

Global patterns in the growth potential of soil bacterial communities

Received: 26 January 2024

Accepted: 9 July 2024

Published online: 11 August 2024

 Check for updates

Ernest D. Osburn^{1,2}✉, Steven G. McBride³, Mohammad Bahram^{4,5,6} & Michael S. Strickland²

Despite the growing catalogue of studies detailing the taxonomic and functional composition of soil bacterial communities, the life history traits of those communities remain largely unknown. This study analyzes a global dataset of soil metagenomes to explore environmental drivers of growth potential, a fundamental aspect of bacterial life history. We find that growth potential, estimated from codon usage statistics, was highest in forested biomes and lowest in arid latitudes. This indicates that bacterial productivity generally reflects ecosystem productivity globally. Accordingly, the strongest environmental predictors of growth potential were productivity indicators, such as distance to the equator, and soil properties that vary along productivity gradients, such as pH and carbon to nitrogen ratios. We also observe that growth potential was negatively correlated with the relative abundances of genes involved in carbohydrate metabolism, demonstrating tradeoffs between growth and resource acquisition in soil bacteria. Overall, we identify macroecological patterns in bacterial growth potential and link growth rates to soil carbon cycling.

Growth is a fundamental property of microbial communities that underpins myriad ecosystem functions^{1,2}. For example, increased microbial growth comes at the cost of microbial resource acquisition (e.g., exoenzyme production)³, thereby influencing decomposition rates of soil carbon (C) sources. Similarly, microbial partitioning of assimilated C resources into growth vs. other processes (e.g., respiration) is a key driver of soil organic matter formation^{4,5}. These tradeoffs highlight key feedbacks among microbial communities, soil C stocks, and global climate, as the decomposition and eventual fate of soil C sources (i.e., storage vs. respiration as CO₂) is ultimately dependent upon microbial growth dynamics⁶. Therefore, to better understand and predict variation in ecosystem C cycling and storage, an improved understanding of the patterns and drivers of microbial growth potential is needed.

Microbial growth is frequently measured in ecological studies using isotope tracing methods, often to quantify microbial C use

efficiency⁴ or to elucidate growth rates of specific microbial groups^{7,8}. Studies employing these methods in soil have demonstrated that microbial growth rates vary among bacterial taxa⁹, with soil depth¹⁰, among ecosystem types¹¹, and in response to many anthropogenic environmental changes, including land use change¹², drought¹³, warming^{13,14}, and fertilization¹⁵. Isotopic methods have the advantage of providing direct measurements of actual in situ microbial growth yield, though the addition of water or exogenous substrates biases isotopic growth estimates⁴. In addition, no study (likely due to low throughput) has applied these methods at the spatial scale required to elucidate macroecological patterns in microbial growth.

Another approach to estimating microbial growth rates involves the use of genomic or metagenomic sequence data^{16–18}. Metagenomic approaches to estimating microbial growth potential, along with other community-aggregated traits^{19,20}, are promising given the increasing prevalence of publicly available sequence data collected from different

¹Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, USA. ²Department of Soil and Water Systems, University of Idaho, Moscow, ID, USA. ³Department of Biology, Radford University, Radford, VA, USA. ⁴Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden. ⁵Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia. ⁶Department of Agroecology, Aarhus University, Slagelse, Denmark.

✉ e-mail: e.osburn@uky.edu

ecosystem types across the globe²¹. An additional advantage of genomic/metagenomic growth estimates is that they do not suffer from biases introduced by substrate or water addition. Estimates of maximum bacterial growth rates from genomes and metagenomes are based on the observation that bacterial growth rates are best predicted by the degree of codon usage bias present within a particular organism or community¹⁶. This is because rapidly-growing bacterial taxa have undergone selection for highly biased codon usage in their highly-expressed genes (e.g., ribosomal protein genes) such that the codons match their cellular tRNA pool, thus enabling rapid production of cellular components required for growth (e.g., ribosomes)¹⁶. Estimates of maximum bacterial growth rates from codon usage statistics have been shown to be highly accurate for both environmental and host-associated bacterial communities^{17,18,22}. Surprisingly, however, these metagenomic growth potential estimates have only very rarely been used for soil bacteria, though one prior study did show that maximum growth rates were higher for soil bacteria in vegetated soils compared with bare soils in an arid ecosystem²³. At the same time, another found no differences in growth rates among Chinese forest soils with different pH²⁴.

Though estimating microbial growth rates is possible using metagenomic methods, we know of no studies that rigorously assess global macroecological patterns in soil bacterial growth potential. This contrasts with other aspects of soil bacterial communities, e.g., taxonomic diversity and functional gene composition, for which some macroecological studies have emerged in recent years^{21,25,26}. These studies have demonstrated that soil bacteria do not exhibit the traditional latitudinal diversity gradients that are typical of macro-organisms (e.g., plants)^{21,25} and that global patterns in the functional attributes of bacterial communities (e.g., antibiotic resistance gene relative abundances and C acquisition gene relative abundances) are strongly associated with ecosystem aridity^{21,26,27}. Similarly, this study aimed to explore macroecological patterns and environmental drivers of soil bacterial growth potential. We also sought to identify global-scale relationships between growth potential and key bacterial C cycle functions. Assessing these relationships is critical because they represent potential mechanistic links between microbial life history and soil C cycling and because the current prevailing theoretical frameworks in microbial ecology predict them. For example, the Y-A-S (i.e., growth yield vs. resource acquisition vs. stress tolerance) framework predicts that the dominant life strategies of microorganisms are subject to tradeoffs. Thus, optimization of growth potential should come at the cost of C resource acquisition potential²⁸. Empirical evidence of these tradeoffs exists for some ecosystems³, though whether or not these tradeoffs manifest at macroecological scales is unknown.

