Metagenomics-informed soil biogeochemical models projected less carbon loss in tropical soils in response to climate warming

Yang song (chopinsong@arizona.edu)
University of Arizona https://orcid.org/0000-0002-3493-3699
Qiuming Yao
Xiaojuan Yao
Joseph Wright
Gangsheng Wang
Terry Hazen
Benjamin Turner
Malak Tfaily
Ljiljana Paša-Tolic
Eric Johnston
Minjae Kim
Konstantinos Konstantinos
Chongle Pan
Melanie Mayes
Oak Ridge National Laboratory https://orcid.org/0000-0001-6368-9210

Article

Keywords: metagenomics, enzyme functional classes, soil biogeochemical model

Posted Date: October 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-918612/v2

Main Manuscript for

Metagenomics-informed soil biogeochemical models projected less carbon loss in tropical soils in response to climate warming

Yang Song[a,b], Qiuming Yao[c,d], , Xiaojuan Yang[b], S. Joseph Wright[g], Gangsheng Wang[b,j], Terry C. Hazen[e,f], Benjamin L. Turner[g], Malak M. Tfaily[h,l], Ljiljana Paša-Tolic[h], Eric R. Johnston[i], Minjae Kim[i], Konstantinos T. Konstantinidis[i], Chongle Pan[c,k], and Melanie A. Mayes[b]

[a] Department of Hydrology and Atmospheric Science, the University of Arizona, Tucson, AZ 85721-0011USA

[b]Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA

[c]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA

[d]Department of Computer Science and Engineering, the University of Nebraska-Lincoln, Lincoln, NE 68588-0115 USA

[e]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA

[f]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996 USA

[g]Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon, Republic of Panama

[h]Environmental Molecular Sciences Laboratory and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352 USA

[i]School of Civil and Environmental Engineering, Georgia Institute of Technology

[j]Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019 USA

[k]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019 USA

[l]Department of Environmental Science, the University of Arizona, Tucson, AZ 85721 USA

*Yang Song and Melanie A. Mayes

Email: chopinsong@arizona.edu and mayesma@ornl.gov

Keywords: metagenomics, enzyme functional classes, soil biogeochemical model

This PDF file includes:

Main Text
Methods
Figures 1 to 7

Abstract (150 words)

A major challenge of quantifying feedback between microbial communities and climate is the vast diversity of microbial communities and the intricacy of soil biogeochemical processes they mediate. We overcome this challenge by simplifying the representation of diverse enzyme functions from metagenomics data. We developed a dynamic allocation scheme for enzyme functional classes (EFCs) based on the premise that microbial communities act to maximize acquisition of limiting resources while minimizing energy expenditure for acquiring unlimited resources. We incorporated this scheme into a biogeochemical model to explicitly represent microbial functional diversity and simulate responses of microbially-mediated soil biogeochemical processes to varying environmental and nutrient conditions. Representing microbial functional diversity and environmental acclimation improved predictions of the stoichiometry of microbial biomass and mitigated the sensitivity of soil organic carbon to warming in nutrient-deficient regions. Our results indicate the importance of microbial functional diversity and environmental acclimation for projecting climate feedbacks of nutrient-limited soils.

Main Text

Microbial communities mediate soil biogeochemical processes that control critical ecosystem functions such as carbon storage and nutrient supply, affecting ecosystem responses and feedbacks to climate change[1]. Understanding and predicting microbial regulatory processes requires a deep knowledge of the composition, function, and adaptability of microbial communities[2]. Community metagenomics provides information about microbial activities, but we still lack an executable strategy to transform this data into insights about microbially-mediated biogeochemical processes[3]. To a large extent, the challenge is due to the tremendous taxonomic diversity of microorganisms and the complexity of the many physical, chemical and biological processes they regulate. As a result, soil biogeochemical models and Earth system models still lack representation of microbial functions. This deficiency has contributed substantially to large uncertainties in predicted soil carbon dynamics[4,5]. Advances in this area depend on the development of practical solutions for identifying and extracting essential information from complex metagenomic datasets, as well as model structures and parameterization schemes capable of assimilating the extracted information.

Meta-omics analyses provide data on the relative abundances of genes, transcripts or proteins in a soil community. Such data inform the taxonomic distribution, structure, and functions of microbial communities[3] and can be used to infer the specific microbes or functional groups that mediate particular biogeochemical processes (e.g., plant litter decomposition, nitrification)[6,7]. However, difficulties arise when an individual or group of microorganisms are responsible for multiple biogeochemical processes[8,9], when several organisms or groups are involved in the same process[7], or when uncertainties exist in assigning gene functions to organisms[10]. These difficulties can be ameliorated by directly incorporating the functions of genes into models and thereby avoiding the use of taxonomy for assuming the functions of individual microbial organisms. This functional gene-based approach has been used to model ocean nitrogen (N) dynamics, cryptic sulfur cycling[11] and denitrification in river sediments[12]. These studies demonstrate the potential of directly linking specific genes controlling enzymatic functions with substrate-specific processes. Further advances are needed to relate microbial functions to complex processes involving an array of substrates and enzymes. However, the realization of these advances will depend on an effective strategy to map an intractable number of genes into a manageable set of representative functional traits.

Multi-omics data can inform responses of microbial communities to environmental perturbation at different time scales. Metatranscriptomics and metaproteomic data inform the instantaneous response of microbial enzymes to environmental perturbation. However, instantaneous environmental response may not reflect sustained ecosystem-level changes [13]. In contrast, metagenomics data, which characterize potential gene functions, can indicate the response of microbial communities to long-term environmental change. For example, our previous study found more genes for phosphorous (P) scavenging in control plots compared to P fertilization plots after 18 years of fertilization, suggesting that the microbial community in the fertilization plots acclimated to P excess by decreasing the relative abundance of genes for producing P-scavenging enzymes [14]. A 10-year warming experiment in a Midwest grassland[15] revealed that warming increased litter input and genes for cellulose degradation and nitrogen cycling. These studies suggest that insights into the acclimation of microbial communities to long-term environmental change can be obtained by quantifying the change in gene abundance for production of specific enzymes in response to changing conditions. Strategies are still needed to upscale variation in the abundance of an enormous number of genes to represent microbial community acclimation to long-term environmental change and the associated impact on ecosystem functioning[16,17].

We tackled these challenges by developing (i) a concept of enzyme functional classes (EFCs) for resolving microbially-mediated biogeochemical processes using metagenomics data, (ii) a solution for upscaling gene abundance to represent the acclimation of microbial communities to changing conditions over the long term; and (iii) an EFC allocation and parameterization scheme that responds to limitations in C, N, P and temperature change through biogeochemical cycling of soil organic matter (SOM). We integrated into a biogeochemical model the EFC concept, the upscaling solution, and the dynamic EFC

allocation and parameterization scheme to represent microbial functional diversity and environmental acclimation. We tested the effects of this novel microbial representation on modeling seasonal soil biogeochemical cycling using data from a long-term P fertilization experiment[18]. We further examined the sensitivity of the projected soil carbon dynamics to this novel microbial representation under future climate scenarios. We found that representing microbial functional diversity and environmental acclimation improves modeling the stoichiometry of microbial communities and reduces projected soil carbon loss in response to warming in tropical forest soils.

Results and Discussion

Applying metagenomics to represent functional diversity and environmental acclimation

We used the abundance of genes encoding for enzyme production as a measure of the potential functional diversity of soil microbial communities. Based on metagenomic measurements, each gene representing the production of an enzyme in the soil was identified with an Enzyme Commission (EC) number. EC numbers hierarchically classify enzymes based on the characteristics of enzyme-catalyzed chemical reactions, such as reaction types, bonds cleaved or formed, reaction centers, cofactors used for catalysis, and substrate specificities[19]. We grouped all EC numbers into EFCs (Fig. 1a) where enzymes within each EFC shared similar functional characteristics. The EFC grouping was based on the chemical composition of the soil substrate that each enzyme acts on, as well as the location of the cleaved chemical bond in the organic substrate (Table S1 and Fig. 1a). No inferences were needed with regards to the identity of the microbial source of the enzyme. Each EFC catalyzes the decomposition of a specific group of chemically-similar organic compounds within SOMs. The decomposition pathway begins with litter residues and proceeds to large polymers and then to oligopolymers, with different EFCs acting along the pathway. Ultimately, the decomposition pathways end with monomers that microbes are capable of assimilating. The decomposition kinetics of each EFC depend on the enzyme composition within each EFC and kinetic parameters of each enzyme. The latter further depends on bacteria : fungi : archaea ratio estimated from metagenomics-informed microbial community composition and the mean kinetic parameters of each enzyme collected for the corresponding microbial class from the biochemical database[20]. The concept of EFC and the corresponding parameterization approach make the representation of complex microbial functions from different community compositions tractable for biogeochemical modeling.

We used the abundance change of gene-encoded enzymes under changing environments as a measure of resource investment of microbial communities. The optimal resource-ratio theory[21-23] indicates that the microbial community can be reconstructed through species competition to maximize the acquisition of limiting nutrients and avoid excessive acquisition of non-limiting nutrients, thereby maintaining overall elemental stoichiometry. This resource optimization strategy for efficient acquisition of limiting resources

is reflected in changing gene abundances of different functional enzymes in response to environmental shifts[14,24,25]. To account for this, we used metagenomic data to estimate the changes in gene abundance for enzymes as measured by the effect size[26,27] in response to soil nutrient availability. We then built a hierarchical model to scale the effect sizes of individual enzymes within an EFC to the effect size of the whole EFC (Fig. 1b). Finally, we integrated the environmental responses of multiple EFCs to quantify the functional allocation among different EFCs at the community level in response to environmental change (Fig. 1c). We define this functional allocation among different EFCs as the environmental acclimation of the soil microbial community.

To implement these new ideas for modeling biogeochemical cycles, we designed a novel framework to accommodate explicit representations of metagenomics-based microbial functioning and their feedbacks to nutrient availability (Fig. 2). This design formed the main architecture of the Continuum Microbial Enzyme Decomposition (CoMEND) model, modified from the original MEND model[28-30]. CoMEND is composed of a single microbial pool, multiple microbially-synthesized EFC pools, SOM pools, and inorganic nutrient pools (Fig. 2). The effects of vegetation on SOM pools and inorganic nutrient pools were considered by inputting observed litterfall to SOM pools and removing estimated plant-nutrient uptake from the inorganic nutrient pools. Plant nutrient uptake is estimated as a function of net primary productivity and observed plant nutrient uptake rate per unit of plant dry biomass. The effects of litter chemistry and structure on soil decomposition processes were parameterized along major decomposition pathways, e.g., lignocellulose inputs were classified as lignin and cellulose litter, and then large biopolymers, oligopolymers, and bioavailable monomers produced from progressive decomposition of the lignin and cellulose. Each SOM pool was classified as either microbially-activated (ASOM), mineral-protected (MSOM), or the adsorbed portion of the ASOM pool (QSOM) (Fig. 2). ASOM is the particulate organic matter associated with sand-sized particles (e.g., particle size > 53 μm) and acted upon by microbes. MSOM is mineral-associated organic matter with particle size < 53 μm. QSOM is differentiated from MSOM and defined as the adsorbed phase of the corresponding ASOM interacting with MSOM through adsorption and desorption[28] (See details in Method section).

