Y-A+S is the new Y-A-S: Updating microbial life history tradeoffs with comparative genomics

Luciana Chavez Rodriguez

luciana.chavezrodriguez@wur.nl

Wageningen University & Research https://orcid.org/0000-0003-1510-6695

José M Murúa

Department of Ecology and Evolutionary Biology, University of California, Irvine

Gianna Marschmann

Lawrence Berkeley National Laboratory https://orcid.org/0000-0002-9065-2023

Katherine L. Shek

University of New Hampshire

Ulas Karaoz

Lawrence Berkeley National Laboratory https://orcid.org/0000-0002-8238-6757

Ashish Malik

University of Edinburgh

Caitlin M Broderick

W.K. Kellogg Biological Station, Michigan State University

Dawson Fairbanks

Funga PBC

Winston E. Anthony

Pacific Northwest National Laboratory

Alicia Clum

US DOE Joint Genome Institute

Amelia R. Nelson Kuhn

Department of Soil and Crop Sciences, Colorado State University

Eoin Brodie

Lawrence Berkeley National Laboratory https://orcid.org/0000-0002-8453-8435

Brittni L. Bertolet

Rutgers University Newark, Department of Earth and Environmental Sciences

Steven Allison

University of California, Irvine https://orcid.org/0000-0003-4629-7842

Article

Keywords: Trait-based ecology, functional diversity, genome size, metagenome-assembled genomes, soil microbial community

Posted Date: May 2nd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6486618/v1

Additional Declarations: There is **NO** Competing Interest.

Y-A+S is the new Y-A-S: Updating microbial life history tradeoffs with comparative genomics

Luciana Chavez Rodriguez [a,b,*], José M. Murúa [b], Gianna L. Marschmann [c], Katherine L. Shek [d], Ulas Karaoz [c], Ashish A. Malik [e], Caitlin M Broderick [f], Dawson Fairbanks [g], Winston E. Anthony [h], Alicia Clum [i], Amelia Nelson Kuhn [j,k], Eoin Brodie [c,l], Brittni L. Bertolet [b,m], Steven D Allison [b,n]

[a] Soil Biology Group, Wageningen University & Research, Wageningen, Gelderland, The Netherlands
[b] Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA
[c] Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[d] University of New Hampshire, Durham, NH, USA
[e] School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FF, United Kingdom
[f] W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI, USA
[g] Funga PBC, Austin, USA
[h] Pacific Northwest National Laboratory, USA
[i] Lawrence Berkeley National Laboratory
[j] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA
[k] Innovative Genomics Institute, University of California, Berkeley
[l] Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA, USA
[m] Rutgers University Newark, Department of Earth and Environmental Sciences, Newark, New Jersey, USA
[n] Department of Earth System Science, University of California Irvine, Irvine, California, USA

* Corresponding author: Luciana Chavez Rodriguez (luciana.chavezrodriguez@wur.nl)

Abstract (150/150)

Trait-based theory, coupled with mechanistic trait-based models, can be used to explore how changes in environmental conditions affect microbial community assembly and the impact on soil biogeochemistry. We used 27214 metagenome-assembled genomes (MAGs) and 5301 soil isolate genomes to investigate tradeoffs between life history strategies encoded in the genome. We used microTrait, a genome-based characterization toolset, to derive trait composition according to the YAS framework, which classifies microorganisms into high-yield (Y), resource-acquisition (A), and stress-tolerance (S) strategists. In contrast to the hypothesized three-way tradeoff, we observed a positive correlation between genome size and functional S and A traits and a negative correlation between genome size and the Y strategy measured as carbon use efficiency (CUE). Microbes with larger genomes favor investment in A and S at the cost of Y. Therefore, genome size, as a master trait, can enhance trait-based models by improving predictions of environmental impacts on soil microorganisms.

Keywords:

Trait-based ecology, functional diversity, genome size, metagenome-assembled genomes, soil microbial community

Introduction (3,555/3,500)

Life history strategies of microorganisms are encoded in their genomes and provide crucial insights into their ecological roles and interactions within ecosystems. This information can be leveraged to constrain trait representation in models to explore their interactions across biological scales[1]. The YAS framework [1] classifies microorganisms and communities into high-yield (Y), resource-acquisition (A), and stress-tolerance (S) strategists, borrowing concepts from Grime's CSR (competitor, stress tolerator, ruderal) framework for plants [2]. Traits associated with the resource-acquisition strategy include the investment in exoenzymes, targeting different nutrient sources, as well as the investment in membrane transporters for the uptake of nutrients [3]. Stress-tolerance traits include physiological and evolutionary mechanisms to respond to a variety of environmental stressors, such as chaperones and proteases against cold, heat, and pH shock stress or osmolyte production against drought [3]. Finally, the high-yield strategy is defined as carbon use efficiency (CUE), as the amount of biomass produced per unit of resource consumed [1]. The main premise is that microorganisms invest in these strategies based on their contemporary environment and evolutionary history with higher investment in any one strategy resulting in tradeoffs with the others.

Microbial trait tradeoffs that align with life history strategies defined by the YAS framework have been previously observed in different contexts. Piton et al. [4], using shotgun metagenomic DNA sequencing, observed trait tradeoffs between responsiveness to changes and the capacity to cycle nutrients. Malik et al. [5] demonstrated that some bacterial metagenome-assembled genomes (MAGs) exhibit trait tradeoffs between specific genes representing the A and S strategies. However, Treseder [6] suggested that a framework like YAS might not be sufficient to integrate all possible microbial ecological strategies. Guseva et al.[7] partially verified this idea by identifying tradeoffs within individual resource-acquisition traits, which are not accounted for in the YAS framework. Broderick et al. [8] did not find tradeoffs between S and A strategies in

metagenomes from across a savannah grassland precipitation gradient. These arguments highlight the critical need to reassess and validate microbial life history strategies to ensure trait theory provides a more accurate and predictive framework for understanding microbial processes [9,10].

The increasing availability of soil MAGs across different environments opens the opportunity to reanalyze microbial life history strategies in soils. Additionally, because microbial genomes (i.e., MAGs and soil isolates) are shaped by evolutionary forces [11], they should display tradeoffs in gene frequencies reflecting different ecological strategies, which could be useful in defining microbial life history strategies. Advances in bioinformatics tools, such as microTrait [3], now facilitate gene annotation and the estimation of functional traits within MAGs, improving our understanding of microbial life history strategies. Estimating microbial traits from MAGs provides a more scalable approach for the functional characterization of microbes because it includes the unculturable portion of the soil microbial community. Moreover, the recent integration of microTrait with the physiological model DEBmicroTrait[12] allows the estimation of kinetic parameters, like CUE, that cannot be derived from genomic information only.

In this work, we used comparative genomics on soil genomes across different soil environments to assess microbial life history strategies. Because trait tradeoffs and their strengths, as well as the implications for life history strategies, will likely be dependent on the set of analyzed genomes, our dataset included 27214 publicly available MAGs and 5301 soil isolates from a diverse range of environments (Tables S1 and S13) to reduce bias. We analyzed MAGs and used the soil isolate data to ground truth our predictions, acknowledging that MAGs vary in quality and their results might not apply to actual microorganisms [13]. Using microTrait and DEBmicroTrait, we functionally classified MAGs and isolates and tested genome-scale tradeoffs predicted by the YAS framework, relating genome size and functional traits. Before testing those relationships, we aggregated MAGs and isolates into functional groups (FGs) based on their traits, as FGs are typically represented in trait-based mechanistic models. The derived FGs are clusters of genomes that are similar in the high-dimensional trait space and will hypothetically represent a guild. Because estimations of certain traits like minimum generation time (mgt) for slow-growing microbes (mgt > 5 hours) via gRodon is less precise due to lower cellular signal of

growth optimization (via high codon usage bias - CUB) [14], we separated FGs into putative slow (mgt > 5 hr) and fast (mgt < 5 hr) growers.