In this study, we use a global dataset of soil metagenomes²¹ to estimate maximum growth rates for each community using a recently developed growth model based on codon usage statistics¹⁸. As an additional metric of growth potential, we estimate the community-averaged 16S rRNA operon number for each metagenome, which has also been shown to be a strong indicator of growth potential in bacteria²⁹. We hypothesize that the global patterns in bacterial growth potential will reflect ecosystem productivity, as highly productive ecosystems should exhibit soil conditions that promote bacterial growth, e.g., high soil moisture and organic matter content. In addition, in accordance with the Y-A-S framework, we predict that bacterial growth potential is negatively correlated with the relative abundances of C cycle resource acquisition genes on a global scale. In this work, we show that maximum growth rates of soil bacterial communities exhibit clear relationships with several environmental variables, including positive associations with forest cover and ecosystem net primary productivity. We also show that bacterial growth potential is negatively associated with C acquisition gene relative abundances, which is in line with the prevailing conceptual frameworks in microbial ecology.

Overall, our work demonstrates that bacterial growth is a key physiological mechanism that drives ecosystem C cycling at a global scale.

Results and discussion

Growth potential of soil bacteria across biomes

We used codon usage statistics to estimate bacterial growth potential in 176 soil metagenomes sampled from 11 terrestrial biomes across six continents (Supplementary Fig. 1). Estimated maximum growth rates exhibited significant variation across biomes, with latitude explaining ~15% of the variation in maximum growth rates (Fig. 1A). The best-supported regression model for maximum growth rates as a function of latitude was quadratic, with higher maximum rates observed at tropical and temperate/boreal latitudes and lower maximum rates in more arid subtropical latitudes (Fig. 1A). Interestingly, a prior analysis of these same metagenomes showed that bacterial taxonomic and functional diversity exhibited opposite latitudinal patterns compared to the patterns for maximum growth rates, i.e., taxonomic and functional diversity were higher in mid-latitudes²¹. This suggests that environmental selection for rapid microbial growth results in a less diverse set of highly competitive taxa that have a more streamlined set of functions. These global patterns are distinct from those observed in plant communities; unlike plants, high growth potential in bacteria is not associated with higher taxonomic diversity. This complements prior studies demonstrating that macroecological patterns in microbial community characteristics are different from the traditional latitudinal gradients seen in macro-organisms^{21,25}.

Based on the latitudinal patterns, it appeared that maximum growth rates were generally higher in forested biomes. To test this, we categorized the biomes as forested vs. non-forested and indeed observed ~12% higher maximum growth rates in forests (Fig. 1B). We observed the same latitudinal patterns and the same effects of forest cover in every iteration of the growth models we ran: in predictions from assembled contigs (Fig. 1), from gene fragments (Supplementary Fig. 2), and in a GC-corrected version of the model (Supplementary Fig. 3) (see “Methods” section). However, this dataset has a limited number of samples from non-forested environments from which to draw firm conclusions. Therefore, we repeated the same growth rate analysis in metagenomes from 42 National Ecological Observatory Network (NEON) sites located across the United States. In the NEON dataset, the effect of forest cover was even more pronounced, with ~31% higher predicted maximum growth rates in forested sites (Supplementary Fig. 4). These results allow us to place global soil bacterial communities within the copiotroph vs. oligotroph life history framework¹⁷, where rapidly-growing communities from forested environments can be considered more copiotrophic while slower growing communities from non-forests are more oligotrophic.

For additional validation, we also assessed community-averaged 16S rRNA gene copy numbers, as organisms adapted to rapid growth are known to have elevated rRNA gene copy numbers in order to boost cellular ribosome numbers needed for replication²⁹. As expected, average 16S rRNA gene copy numbers were significantly positively correlated with maximum growth rates (Fig. 1C). Maximum growth rates were also weakly positively correlated with average genome sizes (Fig. 1D). This contrasts with multiple prior studies that showed no relationship between growth and genome size^{16,30}. This suggests that while higher maximum growth rates and larger genomes tend to co-occur in some soils (i.e., in forested biomes), they may be independently selected for in those environments³⁰. To rigorously assess the relationship between maximum growth rates and metagenome GC content, we used a GC-corrected version of the growth model, as this ensures that the growth model itself has no dependency on GC content. We found a significant negative correlation between maximum growth rates and GC content (Supplementary Fig. 5), which has also been previously shown in soils from arid environments²³. Elevated GC content is a known adaptation to increased thermotolerance in