In the model, our approach of parameterizing SOM decomposition is directly integrated with omics-informed EFCs and therefore can characterize the decomposition rates of diverse soil components. The decomposition of each microbially-activated pool is catalyzed by one specific EFC, which Michaelis-Menten kinetics are determined by collecting corresponding kinetic parameters of each enzyme within an EFC from the BRENDA biochemical database[20] and integrating them based on the relative abundance of each enzyme within an EFC. The decomposition rates of each SOM pool therefore depend on SOM concentration and EFC kinetics which vary with soil temperature. The decomposition rates also depended upon EFC abundance which varied in response to nutrient limitation according to optimal resource-ratio theory. We refer to this as “dynamic EFC allocation” which can potentially alter the stoichiometry of

bioavailable monomers and mitigate nutrient limitations on microbial growth. Thus, the dynamic EFC allocation can simulate environmental acclimation of soil microbial community by altering the functional allocation among different EFCs in responses to environmental change. Overall, the CoMEND framework contains a feedback loop that couples EFC allocation, SOM decomposition, and resource availability for microbial growth. This feedback loop directly links microbial diversity and environmental acclimation with ecosystem functioning[31].

We evaluated this new concept with metagenomics data from control and P-fertilized plots in a 17-year ongoing fertilization experiment in Panama. Due to long-term weathering and high precipitation, Panamanian soils experience chronic P limitation[18]. The annual mean temperature is 26℃ and will increase to 2.6 °C by the year 2100 according to CESM RCP8.5 projection. Microbial P accounts for 70% of biomass P in this tropical forest ecosystem[18], indicating that soil microbial communities may have developed acclimation strategies to immobilize large amounts of P from nutrient-deficient soils and thereby control P cycling through the ecosystem. Our previous studies found higher gene abundances of enzymes for P acquisition, but lower gene abundance of ligninases, in the control soils compared to the P-fertilized soils[14]. Seventeen years of fertilization has significantly increased microbial biomass and litterfall but has not significantly affected the microbial C/P ratio, plant stem growth, or soil carbon stocks[18,32,33]. To test the effect of microbial functional diversity and environmental acclimation on soil biogeochemical cycles, we constructed the model into four versions to represent low, moderate, and high microbial functional diversity as well as environmental acclimation of highly diverse microbial communities, respectively. These four versions were used to test whether the representation of high functional diversity and environmental acclimation of tropical soil microbial community can better predict the microbial stoichiometry, microbial biomass, soil carbon fluxes, soil carbon over the short term and the long term.

Enzyme functional classes (EFCs) in Panamanian soils

We identified 2135 gene-encoded EC numbers (Supplementary Data 1) in the Panamanian soil samples[14]. Among these numbers, 118 participate in SOM decomposition and nutrient mineralization (Figs. S1-S3 and Supplementary Data 2). The application of our EFC grouping strategy (Fig. 1a) resulted in these 118 enzymes being grouped into 22 EFCs relevant to soil C, N, and P cycling (Table S1and Fig. 3c).

Among the 22 EFCs identified, four catalyzed the decomposition of lignocellulose-containing SOM along a continuum beginning with raw litter and ending with bioavailable monomers. The enzymes represented by C_{endo1}, C_{exo}, and C_{oligo} acted on a specific bond location of carbohydrate, collaboratively catalyzing its decomposition from large polymers to bioavailable polymers[34]. C_{endo1} cleaved the internal C-O bonds in carbohydrate residues to produce large carbohydrate polymers. C_{exo} cleaved the terminal C-O bonds in these polymers to produce oligosaccharides. C_{oligo} cleaved the C-O bonds in oligosaccharides to

produce bioavailable monomers (Fig. 3c and Table S1). C_{endo2} was a general EFC for decomposing lignin. It acted on the internal C-C or C-O bonds of different polymerization degree of lignin to produce bioavailable aromatic compounds. Here decomposing C-C bonds within lignin required more energy than decomposing C-O bonds within carbohydrates.

Six EFCs were involved in two decomposition routes of the N-containing SOM. In the first route, three EFCs (N_{endo}, N_{exo1} and N_{oligo1}) cleaved the internal and terminal C-N bonds of proteins and C-N bonds in oligopeptides, respectively, to produce amino acids for microbial assimilation (Fig. 3c and Table S1). In the second route, two EFCs (N_{exo2} and N_{oligo2}) cleaved the terminal C-N bonds of cell walls N components (e.g., chitin, peptidoglycan) and C-N bonds in the products of microbial cell wall decomposition, respectively, to produce bioavailable N-containing monomers (Fig. 3c and Table S1). The C-N bonds in the microbially-assimilated monomers were catalyzed by N_{mono} to release inorganic N into soil. Inorganic N transformation (e.g., nitrification, denitrification, N assimilation, biological N fixation) were catalyzed by four EFCs (N_{inN1}, N_{inN2}, $N_{inN3,}$ and N_{inN4}).

We identified seven EFCs (P_{exo1}, P_{oligo1}, P_{exo2}, P_{oligo2}, P_{mono1}, P_{mono2}, P_{mono3}) for P mineralization and one EFC (P_{inP}) for P immobilization in Panamanian soils. Both biological and biochemical mineralization were indicated by the presence of these EFCs. P_{mono3} catalyzed the biological P mineralization by breaking the phosphoester bonds in microbially-assimilated monomers to satisfy microbial energy needs[35]. P_{mono1}, P_{mono2}, P_{exo1}, P_{oligo1}, P_{exo2} and P_{oligo2} biochemically decomposed and mineralized different forms of organic P compounds (Fig. 3c and Table 1). P_{mono2} required the least energy investment to biochemically release inorganic P from general monophosphates. P_{exo1} and P_{oligo1} were responsible for decomposing nucleic acids while P_{exo2} and P_{oligo2} decomposed phospholipids. These EFCs acted on the phosphoester bonds in corresponding compounds with different degrees of polymerization to release monophosphates for P_{mono2} utilization. P_{mono1} specialized in releasing inorganic P from inositol P, which is the most stable organic P in the soils[36]. The existence of many diverse EFCs for biochemical P mineralization indicates the importance of P deficiency in these soils[35].

Effects of long-term P fertilization on the functional diversity of soil microbial communities

To determine how the functional diversity of soil microbial communities responds to different levels of nutrient availability, we quantified the combined effect size of all gene-encoded enzymes within each EFC in response to 17 years of P fertilization. The combined effect size of an EFC was calculated as the fold change in gene-encoded enzymes abundance within an EFC between control and treatment (e.g., the control and P-fertilized soils) using base 2 logarithms. Here positive effect size indicated P fertilization decreased EFC abundance compared with the control soil. The confidence interval of this effect size for an EFC is estimated as the sum of the sample variation of gene-encoded enzymes within an EFC and the variance among gene-encoded enzymes within an EFC (See details in Method). The gene abundances of

EFCs for oligonucleotide decomposition (P_{oligo1}) and extracellular monophosphate mineralization (P_{mono2}) were 0.50−fold (± 0.09) and 0.43−fold (± 0.09) higher (q-values < 10^{-3}) in the control than P-fertilized plots, respectively (Fig. 4a). This likely reflects P limitation in the control soils, where investment in enzyme production to acquire P is maximized in the absence of P fertilization.

P fertilization decreased the gene abundance of EFCs for decomposing large polymers of carbohydrates and increased the gene abundance of EFC for decomposing lignin. The gene abundances of EFCs for cleaving the internal (C_{endo1}) and terminal C-O bonds (C_{exo}) in large carbohydrate polymers were 0.33-fold ± 0.05 (q-value=0.004) and 0.26-fold± 0.03 (q-value < 10^{-5}) significantly higher in the control than P-fertilized plots, respectively (Fig. 4a). However, the gene abundance of C_{endo2} (for lignin decomposition) was 0.16 ± 0.06-fold lower in the control than P-fertilized plots (Fig. 4a). These results indicate that microbial communities invest enzyme production for acquiring C primarily from large polymers of carbohydrates under P limitations.

P fertilization also decreased the gene abundance of EFCs for decomposing large polymers of N-containing SOM but increased the gene abundance of EFC for inorganic N transformation. The control soils had 0.49± 0.11-fold (q-value < 10^{-5}) and 0.17± 0.04-fold (q-value = 0.02) higher gene abundances of EFCs for cleaving internal (N_{endo1}) and terminal (N_{exo1}) C-N bonds in large protein polymers. For the EFC cleaving the terminal C-O bonds in N components from microbial cell walls (N_{exo2}), the abundance was 0.17± 0.04-fold higher (q-value = 0.04) in the control than P-fertilized plots (Fig. 4a). However, the control plot had a significantly lower (-0.29±0.10-fold with q-value = 0.04) gene abundance of N_{inN1} (responsible for nitrification), than the P-fertilized plots. These findings suggest that microbial communities in the P-limited control plots acquired N primarily from decomposition of protein and N components from microbial cell walls[37].

In summary, we found that long-term P fertilization systematically altered gene abundances of EFCs for acquiring C, N and P to maintain ecosystem-level stoichiometry. Compared with the P-fertilized soils, increased gene-coded enzyme abundances of EFCs for P-containing SOM mineralization in the control plots (Fig. 4b) indicated that soil microbial communities increased enzyme investment to acquire limiting resources. In the meanwhile, significantly decreased C_{lignin} for decomposing lignin (q-value = 0.04) but increased $C_{carbohydrates}$ for decomposing carbohydrates (q-value = 0.002) in the control soils (Fig. 4b) implied that the community favored the simple C substrate (e.g., carbohydrates) to minimize energy expenditure for C acquisition when C is not limited. This efficient resource acquisition strategy is achieved in the model by adjusting gene abundances of EFCs among different functions [14].

Hierarchical modeling of microbial functional diversity and environmental acclimation

A critical question is whether and to which degree it is necessary to represent the functional diversity of soil microbial communities and associated environmental acclimation for accurately simulating soil

biogeochemical cycling. To answer this question, we constructed four versions of CoMEND with different representations of microbial functions. The first three versions represented high (H), moderate (M), and low (L) functional diversity of microbial communities, denoted by $CoMEND_H$, $CoMEND_M$, and $CoMEND_L$, respectively. $CoMEND_H$ included all metagenomics-informed 22 EFCs for SOM decomposition and thus represented the high enzyme functional diversity, with each EFC catalyzing the decomposition of specific chemical compounds along the continuum decomposition pathway from litter residues to bioavailable monomers (Fig. 3c). $CoMEND_M$ ignored the differences in the degree of polymerization and assumed that the EFC with the minimum potential activity, which is the product of its concentration and potential specific kinetic activity, limited the decomposition rates of specific chemical compounds along its decomposition pathway. This assumption reduced the number of EFCs to 15 for SOM decomposition in $CoMEND_M$, representing the moderate functional diversity of microbial communities (Fig. 3b). $CoMEND_L$ didn't distinguish the specific chemical compounds and coarsely classified SOM into lignocellulose-containing SOM, N-containing SOM, and P-containing SOM, with the decomposition of each SOM pools being catalyzed by a cluster of EFCs (Fig. 3a). In total, $CoMEND_L$ contained 11 clusters of EFCs, representing low functional diversity of microbial communities. The structure of $CoMEND_L$ is closer to existing microbial enzyme mediated models (e.g, MEND) with minimal distinction of enzyme functional diversity (two enzymes) and different chemistry of SOM (lignin-like and carbohydrate-like). For all three versions of CoMEND, a fixed allocation fraction of synthesized enzymes was used for each EFC over time, i.e., a static EFC allocation strategy was applied (details in the Methods section). This static EFC allocation strategy assumes no acclimation of soil enzyme functional compositions in response to nutrient availability. As a result, microbial C/P ratios can fluctuate with possible change in microbial community structure in response to nutrient availability. According to previous studies of C/P ratios of microbial communities[38-40], we assume microbial C/P ratio can vary between 9 and 40.