Following the YAS framework, we expected a three-way tradeoff among functional traits characterized by negative relationships between gene-derived A and S traits and between Y and both A and S traits. Additionally, because genomic traits, such as genome size (DNA content of the entire genome) and genomic DNA base composition of guanine-cytosine (GC) content, are often associated with adaptation strategies in the environment [15,16], we analyzed the role of genome-wide traits in relation to the YAS framework. Here, we focused on genome size, as it is related to metabolic capabilities [17–20], energetic capabilities [21,22], including CUE, as well as cellular and ecological microbial strategies [23]. Finally, we discuss the potential implications of the correlations between genome size and other functional traits for trait-based modeling.

Results

Using an iterative pipeline from genes to FGs (see Methods), we derived 1364 FGs (Tables S18, S19) from 27214 MAGs and 5301 soil isolates annotated using microTrait [3]. We chose genes representing A strategy traits, such as genes encoding enzymes for substrate degradation and transporters, and genes encoding S traits, such as for osmolyte production, biofilm formation, and heat and pH tolerance (Table S2-S8). Due to the higher quality and completion of soil isolate genomes, we were able to derive CUE using the DEBmicrotrait model [12] for the Y traits. We also derived quantitative traits like minimum generation time (mgt) and optimal growth temperature (ogt) for both MAGs and isolates from microTrait. Although MAGs varied in quality (see Figure S5), we did not assess the role of quality on the patterns but instead used isolate genomes to validate MAG findings.

Lack of genome-level tradeoffs under the YAS framework

We did not find evidence for YAS tradeoffs (negative correlations) within bacterial MAGs (Figure 1A and Figure S2). Rather, we found positive trait correlations and did not observe trait tradeoffs between or within traits representing the A and S strategies. Stress-tolerance genes like osmolyte production (Table S5) and biofilm formation (Table S6) were positively and strongly correlated. However, consistent tradeoffs were observed between A and S traits and mgt. Stronger negative correlations were observed between mgt and A traits, specifically transporter counts ($r = -0.5$, $p < 0.001$), transporter of amino acids ($r = -0.6$, $p < 0.001$), and genes encoding protein degradation enzymes ($r = -0.5$, $p < 0.001$). Carbohydrate-active enzymes (CAZymes) showed, in general, weaker positive correlations with other A and S traits, and negative correlations with ogt and mgt, similar to the genes encoding biofilm formation. Stronger correlations were found in fast-growing bacterial MAGs (Figure 1A and Table S9) compared to slow-growing bacteria (Figure S2 and Table S10), despite maintaining the same patterns across all traits.

To validate the patterns found with MAGs, we estimated microbial CUE in soil isolates using the DEBmicrotrait model [12]. Similar patterns, although weaker, were found for soil isolates between A-S traits and genome-derived CUE for both fast- (Figure 1B and Table S11) and slow-growing bacteria (Figure S3 and Table S12). It is possible that the patterns found in soil isolates appear weaker because isolates might be less representative of the soil microbiome as they come from the culturable minority. Consistent tradeoffs (negative correlations) were found between A and S traits and CUE, suggesting a potential new framework where SA strategies trade off against Y traits.

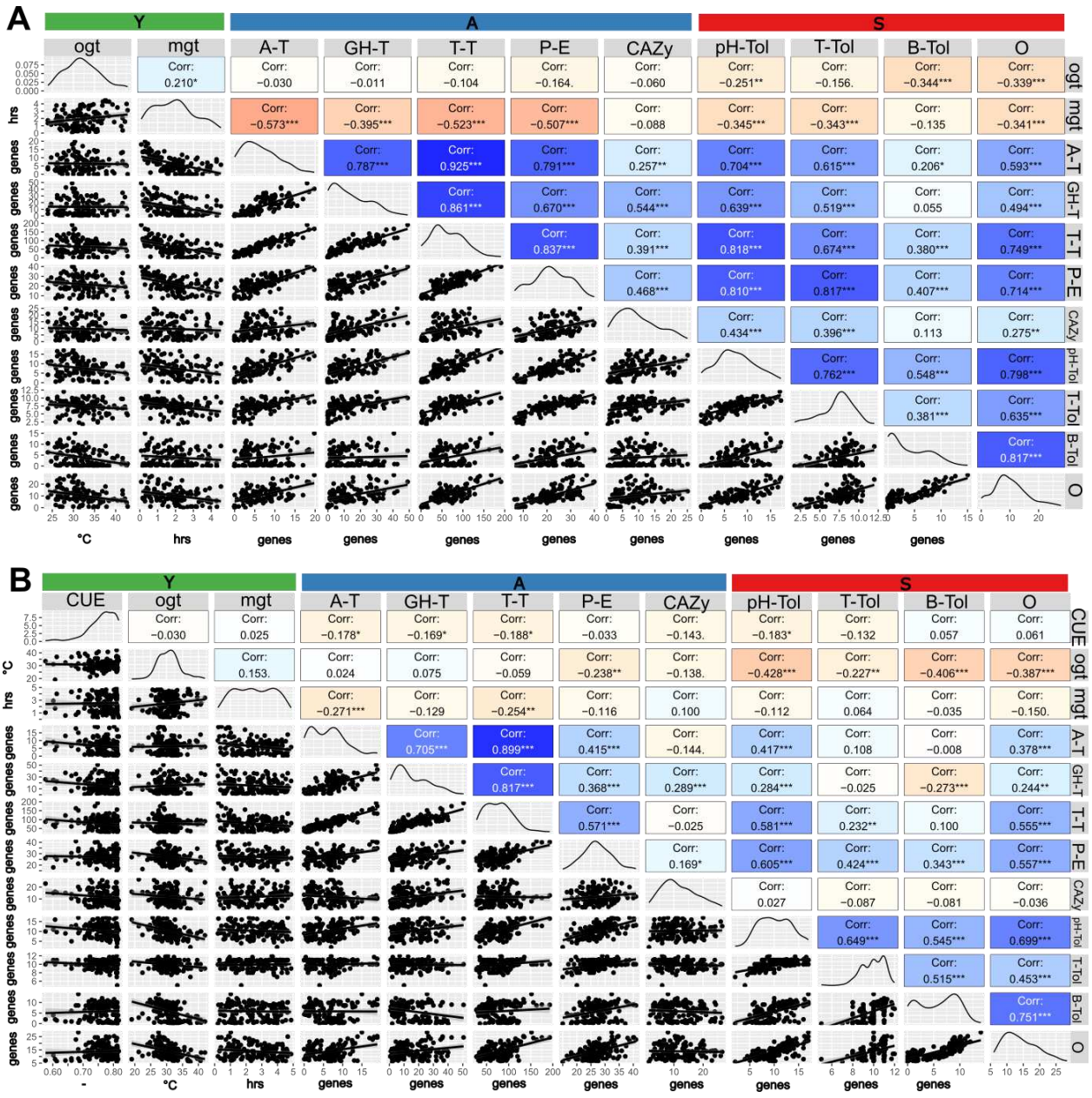

Figure 1. Trait tradeoffs among traits representative of the YAS framework using soil A) MAGs grouped by FGs and B) soil isolates grouped by FGs. We used the minimum generation time (mgt) (considering mgt above 5 hours as slow-growing bacteria), optimal growth temperature (ogt), and estimated yield (CUE) to represent the Y traits. Because of this distinction by mgt and some missing values (NAs) due to the quality of the MAGs and soil isolates, we only plot 148 FGs from MAGs (panel A), and 124 FGs from soil isolates (panel B). Abbreviations: A-T: amino transporters; GH-T: GH transporters; T-T: Total transporters; P-E: Protein enzymes; pH-Tol: pH tolerance; B-Tol: Biofilm; T-Tol: Temperature tolerance; O: Osmolyte.

Relationships between genome size and functional traits

We aimed to derive empirical relationships between genome size and YAS functional traits for potential use with biogeochemical soil models. We only tested linear relationships because the YAS conceptual framework assumes linear tradeoffs between traits. However, in some cases, other empirical relationships might exist (see Discussion).

