncovering Earth's
17 virome. *nature.com*.
- 18 37. Emerson J, Roux S, Brum J, ... BB-N, 2018 undefined. Host-linked soil viral ecology
19 along a permafrost thaw gradient. *nature.com*.
- 20 38. Graham EB, Paez-Espino D, Brislawn C, Neches RY, Hofmockel KS, Wu R, Kyrpidis
21 NC, Jansson JK, Mcdermott JE. 2019. Untapped viral diversity in global soil
22 metagenomes. *Biorxiv* 583997.
- 23 39. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. 2017. Viruses in Soil
24 Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu Rev Virol*
25 4:201–219.
- 26 40. A. Stuart Grandy JCN. 2008. Molecular C dynamics downstream: the biochemical
27 decomposition sequence and its impact on soil organic matter structure and function. *Sci*
28 *Total Environ* 404:297–307.
- 29 41. Deinum B, Sulastri R, ... MZ-... wageningen journal of, 1996 U. 1996. Effects of light
30 intensity on growth, anatomy and forage quality of two tropical grasses (*Brachiaria*
31 *brizantha* and *Panicum maximum* var. *trichoglume*). *Netherlands J Agricultural Sci*
32 44:111–124.
- 33 42. Yaronkaya E, Vershilovskaya I, Poers Y, Alawady AE, Averina N, Grimm B. 2006.
34 Cytokinin effects on tetrapyrrole biosynthesis and photosynthetic activity in barley
35 seedlings. *Planta*.
- 36 43. Cenciani K, Lambais M, ... CC-RB, 2009 U. 2009. Bacteria diversity and microbial
37 biomass in forest, pasture and fallow soils in the southwestern Amazon basin. *R Bras Ci*
38 *Solo* 33:907–916.
- 39 44. Ohno T, Parr TB, Gruselle cile I, Fernandez IJ, Sleighter RL, Hatcher PG. 2014.
40 Molecular Composition and Biodegradability of Soil Organic Matter: A Case Study
41 Comparing Two New England Forest Types. *ACS Publ* 48:7229–7236.
- 42 45. Lee YJ, Narasimhan K, Swarup S. 2013. Enhancement of Plant-Microbe Interactions
43 using Rhizosphere Metabolomics-Driven Approach and its Application in the Removal of
44 Polychlorinated Biphenyls, p. 1191–1198. *In* *Molecular Microbial Ecology of the*
45 *Rhizosphere*. John Wiley and Sons.
- 46 46. Díaz E. 2004. Bacterial degradation of aromatic pollutants: A paradigm of metabolic