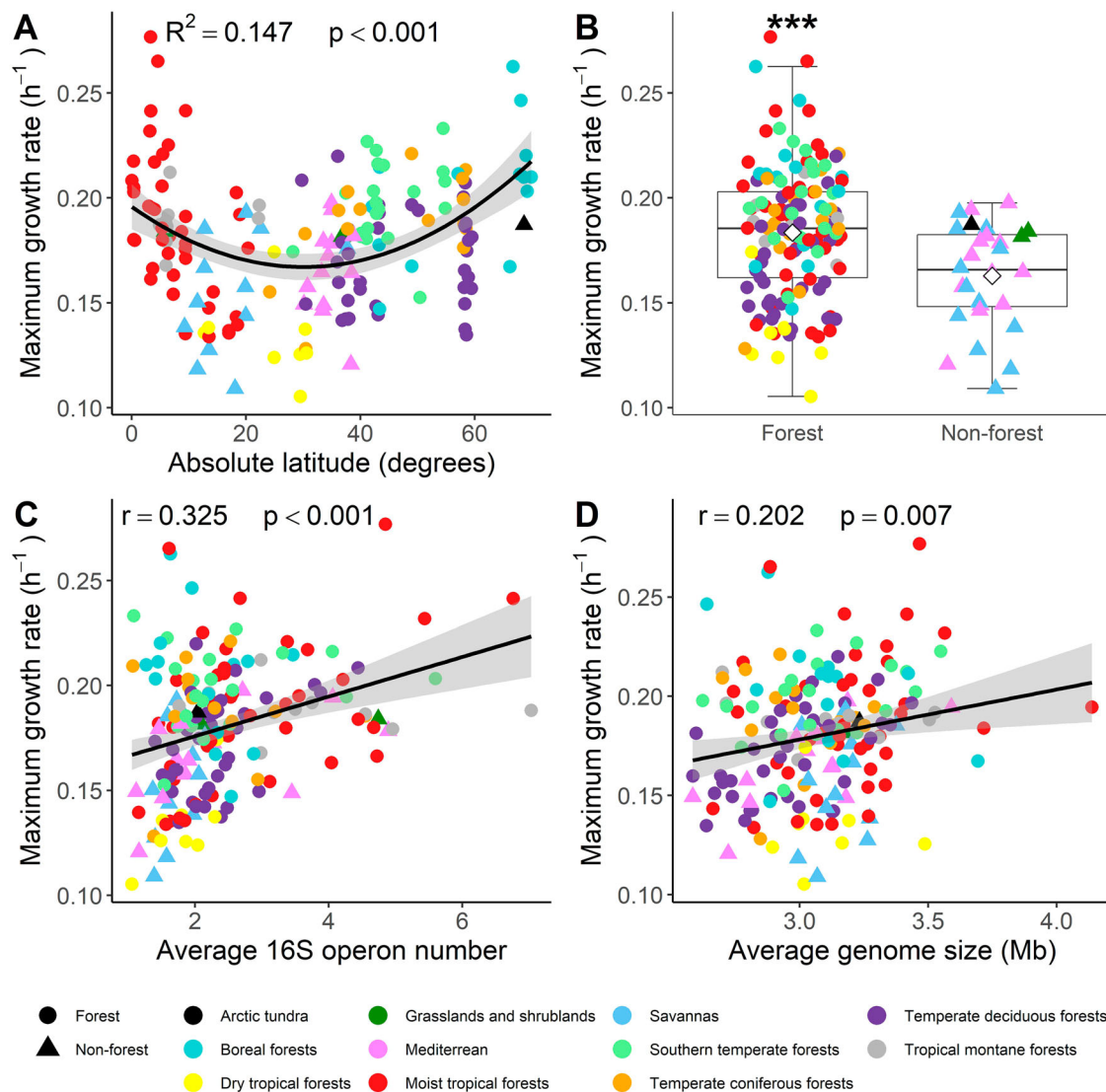


Fig. 1 | Global patterns in the growth potential of soil bacterial communities. Panel (A) shows the best-supported polynomial regression function (quadratic) for growth rates as a function of latitude ($F = 16.1$, $p < 0.001$). Differences in maximum growth rates between forested and non-forested biomes are shown in (B), with asterisks indicating significantly higher maximum growth rates ($F = 11.2$, $p < 0.001$). The box plots in (B) show group medians and interquartile ranges with diamond symbols indicating group means and whiskers representing the upper and lower quartiles plus or minus 1.5 times the interquartile range. The analyses shown in (A)

and (B) were one-tailed (F -tests). Also shown are Pearson correlations between maximum growth rates and average 16S rRNA gene copy number (C) and average genome size (D). The correlation analyses in (C) and (D) were two-tailed. The shaded areas in (A), (B), and (C) represent the 95% confidence interval around the regression line. For all the statistical analyses shown, $n = 176$. Source data are available as a Source Data file and on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

bacteria³¹. Therefore, our results suggest that there is a fundamental trade-off between stress tolerance and growth potential in soil bacteria. This tradeoff is a central component of the growth yield vs. resource acquisition vs. stress tolerance (Y-A-S) microbial life history concept (i.e., ‘yield’ vs. ‘stress tolerance’)²⁸, as well as the older competitor-stress tolerator-ruderal (C-S-R) concept upon which Y-A-S is based³². Our results confirm the existence of this tradeoff and identify the global environmental factors (e.g., forest cover) that influence the balance between stress tolerance and growth potential traits in soil bacteria.

While the ecological patterns in bacterial growth potential we observed appear to be robust, more deeply sequenced metagenomes would provide improved estimates. Whereas the assemblies in this study were not ideal (see ‘Methods’ section), greater sequencing depth would facilitate the assembly of full genomes from the metagenomes³³, which, in turn, would yield more accurate growth

estimates¹⁷. In addition to improved growth prediction accuracy, the estimation of growth rates of individual members of communities would also allow for studies of within-community variation of growth potential across environmental gradients. Given the continually increasing size and availability of sequence datasets, future studies should be able to address these questions. We also note that this study’s sampling scheme was unsuitable for assessing within-biome patterns in bacterial growth potential. Thus, future studies should also assess the scales at which metagenomic life history tradeoffs manifest, i.e., whether these tradeoffs exist within all biomes, only particular biomes, or only manifest at macroecological scales.

Environmental drivers of bacterial growth potential

Predicted maximum growth rates and associated genomic traits were significantly correlated with several climatic and soil variables (Supplementary Fig. 6). Growth metrics were positively correlated with

indicators of productivity – for example, maximum growth rates and 16S rRNA gene copy number were both positively correlated with ecosystem net primary productivity (NPP) (Supplementary Fig. 6). Growth rates were also positively correlated with soil moisture, soil organic C content, and C:N ratios, and negatively correlated with soil pH (Supplementary Fig. 6), all of which are soil properties that vary across productivity gradients. Similar correlations were also present in the metagenomes from the 42 NEON sites, where maximum growth rates were strongly negatively correlated with soil pH (Supplementary Fig. 4). To identify the strongest statistical predictors of bacterial growth potential, we used random forest regression. The random forest regression model explained 43% of the variation in maximum growth rates, and the model results corroborated the correlation analyses, identifying soil pH, soil C:N, and distance to the equator as the strongest predictors of maximum growth rates (Fig. 2). For 16S rRNA gene copy numbers, ecosystem NPP and soil moisture were the strongest predictors (Supplementary Fig. 7).