S-A traits

We observed a significant positive linear relationship between the mean genome size per FG and functional traits related to substrate degradation (genes for CAZymes and protein degradation enzymes; Tables S2 and S3) (Figure 2; $p < 0.001$) across both MAGs and soil isolates. FGs derived from MAGs exhibited stronger linear relationships for fast-growing bacteria, especially for protein degradation genes ($R^2 = 0.60$). These results were opposite (stronger linear relationships for slow-growing bacteria) for the FG of soil isolates, which could be related to a low sample size compared to MAGs or the bias in bacterial species that can be cultured and isolated from soils.

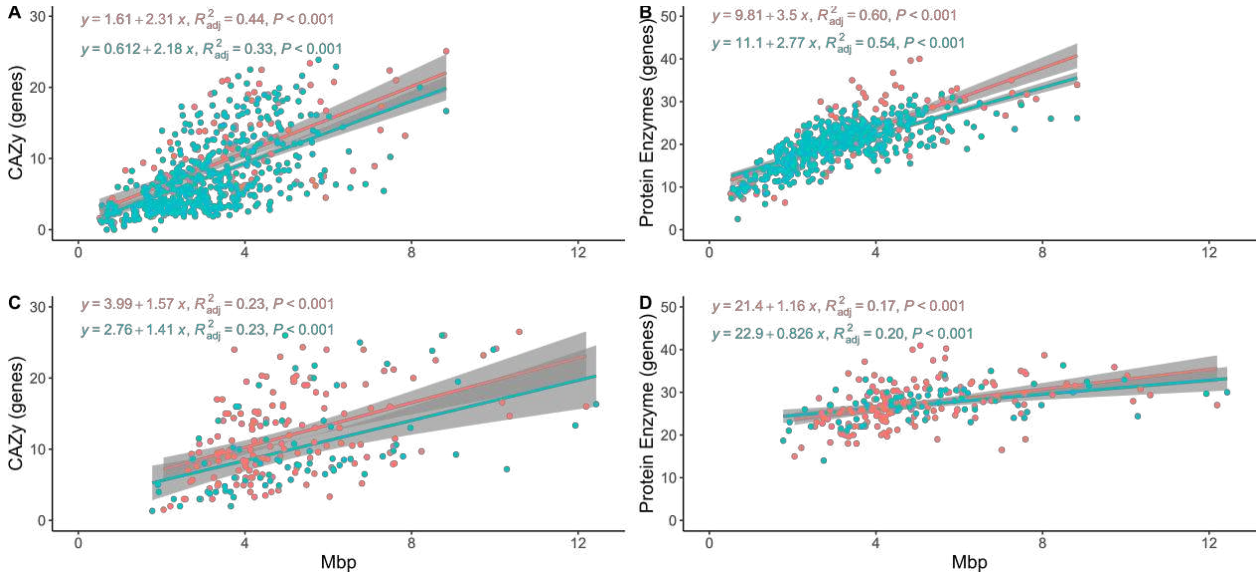

Figure 2. Linear Relationships between mean genome size of functional groups in megabase pairs (Mbp) and total number of unique genes for substrate degradation. Panels A and B correspond to MAG results, and C and D correspond to soil isolates. Red regressions correspond to fast-growing bacteria, and blue regressions correspond to slow-growing bacteria.

Similar to substrate degradation traits, we observed strong positive linear relationships between genome size and total genes encoding transporters (Figure 3), especially for total transporters

and carbohydrate transporters within fast-growing bacteria. Similar trends were found for soil isolates, although the relationships were weaker.

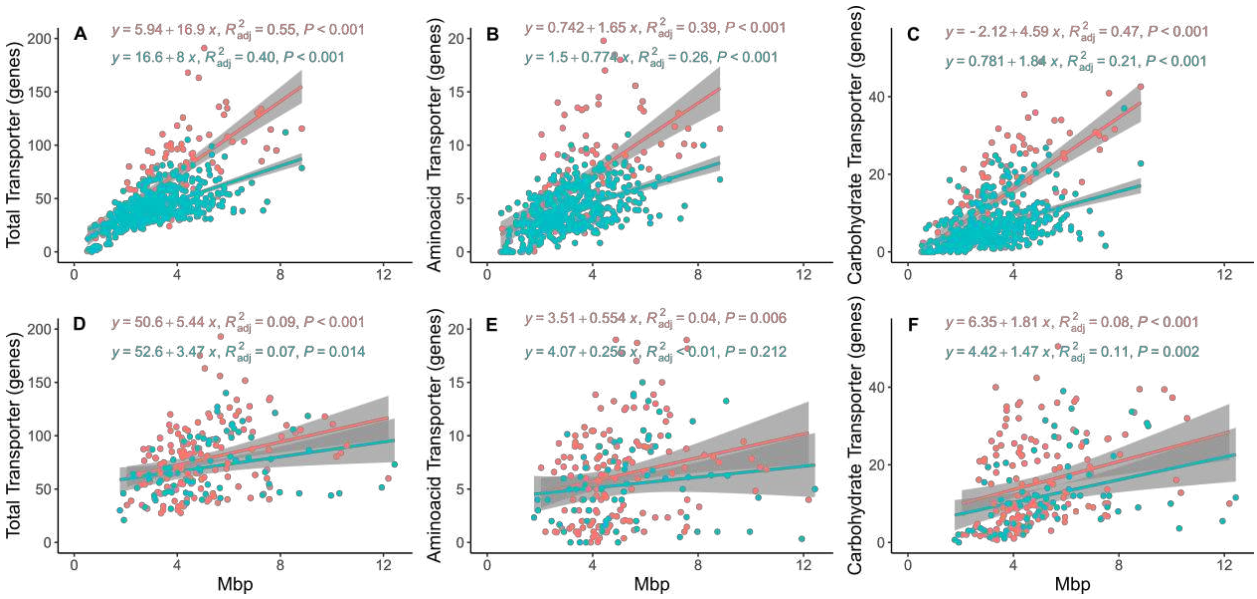

Figure 3. Linear relationships between mean genome size of functional groups in megabase pairs (Mbp) and potential for monomer transport (unique gene counts). Panels A to C correspond to MAG results, and D to F correspond to soil isolates. Panels A and D show genes encoding transporters for amino acids, carbohydrates, vitamins, siderophores, P-compounds, osmolytes, N-compounds, lipids, and ions. Red regressions correspond to fast-growing bacteria, and blue regressions correspond to slow-growing bacteria.

Positive linear relationships were not particularly strong between genome size and S traits (Figure 4), especially for biofilm formation genes. Additionally, MAGs consistently showed stronger positive linear relationships compared to soil isolates, which could indicate that MAGs better capture the gradual adaptation of bacteria to environmental stressors. These findings suggest that genome size might not be a good predictor for some of the stress tolerance traits, such as biofilm formation.