- 1 versatility. *Int Microbiol. Sociedad Espanola de Microbiologia.*
- 2 47. Heinaru E, Viggor S, Vedler E, ... JT-FM, 2001 U. 2001. Reversible accumulation of p-
3 hydroxybenzoate and catechol determines the sequential decomposition of phenolic
4 compounds in mixed substrate cultivations in. *FEMS Microbiol Ecol* 37:79–89.
- 5 48. Atashgahi S, Hornung B, Van Der Waals MJ, Da Rocha UN, Hugenholtz F, Nijssse B,
6 Molenaar D, Van Spanning R, Stams AJM, Gerritse J, Smidt H. 2018. A benzene-
7 degrading nitrate-reducing microbial consortium displays aerobic and anaerobic benzene
8 degradation pathways. *Sci Rep* 8.
- 9 49. Cao B, Bioengineering KL-B and, 2008 U. 2008. Catabolic pathways and cellular
10 responses of *Pseudomonas putida* P8 during growth on benzoate with a proteomics
11 approach. *Biotechnol Bioeng* 101:1297–1312.
- 12 50. Biochemistry I, Ornston LN, Stanier RY. 1966. The Conversion of Catechol and
13 Protocatechuate to g-Ketoadipate by *Pseudomonas putida* THE JOURNAL OF
14 BIOLOGICAL CHEMISTRY.
- 15 51. Neill C, Piccolo M, Steudler P, ... JM-SB and, 1995 U. 1995. Nitrogen dynamics in soils
16 of forests and active pastures in the western Brazilian Amazon Basin. *Soil Biol Biochem*
17 27:1167–1175.
- 18 52. Kido T, Hashizume K, Soda K. 1978. Purification and Properties of Nitroalkane Oxidase
19 from *Fusarium oxysporum* Downloaded from JOURNAL OF BACTERIOLOGY.
- 20 53. Spain J, Hughes J, Knackmuss H. 2000. Biodegradation of nitroaromatic compounds and
21 explosives. CRC Press.
- 22 54. microorganisms JT-B of anaerobic, 1988 undefined. Ecology of denitrification and
23 dissimilatory nitrate reduction to ammonium. researchgate.net.
- 24 55. Melillo J, Steudler P, Feigl B, ... CN-J of, 2001 U. 2001. Nitrous oxide emissions from
25 forests and pastures of various ages in the Brazilian Amazon. *J Geophys Res Atmos*
26 106:34179–34188.
- 27 56. Tiedje JM. 1988. Ecology of denitrification and dissimilatory nitrate reduction to
28 ammonium. *Biol Anaerob Microorg* 717:179–244.
- 29 57. Rastetter EB, Perakis SS, Shaver GR, Ågren GI. 2005. Terrestrial C sequestration at
30 elevated CO₂ and temperature: The role of dissolved organic N loss. *Ecol Appl* 15:71–86.
- 31 58. Borken W, Davidson E, ... KS-SB and, 2006 U. 2006. Effect of summer throughfall
32 exclusion, summer drought, and winter snow cover on methane fluxes in a temperate
33 forest soil. *Soil Biol Biochem* 38:1388–1395.
- 34 59. RAICH JW, SCHLESINGER WH. 1992. The global carbon dioxide flux in soil
35 respiration and its relationship to vegetation and climate. *Tellus B* 44:81–99.
- 36 60. MacDonald NW, Zak DR, Pregitzer KS. 1995. Temperature Effects on Kinetics of
37 Microbial Respiration and Net Nitrogen and Sulfur Mineralization. *Soil Sci Soc Am J*
38 59:233–240.
- 39 61. Malhi S, McGill W, Biochemistry MN-SB and, 1990 U. 1990. Nitrate losses in soils:
40 effect of temperature, moisture and substrate concentration. *Soil Biol Biochem* 22:733–
41 737.
- 42 62. Crill PM, Bartlett KB, Harriss RC, Gorham E, Verry ES, Sebacher DI, Madzar L, Sanner
43 W. 1988. Methane flux from Minnesota Peatlands. *Global Biogeochem Cycles* 2:371–384.
- 44 63. Strauss EA, Lamberti GA. 2000. Regulation of nitrification in aquatic sediments by
45 organic carbon. *Limnol Oceanogr*.
- 46 64. Knoepp J, Soils WS-B and F of, 2002 undefined. 2002. Using soil temperature and