Given that acidic soils are generally understood to be stressful environments for soil microorganisms, it seems unlikely that increasingly acidic soil pH in itself would promote rapid bacterial growth, as is suggested by our random forest model (Fig. 2). Indeed, prior work has shown that within specific biomes, lower soil pH tends to inhibit microbial growth (grasslands)³ or does not influence growth rates (forests)²⁴. Therefore, we suggest that the strong association of low soil pH with high growth potential we observed on a global scale is probably due to low soil pH simply being a highly consistent indicator of high productivity rather than any direct positive influence of low pH on growth. Soil pH was indeed significantly correlated with nearly all indicators of ecosystem productivity (e.g., NPP, MAP, distance to equator) (Supplementary Fig. 8). Highly productive ecosystems then have other soil characteristics that are more likely to directly promote bacterial growth potential, e.g., high moisture and organic carbon content, both of which were positively correlated with growth potential (Supplementary Fig. 6). To support this conclusion, we ran an alternative random forest model for maximum growth rates that excluded soil pH as a candidate environmental predictor. The alternative model still explained 38% of the variation in maximum growth

rates (compared with 43% for the model that includes soil pH) and identified soil C:N, distance to equator, and soil moisture as the most important predictors (Supplementary Fig. 9). Thus, while we cannot rule out the possibility that soil pH plays a direct role in driving bacterial growth potential, our results do show that indicators of productivity other than soil pH perform similarly in accounting for variation in maximum bacterial growth rates. Overall, these results support our first hypothesis, demonstrating that the productivity of soil bacterial communities mirrors the productivity of the ecosystems that they inhabit. This finding demonstrates that through a chain of energy and nutrient transfer, the productivity of an ecosystem drives the productivity of all organisms within that ecosystem, including microorganisms. While there are surprisingly few studies investigating patterns and environmental drivers of bacterial growth across biomes^{8,11}, this conclusion is in agreement with experimental work that has demonstrated positive links between plant productivity and microbial growth^{34,35}. Our results are also in agreement with a previous study from an arid ecosystem that showed higher bacterial growth rates in more productive vegetated soils than in bare soils²³ as well as a study on marine bacteria that showed higher maximum growth rates at the ocean surface than in the less productive subsurface¹⁸. Prior work has also demonstrated positive links between ecosystem productivity and key soil health indicators³⁶, which suggests that bacterial growth potential may also be positively associated with overall soil health.

Overall, our results suggest that ecosystem productivity is a key global driver of bacterial growth potential, which, in turn, was highest in both tropical forests and temperate/boreal forests. This result seems inconsistent with the prevailing paradigm in ecology that ecosystem productivity is much higher in tropical latitudes than in temperate/boreal latitudes³⁷. However, in contrast to the historical paradigm, some recent studies have shown that temperate/boreal forests may actually have productivity similar to that of tropical forests^{37,38}, at least during the growing season³⁹. Therefore, ecosystem productivity and associated soil properties are plausible global drivers of bacterial growth potential. It is also possible that other characteristics of forest ecosystems, potentially related to ecosystem productivity but not productivity per se (e.g., high plant biomass accumulation), play a role in driving high bacterial growth potential, which should be investigated in future studies.

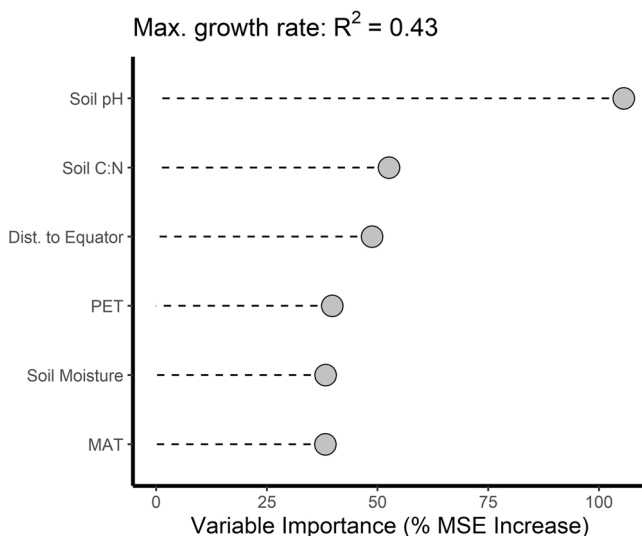


Fig. 2 | Random forest regression model showing the strongest environmental predictors of bacterial community-averaged maximum growth rates. The ‘importance’ of each variable was quantified as the increase in model error (mean squared error, ‘MSE’) when each respective variable was randomly shuffled across the dataset. For the random forest model, $n = 176$ samples. Source data are available as a Source Data file and on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

Relationships between growth potential and functional genes

Our second hypothesis was that bacterial growth metrics would be associated with the potential C cycle functions of the bacterial communities, with the expectation that higher growth potential would come at the cost of reduced potential for C resource acquisition. Here, we focus on the relative abundances of genes encoding enzymes involved in carbohydrate metabolism and transport (eggNOG COG ‘G’) and energy production and conversion (eggNOG COG ‘C’), as these genes are key drivers of microbial C cycling in soil. We found that maximum growth rates were positively correlated with the relative abundances of energy production and conversion genes (Fig. 3A) and negatively correlated with relative abundances of carbohydrate metabolism and transport genes (Fig. 3B). We observed the same correlations between 16S rRNA gene copy numbers and the two gene categories (Supplementary Fig. 10). In addition, similar to the results for eggNOG carbohydrate metabolism/transport genes, relative abundances of genes encoding carbohydrate-active enzymes (CAZymes) were negatively correlated with both maximum growth rates and 16S rRNA gene copy numbers (Supplementary Fig. 11). These results support our second hypothesis and are similar to prior studies that also found enrichment of carbohydrate metabolism genes in more oligotrophic (i.e., slower growing) bacterial communities²³ as well as prior work that has shown rapidly growing communities to have reduced investment into resource acquisition (i.e., extracellular enzymes)³. This tradeoff between bacterial growth potential and