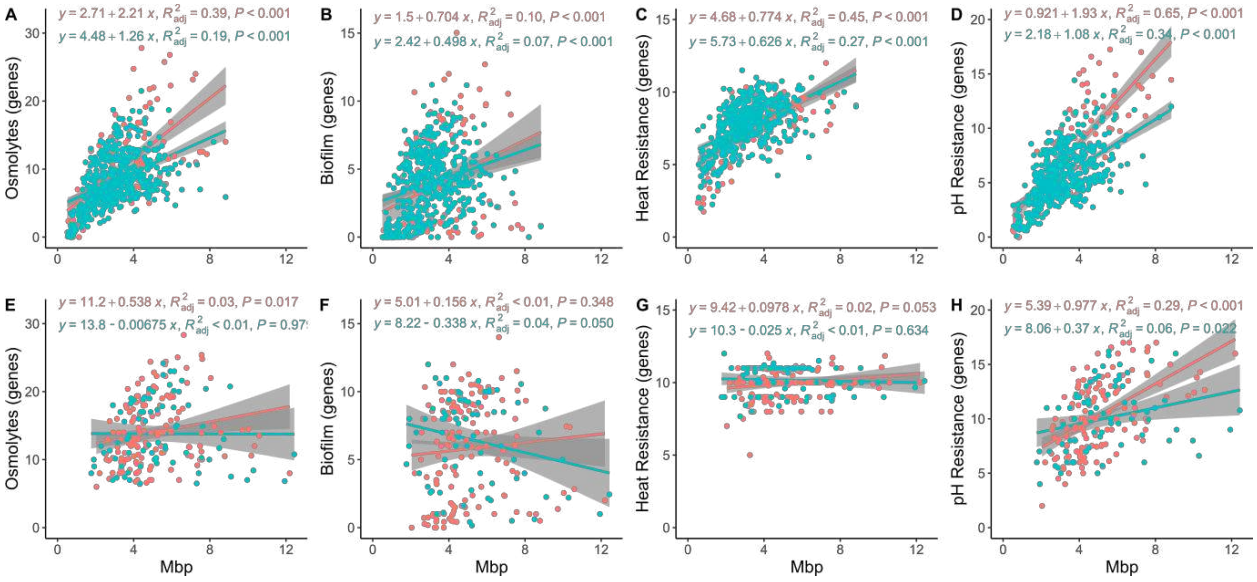

Figure 4. Linear relationships between mean genome size of functional groups in megabase pairs (Mbp) and potential for stress tolerance (unique gene counts). Panels A to D correspond to MAG results, and E to H correspond to soil isolates. Red regressions correspond to fast-growing bacteria, and blue regressions correspond to slow-growing bacteria.

Y traits

We observed a significant but weak negative linear relationship between genome size and yield (Figure 5C) only for fast-growing FGs derived from soil isolates ($p = 0.041$ and $R^2 = 0.02$). Similar patterns were also observed when soil isolates were grouped by different taxonomic groups (Figure S4), especially at the phylum, family, and genus levels.

We observed inconclusive results between genome size and mgt. mgt is not a yield trait but can correlate with CUE [12]. For both MAGs and soil isolates, our results suggest a weak negative linear relationship between genome size and mgt for fast-growing bacteria (Figure 5A), where genomic predictors (codon usage bias) are better indicators of growth rate [14]. In contrast, although not statistically significant, slow-growing bacteria tend to grow slower with increased genome size for both MAGs and isolates (Figure 5A and 5B).

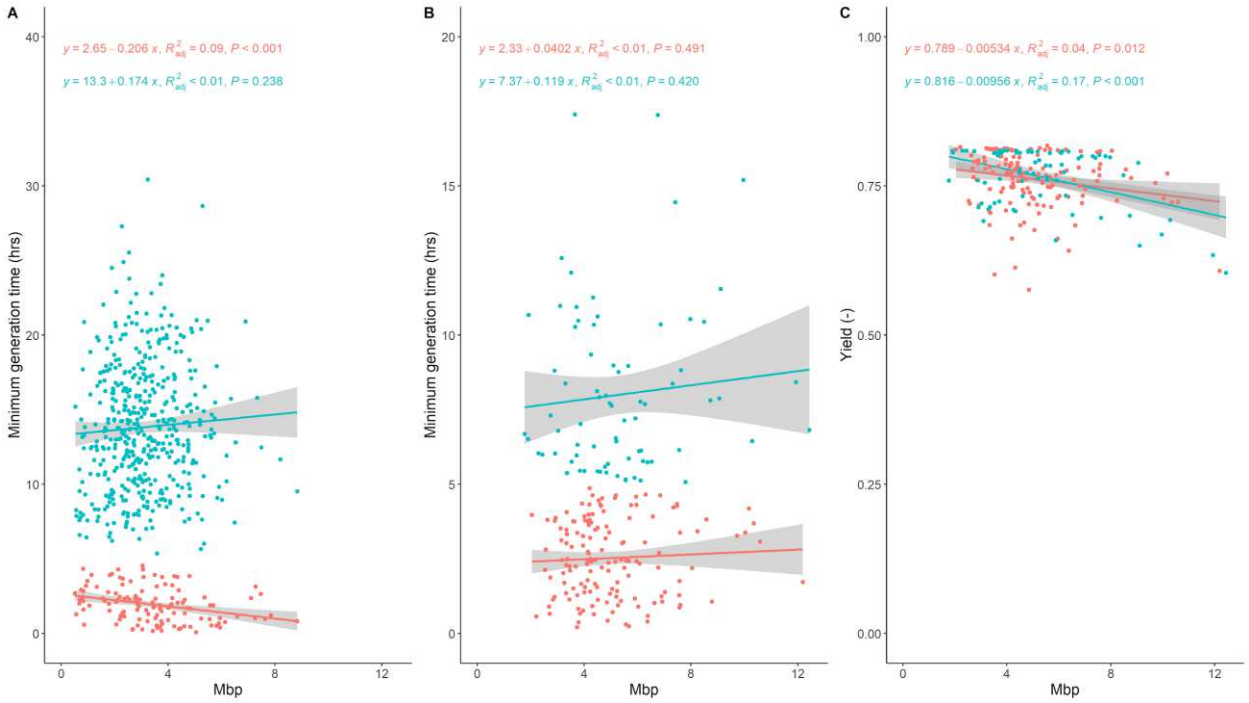

Figure 5. Linear relationships between mean genome size of functional groups in megabase pairs (Mbp) and minimum generation time (mgt) and yield. Panel A corresponds to MAG results, and B and C correspond to soil isolates. Red regressions correspond to fast-growing bacteria, and blue regressions correspond to slow-growing bacteria. Yield estimates (CUE) could not be made for MAGs.

Life history traits

Optimal growth temperature is an essential trait that modulates microbial growth and other physiological processes [3,24]. We observed that optimal growth temperature was negatively correlated with genome size independently of whether the FGs were from MAGs or isolates or whether they are from fast or slow-growing bacteria (Figure 6 A and B). Our results suggest that smaller genomes can cope better with higher temperatures up to 60°C [25–27].

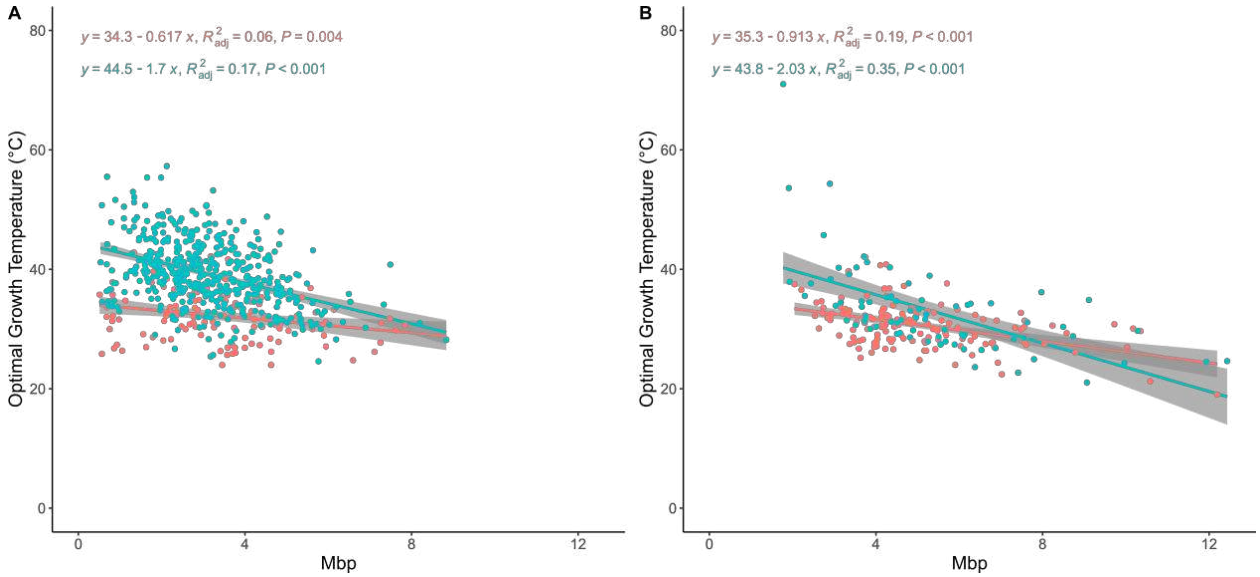

Figure 6. Linear relationships between mean genome size of functional groups in megabase pairs (Mbp) and optimal growth temperature. Panel A corresponds to MAG results and B to soil isolates. Red regressions correspond to fast-growing bacteria, and blue regressions correspond to slow-growing bacteria.