- 1 moisture to predict forest soil nitrogen mineralization. Springer 36:177–182.
- 2 65. Paul K, Polglase P, ... AO-E journal of, 2003 U. 2003. Defining the relation between soil
3 water content and net nitrogen mineralization. Eur J Soil Sci 54:39–47.
- 4 66. Mcgrath DA, Smith CK, Gholz HL, De F, Oliveira A. 2001. Effects of Land-Use Change
5 on Soil Nutrient Dynamics in Amazonia. Ecosystems 4:625–645.
- 6 67. Mirza BS, Rodrigues JLM. 2012. Development of a direct isolation procedure for free-
7 living diazotrophs under controlled hypoxic conditions. Appl Environ Microbiol 78:5542–
8 5549.
- 9 68. Nobre CA, Sampaio G, Borma LS, Castilla-Rubio JC, Silva JS, Cardoso M. 2016. Land-
10 use and climate change risks in the amazon and the need of a novel sustainable
11 development paradigm. Proc Natl Acad Sci U S A 113:10759–10768.
- 12 69. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues
13 C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J,
14 Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White
15 BA, Rohwer F. 2008. Functional metagenomic profiling of nine biomes. Nature 452:629–
16 632.
- 17 70. Uchiyama J, Rashel M, Matsumoto T, Research YS-V, 2009 U. 2009. Characteristics of a
18 novel Pseudomonas aeruginosa bacteriophage, PAJU2, which is genetically related to
19 bacteriophage D3. Virus Res 139:131–134.
- 20 71. Usme-Ciro J, Paredes A, ... DW-E infectious, 2017 U. 2017. Detection and molecular
21 characterization of zoonotic poxviruses circulating in the Amazon region of Colombia,
22 2014. Emerg Infect Dis 23:649–653.
- 23 72. Trindade G, Emerson G, ... DC-E infectious, 2007 U. 2007. Brazilian vaccinia viruses
24 and their origins. Emerg Infect Dis 13:965–972.
- 25 73. Zug R, Hammerstein P. 2012. Still a host of hosts for Wolbachia: Analysis of recent data
26 suggests that 40% of terrestrial arthropod species are infected. PLoS One 7.
- 27 74. Bordenstein SR, Marshall ML, Fry AJ, Kim U, Wernegreen JJ. 2006. The tripartite
28 associations between bacteriophage, Wolbachia, and arthropods. PLoS Pathog 2:384–393.
- 29 75. Joseph Kamtchum-Tatuene, Benjamin L. Makepeace, Laura Benjamin, Matthew Baylis
30 and TS, Purpose. 2017. The potential role of Wolbachia in controlling the transmission of
31 emerging human arboviral infections. Curr Opin Infect Dis 30:108–116.
- 32 76. Fornace KM, Abidin TR, Alexander N, Brock P, Grigg MJ, Murphy A, William T, Menon
33 J, Drakeley CJ, Cox J. 2016. Association between landscape factors and spatial patterns of
34 Plasmodium knowlesi infections in Sabah, Malaysia. Emerg Infect Dis 22:201–208.
- 35 77. Delmont TO, Prestat E, Keegan KP, Faubladier M, Robe P, Clark IM, Pelletier E, Hirsch
36 PR, Meyer F, Gilbert JA, Le Paslier D, Simonet P, Vogel TM. 2012. Structure, fluctuation
37 and magnitude of a natural grassland soil metagenome. ISME J 6:1677–1687.
- 38 78. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA,
39 Wall DH, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial
40 communities and their functional attributes. Proc Natl Acad Sci U S A 109:21390–21395.
- 41 79. Roux S, Enault F, Hurwitz B, PeerJ MS-, 2015 undefined. VirSorter: mining viral signal
42 from microbial genomic data. peerj.com.
- 43 80. Bolduc B, Jang H Bin, Doucier G, You ZQ, Roux S, Sullivan MB. 2017. vConTACT: An
44 iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria.
45 PeerJ 2017.
- 46 81. Rampelli S, Soverini M, Turrone S, Quercia S, Biagi E, Brigidi P, Candela M. 2016.

- 1 ViromeScan: A new tool for metagenomic viral community profiling. *BMC Genomics* 17.
2 82. Ajami NJ, Wong MC, Ross MC, Lloyd RE, Petrosino JF. 2018. Maximal viral
3 information recovery from sequence data using VirMAP. *Nat Commun* 9:3205.
4 83. Tithi SS, Aylward FO, Jensen R V., Zhang L. 2018. FastViromeExplorer: A pipeline for
5 virus and phage identification and abundance profiling in metagenomics data. *PeerJ*
6 2018:e4227.
7 84. INPE. 2011. Program for the Estimation of Amazon Deforestation (Projeto PRODES
8 Digital). Brazilian Nat Inst Sp Res.
9 85. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez
10 A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server
11 - A public resource for the automatic phylogenetic and functional analysis of
12 metagenomes. *BMC Bioinformatics* 9.
13 86. Kanehisa M, Research SG-N acids, 2000 U. 2000. KEGG: kyoto encyclopedia of genes
14 and genomes. *Nucleic Acids Res* 28:27–30.
15 87. Pruitt K, Tatusova T, Research DM-N acids, 2007 U. 2007. NCBI reference sequences
16 (RefSeq): a curated non-redundant sequence database of genomes, transcripts and
17 proteins. *Nucleic Acids Res* 35:D61–D65.
18 88. Gregory Caporaso J, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,
19 Fierer N, Gonzalez Peña A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,
20 Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J,
21 Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight
22 R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Publ*
23 *Gr* <https://doi.org/10.1038/NMETH.F.303>.
24 89. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in
25 R language. *Bioinforma Appl NOTE* 20:289–290.
26 90. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci*.
27 Opulus Press AB.
28 91. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion
29 for RNA-seq data with DESeq2. *Genome Biol* 15.
30 92. Liaw A, Wiener M. 2001. Classification and Regression by
31 RandomForestresearchgate.net.
32
33

34

35 **FIGURE LEGENDS**

36

37 **FIG. 1. Response of forest-to-pasture conversion to bacterial (A-B), archaeal (C-D) and**
38 **protein-coding gene (E-F) alpha and beta diversities in soil metagenomes.** Shannon index
39 and Bray-Curtis similarity were used for estimating alpha (n = 5), and beta diversities (n = 10 for

1 pairwise similarity), respectively, where mean values of diversity were depicted. Error bars
2 represent standard error of the mean (S.E.M). Statistical significance was calculated using Mann-
3 Whitney test with 1,000 permutations. Symbols *, and ** indicate significance values of $P <$
4 0.05 and $P < 0.01$, respectively.