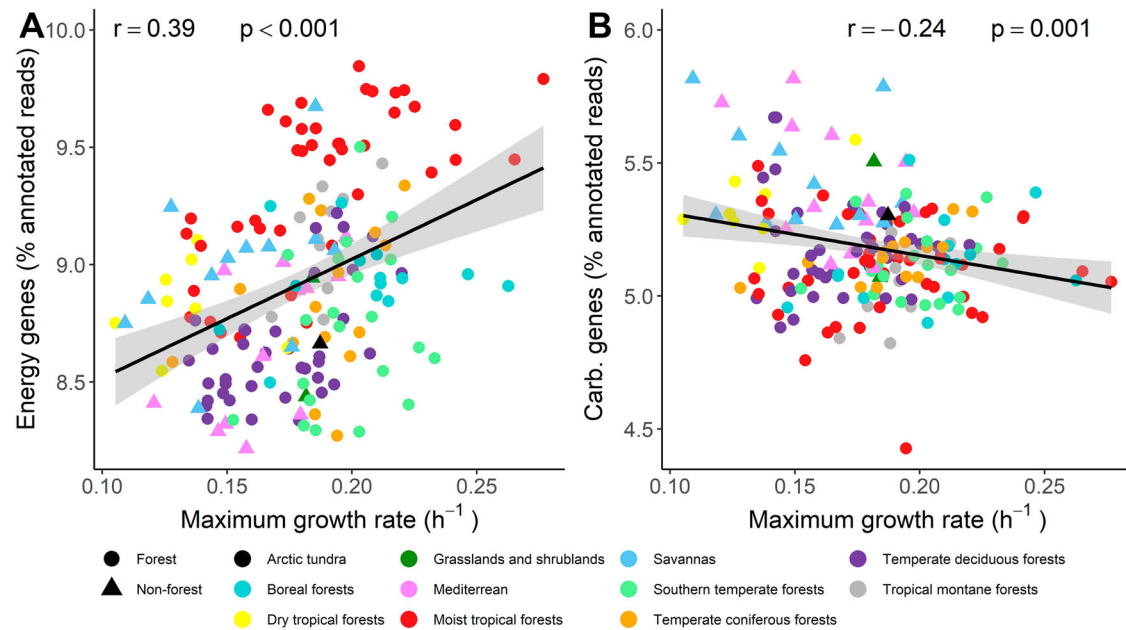


Fig. 3 | Pearson correlations between maximum growth rates of bacterial communities and bacterial C cycle functional gene relative abundances. Shown are correlations between growth rates and the relative abundances of energy production and conversion genes (eggNOG COG category 'C') (A) and between growth rates and the relative abundances of carbohydrate transport and metabolism genes (eggNOG category 'G') (B). The correlation analyses on both (A) and (B)

were two-tailed and $n = 176$. The shaded areas in (A) and (B) represent the 95% confidence interval around the regression lines. Functional gene relative abundances are expressed as the percentage of annotated reads assigned to the respective categories. Correlations between the two categories and 16S rRNA gene copy numbers are shown in Supplementary Fig. S10. Source data are available as a Source Data file and on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

resource acquisition potential is another fundamental aspect of the Y-A-S microbial life history framework (i.e., 'yield' vs. 'acquisition')²⁸. Again, our results confirm the existence of the tradeoff and identify ecosystem productivity as the key environmental gradient along which the tradeoff operates.

In addition, the positive association between bacterial growth potential and energy production/conversion genes we observed suggests that high growth potential is associated with the conversion of assimilated C resources to energy as opposed to biomass accrual. Another recent study in soils also associated copiotrophic bacterial communities with greater energy metabolism²⁴, and prior culture-based studies have similarly demonstrated a negative relationship between maximum growth rates and bacterial C use efficiency²⁹. It is possible that rapid bacterial growth and turnover will, in turn, promote soil organic matter formation via increased production of microbial residues⁵. Alternatively, the reduction in C use efficiency that likely accompanies increased growth rates could reduce rates of soil organic matter formation due to increased respiratory loss of C⁴⁰. Future research should attempt to disentangle these potential C cycle pathways in relation to bacterial growth potential. It should also be noted that some of the relationships between metagenomic traits we observed are at odds with prior work on marine bacteria, which showed faster-growing organisms to be enriched in carbohydrate metabolism genes and slower-growing communities to be enriched in energy metabolism genes¹⁷, the opposite of what we observed. The reason for the discrepancy is not clear, though it is possible that marine vs. soil environments impose different selective pressures on bacteria that lead to different associations between metagenomic traits. Given this discrepancy, it is not clear whether Y-A-S tradeoffs will be universally observed across all microbial habitats. Identifying the similarities and differences in microbial tradeoffs among different microbial habitats represents another important area of future research.

The results described above are corroborated and summarized by our structural equation model (SEM), which illustrates the links

among climate, soil properties, bacterial growth potential, and C cycle functions (Fig. 4). The SEM reflects our hypothesis that high bacterial growth potential is unlikely to be mechanistically linked to low soil pH, i.e., pH was not included in the a priori SEM. The SEM also reflects our hypothesis that productivity is an important global control over bacterial growth potential and thus includes NPP as a candidate variable. Though NPP was not among the strongest predictors of maximum growth rates in the random forest model (Fig. 2), it is reasonable to include NPP in the SEM given our hypotheses and given that NPP was significantly correlated with maximum growth rates (Supplementary Fig. 6) and was also the strongest predictor of 16S rRNA gene copy number (Supplementary Fig. 7). The SEM supports our hypothesis that bacterial growth potential is driven primarily by soil properties such as moisture and C availability, which, in turn, are controlled by productivity (i.e., NPP, distance to equator) (Fig. 4). Growth potential is then significantly associated with C cycle functions, with a negative effect of maximum growth rates on carbohydrate metabolism genes and a positive effect of maximum growth rates on energy production/conversion genes (Fig. 4). Because we cannot rule out the possibility that soil pH plays a direct role in driving growth potential, we ran an alternative SEM that includes pH, which produced very similar results (Supplementary Fig. 12). Overall, our results identify the environmental conditions that select for high bacterial growth potential and suggest that variation in bacterial growth potential among biomes will drive varying rates of decomposition and C storage in soils from different environments. These results complement a prior study using the same metagenomes, which described life history trait dimensions on the basis of functional gene relative abundances and genomic traits²⁶. That study found that arid biomes were associated with bacterial communities with small genomes and increased relative abundance of C acquisition genes²⁶. Other studies have similarly found that global patterns in the taxonomic composition of bacterial communities are strongly associated with ecosystem aridity vs. productivity²⁷. Our results demonstrate that those macroecological