Discussion

Redefining the YAS framework at the genome scale

Microbial life history strategies are crucial to understanding microbial mechanisms behind important biogeochemical processes in soils. In this regard, the YAS framework has provided an alternative to characterize soil microorganisms, encompassing key evolutionary and physiological traits[1]. However, in our study, we do not find evidence of tradeoffs between the proposed dominant microbial life history strategies within the YAS framework, using a broad dataset of 27214 soil MAGs and validated with 5377 soil isolates and a comprehensive set of genes. We observe positive correlations between the A (resource acquisition) and S (stress tolerance) strategies, indicating no tradeoff between them. There are tradeoffs between A and Y strategies and between S and Y. These results are consistent among FGs of both MAGs and isolates, with slight differences in the strength of the relationships, especially for stress-related traits. We decided to do a separate analysis for both MAGs and isolates, as MAGs vary in quality and completeness; therefore, an interpretation based on MAGs might not be transferable to the whole microbial community[13]. However, MAGs represent a bigger portion of the microbiome as they include non-culturable microorganisms. Through a broader representation

of the soil microbial community and aggregating them into functional groups, we might be able to reduce the bias and potential error compared to when using MAGs only for taxonomic characterization of microbial taxa.

Our results call for a new interpretation of the widely used YAS framework and suggest a new Y-A+S framework, where A and S strategies do not trade off but correlate positively, especially in bacteria with bigger genomes. Larger genomes have a higher metabolic potential to perform and adapt to many environmental conditions but at higher metabolic costs due to the maintenance of larger genomes [22]. This tradeoff is observed via a negative correlation between AS traits and yield. The fact that genome size is negatively related to yield supports the argument that larger genomes (with more traits) have higher metabolic costs. Thus, the more genes a genome accumulates, the higher the metabolic cost to maintain them, ultimately reducing yield [4]. The Y-A+S framework would imply that large genome microbes (generalists) could be more prevalent in environments that are more variable, stressful, and/or rich in complex resources [28,29]. In contrast, smaller genome microbes (specialists) might be favored in more stable, low-resource environments with primarily simple resources available [15,30,31] through genomic streamlining [32].

Prokaryotic species with large genome sizes are more versatile, have higher metabolic capabilities [33,34] and adaptive plasticity[4,35,36], potentially have a large habitat breadth [37], and have high ecological diversification potential [38]. In support of this framework, Lauro et al. [39] predicted marine microbes with small genomes are predominant in oligotrophic environments, whereas species with larger genomes are more abundant in copiotrophic environments. Similar results were observed in deep ocean environments where high hydrostatic pressure and cold temperatures in combination with high nutrients select for large genome sizes in prokaryotic species [40]. In fertilized grasslands, copiotrophic bacterial taxa tend to have smaller genomes due to specialization on the high abundance of simple forms of P and N [41]. Finally, bacterial pathogens and symbionts often have small genomes, potentially due to genome streamlining as simple C and N sources can be accessed from the hosts [33,42]. In the gut microbiome, where bacteria are in a competitive environment rich in nutrients, genome sizes tend to be twice as high as genome sizes in the microbiomes of other body parts [43].

Our work relies on comparative genomics as it can help reveal tradeoffs in overall metabolic capacity based on evolutionary shifts in gene potential. However, we acknowledge that trait

tradeoffs might be more apparent at the proteomic level (gene expression) or might manifest in other ways that are not fully captured at the genome level (i.e., physiological constraints). Nevertheless, some proteomic studies have found trait tradeoffs that might support the results found in this study. Proteomic data have helped reveal tradeoffs between microbial growth and traits related to microbial adaptability, survival, and stress tolerance [44]. Adaptability to nutrient fluctuations can be reflected by an increase in synthesis of guanosine tetra- and penta-phosphate under high nutrient concentrations, allowing adaptation to low nutrients at the cost of cell growth due to inhibition of ribosome synthesis [45,46]. Likewise, growth capacities, defined as rate and yield, trade off against survival traits and stress responses in *Bacillus subtilis* due to proteomic constraints [47]. Proteomic studies so far have focused mainly on binary tradeoffs between growth capabilities vs. A and S-related traits. Further studies are needed to investigate whether A and S tradeoffs exist at the proteomic level.

Relevance of genome size for traits and tradeoffs

We observed that genome size explains the variance in many functional traits relevant to modeling microbial-driven carbon and nutrient cycling, like protein degradation capacity, pH tolerance, yield, and optimal growth temperature. Thus, genome size can be used to derive empirical relationships with functional traits in biogeochemical soil models.

It is not surprising that large genomes are likely to have more functional traits and more genes per trait [48], and several studies have supported using genome size to predict functional traits. For example, Kelly [49] found that larger genomes have a high protein degradation capacity. Smaller genomes are typically selected in acidic environments, whereas larger genomes are found in alkaline environments [50,51]. In resource-rich soils with relatively neutral pH, small genomes are selected [19]. Sabath et al. [52] showed that high temperatures select for smaller genomes in microbes to reduce maintenance costs. Specifically for soils, Sørensen et al. [27] showed a negative linear relationship between genome size and soil temperature across a gradient of temperate-to-thermal soils. Genome size can also predict microbial activity, i.e., growth rate, discriminating between slow-growing bacteria with bigger genomes and fast-growing bacteria with small genomes [20].

With other functional traits, genome size does not always show strong, linear relationships. For example, we observed a weak linear relationship with biofilm formation capacity, possibly due to different successional processes that form biofilms, which would imply time-dependent tradeoffs [54]. Moreover, the strength of the relationship between genome size and other traits depends on the number of genes needed for a particular stress response. If a relatively small gene set is required, we might expect a weaker relationship. This may be the case for biofilm formation, where we found fewer stress-related genes. However, for other traits like pH resistance traits, the linear relationship with genome size is strong despite a lower number of related genes. mgt and genome size are almost independent across many bacteria species [48], possibly because minimum generation time is influenced by factors besides genomic characteristics, particularly environmental conditions [1]. We only observed a weak negative linear relationship between genome size and mgt in fast-growing bacteria, where the genomic predictions of mgt tend to be more robust [14]. Similarly, we did not observe strong linear relationships between genome size and yield, but these relationships may be hard to detect due to measurement limitations. Our estimate of CUE was based only on a generic metabolism and carbon source (glucose) using the DEBmicrotrait model [12]. Finally, the weak linear relationships seen among genome size and other functional traits could be explained by the greater gene multifunctionality that can be observed in small genome size FGs, where the loss of functional genes is partly compensated by other genes [55].

We observed that the capacity to degrade proteins (i.e., genes encoding protein degradation enzymes) correlated clearly with other functional traits and could thus serve as a potential master trait for defining microbial strategies. However, the definition of a functional trait from genetic information depends on gene annotation tools, which are biased toward known genes and functions [56]. Genome size, on the other hand, is a genomic trait that does not depend on gene annotation and can provide more reliable predictions. Additionally, genome size can be and has been estimated based on more available information like taxonomy [57] or environmental variables such as land cover and precipitation seasonality [58]. Bioinformatics tools that predict function independently of protein annotation [59] could help verify the relationships observed in this work. Finally, GC content and gene richness also correlate with other functional traits [17]. From these three genomic traits, genome size remains the simplest to measure and, thus, has the highest potential to serve as a master trait underlying microbial strategies represented in biogeochemical models.