The fourth version ($CoMEND_{HD}$) was $CoMEND_H$ plus the scheme of dynamic enzyme allocation in response to resource availability, and thus capable of simulating the functional allocation among different EFCs in responses to environmental change, i.e., environmental acclimation of the microbial communities. The dynamic enzyme allocation strategy in $CoMEND_{HD}$ optimized the enzyme investment strategy of the microbial community for mitigating resource limitation and was calibrated by comparing the metagenomics-informed and simulated effect sizes of EFCs in response to P fertilization (details in the Methods section). The model captured the metagenomics-informed effect sizes of EFCs to P fertilization (Fig. S4), indicating that $CoMEND_{HD}$ can simulate change in enzyme investment for acquiring limited resources. The Willmott Index (WI) of agreement[41] was 0.49 for all EFCs and 0.67 for the EFCs with statistically significant effect sizes in response to P fertilization (Fig. S4). Higher WI implies better performance[41].

We first ran each of the four versions of CoMEND over the period of 1998-2014 in the control and P-fertilized soils. Since the Panamanian soils are not limited by N availability[18], our historical comparisons focused on the effects of microbial functional diversity on soil C and P dynamics. To test how representation of functional diversity of enzyme composition impacts soil carbon dynamics at longer time scales, we compared the future soil carbon dynamics projected over the year 2015-2100 by $CoMEND_L$, $CoMEND_M$, $CoMEND_H$. To test the effect of environmental acclimation on long-term soil carbon dynamics, we compared the same projection by $CoMEND_H$ with $CoMEND_{HD}$. Each model was first driven by the Community Earth System Model (CESM) large assemble projection of litter input, plan P uptake, 10 cm depth of soil temperature and the detrended soil moisture under the RCP8.5 scenario (Simulation I)[42], and then by the detrended soil temperature and moisture (Simulation II, see supplementary information (SI) Method S1), respectively. No nutrient fertilization was applied for the two simulations. Simulation I represented the projected SOM under changing soil temperature, whereas Simulation II served as the control simulation without considering trends in either temperature or soil moisture. The difference in the modeled SOM between the two simulations isolated the projected responses of soil carbon dynamics to climate change.

Results of the hierarchical modeling

We compared the effect of high, moderate, and low microbial functional diversity and environmental acclimation on (a) microbial stoichiometry, (b) microbial carbon, (c) microbial respiration, and (d) short- and long-term soil carbon dynamics. In summary (Fig. 5), low functional diversity of the modeled microbial community ($CoMEND_L$) is sufficient to capture microbial carbon, microbial CO_2 fluxes, and short-term soil organic carbon (SOC) in both the P-fertilized and the control soils, and microbial stoichiometry in the P-fertilized soils. However, representation of high functional diversity and environmental acclimation ($CoMEND_{HD}$) is required to capture microbial stoichiometry and to mitigate projected soil carbon loss to climate warming in the P-limited control soils.

The effect of microbial functional diversity and environmental acclimation on microbial stoichiometry. This effect is assessed by comparing observed microbial C/P ratios with the corresponding simulations by $CoMEND_H$, $CoMEND_M$, $CoMEND_L$, and $CoMEND_{HD}$ over the year 2006-2007[18] in the P-fertilized soils and the control soils with P deficiency (Fig. 6a). To further interpret the outcomes of this comparison, we also compared modeled C/P ratios of decomposition-released bio-available organic matter (Fig. 6b), and monophosphate mineralization rates (Fig. 6c) by $CoMEND_H$, $CoMEND_M$, $CoMEND_L$, and $CoMEND_{HD}$. To explain how enzyme functional diversity results in a difference in modeled decomposition rates among $CoMEND_H$, $CoMEND_M$, $CoMEND_L$, we compared the catalyzation capacity of the corresponding EFC within each version of the model (Fig. 6e-f). Here the catalyzation capacity of an EFC refers to the Michaelis-Menten equation calculated turnover rates of each EFC-catalyzed SOM pool and how they vary

with the availability of corresponding SOM pool. The shaded area under each curve indicates the range of turnover rates of each EFC-catalyzed SOM over the year 2006-2007 according to the availability of corresponding SOM pool (Fig. 6e-f).

The effect of microbial functional diversity and environmental acclimation on microbial stoichiometry depends on the soil nutrient status. Comparing modeled C/P ratios in the P-fertilized and the control soils by four versions of the model (Fig 6a), there is no significant difference in the modeled C/P ratios of microbial communities in the P-fertilized soils, but significant differences in the modeled C/P ratios in the control soils over the year 2006-2007 among $CoMEND_H$, $CoMEND_M$, $CoMEND_L$, and $CoMEND_{HD}$, with closer agreement with observed values for complex representations of microbial communities in $CoMEND_{HD}$. This result indicates that representation of microbial functional diversity and environmental acclimation is particularly important for capturing microbial stoichiometry in the nutrient-deficient soils.

Representation of high functional diversity mitigates overestimated microbial C/P ratios in the control soils. $CoMEND_L$ significantly overestimated the C/P ratio in the microbial community with the WI of -0.73 (Fig. 6a). This model bias was gradually improved with increased representation of microbial functional diversity in $CoMEND_M$ (WI = -0.55) and $CoMEND_H$ (WI=-0.40) (Fig. 6a). This improvement is due to the fact that increased enzyme functional diversities in $CoMEND_H$ and $CoMEND_M$ decrease the C/P ratios of decomposition-released bioavailable organic matter (R_{SOC}/R_{SOP}) (Fig. 6b) and also increase the mineralization of monomer P (Fig. 6c). Modeled higher P availability for microbial assimilation by $CoMEND_H$ and $CoMEND_M$ is because both versions of CoMEND are able to partition different catalyzation capacity of different EFCs for decomposing organic carbon and phosphorus (Fig. 6d-f). For example, comparing modeled turnover rates (TRs) over 2006 – 2007 year for represented lignocellulose pools (Fig. 6d), diester P pools (Fig. 6e) and monomer P pools (Fig. 6f) in $CoMEND_H$, $CoMEND_M$, and $CoMEND_L$, respectively, we find that the carbohydrates ($C_{carbohydrates}$) pool in $CoMEND_M$ and the oligopolymer carbon (C_{oligo}) pool in $CoMEND_H$ usually have lower TRs than the lignocellulose ($C_{Lignocellulose}$) pool in $CoMEND_L$ (Fig. 6d). However, the nucleotides pool ($P_{nucleicacids}$) in $CoMEND_M$ and the oligonucleotides (P_{oligo1}) pool in $CoMEND_H$ have higher TRs than the total diester P pool in $CoMEND_L$ (Fig. 6e). These differences lead to the release of more bioavailable organic P, but less bioavailable organic C, and thus explain smaller C/P ratios of bioavailable organic matter in $CoMEND_H$ and $CoMEND_M$ than in $CoMEND_L$ (Fig. 6b). Moreover, both $CoMEND_H$ and $CoMEND_M$ differentiate the inositol P pool ($P_{mono1EFC15}$ and $P_{mono1EFC22}$) and the non-inositol monophosphate pool ($P_{mono2EFC15}$ and $P_{mono2EFC22}$). The latter has higher concentration and TRs (Fig. 6f) due to more release of nucleotides from $P_{nucleicacids}$ pool in $CoMEND_H$ and the P_{oligo1} in $CoMEND_M$ (Fig. 6e), compared with $CoMEND_L$. Consequently, $CoMEND_H$ and $CoMEND_M$ simulate higher mineralization rates of total monomer P than $CoMEND_L$ (Fig. 6c).

373 Compared with $CoMEND_H$, the dynamic enzyme allocation in $CoMEND_{HD}$ further improved the
374 simulations of microbial stoichiometry during the year 2006-2007 in the control Panamanian soils
375 (WI=0.12). The modeled microbial C/P ratios by $CoMEND_{HD}$ are consistent with observations from the P-
376 deficient control soils (Fig. 6a). This improvement is due to the capability of the dynamic enzyme
377 allocation scheme to simulate increased enzyme allocation for organic P decomposition and mineralization
378 under P-limited condition, as indicated by modeled lower C/P ratios of decomposition-released bio-
379 available organic matter (R_{SOC}/R_{SOP}) (Fig. 6b) and higher monophosphate mineralization rates (Fig. 6c) in
380 $CoMEND_{HD}$ than in $CoMEND_H$. This improved simulation of microbial C/P ratios in the P-deficient soils
381 makes it possible for $CoMEND_{HD}$ to model the relative homeostatic C/P ratios of soil microbial
382 communities under changing nutrient availability (Fig. 6a). Consistent with observations [14,18,32], both
383 $CoMEND_H$ and $CoMEND_{HD}$ parameterized a marginally significant increase in plant litter input but
384 unchanged bacteria-dominated microbial communities in the P-fertilized simulation, compared with the
385 control simulation. Thus, the capture of homeostatic C/P ratios of soil microbial communities in the P-
386 fertilized and the control soils by $CoMEND_{HD}$ indicate that enzyme dynamics in response to nutrient
387 availability is critical for maintaining homeostatic stoichiometry of a microbial community without
388 structural change. Also, $CoMEND_{HD}$ can use the same parameters to simulate control and P-fertilized
389 soils, simplifying the parameterization needs when the model is applied to different sites with diverse
390 nutrient availability.

391

392 ***The effect of microbial functional diversity and environmental acclimation on short-term and long-term***
393 ***soil carbon dynamics***. All four versions of the model are able to capture microbial carbon stocks during
394 2006-2007 and microbial CO_2 fluxes in an incubation experiment in both control and P-fertilized soils
395 (Fig.7a-b and d-e). The WI of agreement between measured microbial carbon and the simulated microbial
396 carbon by four versions of the model is around 0.52-0.55 in the P-fertilized soils and 0.43-0.46 in the
397 control soils. All four versions of the model simulated around 18% higher microbial carbon stock in the P
398 fertilized soils than in the control soils, which is close to observed 20% higher microbial carbon stock. This
399 result indicates that higher microbial carbon stock with P fertilization mainly results from significant
400 increase in plant litter input[32], but not enzyme dynamics of microbial communities. Modeled microbial CO_2
401 fluxes fall within the measurement error of observed values, and $CoMEND_H$ and $CoMEND_{HD}$ can better
402 capture the CO_2 fluxes at the beginning of the incubation experiment than $CoMEND_M$ and $CoMEND_L$, as
403 indicated by increased WI values (Fig. 7b, e). The total soil organic carbon stocks (SOC) from field
404 measurements from 2006 to 2014 was modeled by $CoMEND_L$, $CoMEND_H$, and $CoMEND_{HD}$ and fall
405 within the measurement error of observed SOC stocks, whereas the WI agreement between observed and
406 simulated SOC stocks by $CoMEND_H$ and $CoMEND_{HD}$ are higher than the corresponding simulation by
407 $CoMEND_L$ (Fig. 7c, f). $CoMEND_M$ underestimated SOC stocks in the P-fertilized soils (Fig. 7c) because

408 $CoMEND_M$ doesn't differentiate the carbohydrates with different degrees of polymerization, and thus
409 $CoMEND_M$ may overestimate available organic carbon for $C_{carbohydrates}$.
410 Although $CoMEND_L$, $CoMEND_H$, and $CoMEND_{HD}$ have the ability to capture short-term microbial
411 dynamics and field SOC dynamics, their differences in simulating SOC dynamics are distinct over
412 timeframes of interest for climate projections. Representation of high functional diversity of microbial
413 communities in $CoMEND_H$ reduces the projected SOC loss in response to climate warming in the
414 Panamanian soils. Climate change under the RCP8.5 scenario will increase soil temperature to 2.6 °C by
415 the year 2100 (Fig. 7g). Driven by this warming trend, all versions of the model simulate a loss of SOC due
416 to decreased microbial carbon use efficiency, increased microbial maintenance activity, increased enzyme
417 kinetic activity as well as the acceleration of desorption/adsorption of soil substrates. In the P-fertilized
418 soils, $CoMEND_L$, $CoMEND_M$ and $CoMEND_H$ projected SOC loss of 4.0%, 3.8% and 2.8%, respectively,
419 by the year 2100 (Fig. 7h). The corresponding loss of SOC in control soils is 4.0%, 3.5% and 2.2%,
420 respectively (Fig. 7i). $CoMEND_H$ projects reduced SOC loss because observed EFCs have different
421 temperature sensitivities (Fig. S5a-b). Increased representation of enzyme functional diversity enables
422 $CoMEND_H$ to model diverse temperature responses of soil microbial communities to climate change,
423 leading to mitigated SOC loss in response to soil warming.
424 Representation of environmental acclimation of soil enzyme composition in $CoMEND_{HD}$ reduces the
425 projected SOC loss in response to soil warming in the control soils; however, this acclimation effect
426 gradually weakens over the time. $CoMEND_{HD}$ predicts SOC loss of 0.5% from 2025 to 2060, almost half of
427 the SOC loss projected by $CoMEND_H$ over the same period (Fig.7i). The reduced projected loss by
428 $CoMEND_{HD}$ is because P-deficiency in the control soils increases the microbial production of P-acquisition
429 enzymes, which usually have lower temperature sensitivity than C-acquisition enzymes (Fig. S5a-b),
430 leading to decreased SOC loss in response to warming. It is noticeable that this mitigation of SOC loss
431 tends to decrease after the year 2060 when dynamic enzyme allocation is not enough to mitigate P
432 limitation on microbial growth.