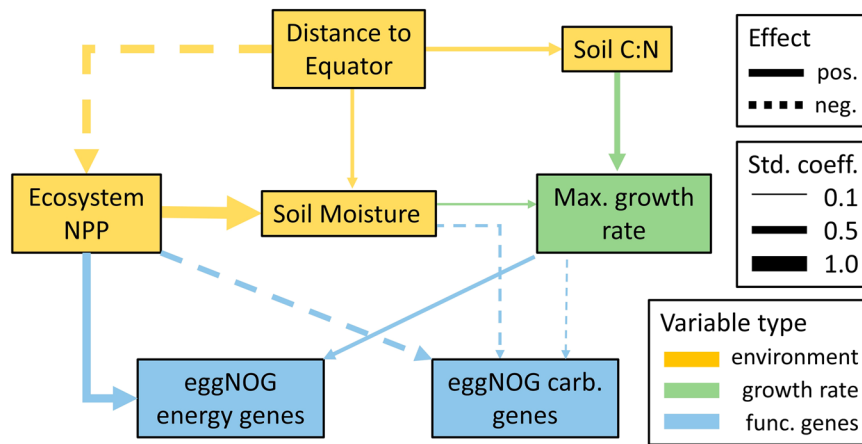


Fig. 4 | Structural equation model demonstrating links among climate variables, soil variables, bacterial community growth potential, and relative abundances of bacterial C cycle functional genes. The path diagram shows standardized path coefficients, and only statistically significant paths are shown. The SEM explains 43% of the variation in energy production/conversion genes, 21%

of the variation in carbohydrate transport/metabolism genes, and 21% of the variation in maximum growth rates. All fit indices indicated adequate fit of the model (CFI > 0.99, SRMR and RMSEA both < 0.05). For the structural equation model, $n = 176$ samples. Source data are available as a Source Data file and on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

patterns in the functional and taxonomic characteristics of bacterial communities are fundamentally linked to the growth potential of those communities.

Concluding remarks

Growth potential is a fundamental property of microbial communities that control rates of C cycling and storage in terrestrial ecosystems. We demonstrate that the growth potential of soil bacterial communities varies among terrestrial biomes, with high bacterial growth potential associated with highly productive forest ecosystems. Further, variations in growth metrics were associated with variations in microbial C cycle gene relative abundances, with an apparent tradeoff between growth potential and resource acquisition potential. Importantly, the metagenomic approach we used only estimates maximum growth rates and not actual in situ growth, which can be substantially slower than maximum rates¹¹. Thus, future studies should use empirical methods (e.g., isotope tracing) to validate the patterns we observed. Future studies should also validate the results observed here using more deeply sequenced metagenomes, which would allow for a more accurate estimation of growth potential in bacterial genomes assembled from the metagenomes¹⁷. Regardless, the macroecological patterns in bacterial growth potential we identify are a substantial contribution to bacterial life history theory, confirming the existence of hypothesized bacterial life history tradeoffs (i.e., growth vs. stress tolerance and growth vs. resource acquisition) as well as identifying the environmental gradients along which those tradeoffs manifest. These tradeoffs and their associated environmental drivers, in turn, determine the relative dominance of different soil C cycle pathways. Thus, our work establishes a microbial physiological framework with which to better understand the decomposition and storage of soil C at a global scale.

Methods

Sample collection

For this study, we analyzed a dataset of 176 metagenomes generated from soils from 11 terrestrial biomes across six continents (Supplementary Fig. 1). All soil samples were collected and shipped in accordance with local and international laws, as described previously⁴¹. The soil samples were analyzed for chemical properties, including pH, total N, organic C, total phosphorus (P), calcium (Ca), potassium (K), and magnesium (Mg) contents²¹. Temperature and precipitation data for each soil sampling location were obtained from the WorldClim

database, while estimates of potential evapotranspiration (PET) and net primary productivity (NPP) were obtained from the Atlas of the Biosphere²¹.

Metagenomic analyses

DNA was extracted from 2 g of each soil sample using the PowerMax soil DNA isolation kit (MoBio). DNA was prepped for shotgun metagenomic sequencing using the Illumina TruSeq Nano HT kit and sequenced on the Illumina HiSeq 2500 platform (2 × 250 bp)²¹. We downloaded the raw metagenomic sequence reads from NCBI accession number PRJEB18701. We quality-filtered the raw reads using trimmomatic (version 0.39)⁴² and assigned taxonomy to the quality-filtered reads using kraken2 (version 2.1.3)⁴³ with the Refseq genomes for bacteria, archaea, viruses, fungi, and protists as reference databases. To avoid biases from varying relative abundances of eukaryotic sequences among ecosystems (Supplementary Fig. 13), we conducted all downstream analyses using only the metagenomic reads identified as bacterial in origin.

We estimated the average minimum doubling times of the bacterial communities using gRodon2¹⁷, which works by quantifying codon usage bias in highly expressed ribosomal protein genes while also accounting for the relative coverages of genes in the community. The method has previously been shown to be most accurate when genes are predicted from assembled contigs¹⁸. Therefore, we assembled contigs using megahit (version 1.1.3)⁴⁴ and predicted coding regions in the contigs using prodigal (version 2.6.3)⁴⁵. We then identified ribosomal protein genes from the coding regions by using blastn (version 2.9.0)⁴⁶ with a previously constructed database of ribosomal protein gene sequences as the reference database¹⁶. Relative coverages of the genes were determined by aligning the sequence reads to the predicted coding regions using bwa mem (version 0.7.17)⁴⁷ and quantifying gene coverages using samtools (version 1.9)⁴⁸.