Finally, although it is unclear if non-linear functions relating genome size and functional traits might explain more of the observed variance, many studies also follow a parsimonious approach and use linear functions to predict traits using genome size [53,30,19,60,29]. On the other hand, Sela et al. [35] proposed using power laws with linear-like scaling exponents to model the evolution of different functional traits from genome size. These models incorporate the fitness costs of gene deletion/acquisition and genomic plasticity (likelihood of acquiring a new gene). Such approaches are suggested to be tested against the simpler linear model estimates in this study when integrated into a mechanistic biogeochemical model to test their performance. For modeling purposes, focusing on parsimonious functions could be beneficial for further calibration and validation purposes.

Implications for biogeochemical modeling

Our findings take us a step closer to understanding the metabolic potential of soil microbes and combined with omics data can be applicable to model validation [63]. Typical soil biogeochemical trait-based models characterize microbial taxa through functional traits (capacity to degrade substrates, monomer transport, and stress resistance) derived from uniform distributions [64] but do not explicitly represent genome size. From such models, we can expect hypothetical taxa with more enzymes, transporters, stress tolerance, etc., to be analogous to large-genome taxa. Moreover, cheaters with low stress tolerance would be analogous to small-genome microbes. For model validation, these models could run under different conditions to see what tradeoffs emerge when genome size is considered implicitly. For example, simulations with stable moisture and lower polymer diversity should favor cheaters with low stress tolerance, whereas many diverse polymers and variable moisture should favor the large-genome taxa.

Our work highlights that there is still a mismatch between the current soil model structure (inputs and outputs) and genomic data but presents an avenue to obtain at least some of the parameters through empirical relationships between genome size and functional traits. Nevertheless, alternative model formulations might need to be explored to better exploit omics data. Given that the mgt can be partially derived from rRNA gene copies [20], models can incorporate Monod kinetics, which relies on the mgt instead of Michaelis-Menten kinetics. Additionally, surrogate models [62] can help produce reduced and more manageable model

versions of current models to allow inverse modeling calibration efforts against biogeochemical data. Finally, other omics data, like metatranscriptomics, have the potential to indicate microbial activity in soils [20]. However, we still need to analyze how well these other data types can be integrated into existing models.

Methods

Data description

Our dataset includes bacterial MAGs from two field sites. The first site is a 1-year post-fire coniferous forest in Colorado and Wyoming, USA, with a gradient of burn severity (low, moderate, and high severity), from which 635 MAGs were assembled from soil metagenomes [65]. The second is the Loma Ridge Global Change Experiment in southern California, USA. From this site, 533 MAGs were assembled from litter samples of grassland and shrubland litter bags collected between 2017 and 2019 at four different time points [5].

Given the increasing number of publicly available MAGs across different environments, we selected publicly available MAGs on the Integrated Microbial Genomes and Microbiomes (IMG/MER) database that were categorized as ‘soil’ and ‘host associated: rhizosphere’ MAGs according to their GOLD ecosystem classification [66,67]. We only targeted unrestricted MAGs with medium (≥50% completion, <10% contamination) and high (>90% completion, <5% contamination) quality [68], resulting in 26046 MAGs, representing 3124 metagenomes. Individual MAGs and their annotations with associated metadata were downloaded to the University of Massachusetts high-performance computing cluster, Unity, using Globus. The combined dataset of 27214 MAGs allowed us to conduct a more comprehensive comparative genomic analysis across different environmental conditions (Table S13).

The MAGs used here vary in completeness and contamination. Incomplete MAGs may have fewer genes and smaller genome sizes, leading to potential bias and spurious correlations. Therefore, we downloaded 5301 publicly available whole-genome sequences from soil isolates from the JGI - IMG (Joint Genome Institute Integrated Microbial Genomes System) portal (https://img.jgi.doe.gov/m) to compare our results (Table S1). The data was accessed on May 15, 2024.

All data used in this study were publicly available.

MAGs to emergent functional groups pipeline

As we aimed for functional trait-based analysis, we developed an iterative pipeline to group our MAGs and whole-genomes into different functional groups (FGs) based on their functional capacity (Figure 7).

Step 1. microTrait annotation:

For a pipeline iteration, we started with a metagenome sample, binned and assembled using typical de novo assembly methods to yield MAGs [69,5]. These MAGs were annotated using the computational R tool microTrait [3]. We used microTrait only for gene annotation and later grouped the genes into traits following the YAS framework (Tables S13, S14, S15, S16). The pipeline is meant to be iterative, so as more metagenomes become available, the pipeline can be rerun to update the number of FGs.

Step 2. Gene selection:

microTrait looks for a total of 2298 different genes within the genomes. We fit a linear regression between MAG abundance and gene counts for each sample/metagenome to select genes that were strongly associated with site and treatment (eq. 1). We checked for perfect correlation and multicollinearity among genes using the "cor_select" and "collinear" functions from the "collinear" package (v1.1.1)[70]. Finally, we used the stepwise forward regression for the final selection of the genes, using the function "ols_step_forward_p" from the "olsrr" package (v0.6.0)[71].

$$MAG_{abundance} = f(genes_{metagenome}) \quad \text{eq. 1}$$

Among the 26046 MAGs from IMG, some came from projects where few MAGs were assembled, making the gene selection based on MAG abundance step meaningless. Therefore, we only applied this step to projects for which at least 20 MAGs were assembled, accounting for about 50% of the IMG MAGs. We did not apply this step for isolates because relative abundance data was unavailable.

Finally, we combined all the genes selected from each metagenome sample (963 genes were selected) to perform the functional grouping (Table S17).

Step 3: Functional grouping:

Based on our reduced MAG-to-genes matrix, we derived the number of FGs. We first used the function “parDist” from the R package “parallelDist” (v0.2.6) [72] to obtain the distance matrix using the *Jaccard* method for binary data. We used the “hclust“ function from the R package “Stats“ (v4.3.1) [73] and the "ward.D2" method to generate a dissimilarity matrix. To select the maximum number of different FGs, we performed a pairwise multilevel comparison using the “pairwise.adonis” function from the R package “pairwiseAdonis” (v0.4.1) [74], setting an $\alpha = 0.05$ and 999 permutations. This step was done iteratively until at least one pairwise interaction resulted in a p-value bigger than 0.05 [75].

Our method was computationally too expensive to be applied to our full MAG-to-genes matrix of 27214 MAGs. Therefore, we derived a curve to extrapolate the potential number of FGs for the entire matrix (Figure S1). We randomly selected 100, 200, 500, 750, 1000, 1500, 2000, 3500, and 5000 MAGs to obtain the maximum number of FGs for each sample size. We repeated the process with 12 different samples per sample size and used the mean value of their FGs to fit a linear model.