5
6
7 **FIG. 2. Relationships of protein-coding gene compositions with bacterial and archaeal**
8 **species compositions across forest and pasture metagenomes.** The relationships of
9 compositional similarities are visualized using the principal coordinate loadings for the first axes
10 (PCo1). Protein-coding genes demonstrated significantly positive correlations with bacterial (A)
11 and archaeal (B) community compositions. Vector fitter principal coordinate analysis showed
12 that bacterial (C) and archaeal (D) communities have distinct clustering patterns by soil types,
13 where vectors represent the association of the functional characteristics at KEGG level 2. The
14 vectors are shown with arrows pointing to the direction of an increase in the gradient for the
15 corresponding variable, and with length proportional to the correlation between ordination and
16 variable. The circles (forest, green; pasture, red) represent the relative positions of each soil
17 taxonomic community. The significances (P -values) of the vectors were calculated based on 999
18 random permutations of the data. Here, only functional characteristics that were estimated to be
19 significantly ($P < 0.01$) associated with PCo1 are shown. Bray-Curtis metric was used for
20 estimating distances between different samples.

21
22 **FIG. 3. Heat map comparison of Z-scores calculated from protein-coding gene abundances**
23 **between forest and pasture metagenomes.** Only subset of genes encoding KEGG ECs

1 involved in nutrient cycling is shown and a complete list of all discriminatory protein-coding
2 genes is shown in **Table ST2**. Only genes that passed both selection criteria for DESeq2 (BH-
3 adjusted $P < 0.05$) and Random Forest (importance score > 0.0001) methods were considered.

4
5 **FIG. 4. Metabolic map for glycolysis/gluconeogenesis and the tricarboxylic acid (TCA)**
6 **cycle based on differentially abundant genes observed in forest and pasture metagenomes.**

7 Only genes that passed both selection criteria for DESeq2 (BH-adjusted $P < 0.05$) and Random
8 Forest (importance score > 0.0001) methods are represented. Arrows indicate the enzyme-
9 mediated steps of the pathway, with KEGG EC numeric classification representing the reaction
10 being catalyzed. A green arrow means that the abundance of a gene encoding KEGG EC enzyme
11 was higher in the forest, while red means higher in the pasture. Black arrows indicate genes with
12 similar abundances between forest and pasture metagenomes.

13
14 **FIG. 5. Metabolic maps for the nitrogen (A) and methane (B) metabolisms based on**
15 **differentially abundant genes observed in forest and pasture metagenomes.** Only genes that

16 passed both selection criteria for DESeq2 (BH-adjusted $P < 0.05$) and Random Forest
17 (importance score > 0.0001) methods are represented. Arrows indicate the enzyme-mediated
18 steps of the pathway, with KEGG EC numeric classification representing the reaction being
19 catalyzed. A green arrow means that the abundance of a gene encoding KEGG EC enzyme was
20 higher in the forest, while red means higher in the pasture. Black arrows indicate genes with
21 similar abundances between forest and pasture metagenomes.

22
23 **FIG. 6. Virome community structure in the metagenomes obtained from the Amazon forest**

1 **(green) and pasture (red) soils.** (A) Ordination plot from principal coordinate analysis showed
2 that forest viral composition is distinct from that of pasture. (B) Alpha diversity as estimated by
3 Shannon index is significantly higher in pasture compared to forest. (C) Relative abundances of
4 viral communities are estimated at family level in forest and pasture metagenomes. Error bars
5 represent standard error of the mean (S.E.M). Symbols * and ** indicate significance values of P
6 < 0.05 and $P < 0.01$, respectively, which were calculated using Mann-Whitney test with 1,000
7 permutations.

8

9

10 **SUPPLEMENTAL MATERIAL**

11

12 **FIG. SF1. Rarefaction curves of bacterial (A), archaeal (B), viral (C) taxonomic and**
13 **protein-coding (KO) (D) gene distributions of metagenomes obtained from Amazon forest**
14 **and pasture soil samples.** Species richness was used as a function of sequencing depth for forest
15 and pasture samples. Green and red lines indicate forest and pasture samples, respectively.