For the gRodon models, we used the 'metagenome_V2' (or 'MMv2') version of the software. MMv2 is an updated version of the original gRodon model (MMv1), which toggles between MMv1 and a GC bias-corrected version of the model (MMBC) depending on the consistency of codon usage in the ribosomal protein genes in a metagenome. Further information on the mechanics of the model can be found in the original gRodon studies^{17,18}. For our samples, consistency values were always below the toggle threshold, so the original MMv1 was applied. We also analyzed our samples with MMBC to ensure that the patterns we observed were not simply reflecting any

GC dependency in the model (mode = 'meta_testing', bg = 'individual'). MMBC also allowed us to rigorously assess relationships between growth rates and metagenome GC content. However, we focus on the results from the recommended MMv2 (really MMv1) results, as the tradeoff between GC bias in MMv1 vs. the higher variance in the MMBC model favors the MMv1 predictions for our samples¹⁸. Because the assemblies of these metagenomes were poor (mean contig N50 = 508 bp), we also repeated the above growth rate analysis using gene fragments predicted directly from the sequence reads using FragGeneScanRs⁴⁹. The predicted growth rates from all the gRodon model variations we ran were well correlated and exhibited the same global patterns (Supplementary Fig. 2, 3), though in the main text, we present the MMv2 model results using genes predicted from contigs, as recommended by the gRodon developers¹⁸. The model outputs community-averaged minimum doubling times (in hours), which we converted to maximum growth rates (i.e., cell doublings h⁻¹) by taking the reciprocal of the doubling times. For all growth rate analyses, we used the default minimum gene length of 240 bp and did not include the optional correction for optimal growth temperature. For additional validation of the ecological patterns in growth rates we observed, we repeated the growth rate analysis using metagenomes from 42 National Ecological Observatory (NEON) sites across the United States. Sample information for the NEON metagenomes is provided in our figshare repository: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

To estimate the average genome size and the effective number of genomes within each metagenome, we used MicrobeCensus (version 1.1.1)⁵⁰. To estimate the average 16S rRNA gene copy number in each community, we first curated a custom bacterial 16S database using RESCRIPt (version 2023.5.0)⁵¹. Our custom database is based on the SILVA SSU database (version 138.1)⁵² but with all sequences belonging to eukaryotes, archaea, chloroplasts, mitochondria, or unclassified organisms removed. We also removed any sequences with 'n' bases and sequences that were shorter than 1200 bp. We then dereplicated the sequences at a 99% identity threshold, resulting in a 16S database containing 323,218 high-quality sequences. We then identified bacterial 16S genes within the metagenomes by aligning the sequence reads to the curated 16S database using bwa mem⁴⁷. This approach is similar to that of riboPicker⁵³ but with an updated 16S database and an updated alignment algorithm. We then quantified the number of aligned base pairs in each metagenome using samtools⁴⁸ and quantified the effective number of 16S regions in each metagenome by dividing the number of aligned base pairs by 1542 (the length of 16S in *E. coli*). The average 16S gene copy number for each metagenome was then calculated as the number of 16S regions divided by the effective number of genomes present⁵⁴. We assigned functional gene categories to the sequence reads using eggNOG-mapper (version 2.1.3, 'DIAMOND' mode) with a reference database of bacterial eggNOG COGs⁵⁵. Functional gene relative abundances were calculated as the percentage of annotated reads (as opposed to the percentage of total reads) to account for biases resulting from systematic variation in annotation success among samples and environments²⁶. Full details on all metagenomic analyses, including bioinformatics scripts, are provided in the figshare repository: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

Statistical analyses

All statistical analyses were conducted in ref. 56. We analyzed latitudinal patterns in growth rates using polynomial regression, with the best-supported regression model selected using AICc. We assessed the effects of forest cover on growth rates using linear models or generalized linear models (Gamma distribution, log-link function) when linear models did not meet assumptions of normality of residuals. Polynomial regression and linear models were analyzed using *F* statistics (or χ^2 statistics in the case of generalized linear

models). We determined relationships among variables using Pearson correlation. To identify environmental drivers of growth rates, we used random forest regression in the randomForest R package⁵⁷. We quantified the 'importance' of the environmental variables in accounting for variation in the growth rates by determining the increase in model error (mean squared error, 'MSE') when each respective variable was randomly shuffled across the dataset. We identified putative mechanistic links between environmental variables, bacterial growth metrics, and functional genes using structural equation modeling (SEM) in the lavaan R package⁵⁸. We assessed SEM fit by bootstrapping estimates of the comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA). All of these metrics indicated adequate fit for our model: CFI was >0.99 while SRMR and RMSEA were both <0.05.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequence data are available at NCBI accession number PRJEB18701. The data used for this study are derived from previously collected soil samples⁴¹, and the soil and environmental data were generated as previously described²¹. All soil and environmental data are available on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>. Source data are also available with this paper. Source data are provided with this paper.

Code availability

Statistical analysis scripts and bioinformatics scripts are available on figshare: <https://doi.org/10.6084/m9.figshare.24498484.v3>.

References

1. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).
2. Hall, E. K. et al. Understanding how microbiomes influence the systems they inhabit. *Nat. Microbiol.* **3**, 977–982 (2018).
3. Malik, A. A., Puissant, J., Goodall, T., Allison, S. D. & Griffiths, R. I. Soil microbial communities with greater investment in resource acquisition have lower growth yield. *Soil Biol. Biochem.* **132**, 36–39 (2019).
4. Geyer, K. M., Dijkstra, P., Sinsabaugh, R. & Frey, S. D. Clarifying the interpretation of carbon use efficiency in soil through methods comparison. *Soil Biol. Biochem.* **128**, 79–88 (2019).
5. Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Deneff, K. & Paul, E. The microbial efficiency-matrix stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Change Biol.* **19**, 988–995 (2013).
6. Manzoni, S., Taylor, P., Richter, A., Porporato, A. & Ågren, G. I. Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. *N. Phytol.* **196**, 79–91 (2012).
7. Hungate, B. A. et al. Quantitative microbial ecology through stable isotope probing. *Appl. Environ. Microbiol.* **81**, 7570–7581 (2015).
8. Rousk, J. & Bååth, E. Growth of saprotrophic fungi and bacteria in soil. *FEMS Microbiol. Ecol.* **78**, 17–30 (2011).
9. Morrissey, E. M. et al. Evolutionary history constrains microbial traits across environmental variation. *Nat. Ecol. Evol.* **3**, 1064–1069 (2019).
10. Spohn, M., Klaus, K., Wanek, W. & Richter, A. Microbial carbon use efficiency and biomass turnover times depending on soil depth – Implications for carbon cycling. *Soil Biol. Biochem.* **96**, 74–81 (2016).