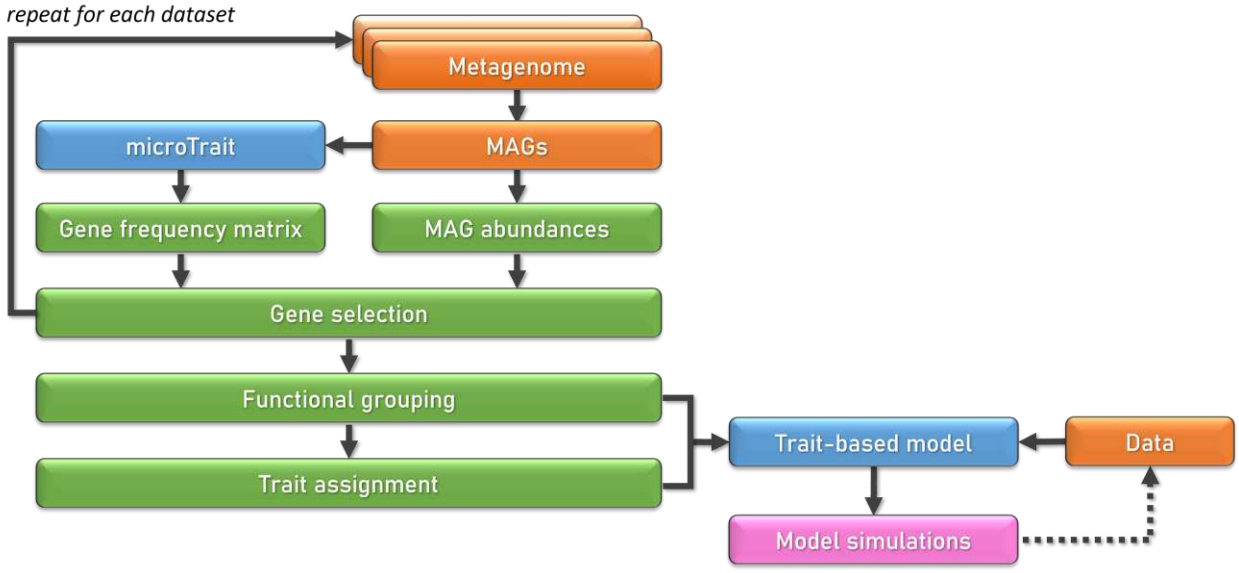

Figure 7. Pipeline from metagenomes to functional groups.

YAS aggregated traits and tradeoffs

Model Traits

Based on the YAS framework [1], we chose functional traits representing resource acquisition (A), such as genes encoding carbohydrate and protein degrading enzymes and monomer transporters, and genes encoding stress tolerance traits (S), such as osmolyte production, biofilm formation, and heat and pH tolerance (Table S2-S8). The selected traits represented the A and S strategies of the framework.

Yield cannot be directly extracted from microbial genomes. For example, microTrait [3] infers resource use (RU) traits related to energy-generating traits that could be used as a proxy for YAS's yield strategy. The nature of these RU traits in microTrait is binary, so their interpretation as a quantitative parameter is limited. Therefore, we estimated CUE (Table S23) using the DEBmicrotrait model [12]. Due to higher quality and completion, we ran DEBmicrotrait only on the 5301 soil isolates, assuming glucose as substrate and an initial concentration of 0.025 mgC/g soil. We only parameterized the DEBmicrotrait model with total transporters and carbohydrate degradation gene copies.

We also derived quantitative traits like the minimum generation time (mgt) and optimal growth temperature (ogt) from microTrait, which are related to microbial growth and adaptation to different environmental niches [3]. We estimated mgt for all our MAGs using microTrait. microTrait bases this estimation on ribosomal RNA aided by the gRodon R package. However, estimations of mgt for slow-growing microbes (mgt > 5 hours) via gRodon is less precise due to lower cellular signal of growth optimization (via high codon usage bias - CUB) [14], so we separated MAGs into putative slow (mgt > 5 hr) and fast (mgt < 5 hr) growers to plot against genome size.

Trait aggregation process

Because microTrait generally produces gene count numbers per trait (Tables S20, S21), we expressed traits as total genes per YAS trait category and functional group (eq. 2).

$$MT = \sum_{i}^{n} T_j \quad \text{eq. 2}$$

where *i* is the related gene per trait *T* category *j*, and *n* is the total genes per trait category.

A similar analysis was performed with complete genomes from soil isolates (Table S22) to compare the robustness of the MAG results with complete genomes.

Trait tradeoffs

We tested for correlations among all YAS traits to test for potential trait tradeoffs. We used the function "ggpairs" from the package "GGally" (v2.2.1) [76] with FGs from both MAGs and soil isolates.

Genome size and YAS trait empirical relationships

We tested for linear relationships between the mean genome size of functional groups and aggregated traits calculated using eq 2 with FGs from both MAGs and soil isolates, assuming that genome size is the driving variable of YAS functional traits. We tested only for linear relationships, following the parsimony principle.

$MT = f(genome\ size)$ eq. 3

where MT are the aggregated traits calculated by eq.2.

Finally, and only for yield, we tested mean genome size and aggregated yield at different taxonomic levels.

Supplementary information.

Supporting Tables and Figures can be found in the following document: "Y-A+S is the new YAS_SI.doc"

Acknowledgment

(A portion of) These data were produced by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337; operated under Contract No. DE-AC02-05CH11231) in collaboration with the user community.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- **Funding**: This work is supported by the U.S. Department of Energy Office of Science, through the Genomic Science Program in the Office of Biological and Environmental Research (BER). The national laboratory partners are operated under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), and DE-AC05-76RL01830 (PNNL). This work was also supported Schmidt Sciences under the CALIPSO project and the U.S. Department of Energy Office of Science, through the Genomic Science Program, Biological and Environmental Research (BER), under award DE-SC0023127 and US Department of Energy (DOE), Office of Biological and Environmental Research, Genomic Science Program (GSP) LLNL 'Microbes Persist' Soil Microbiome Scientific Focus Area SCW1632. This research was performed at Lawrence Berkeley National Laboratory under the auspices of the US Department of Energy contract number DE-AC02-05CH11231.

- **Competing interests**: The authors declare no competing interests.
- **Ethics approval and consent to participate**: Not applicable.
- **Consent for publication**: Not applicable.
- **Data availability**: Soil MAGs and soil isolates from the JGI-IMG portal https://img.jgi.doe.gov/m. The data to reproduce this work will be uploaded to Zenodo/figshare after acceptance of the manuscript.
- **Materials availability**: Not applicable.
- **Code availability**: All scripts to reproduce the work in this paper are found in the following GitHub repository https://github.com/Luciana-cloud/microtrait_DEMENT
- **Author contribution**:
  LCR: Conceptualization, Data acquisition, Formal analysis, Visualization, Writing - Original Draft; **JMM**: Initial discussions, Visualization, Writing - Review & Editing; **GLM**: Initial discussions, Formal analysis, Review & Editing; **KLS**: Initial discussions, Data acquisition, Writing - Review & Editing, **UK**: Initial discussions, Formal analysis, Review & Editing; **AAM**: Initial discussions, Data acquisition, Review & Editing; **CMB**: Initial discussions, Data acquisition, Review & Editing; **DF**: Initial discussions, Data acquisition, Review & Editing; **WEA**: Initial discussions, Data acquisition, Review & Editing; **AC**:

Initial discussions, Formal analysis, Data acquisition, Review & Editing; **ARN**: Initial discussions, Data acquisition, Supervision, Review & Editing; **EB**: Initial discussions, Supervision, Review & Editing; **BLB**: Initial discussions, Supervision, Review & Editing; **SDA**: Conceptualization, Funding acquisition, Supervision, Writing - Review & Editing.

References

1. 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).
2. 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).
3. Karaoz, U. & Brodie, E. L. microTrait: A Toolset for a Trait-Based Representation of Microbial Genomes. *Front. Bioinforma.* **2**, 918853 (2022).
4. Piton, G. *et al.* Life history strategies of soil bacterial communities across global terrestrial biomes. *Nat. Microbiol.* **8**, 2093–2102 (2023).
5. Malik, A. A. *et al.* Bacterial population-level trade-offs between drought tolerance and resource acquisition traits impact decomposition. *ISME J.* wrae224 (2024) doi:10.1093/ismejo/wrae224.
6. Treseder, K. K. Ecological strategies of microbes: Thinking outside the triangle. *J. Ecol.* **111**, 1832–1843 (2023).
7. Guseva, K. *et al.* Bacteria face trade-offs in the decomposition of complex biopolymers. *PLOS Comput. Biol.* **20**, e1012320 (2024).
8. Broderick, C. M. *et al.* Long-term climate establishes functional legacies by altering microbial traits. *ISME J.* **19**, wraf005 (2025).
9. Ferenci, T. Trade-off Mechanisms Shaping the Diversity of Bacteria. *Trends Microbiol.* **24**, 209–223 (2016).
10. Knight, C. G. *et al.* Soil microbiomes show consistent and predictable responses to extreme

events. *Nature* (2024) doi:10.1038/s41586-024-08185-3.