433 In summary, our scenario analysis suggests that the projected responses of SOC stocks to climate
434 warming are highly sensitive to model structures in representing functional diversity and environmental
435 acclimation of microbial communities. Such representation is important for mitigating uncertainty in the
436 projected SOC dynamics in soils with resource limitation in a warming climate.

437 **Conclusion and future directions**

438 The increasing availability of environmental metagenomic analyses offers new opportunities to
439 elucidate microbial processes and their mediation of ecosystem functioning in models of biogeochemical
440 cycles. Such critical quantitative insights so far have remained elusive because of the complexities of

441 metagenomic data and the large gaps between gene and ecosystem scales. Our study demonstrates a
442 solution to bridge this gap.

443 Our strategy to represent enzymatic activities is similar to the biochemical model of photosynthesis[43]
444 which has been successfully applied in ESMs. However, unlike the process of photosynthesis where only
445 one enzyme (Rubisco) plays a dominant role in CO_2 assimilation, a vast number of enzymes play equally
446 important roles in decomposing complex SOM to release vital nutrients, making the explicit representation
447 of individual enzymatic activities unrealistic within a modeling framework. We overcome this challenge by
448 applying an EFC approach which groups enzymes into functionally similar classes along continuum
449 depolymerization pathways[34] beginning with litter residues and ending with bioavailable monomers. The
450 joint use of metagenomic datasets with growing biochemical databases (e.g., BRENDA) enables such
451 grouping. A clear advantage of the approach is that taxonomic specification of the microbial community,
452 which can change radically over spatial scales, becomes unnecessary.

453 Our study found that EFCs are dynamic with consequences regarding the decomposition of SOM. The
454 dynamics of EFCs reflect the adaptation of microbial communities to changing environmental conditions.
455 EFC dynamics can be predicted based on the strategy that investment in enzymes for acquisition of the
456 limiting nutrients is maximized while enzyme production for non-limiting nutrients is minimized. This
457 strategy allows an explicit representation of microbial functional acclimation to environmental perturbation
458 for modeling carbon and nutrient cycles. The integration of the EFC concept with this optimization strategy
459 for acquiring limiting resources provides a logical avenue for harnessing the power of metagenomics data
460 to improve soil biogeochemical models.

461 Instead of representing diverse microbial community composition, CoMEND is the first
462 metagenomics-informed soil biogeochemical model that represents enzyme functional diversity of
463 microbial communities and dynamically predicts the acclimation of enzyme functional composition of
464 microbial communities to environmental perturbation (e.g., fertilization). CoMEND explicitly considers
465 efficient resource acquisition of the microbial community and the feedbacks between enzyme functional
466 allocation of microbial communities, SOM decomposition, and resource availability. Testing and applying
467 this novel model in Panamanian soils demonstrated that CoMEND with low functional diversity
468 representation can be used to model short-term microbial and SOC dynamics in nutrient-abundant soils,
469 whereas representing high functional diversity and environmental acclimation of microbial communities is
470 important for SOC projection under nutrient-limited conditions under a changing climate over the long
471 term.

472 Our study identifies potential advances along several promising directions for future endeavors. First,
473 developed CoMEND with hierarchal level of representation of microbial functional diversity provides
474 flexibility to choose models for different study needs. For the studies projecting long term carbon dynamics

in nutrient-limited soils, $CoMEND_{HD}$ can integrate omics data to represent functional diversity and environmental acclimation of microbial communities and their impacts on SOC dynamics. Second, the concept and analytical framework developed here are sufficiently general to be applicable to other -omics measurements (e.g., metatranscriptome, metaproteome). These techniques can be used to measure instant or short-term response of microbial-mediated biogeochemical processes to environmental change, which has not been considered in the current metagenomics-informed CoMEND model. Prediction of actual microbial community functions would require an integration of multiple types of -omics measurements into a single modeling framework. This model represents the first step to incorporating multiple -omics measurements into models. Finally, applying the approach developed here in the Panama soils simulation give us confidence to apply our model to simulate other highly weathered and P-limited tropical soils. To test the performance of the presented model in other region, more experiment-CoMEND integration studies are needed. Future studies could establish the biogeography of microbial enzyme functional composition and environmental adaptation strategy from large-scale -omics databases (e.g., Earth Microbiome Project)[44]. Especially, soil omics data from experiments with both soil temperature, soil moisture, and nutrient treatments are needed to parameterize environmental adaptation strategy in response to combined nutrient and moisture change. Also, long-term and high sampling frequency of omics data are needed to parameterize how environmental acclimation strategies of microbial communities vary with time. As most of identified 118 enzymes in Panamanian sites involve the decomposition of common soil C, N, and P substrates, we hypothesize that most of these enzymes can be found in other regions also, but their relative abundance may vary with environmental condition. This hypothesis could be tested in future large-scale -omics database analysis. These efforts will pave the way for explicit representation of microbial processes in ESMs to better predict interactions between the soil microbial community and vegetation.

Methods

Metagenomic data source

The metagenomic data were measured on soil samples collected from P-deficient control plots and P-fertilized plots in December 2014 on the Gigante Peninsula, Republic of Panama[14]. This site (9.1°N, 79.84°W), managed by the Smithsonian Tropical Research Institute, is underlain by Oxisol soils and has alternating wet (May-November) and dry seasons (December - April)[18]. Since 1998, two control plots had not received any fertilization while two P-fertilized plots received 50 kg P ha^{-1} $year^{-1}$ as triple superphosphate applied in four equal doses each year[18,32]. Metagenomic data were attained through deep sequencing of extracted genomic DNA from four soil samples collected from two control plots and two P-fertilized plots. The gene abundance of an EC number was quantified in the metagenomes based on normalized base-pair counts of reads mapped onto proteins annotated with this EC number. The difference in the gene abundance of an EC number between the control and P-fertilized plots was expressed as the

fold change of the former relative to the latter. The statistical significance of this specific gene abundance difference between the control and P-fertilized plots was examined through differential analysis, performed using likelihood ratio tests with Benjamini-Hochberg multi-comparison corrections in the edge R package[14]. Statistically significant differences in gene abundances between control and P-fertilized plots were identified using optimized false discovery rates (i.e., q values) <0.05 and fold changes of gene abundance ≥ 1.2. More details about metagenomic analyses were described in our previous study[14].

Grouping of enzyme functional classes (EFCs)

We applied both bottom-up and top-down considerations to classify enzymes into EFCs in the tropical soils. We narrowed the full set of metagenomics data, which consisted of 2135 different enzymes encoded by genes, down to a list of 118 EC numbers that were related to SOM decomposition, mineralization, and inorganic N and P transformation. This list was created by querying properties of each EC number from biochemical databases, such as BRENDA[20] and MetaCyc[45], and querying enzyme available SOM compounds by collecting reported SOM and inorganic N and P components at Panama site[46-48] and performing SOM analysis with Electrospray Ionization Fourier Transformed Ion Cyclotron Resonance Mass Spectrometry (ESI-FTICR MS), which can inform specific soil chemical components, such as lignin-like, carbohydrate-like, and protein-like SOM[49,50] (See details of ESI-FTICR MS analysis in SI Method S2). Properties considered in this bottom-up process included the functions, reactions, involved pathways, and substrates that the enzymes catalyze. As the number of different enzymes identified in the bottom-up process was too great to warrant individual representation in ecosystem-scale applications, a top-down process was then applied to classify the identified enzymes into groups based on functional similarities, i.e., EFCs. Functional similarities considered included:

- The chemical compound acted on by the enzyme, i.e., EC number
- The location of the cleaved bond in this compound (e.g., the internal or terminal bonds of large polymers, or the bonds within oligopolymers)
- The degree of polymerization of products along the continuum decomposition pathway from litter residues to bioavailable polymers
- Specific transformation pathways for inorganic nutrients.

The application of these top-down considerations logically resulted in the identification of 22 EFCs. These 22 EFCs fell into four broad clusters:

- Cluster for the decomposition of lignocellulose-containing SOM (e.g., carbohydrates, lignin)
- Cluster for the decomposition of N-containing SOM (e.g., proteins, N components from microbial cell walls)
- Cluster for the decomposition of P-containing SOM decomposition pathway (e.g. nucleic acids, phospholipids, and inositol P decomposition)

- Cluster for nitrification, denitrification, N assimilation, N fixation, P immobilization, N mineralization, and biological P mineralization and biochemical P mineralization for phytate and general monophosphate.

Integrative analysis of EFC-specific effect size

We estimated the response of an EFC *j* to P availability by calculating the combined effect size for all I_j gene-encoded enzymes within the EFC *j*[26,27]. Effect size measures differential gene-encoded enzyme abundance between control and treatment (e.g., the control and P-fertilized soils) and is defined as the $\log_2$ fold change (EF_{ij})[14,27].

$$EF_{ij} = \log_2\left(E_{c_{ij}}/E_{t_{ij}}\right), \quad (1)$$

where $E_{c_{ij}}$ and $E_{f_{ij}}$ are the means of abundance of gene-encoded enzyme *i* (*i*=1, 2… I_j) within EFC $_j$ (*j*=1, 2,……, 22) in the control and P-fertilized samples, respectively.

We built a hierarchical effect size model to calculate the combined effect size for all EC numbers in an EFC *j* ($\overline{EF}_j$)[26,27].

$$\begin{cases} EF_{ij} = EF_j + \varepsilon_{ij}, \ \ \varepsilon_{ij} \sim N(0, s_{ij}^2) \\ EF_j = \overline{EF}_j + \delta_j, \ \ \delta_j \sim N(0, {\tau_j}^2) \end{cases}, \quad (2)$$

where $\varepsilon_{ij} \sim N(0, s_{ij}^2)$ is the sample variance of gene-encoded enzyme i within an EFC j and $\delta_j \sim N(0, {\tau_j}^2)$ is the variance among gene-encoded enzymes within an EFC *j*. s_{ij}^2 is the sample variation of the log fold-change EF_{ij} given by[27]:

$${s_{ij}}^2 = \frac{1}{n_t}\frac{s_{t_{ij}}^2}{E_{t_{ij}}^2} + \frac{1}{n_c}\frac{s_{c_{ij}}^2}{E_{c_{ij}}^2}, \quad (3)$$

where n_t and n_c are sample sizes in the P-fertilized and control soils, respectively ($n_t = n_c$ =2). $s_{t_{ij}}^2$ and $s_{c_{ij}}^2$ are the variance of abundance of enzyme i within an EFC *j* in the P-fertilized and control soils, respectively.