16

17 **FIG. SF2. Relative abundances of bacterial (A) and archaeal (B) communities at phylum**
18 **level in the metagenomes of Amazon forest (green) and pasture (red).** Taxa that are shown in
19 each of the bacterial and archaeal domains represent over 98% of their taxonomic sequences.
20 The most abundant phylum *Proteobacteria* is broken down into five classes. Error bars represent
21 standard error of the mean (S.E.M). P -value is calculated using Mann-Whitney test with 1,000
22 permutations and symbols * and *** indicate significance values of $P < 0.05$ and $P < 0.001$,
23 respectively.

24

1 **FIG. SF3. Relative abundances of functional categories at KEGG level 2 in the**
2 **metagenomes obtained from the Amazon forest (green) and pasture (red).** Error bars
3 represent standard error of the mean (S.E.M). Symbols * and *** indicate significance values of
4 $P < 0.05$ and $P < 0.001$, which were calculated using Mann-Whitney test with 1,000
5 permutations.

6
7 **FIG. SF4. KEGG pathway map of differentially abundant genes observed in forest and**
8 **pasture metagenomes for amino acid metabolism of (A) oxaloacetate acid, (B) pentose**
9 **phosphate pathway, (C) pyruvate, and (D) alpha-ketoglutarate.** Only genes that passed both
10 selection criteria for DESeq2 (BH-adjusted $P < 0.05$) and Random Forest (importance score $>$
11 0.0001) methods are represented. Arrows indicate the enzyme-mediated steps of the pathway,
12 with KEGG EC numeric classification representing the reaction being catalyzed. A green arrow
13 means that the abundance of a gene encoding KEGG EC enzyme was higher in the forest, while
14 red means higher in the pasture. Black arrows indicate genes with similar abundances between
15 forest and pasture metagenomes.

16 **FIG. SF5. KEGG pathway map for xenobiotic metabolism involved in benzoate, dioxin,**
17 **and polycyclic aromatic hydrocarbon degradation.** Only genes that passed both selection
18 criteria for DESeq2 (BH-adjusted $P < 0.05$) and Random Forest (importance score $>$ 0.0001)
19 methods are represented. Arrows indicate the enzyme-mediated steps of the pathway, with
20 KEGG EC numeric classification representing the reaction being catalyzed. A green arrow
21 means that the abundance of a gene encoding KEGG EC enzyme was higher in the forest, while
22 red means higher in the pasture. Black arrows indicate genes with similar abundances between
23 forest and pasture metagenomes.

1
2 **FIG. SF6. Relationships of protein-coding (KO) gene compositions with soil**
3 **physicochemical factors.** Vector fitted principal coordinate analysis of protein-coding genes
4 with vectors representing different soil factors. Each vector points to the direction of an increase
5 in the gradient for the corresponding variable and length proportional to the correlation between
6 ordination and variable. The circles (green, forest; red, pasture) represent the relative positions of
7 protein-coding genes observed in a sample. The significances (P -values) of the vectors were
8 calculated based on 999 random permutations of the data. Only soil physicochemical factors that
9 were estimated to be significantly ($P < 0.01$) associated with PCo1 are shown. Bray-Curtis
10 metric was used for estimating distances between different samples.

11
12 **FIG. SF7. Relationships of viral composition with bacterial (A), archaeal (B), and protein-**
13 **coding gene (C) compositions across forest and pasture metagenomes.** The relationships of
14 compositional similarities are visualized using the principal coordinate loadings for the first axes
15 (PCo1). The circles (green, forest; red, pasture) represent the relative positions of each soil
16 taxonomic community. Bray-Curtis metric was used for estimating distances between different
17 samples.

18
19 **TABLE ST1. A complete list of bacterial (sheet A) and archaeal (sheet B) lineages at**
20 **species level that discriminate the metagenomes between forest and pasture soils.** DESeq2
21 (BH-adjusted $P < 0.05$) and Random Forest (importance score > 0.0001) methods were
22 employed to calculate the discriminatory prokaryotic and viral species between forest and
23 pasture metagenomes. Random Forest is a supervised classifier that classifies a subset of protein-

1 coding genes that are discriminatory between ecosystems. This classifier calculates an
2 importance score for each feature (i.e., protein-coding gene) by estimating the increase in
3 generalized error when feature is ignored. In our analysis, we considered 0.0001 as the cutoff for
4 importance score. On the other hand, DESeq2 employs negative binomial distribution to identify
5 the differentially abundant protein-coding genes, where \log_2FC value is calculated to
6 demonstrate the degree of variations in abundances for each protein-coding gene. Here, the
7 positive and negative \log_2FC values indicate the genes that are differentially abundant in forest
8 and pasture metagenomes, respectively. The $lfcSE$ stands for standard error of \log_2FC estimate.
9 P -value is calculated by the Wald statistic, which is the ratio of \log_2FC and its standard error. P -
10 values were adjusted for multiple comparisons using Benjamini-Hochberg False Discovery Rate,
11 and adjusted $P < 0.05$ was used for statistical significance.