11. Caro, T. A., McFarlin, J., Jech, S., Fierer, N. & Kopf, S. Hydrogen stable isotope probing of lipids demonstrates slow rates of microbial growth in soil. *Proc. Natl. Acad. Sci. USA* **120**, e2211625120 (2023).
12. Malik, A. A. et al. Land use driven change in soil pH affects microbial carbon cycling processes. *Nat. Commun.* **9**, 3591 (2018).
13. Rousk, J., Smith, A. R. & Jones, D. L. Investigating the long-term legacy of drought and warming on the soil microbial community across five European shrubland ecosystems. *Glob. Change Biol.* **19**, 3872–3884 (2013).
14. Hagerty, S. B. et al. Accelerated microbial turnover but constant growth efficiency with warming in soil. *Nat. Clim. Change* **4**, 903–906 (2014).
15. Spohn, M. et al. Soil microbial carbon use efficiency and biomass turnover in a long-term fertilization experiment in a temperate grassland. *Soil Biol. Biochem.* **97**, 168–175 (2016).
16. Vieira-Silva, S. & Rocha, E. P. C. The systemic imprint of growth and its uses in ecological (Meta)genomics. *PLOS Genet.* **6**, e1000808 (2010).
17. Weissman, J. L., Hou, S. & Fuhrman, J. A. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci. USA* **118**, e2016810118 (2021).
18. Weissman, J. L., Peras, M., Barnum, T. P. & Fuhrman, J. A. Benchmarking community-wide estimates of growth potential from metagenomes using codon usage statistics. *mSystems* **7**, e00745–22 (2022).
19. Fierer, N., Barberán, A. & Laughlin, D. C. Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities. *Front. Microbiol.* **5**, <https://doi.org/10.3389/fmicb.2014.00614> (2014).
20. Barberán, A., Fernández-Guerra, A., Bohannan, B. J. M. & Casamayor, E. O. Exploration of community traits as ecological markers in microbial metagenomes. *Mol. Ecol.* **21**, 1909–1917 (2012).
21. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
22. Long, A. M., Hou, S., Ignacio-Espinoza, J. C. & Fuhrman, J. A. Benchmarking microbial growth rate predictions from metagenomes. *ISME J.* **15**, 183–195 (2021).
23. Chen, Y., Neilson, J. W., Kushwaha, P., Maier, R. M. & Barberán, A. Life-history strategies of soil microbial communities in an arid ecosystem. *ISME J.* **15**, 649–657 (2021).
24. Wang, C. et al. Bacterial genome size and gene functional diversity negatively correlate with taxonomic diversity along a pH gradient. *Nat. Commun.* **14**, 7437 (2023).
25. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **103**, 626–631 (2006).
26. Piton, G. et al. Life history strategies of soil bacterial communities across global terrestrial biomes. *Nat. Microbiol.* **8**, 2093–2102 (2023).
27. Delgado-Baquerizo, M. et al. A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
28. Malik, A. A. et al. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J.* **14**, 1–9 (2020).
29. Roller, B. R. K., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 1–7 (2016).
30. Westoby, M. et al. Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecol. Evol.* **11**, 3956–3976 (2021).
31. Musto, H. et al. Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Lett.* **573**, 73–77 (2004).
32. Grime, J. P. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am. Nat.* **111**, 1169–1194 (1977).
33. Ma, B. et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. *Nat. Commun.* **14**, 7318 (2023).
34. Bradford, M. A., Keiser, A. D., Davies, C. A., Mersmann, C. A. & Strickland, M. S. Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. *Biogeochemistry* **113**, 271–281 (2013).
35. Prommer, J. et al. Increased microbial growth, biomass, and turnover drive soil organic carbon accumulation at higher plant diversity. *Glob. Change Biol.* **26**, 669–681 (2020).
36. Romero, F. et al. Soil health is linked to primary productivity across Europe. Preprint at <https://doi.org/10.1101/2023.10.29.564603> (2023).
37. Huston, M. A. & Wolverton, S. The global distribution of net primary production: resolving the paradox. *Ecol. Monogr.* **79**, 343–377 (2009).
38. Šimová, I. & Storch, D. The enigma of terrestrial primary productivity: measurements, models, scales and the diversity–productivity relationship. *Ecography* **40**, 239–252 (2017).
39. Gillman, L. N. et al. Latitude, productivity and species richness. *Glob. Ecol. Biogeogr.* **24**, 107–117 (2015).
40. Tao, F. et al. Microbial carbon use efficiency promotes global soil carbon storage. *Nature* **618**, 981–985 (2023).
41. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
42. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
43. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
44. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
45. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
46. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
48. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
49. Van der Jeugt, F., Dawyndt, P. & Mesuere, B. FragGeneScanRs: faster gene prediction for short reads. *BMC Bioinform.* **23**, 198 (2022).
50. Nayfach, S. & Pollard, K. S. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**, 51 (2015).
51. li, M. S. R. et al. RESCRIPt: Reproducible sequence taxonomy reference database management. *PLOS Comput. Biol.* **17**, e1009581 (2021).
52. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

53. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).
54. Pereira-Flores, E., Glöckner, F. O. & Fernandez-Guerra, A. Fast and accurate average genome size and 16S rRNA gene average copy number computation in metagenomic data. *BMC Bioinform.* **20**, 453 (2019).
55. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
56. R. Core Development Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2019).
57. Cutler, F. original by L. B. and A. & Wiener, R. port by A. L. and M. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. (2022).
58. Rosseel, Y. et al. lavaan: Latent Variable Analysis. (2019).

Acknowledgements

This work was supported by NSF award #1845417, awarded to M.S.S., USDA NIFA award #2020-67034-41310 to S.G.M., and the Swedish Research Council (VR; Grant 2021-03724) to M.B. This manuscript uses data products generated by the National Ecological Observatory Network (NEON), a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. We also thank J. L. Weissman for fielding questions regarding the gRodon model results.

Author contributions

E.D.O. and S.G.M. conceived the study. E.D.O. conducted all bioinformatics and statistical analyses. M.B. consulted on data acquisition and handling. M.S.S. provided project oversight. All authors contributed to writing and revising the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50382-1>.

Correspondence and requests for materials should be addressed to Ernest D. Osburn.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024