The effect size model interprets that the difference in the effect size of all gene-encoded enzymes in an EFC *j* is either from sampling error alone (${\tau_j}^2 = 0$, the fixed-effects model) or from both sampling error and the variance among gene-encoded enzymes within the EFC *j* (${\tau_j}^2 \neq 0$, the random-effects model).

We used the Cochran's homogeneity test[51,52] to examine the null hypothesis that the variance among gene-encoded enzymes within the EFC *j* is equal to zero.

$$Q_j = \sum_i^{I_j} w_{ij} \left(EF_{ij} - \widehat{\overline{EF}}_j\right)^2, \quad (4)$$

where $w_{ij} = s_{ij}^{-2}$, and $\widehat{\overline{EF}}_j = \frac{\sum_{i=1}^{I_j} w_{ij} EF_{ij}}{\sum_{i=1}^{I_j} w_{ij}}$ is the weighted least square estimator of the combined effect size when a zero ${\tau_j}^2$ is assumed. The null hypothesis is accepted when Q_j follows a Chi-square (χ_{I-1}) distribution and yields a p-value larger than 5% (within 95-percentile). Otherwise, ${\tau_j}^2$ and $\widehat{\overline{EF}}_j$ are estimated for a random-effects model as follows[26].

$$\tau_j^2 = max\left\{0, \frac{(Q_j - (I-1))}{(\sum w_{ij} - (\frac{\sum {w_{ij}}^2}{\sum w_{ij}}))}\right\}, \quad (5)$$

$$\widehat{\overline{EF}}_j = \frac{\sum_{i=1}^{I_j} {w_{ij}}^R EF_{ij}}{\sum_{i=1}^{I_j} {w_{ij}}^R}, \quad (6)$$

where ${w_{ij}}^R = (s_{ij}^2 + \tau_j^2)^{-1}$. The variances for both the fixed-effects model and random-effects model are defined as $Var\left(\widehat{\overline{EF}}_j\right) = \frac{\sum_{i=1}^{I_j} {w_{ij}}^R}{\left(\sum_{i=1}^{I_j} {w_{ij}}^R\right)^2}$, where for the fixed-effects model, τ_j^2 in ${w_{ij}}^R$ is zero.

Finally, we evaluated the statistical significance of the effect size of an EFC j by calculating the p-value of the adjusted Z statistics[26].

$$Z_{adj_j} = \frac{\widehat{\overline{EF}}_j}{(s_0 + \sqrt{Var(\widehat{\overline{EF}}_j)})}, \quad (7)$$

here s_0 is the quantile of the EFC-wise standard errors. We further calculated the adjusted p-value (q-value) by estimating the false discovery rates (FDR)[53]. The effect size of an EFC j is considered statistically significant with the q-value <0.05.

The CoMEND model

To enable representation of metagenomics-informed enzyme functional dynamics and the continuum SOM decomposition cascade[34], we developed the Continuum Microbial Enzyme Decomposition (CoMEND) model based upon the MEND model[28-30]. The original MEND model uses Michaelis-Menten kinetics with three generic enzyme pools to simulate enzyme-mediated decomposition processes of three physically-defined SOM pools[28]. In contrast, the CoMEND model contains 22 metagenomics-informed EFCs and uses chemical compositions and degrees of polymerization to hierarchically reclassify physically-defined SOM and inorganic nutrient pools. The first hierarchical level separates the chemistry of the raw organic inputs into two lignocellulose-containing (carbohydrate and lignin), two N-containing (protein and cell wall components), and three P-containing (nucleic acids, phospholipids and inositol P) pools. The second hierarchical level considers the degrees of polymerization, resulting in the SOM pools being further classified into residue pool (ROM), large polymer pool (LOM), oligopolymer pool (OOM), and monomer pool (MOM). Together these two hierarchies lead to four lignocellulose-containing, five N-

containing, five P-containing SOM pools and one bioavailable monomer pool (Fig. 2 and Supplementary Data S3). The third hierarchical level considers physical factors and classifies each SOM pool into microbial-activated (ASOM), mineral-protected (MSOM), and adsorbed phase (QSOM). Only the ASOM pools are catalyzed by the corresponding EFCs. The microbial-activated monomer pool (AMOM) is similar to the dissolved organic matter (DOM) pool in the original MEND model, which was defined as being potentially available for microbial uptake. The details of fluxes and dynamics of each SOM can be found in the SI Method S3.

Parameterization of the dynamic EFC allocation scheme for resource acquisition

The original MEND model considered the modification of environmental factors on reaction rates (Table S2), but not microbial acclimation to environmental perturbation. In the CoMEND model, we develop a metagenomics-informed dynamic EFC allocation scheme to parameterize adaptive microbial responses to environmental perturbation. This scheme assumes that the allocation of microbially-synthesized enzymes to each EFC varies with the availability of C, nutrients and soil water in order to maximize the acquisition of limiting resources and minimize energy consumption and osmotic stress. The fractions of microbially synthesized enzymes allocated to 22 EFC (f_{C_i} , f_{N_j} , f_{P_k}) are calculated as a function of the limitation factors of C, N, P, and soil water and the sensitivity of specific EFC allocation to each limitation factor, which depend on metagenomics-informed enzyme synthesis and allocation among nutrient acquisition, energy investment, and osmolyte synthesis and can be validated by integrated effect size of each EFC in response to nutrient and water stresses. The parameterization details of the dynamic EFC allocation scheme can be found in the SI Method S4.

Model parameterization

The CoMEND model had two types of parameters: the general kinetics parameters and the site-specific parameters. The kinetics parameters for a given EFC (Supplementary Data S4) were the maximum specific enzyme activity (Vd) and half-saturation constant (Ks) at the reference temperature and optimum pH, activation energy (Ea), optimal soil pH (pH_{opt}), and sensitivity to soil pH (pH_{sen}). We identified these parameters by collecting around 4900 kinetics parameters for 118 EFCs-involved enzymes from the BRENDA biochemical database[20]. To differentiate kinetics parameters among different microbial kingdoms, the kinetic parameters of each enzyme and their variability among isoenzymes were quantified by estimating the mean value and standard deviation of each enzyme's data collected from bacteria, fungi or archaea respectively and then calculating their weighted mean based on omics-informed relative ratios of bacteria, fungi or archaea[14]. Then, the kinetic parameters of each EFC and their variability were estimated by calculating weighted mean and weighted standard deviation of the corresponding kinetic parameters of all enzyme classification (EC) in each EFC (SI Method S5). The weighted factor was the relative gene abundance of each EC number within the EFC (Supplementary Data S2). The site-specific parameters (e.g.,

the sensitivity parameters of enzyme allocation to resource availability, rate constant of inorganic P conversion, etc. Supplementary Data S5) were optimized using the SCE (Shuffled Complex Evolution) algorithm[29,54] (Method S6 in SI). To avoid over-fitting and make sure the tested differences among $CoMEND_L$, $CoMEND_M$, $CoMEND_H$ and $CoMEND_{HD}$ result from model structure differences rather than model parameters optimization, we only calibrated these site-specific parameters for $CoMEND_{HD}$ and applied them in the other version of CoMEND. Other site-specific parameters were determined based on literature research (Supplementary Data S6).

Model initialization and case simulation

To determine how metagenomics-informed functional diversity and environmental acclimation of microbial communities affect the simulation of microbial activities and the soil decomposition processes mediated by them, we constructed four versions of the CoMEND model, labeled $CoMEND_L$, $CoMEND_M$, $CoMEND_H$ and $CoMEND_{HD}$, respectively. $CoMEND_L$, $CoMEND_M$, and $CoMEND_H$ represented low (11 EFCs), moderate (15 EFCs), and high (22 EFCs) functional diversity of microbial communities, with fixed enzyme allocation and thus no environmental acclimation. $CoMEND_{HD}$ was the complete version of CoMEND with high functional diversity and dynamic enzyme allocation strategy. These four versions were used to simulate the soil organic matter and microbial biomass in one P-fertilized plot and one control plots at the long-term (17 years) fertilization site in Panama. For each simulation, the pool sizes of SOM, inorganic N and P, and microbial biomass were initialized with soil measurements from 1998 representing the initial soil physical and chemical properties before fertilization[18,47,48,55] (Supplementary Data S7). Then two simulations were conducted for each version of the model with and without P fertilization from the year 1998 to 2014, respectively and driven by observed hourly soil temperature, hourly soil moisture, and monthly litter input and plant P uptake rate. Built upon these simulations, we further use CESM-projected climate forcing, litter input, and plant P uptake rate to drive simulations over the period of 2015-2100. Details about the input data can be found in the Method S7-S8 of the SI.

Soil incubation experiment

To evaluate interactive effects of microbial communities and environments on the soil CO_2 emission, we performed lab soil incubation experiments at 26^{o}C and sampled soil water content of the control and P-fertilization plots, respectively. One hundred grams equivalent dry weight of each soil sample was used and the incubation had three replicates. Each of the control and P-fertilized plots were incubated with no treatment to represent dry conditions, and 20 ml of Milli-Q water was added to represent the normal moisture level in Panama soils. Each incubation lasted for 120 hours, and the CO_2 emission was measured with a MicroOxymax respirometer (Columbus Instruments International, Columbus, OH, USA).

Model evaluation

We used the refined Willmott's index (WI)[41] to quantify the degree to which the model simulations agreed with the observed enzyme allocation and microbial CO_2 emission from laboratory incubations of the 2014 soil samples, and microbial biomass, C/P ratio, and soil carbon stock[18]. The WI index was calculated in Eq. 8.

$$WI = \begin{cases} 1 - \frac{\sum_{i=1}^{N}|M_i - O_i|}{2\sum_{i=1}^{N}|O_i - \bar{O}|} & if\ \sum_{i=1}^{N}|M_i - O_i| \le 2\sum_{i=1}^{N}|O_i - \bar{O}| \\ \frac{2\sum_{i=1}^{N}|O_i - \bar{O}|}{\sum_{i=1}^{N}|M_i - O_i|} - 1 & if\ \sum_{i=1}^{N}|M_i - O_i| > 2\sum_{i=1}^{N}|O_i - \bar{O}| \end{cases}, \quad (8)$$

where M_i and O_i represent the modeled and observed data, respectively. $\bar{O}$ is the mean of observed data. *N* is the number of the paired observed and modeled data. WI index varies from -1 to 1. The value of 1 indicates perfect agreement between the modeled and observed values, while -1 indicates either a lack of agreement between the model and observation or insufficient variation in observations to adequately test the model.

In addition, we performed a one-way analysis of variance (ANOVA) test to assess whether there were significant differences among observations and simulations by $CoMEND_L$, $CoMEND_M$, $CoMEND_H$, and $CoMEN_{HD}$. Analyzed variables includes the C/P ratios of microbial communities, the C/P ratios of decomposition-released bio-available organic matter ($R_{SOC/SOP}$), and monophosphate mineralization rates ($R_{monomerP}$), The significant differences in data among observations and different versions of the model were defined when the p-value of the ANOVA test was less than the 0.05 significance level. The Dunn and Sidak significant difference post hoc was performed to identify which two categories (e.g., observation vs simulation, or two simulation by different version of the model) for the tested variables were significantly different from each other, respectively. All statistical analyses were performed with Matlab 2019b.

Acknowledgments

This work is financially supported by the U.S. Department of Energy (DOE) Office of Biological and Environmental Research through an Early Career Award to MAM. We thank Dr. Lianhong Gu for his editing and constructive comments, which helped us to improve the manuscript. Oak Ridge National Laboratory is managed by the University of Tennessee-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. DOE.

References

1 Deutsch, C. A. *et al.* Impacts of climate warming on terrestrial ectotherms across latitude. *P Natl Acad Sci USA* **105**, 6668-6672, doi:10.1073/pnas.0709472105 (2008).