12

13 **TABLE ST2. A complete list of protein-coding genes at KEGG level 4 that discriminate the**
14 **metagenomes between forest and pasture soils.** DESeq2 and Random Forest methods were
15 employed to calculate the protein-coding genes that differentiate between forest and pasture
16 metagenomes.

17

18 **TABLE ST3. A complete list of viral lineages at species level that discriminate the**
19 **metagenomes between forest and pasture soils.** DESeq2 (BH-adjusted $P < 0.05$) and Random
20 Forest (importance score > 0.0001) methods were employed to calculate the discriminatory
21 prokaryotic and viral species between forest and pasture metagenomes.

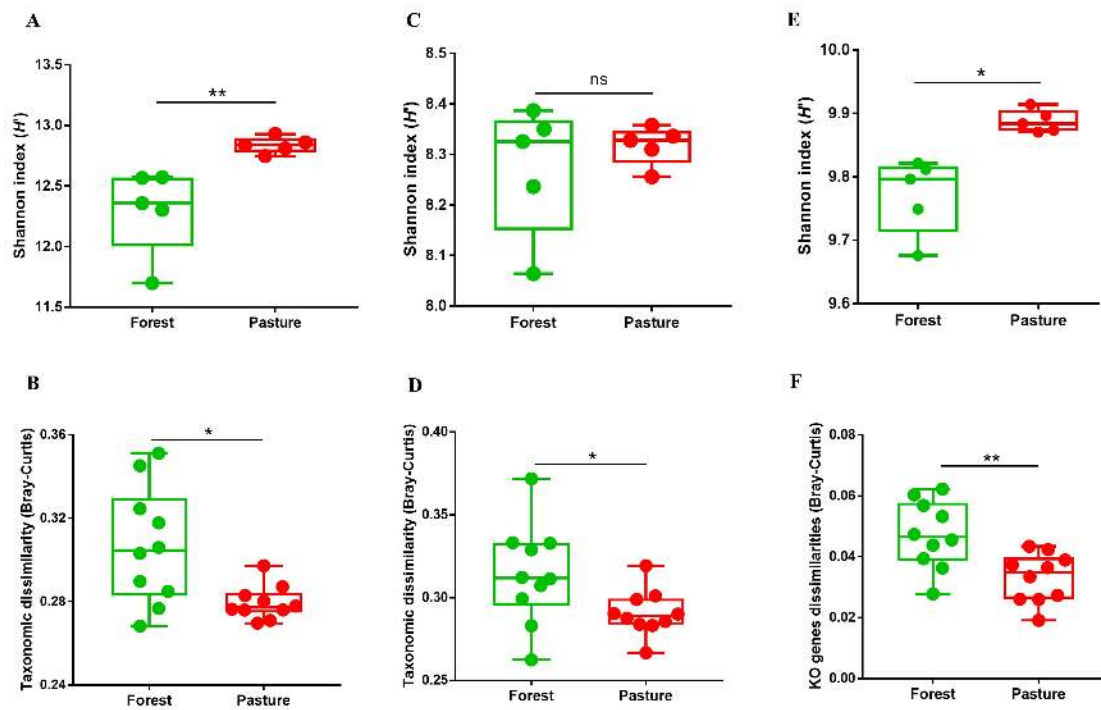


Figure 1

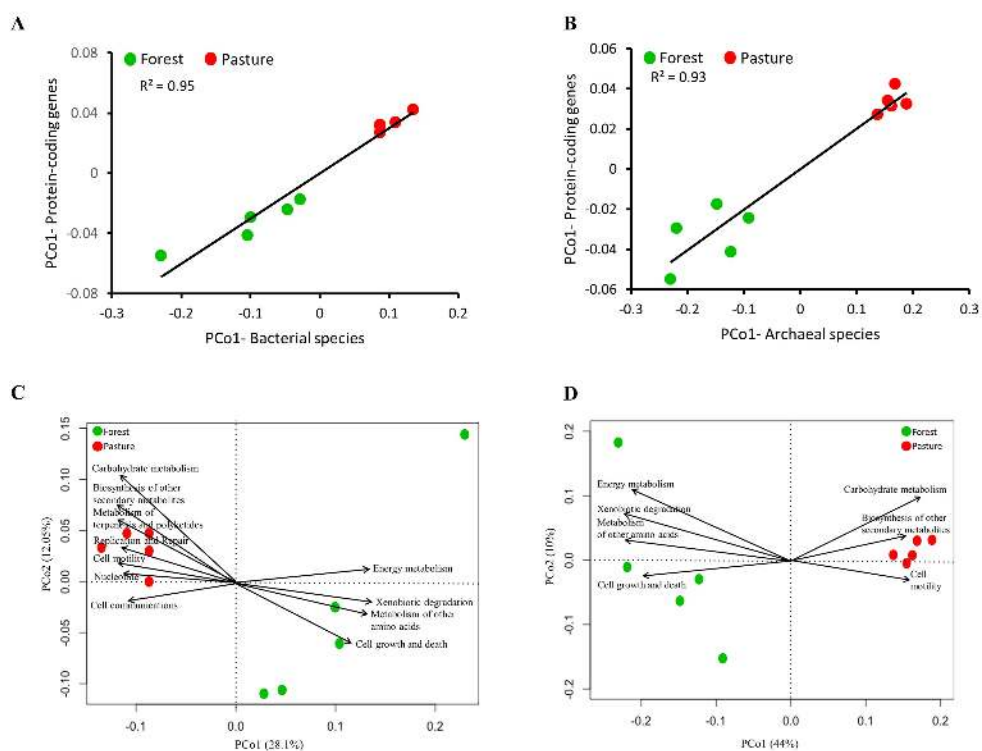


Figure 2

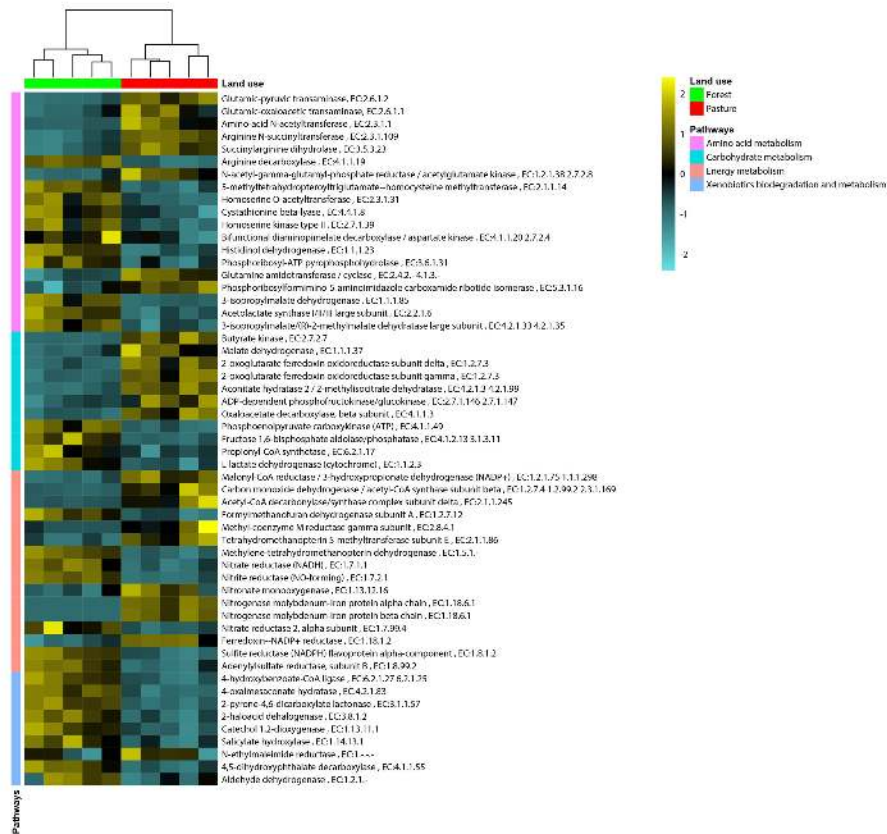


Figure 3

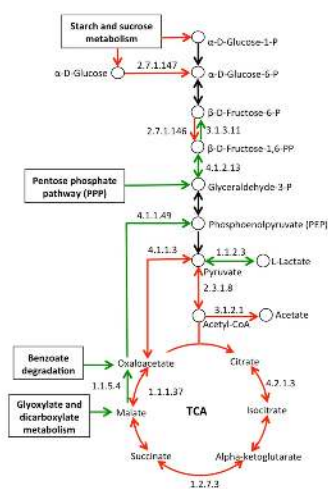


Figure 4

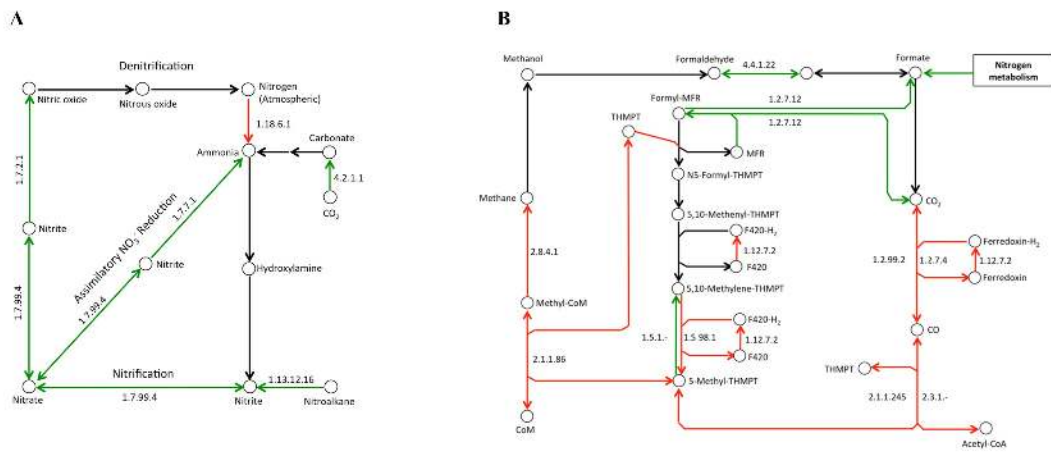


Figure 5

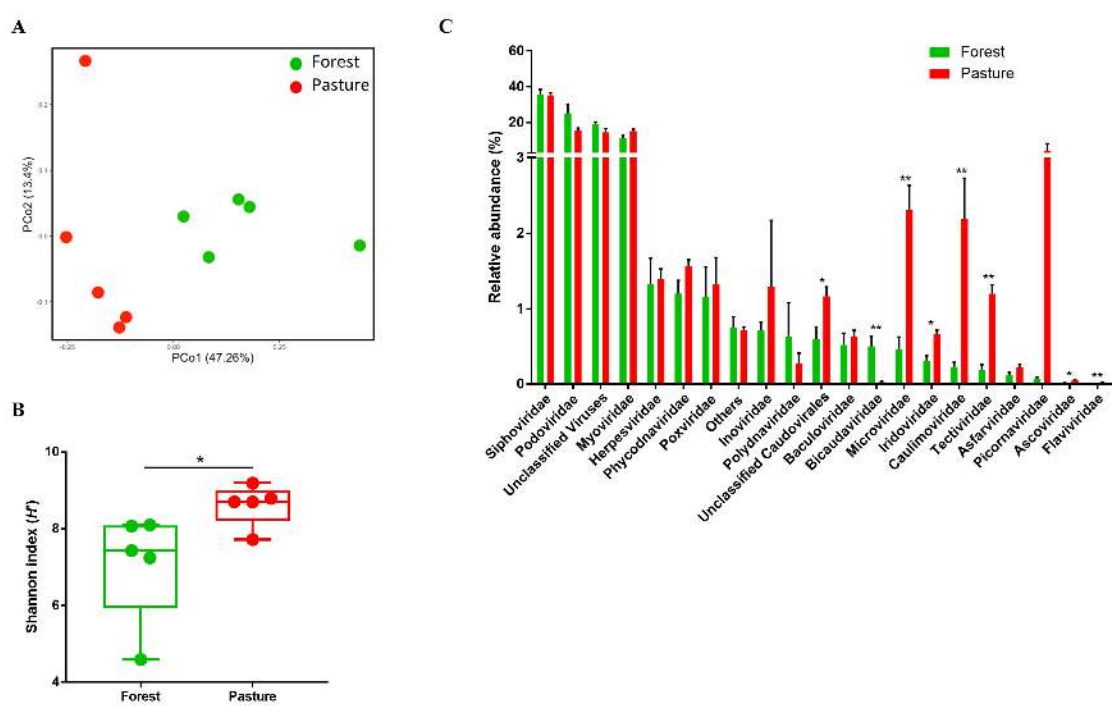


Figure 6