2 Schimel, J. P. & Schaeffer, S. M. Microbial control over carbon cycling in soil. *Frontiers in Microbiology* **3**, doi:ARTN 348 10.3389/fmicb.2012.00348 (2012).

3 Trivedi, P., Anderson, I. C. & Singh, B. K. Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. *Trends Microbiol* **21**, 641-651, doi:10.1016/j.tim.2013.09.005 (2013).

4 Todd-Brown, K. E. O. *et al.* Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* **11**, 2341-2356, doi:10.5194/bg-11-2341-2014 (2014).

5 Todd-Brown, K. E. O. *et al.* Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **10**, 1717-1736, doi:10.5194/bg-10-1717-2013 (2013).

6 Allison, S. D. A trait-based approach for modelling microbial litter decomposition. *Ecol Lett* **15**, 1058-1070, doi:10.1111/j.1461-0248.2012.01807.x (2012).

7 Bouskill, N. J., Tang, J., Riley, W. J. & Brodie, E. L. Trait-based representation of biological nitrification: model development, testing, and predicted community composition. *Front Microbiol* **3**, 364, doi:10.3389/fmicb.2012.00364 (2012).

8 Zehnder, A. J. B. & Brock, T. D. Methane Formation and Methane Oxidation by Methanogenic Bacteria. *J Bacteriol* **137**, 420-432 (1979).

9 McGlynn, S. E. Energy Metabolism during Anaerobic Methane Oxidation in ANME Archaea. *Microbes Environ* **32**, 5-13, doi:10.1264/jsme2.ME16166 (2017).

10 Chai, J. J., Kora, G., Ahn, T. H., Hyatt, D. & Pan, C. Functional phylogenomics analysis of bacteria and archaea using consistent genome annotation with UniFam. *Bmc Evol Biol* **14**, doi:ARTN 207 10.1186/s12862-014-0207-y (2014).

11 Reed, D. C., Algar, C. K., Huber, J. A. & Dick, G. J. Gene-centric approach to integrating environmental genomics and biogeochemical models. *P Natl Acad Sci USA* **111**, 1879-1884, doi:10.1073/pnas.1313713111 (2014).

12 Song, H. S. *et al.* Regulation-Structured Dynamic Metabolic Model Provides a Potential Mechanism for Delayed Enzyme Response in Denitrification Process. *Frontiers in Microbiology* **8**, doi:ARTN 1866 10.3389/fmicb.2017.01866 (2017).

13 Jansson, J. K. & Hofmockel, K. S. The soil microbiome - from metagenomics to metaphenomics. *Curr Opin Microbiol* **43**, 162-168, doi:10.1016/j.mib.2018.01.013 (2018).

14 Yao, Q. *et al.* Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. *Nat Ecol Evol* **2**, 499-509, doi:10.1038/s41559-017-0463-5 (2018).
15 Luo, C. *et al.* Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. *Appl Environ Microbiol* **80**, 1777-1786, doi:10.1128/AEM.03712-13 (2014).
16 Myrold, D. D. & Nannipieri, P. in *Omics in soil science* (eds Paolo Nannipieri, Giacomo Pietramellara, & Giancarlo Renella) 179-187 (Caister Academic Press, 2014).
17 Trivedi, P. *et al.* Microbial regulation of the soil carbon cycle: evidence from gene-enzyme relationships. *Isme J* **10**, 2593-2604, doi:10.1038/ismej.2016.65 (2016).
18 Turner, B. L. & Joseph Wright, S. The response of microbial biomass and hydrolytic enzymes to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. *Biogeochemistry* **117**, 115-130, doi:10.1007/s10533-013-9848-y (2014).
19 Cuesta, S. M., Rahman, S. A., Furnham, N. & Thornton, J. M. The Classification and Evolution of Enzyme Function. *Biophys J* **109**, 1082-1086, doi:10.1016/j.bpj.2015.04.020 (2015).
20 Scheer, M. *et al.* BRENDA, the enzyme information system in 2011. *Nucleic Acids Research* **39**, D670-D676, doi:10.1093/nar/gkq1089 (2011).
21 Tilman, D. The Resource-Ratio Hypothesis of Plant Succession. *Am Nat* **125**, 827-852, doi:Doi 10.1086/284382 (1985).
22 Smith, V. H., Graham, D. W. & Cleland, D. D. Application of Resource-Ratio Theory to Hydrocarbon Biodegradation. *Environmental Science & Technology* **32**, 3386-3395, doi:10.1021/es9805019 (1998).
23 Malik, A. A. *et al.* Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *The ISME Journal* **14**, 1-9, doi:10.1038/s41396-019-0510-0 (2020).
24 Bouskill, N. J. *et al.* Belowground Response to Drought in a Tropical Forest Soil. II. Change in Microbial Function Impacts Carbon Composition. *Frontiers in Microbiology* **7**, doi:ARTN 323
10.3389/fmicb.2016.00323 (2016).
25 Bouskill, N. J. *et al.* Belowground Response to Drought in a Tropical Forest Soil. I. Changes in Microbial Functional Potential and Metabolism. *Frontiers in Microbiology* **7**, doi:ARTN 525
10.3389/fmicb.2016.00525 (2016).
26 Hu, P. Z., Greenwood, C. M. T. & Beyene, J. Integrative analysis of multiple gene expression profiles with quality-adjusted effect size models. *Bmc Bioinformatics* **6**, doi:Artn 128
10.1186/1471-2105-6-128 (2005).

27 Hu, P. Z., Greenwood, C. M. T. & Beyene, J. Using the ratio of means as the effect size measure in combining results of microarray experiments. *Bmc Syst Biol* **3**, doi:Artn 106 10.1186/1752-0509-3-106 (2009).
28 Wang, G. S., Post, W. M. & Mayes, M. A. Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. *Ecol Appl* **23**, 255-272, doi:Doi 10.1890/12-0681.1 (2013).
29 Wang, G. S. *et al.* Microbial dormancy improves development and experimental validation of ecosystem model. *Isme J* **9**, 226-237, doi:10.1038/ismej.2014.120 (2015).
30 Wang, G. S., Mayes, M. A., Gu, L. H. & Schadt, C. W. Representation of Dormant and Active Microbial Dynamics for Ecosystem Modeling. *Plos One* **9**, doi:ARTN e89252 10.1371/journal.pone.0089252 (2014).
31 Krause, S. *et al.* Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* **5**, 251, doi:10.3389/fmicb.2014.00251 (2014).
32 Wright, S. J. *et al.* Potassium, phosphorus, or nitrogen limit root allocation, tree growth, or litter production in a lowland tropical forest. *Ecology* **92**, 1616-1625 (2011).
33 Turner, B. L., Yavitt, J. B., Harms, K. E., Garcia, M. N. & Wright, S. J. Seasonal changes in soil organic matter after a decade of nutrient addition in a lowland tropical forest. *Biogeochemistry* **123**, 221-235, doi:10.1007/s10533-014-0064-1 (2015).
34 Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60-68, doi:10.1038/nature16069 (2015).
35 Bunemann, E. K. Assessment of gross and net mineralization rates of soil organic phosphorus - A review. *Soil Biol Biochem* **89**, 82-98, doi:10.1016/j.soilbio.2015.06.026 (2015).
36 Turner, B. L. Resource partitioning for soil phosphorus: a hypothesis. *J Ecol* **96**, 698-702, doi:10.1111/j.1365-2745.2008.01384.x (2008).
37 Turner, B. L. & Wright, S. J. The response of microbial biomass and hydrolytic enzymes to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. *Biogeochemistry* **117**, 115-130, doi:10.1007/s10533-013-9848-y (2014).
38 Mouginot, C. *et al.* Elemental stoichiometry of Fungi and Bacteria strains from grassland leaf litter. *Soil Biology and Biochemistry* **76**, 278-285, doi:10.1016/j.soilbio.2014.05.011 (2014).
39 Griffiths, B. S., Spilles, A. & Bonkowski, M. C:N:P stoichiometry and nutrient limitation of the soil microbial biomass in a grazed grassland site under experimental P limitation or excess. *Ecological Processes* **1**, 6, doi:10.1186/2192-1709-1-6 (2012).

40 Xu, X. *et al.* Convergence of microbial assimilations of soil carbon, nitrogen, phosphorus, and sulfur in terrestrial ecosystems. *Sci Rep* **5**, 17445, doi:10.1038/srep17445 (2015).
41 Willmott, C. J., Robeson, S. M. & Matsuura, K. A refined index of model performance. *Int J Climatol* **32**, 2088-2094, doi:10.1002/joc.2419 (2012).
42 Kay, J. E. *et al.* The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society* **96**, 1333-1349, doi:10.1175/bams-d-13-00255.1 (2015).
43 Farquhar, G. D., Caemmerer, S. V. & Berry, J. A. A Biochemical-Model of Photosynthetic Co2 Assimilation in Leaves of C-3 Species. *Planta* **149**, 78-90, doi:Doi 10.1007/Bf00386231 (1980).
44 Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457-+, doi:10.1038/nature24621 (2017).
45 Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44**, D471-D480, doi:10.1093/nar/gkv1164 (2016).
46 Darch, T. *et al.* Assessment of bioavailable organic phosphorus in tropical forest soils by organic acid extraction and phosphatase hydrolysis. *Geoderma* **284**, 93-102, doi:10.1016/j.geoderma.2016.08.018 (2016).
47 Mirabello, M. J. *et al.* Soil phosphorus responses to chronic nutrient fertilisation and seasonal drought in a humid lowland forest, Panama. *Soil Research* **51**, 215-221, doi:https://doi.org/10.1071/SR12188 (2013).
48 Schwendenmann, L. & Pendall, E. Effects of forest conversion into grassland on soil aggregate structure and carbon storage in Panama: evidence from soil carbon fractionation and stable isotopes. *Plant Soil* **288**, 217-232, doi:10.1007/s11104-006-9109-0 (2006).
49 Tfaily, M. M. *et al.* Advanced Solvent Based Methods for Molecular Characterization of Soil Organic Matter by High-Resolution Mass Spectrometry. *Analytical Chemistry* **87**, 5206-5215, doi:10.1021/acs.analchem.5b00116 (2015).
50 Tfaily, M. M. *et al.* Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Analytica Chimica Acta* **972**, 54-61, doi:10.1016/j.aca.2017.03.031 (2017).
51 Cochran, W. G. The combination of estimates from different experiments. *Biometrics* **10**, 101-129 (1954).
52 Choi, J. K., Yu, U., Kim, S. & Yoo, O. J. Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* **19**, i84-i90, doi:10.1093/bioinformatics/btg1010 (2003).
53 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).

54 Duan, Q. Y., Sorooshian, S. & Gupta, V. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resour Res* **28**, 1015-1031, doi:Doi 10.1029/91wr02985 (1992).

55 Sayer, E. J. *et al.* Variable Responses of Lowland Tropical Forest Nutrient Status to Fertilization and Litter Manipulation. *Ecosystems* **15**, 387-400, doi:10.1007/s10021-011-9516-9 (2012).

Figures and Tables

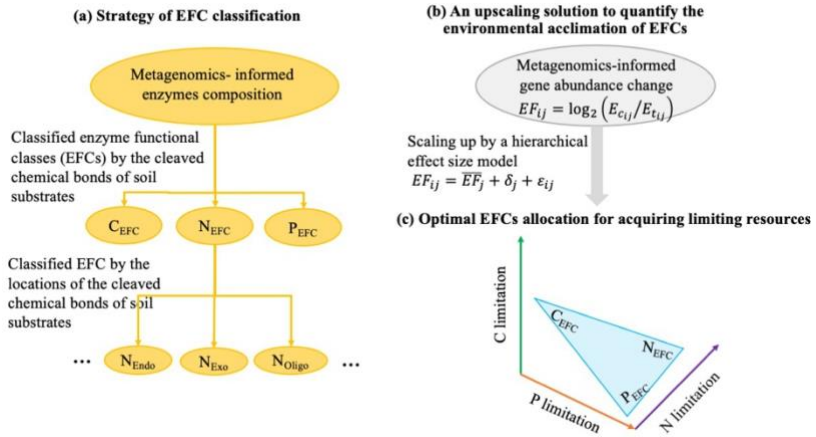


Figure 1. Strategies for modeling functional diversity and environmental acclimation of soil microbial communities. (a) Classification of enzyme functional classes (EFCs); (b) upscaling environmental acclimation of microbial communities from the gene to community level; (c) Optimal EFCs allocation for acquiring limiting resources. Here C_{EFC}, N_{EFC}, and P_{EFC} represent EFCs for decomposing lignocellulose, N-containing soil organic matter (SOM), and P-containing SOM, respectively. N_{Endo}, N_{Exo}, N_{Oligo} indicate the EFC that acts on the internal chemical bond of N-containing SOM residues, terminal bonds of large polymer N-containing SOM, and chemical bonds in the oligopolymers of N-containing SOM. The angle size of the triangle indicates the tradeoff of enzyme production between C_{EFC}, N_{EFC}, and P_{EFC}. The C, N, and P limitation increases along the direction of each axis.

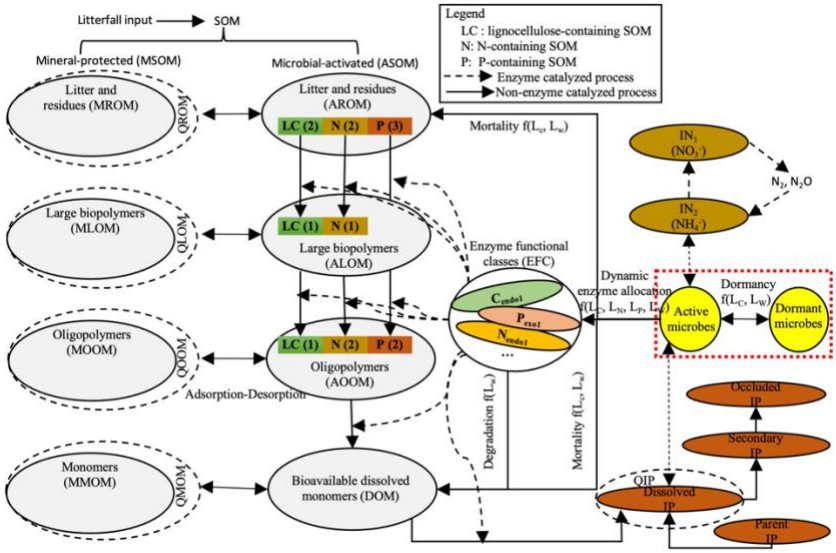
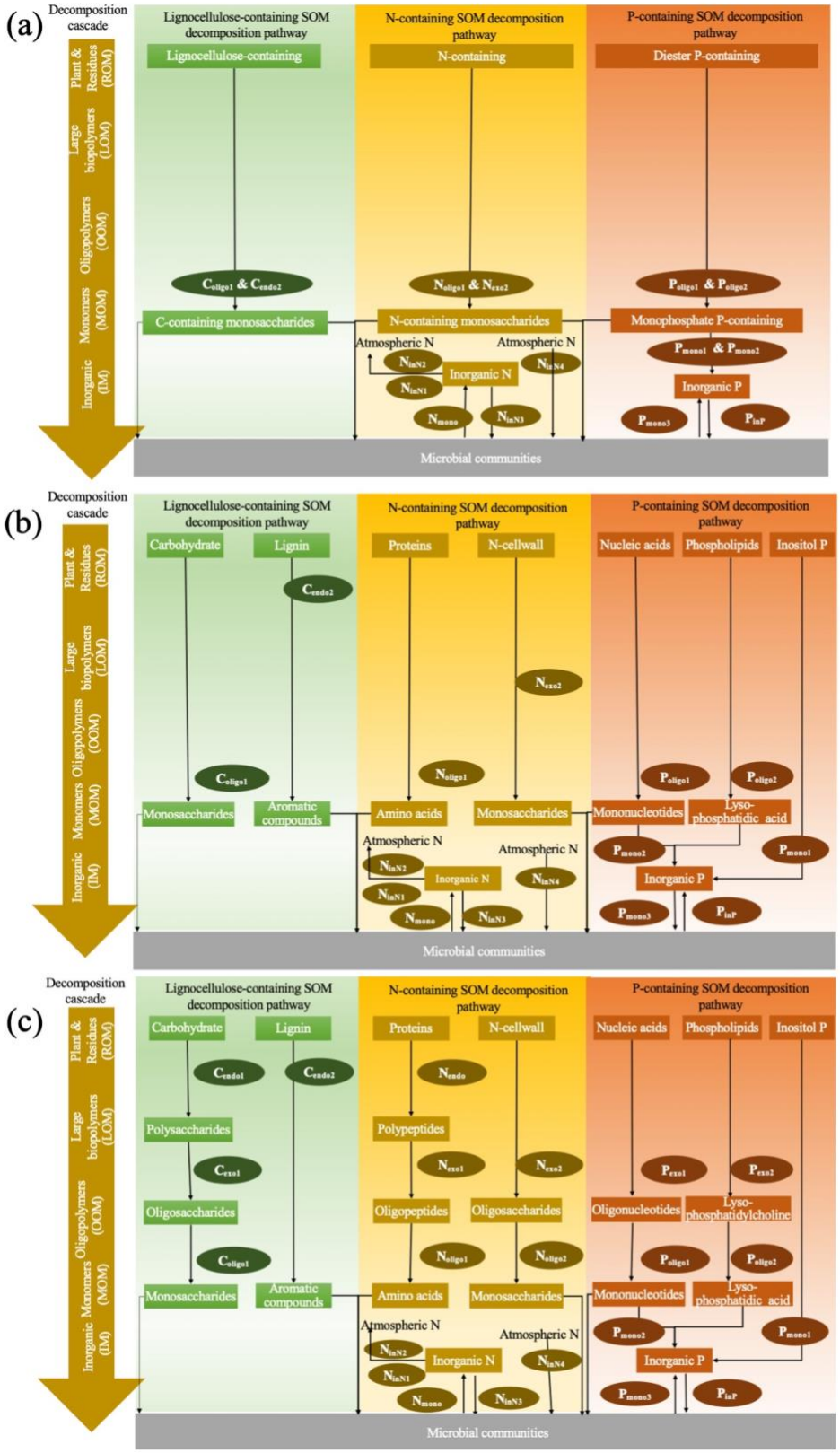
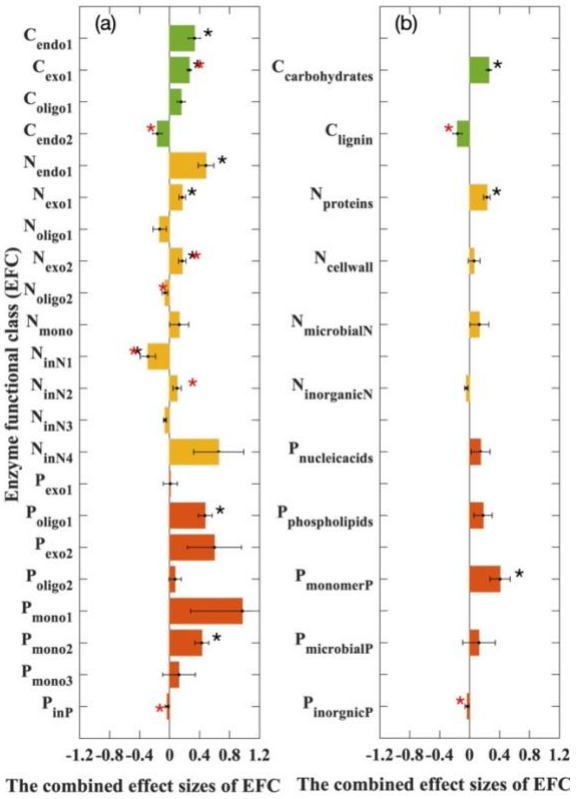


Figure 2. Conceptual diagram of the Continuum Microbial Enzyme Decomposition (CoMEND) model. The soil organic matter (SOM) is partitioned following the continuum decomposition from the microbially-activated residue pool (AROM) to large polymer pool (ALOM), to oligopolymer pool (AOOM) and finally to bioavailable dissolved monomer pool (DOM). Each microbially-activated SOM pool has its corresponding adsorbed phase pool (e.g. QROM, QLOM, QOOM, QMOM) and mineral-protected pool (e.g. MROM, MLOM, MOOM, MMOM). Each SOM pool is classified as lignocellulose-containing, N-containing and P-containing SOM according to the central pathways involved in releasing bioavailable C, N, and P. The value in the parentheses indicates the number of sub-pools in each lignocellulose-containing, N-containing, and P-containing SOM pool. L_C, L_N, L_P, and L_W represent the limitation factors for C, N, P, and soil water. The closer their values are to 1, the stronger their limitations of C, N, P and soil water.


(a)
Decomposition cascade
Plant & Residues (ROM)
Large biopolymers (LOM)
Oligopolymers (OOM)
Monomers (MOM)
Inorganic (IM)
Lignocellulose-containing SOM decomposition pathway
Lignocellulose-containing
N-containing SOM decomposition pathway
N-containing
P-containing SOM decomposition pathway
Diester P-containing
C-containing monosaccharides
N-containing monosaccharides
Monophosphate P-containing
Atmospheric N
Atmospheric N
Inorganic N
Inorganic P
Microbial communities
(b)
Decomposition cascade
Lignocellulose-containing SOM decomposition pathway
Carbohydrate
Lignin
N-containing SOM decomposition pathway
Proteins
N-cellwall
P-containing SOM decomposition pathway
Nucleic acids
Phospholipids
Inositol P
Monosaccharides
Aromatic compounds
Amino acids
Monosaccharides
Mononucleotides
Lyso-phosphatidic acid
Atmospheric N
Atmospheric N
Inorganic N
Inorganic P
Microbial communities
(c)
Decomposition cascade
Lignocellulose-containing SOM decomposition pathway
Carbohydrate
Lignin
N-containing SOM decomposition pathway
Proteins
N-cellwall
P-containing SOM decomposition pathway
Nucleic acids
Phospholipids
Inositol P
Polysaccharides
Polypeptides
Oligosaccharides
Oligopeptides
Oligosaccharides
Oligonucleotides
Lyso-phosphatidylcholine
Monosaccharides
Aromatic compounds
Amino acids
Monosaccharides
Mononucleotides
Lyso-phosphatidic acid
Atmospheric N
Atmospheric N
Inorganic N
Inorganic P
Microbial communities

901

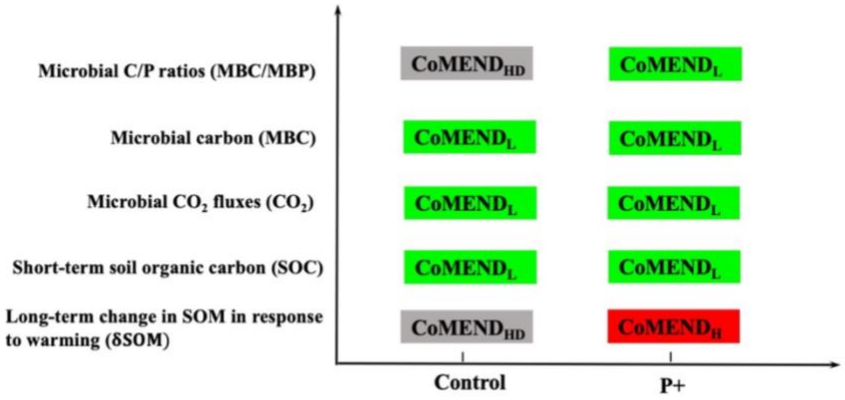
Figure 3. Metagenomics-informed enzyme functional classes (EFCs) and the cascade of soil organic matter (SOM) decomposition catalyzed by the EFCs in three versions of CoMEND: (a) $CoMEND_L$; (b) $CoMEND_M$, and (c) $CoMEND_H$. $CoMEND_H$ included all 22 metagenomics-informed EFCs for decomposing diverse SOM with different degrees of polymerization (e.g., polysaccharides vs. oligosaccharides) and energy expenditures for decomposition (e.g., complex lignin vs. simple carbohydrates), and thus represented high enzyme functional diversity. $CoMEND_M$ only included 15 EFCs for diverse SOM with different energy expenditure for decomposition and thus represented moderate functional diversity of microbial community. $CoMEND_L$ did not distinguish SOM complexity, included just 11 EFCs, and thus represented low functional diversity of microbial community. $CoMEND_{HD}$ added environmental acclimation of microbial communities to $CoMEND_H$. The definition of each EFC can be found in Table S1.



915

916 **Figure 4.** The combined effect sizes of soil enzyme functional classes (EFCs) between the control and the
917 P-fertilized soils. The green, yellow, and red bars represent EFCs for decomposing lignocellulose-
918 containing, N-containing, and P-containing SOM, respectively. In the P-limited control soils, microbial
919 communities increased enzyme investment to acquire limiting P resources and favored simple C substrate
920 (e.g., carbohydrates), instead of complex lignin, to minimize energy expenditure for C acquisition.: (a)
921 EFC is classified based on both the chemical compounds and the locations of the cleaved chemical bonds
922 of the organic substrate that each EC number acts on; (b) EFC is classified based on the chemical
923 compounds of soil substrate that each EC number acts on. The error bar represents the standard deviation
924 of the combined effect size of each EFC. Black stars (*) indicate statistically significant differences in the
925 EFC between the control and P-fertilized soils (q-value <0.05). Red stars (*) indicate a homogenous effect
926 size for all EC numbers within an EFC.

927



928

929 Figure 5. A summary of the effect of high, moderate, and low microbial functional diversity and
930 environmental acclimation on microbial C/P ratios (MBC/MBP), microbial carbon (MBC), microbial CO_2
931 fluxes (CO_2), and short-term and long-term soil organic matter (SOM) dynamics. Here we indicate this
932 effect by comparing the performance of four versions of CoMEND and listing the version of the model that
933 can better capture each tested variable and has the minimum complexity of model parameterization. The
934 four versions of the model with different complexity of model parameterization include: (1) $CoMEND_H$
935 included all 22 metagenomics-informed EFCs for decomposing diverse SOM with different degrees of
936 polymerization (e.g., polysaccharides vs. oligosaccharides) and energy expenditures for decomposition
937 (e.g., complex lignin vs. simple carbohydrates), and thus represented high enzyme functional diversity.
938 $CoMEND_M$ only included 15 EFCs for diverse SOM with different energy expenditure for decomposition
939 and thus represented moderate functional diversity of microbial community. $CoMEND_L$ did not distinguish
940 SOM complexity, included just 11 EFCs, and thus represented low functional diversity of microbial
941 community. $CoMEND_{HD}$ added environmental acclimation of microbial communities to $CoMEND_H$. The
942 definition of each EFC can be found in Table S1. Here Control and P+ represent the P-limited Control plots
943 and the P-fertilized plots, respectively

944

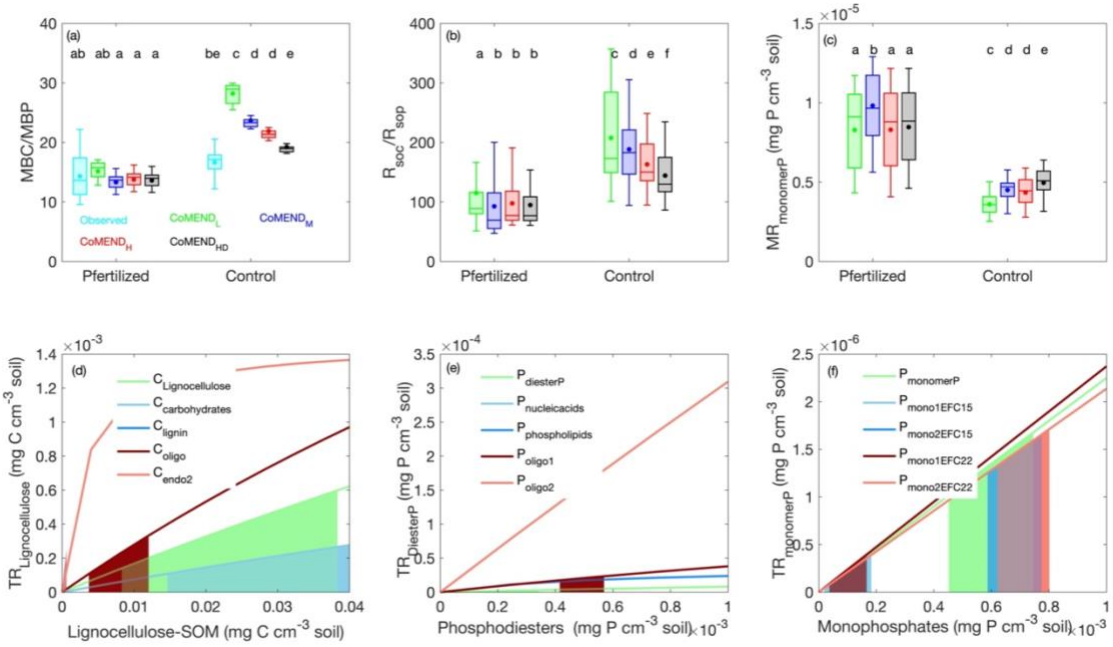
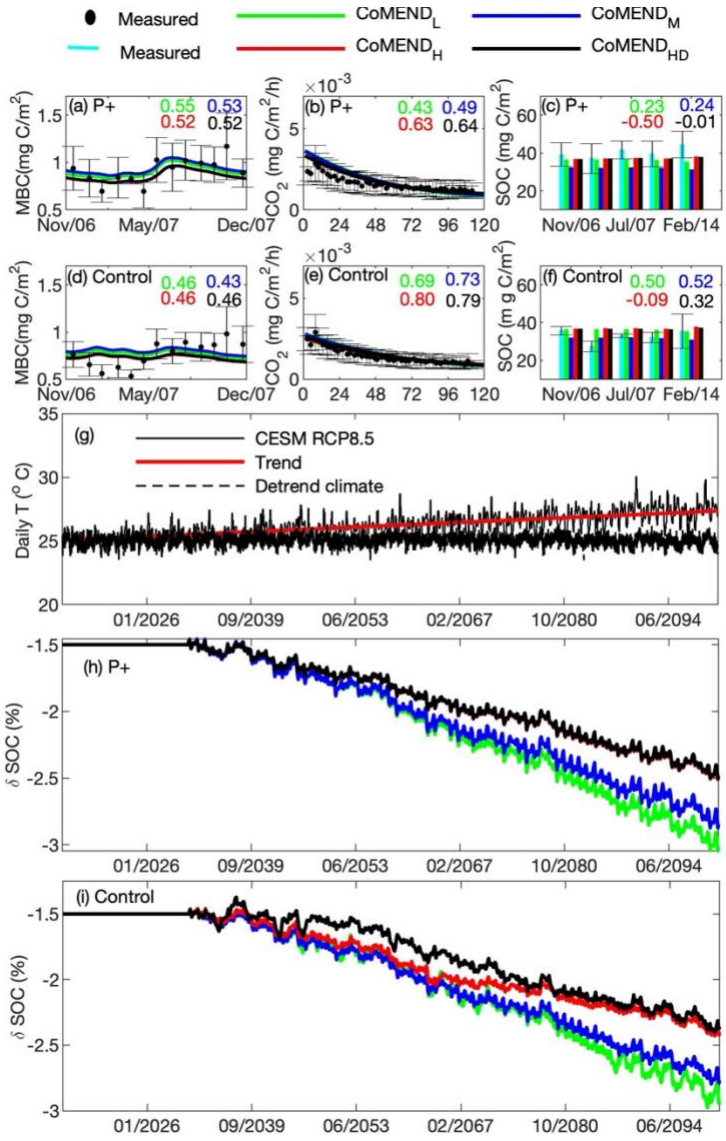


Figure 6. The effects of microbial functional diversity and environmental acclimation on modeled (a) microbial C/P ratios (MBC/MBP) (data number, N=13); (b) C/P ratios of decomposition-released bio-available organic matter (R_{SOC}/R_{SOP}) (N=395); (c) monophosphates mineralization rates ($MR_{monomerP}$) (N=395); and the turnover rates (TR) of each EFC-catalyzed (d) lignocelluloses, (e) phosphodiesters, and (f) monophosphates in response to the availability of the corresponding soil substrates during the year 2006-2007. Four versions of the CoMEND model are compared here: $CoMEND_H$ included all metagenomics-informed 22 EFCs for SOM decomposition and thus represented high enzyme functional diversity. $CoMEND_M$ only included 15 EFCs for SOM decomposition and thus represented moderate functional diversity of microbial community. $CoMEND_L$ included 11 clusters of EFCs and represented low functional diversity of microbial community. Build upon the version of $CoMEND_H$, $CoMEND_{HD}$ version refers to the model with high microbial functional diversity and the representation of environmental acclimation of microbial communities. $CoMEND_{HD}$ assumed that enzyme synthesis and allocation to 22 EFCs is dynamically adjusted in response to resource availability. Here the letters above each box in subplots a-c indicate the statistical differences (Dunn and Sidak, $p<0.05$) among observation, $CoMEND_L$, $CoMEND_M$, $CoMEND_H$ and $CoMEND_{HD}$. In the subplots d-f, $C_{Lignocellulose}$, $P_{diesterP}$, and $P_{monomerP}$ are EFCs considered in $CoMEND_L$; $C_{carbohydrates}$, C_{lignin}, $P_{nucleicacids}$, $P_{phospholipids}$, $P_{mono1EFC15}$ and $P_{mono2EFC15}$ are EFCs considered in $CoMEND_M$, and $P_{mono1EFC22}$ and $P_{mono2EFC22}$ are EFCs considered in $CoMEND_H$ and $CoMEND_{HD}$. The colored shaded area under each curve in (d-f) indicates the range of turnover rates of each EFC-catalyzed SOM over the year 2006-2007.



966

967 **Figure 7.** The effects of microbial functional diversity and environmental acclimation on (a and d)
968 microbial carbon over the year 2006-2007 (data number, N=13); (b, e) incubated microbial respiration rate
969 (N=43); (c, f) soil organic carbon (SOC) stocks over the year 2006-2014 (N=5); and (g) temperature change
970 trends projected by the Community Earth System Model (CESM) under the high-emission scenario of
971 representative concentration pathway (RCP8.5) and (h-i) its impact on projected SOC over the year 2015-
972 2100 in the Panamanian soils. Here P+ and Control refers to the P-fertilized soils and the control soils. The
973 soils are above 10cm of soil depth. The colored numbers in a-f are the Willmott index of agreement (WI)
974 between measurements and simulation by CoMEND$_L$, CoMEND$_M$, CoMEND$_H$, and CoMEND$_{HD}$,
975 respectively.

976

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SongetalSI.pdf
- SIGuide.pdf
- SupplementaryDataS1.xls
- SupplementaryDataS2.xlsx
- SupplementaryDataS3.xlsx
- SupplementaryDataS4.xlsx
- SupplementaryDataS5.xlsx
- SupplementaryDataS6.xlsx
- SupplementaryDataS7.xlsx