11. Charlesworth, B. The origin of genomes — not by natural selection? *Curr. Biol.* **18**, R140–R141 (2008).
12. Marschmann, G. L. *et al.* Predictions of rhizosphere microbiome dynamics with a genome-informed and trait-based energy budget model. *Nat. Microbiol.* **9**, 421–433 (2024).
13. Meziti, A. *et al.* The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Appl. Environ. Microbiol.* **87**, e02593-20 (2021).
14. 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.* **118**, e2016810118 (2021).
15. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
16. Shenhav, L. & Zeevi, D. Resource conservation manifests in the genetic code. *Science* **370**, 683–687 (2020).
17. Beier, S., Werner, J., Bouvier, T., Mouquet, N. & Violle, C. Trait-trait relationships and tradeoffs vary with genome size in prokaryotes. *Front. Microbiol.* **13**, 985216 (2022).
18. Rodríguez-Gijón, A. *et al.* Linking prokaryotic genome size variation to metabolic potential and environment. *ISME Commun.* **3**, 25 (2023).
19. 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).
20. Chuckran, P. F. *et al.* Codon bias, nucleotide selection, and genome size predict *in situ* bacterial growth rate and transcription in rewetted soil. Preprint at https://doi.org/10.1101/2024.06.28.601247 (2024).
21. Saifuddin, M., Bhatnagar, J. M., Segrè, D. & Finzi, A. C. Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat. Commun.* **10**, 3568 (2019).

22. Xu, Q. *et al.* Active microbial population dynamics and life strategies drive the enhanced carbon use efficiency in high-organic matter soils. *mBio* **15**, e00177-24 (2024).
23. Konstantinidis, K. T. & Tiedje, J. M. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci.* **101**, 3160–3165 (2004).
24. Sauer, D. B. & Wang, D.-N. Predicting the optimal growth temperatures of prokaryotes using only genome derived features. *Bioinformatics* **35**, 3224–3231 (2019).
25. Musto, H. *et al.* Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem. Biophys. Res. Commun.* **347**, 1–3 (2006).
26. Haverkamp, T. H. A. *et al.* Thermosipho spp. immune system differences affect variation in genome size and geographical distributions. *Genome Biol. Evol.* (2018) doi:10.1093/gbe/evy202.
27. Sorensen, J. W., Dunivin, T. K., Tobin, T. C. & Shade, A. Ecological selection for small microbial genomes along a temperate-to-thermal soil gradient. *Nat. Microbiol.* **4**, 55–61 (2018).
28. Bentkowski, P., Van Oosterhout, C. & Mock, T. A Model of Genome Size Evolution for Prokaryotes in Stable and Fluctuating Environments. *Genome Biol. Evol.* **7**, 2344–2351 (2015).
29. Dewar, A. E., Hao, C., Belcher, L. J., Ghoul, M. & West, S. A. Bacterial lifestyle shapes pangenomes. *Proc. Natl. Acad. Sci.* **121**, e2320170121 (2024).
30. Chuckran, P. F., Hungate, B. A., Schwartz, E. & Dijkstra, P. Variation in genomic traits of microbial communities among ecosystems. *FEMS Microbes* **2**, xtab020 (2022).
31. Von Meijenfeldt, F. A. B., Hogeweg, P. & Dutilh, B. E. A social niche breadth score reveals niche range strategies of generalists and specialists. *Nat. Ecol. Evol.* **7**, 768–781 (2023).
32. Giovannoni, S. J. *et al.* Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* **309**, 1242–1245 (2005).
33. Moran, N. A. Microbial Minimalism. *Cell* **108**, 583–586 (2002).

34. Guieysse, B. & Wuertz, S. Metabolically versatile large-genome prokaryotes. *Curr. Opin. Biotechnol.* **23**, 467–473 (2012).
35. Sela, I., Wolf, Y. I. & Koonin, E. V. Selection and Genome Plasticity as the Key Factors in the Evolution of Bacteria. *Phys. Rev. X* **9**, 031018 (2019).
36. Salwan, R. & Sharma, V. Genomics of prokaryotic extremophiles to unfold the mystery of survival in extreme environments. *Microbiol. Res.* **264**, 127156 (2022).
37. Barberán, A. *et al.* Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol. Lett.* **17**, 794–802 (2014).
38. Cordero, O. X. & Hogeweg, P. The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci.* **106**, 21748–21753 (2009).
39. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci.* **106**, 15527–15533 (2009).
40. Ngugi, D. K. *et al.* Abiotic selection of microbial genome size in the global ocean. *Nat. Commun.* **14**, 1384 (2023).
41. Leff, J. W. *et al.* Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl. Acad. Sci.* **112**, 10967–10972 (2015).
42. Fijarczyk, A., Hessenauer, P., Hamelin, R. C. & Landry, C. R. Lifestyles shape genome size and gene content in fungal pathogens. Preprint at https://doi.org/10.1101/2022.08.24.505148 (2022).
43. Moeller, A. H. Genomic Expansions in the Human Gut Microbiome. *Genome Biol. Evol.* **13**, evab156 (2021).
44. Zhu, M. & Dai, X. Shaping of microbial phenotypes by trade-offs. *Nat. Commun.* **15**, 4238 (2024).
45. Wu, C. *et al.* Enzyme expression kinetics by Escherichia coli during transition from rich to minimal media depends on proteome reserves. *Nat. Microbiol.* **8**, 347–359 (2023).
46. Zhu, M. & Dai, X. Stringent response ensures the timely adaptation of bacterial growth to

686 nutrient downshift. *Nat. Commun.* **14**, 467 (2023).
687 47. Zhu, M. *et al.* A fitness trade-off between growth and survival governed by Spo0A-mediated
688 proteome allocation constraints in *Bacillus subtilis*. *Sci. Adv.* **9**, eadg9733 (2023).
689 48. Westoby, M. *et al.* Cell size, genome size, and maximum growth rate are near-independent
690 dimensions of ecological variation across bacteria and archaea. *Ecol. Evol.* **11**, 3956–3976
691 (2021).
692 49. Kelly, S. The Amount of Nitrogen Used for Photosynthesis Modulates Molecular Evolution in
693 Plants. *Mol. Biol. Evol.* **35**, 1616–1625 (2018).
694 50. Chuckran, P. F. *et al.* Edaphic controls on genome size and GC content of bacteria in soil
695 microbial communities. *Soil Biol. Biochem.* **178**, 108935 (2023).
696 51. Cortez, D., Neira, G., González, C., Vergara, E. & Holmes, D. S. A Large-Scale Genome-
697 Based Survey of Acidophilic Bacteria Suggests That Genome Streamlining Is an Adaption
698 for Life at Low pH. *Front. Microbiol.* **13**, 803241 (2022).
699 52. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth Temperature and Genome Size in
700 Bacteria Are Negatively Correlated, Suggesting Genomic Streamlining During Thermal
701 Adaptation. *Genome Biol. Evol.* **5**, 966–977 (2013).
702 53. Li, J. *et al.* Predictive genomic traits for bacterial growth in culture versus actual growth in
703 soil. *ISME J.* **13**, 2162–2172 (2019).
704 54. Niederdorfer, R., Besemer, K., Battin, T. J. & Peter, H. Ecological strategies and metabolic
705 trade-offs of complex environmental biofilms. *Npj Biofilms Microbiomes* **3**, 21 (2017).
706 55. Kelkar, Y. D. & Ochman, H. Genome Reduction Promotes Increase in Protein Functional
707 Complexity in Bacteria. *Genetics* **193**, 303–307 (2013).
708 56. Maxeiner, S., Krasteva-Christ, G. & Althaus, M. Pitfalls of using sequence databases for
709 heterologous expression studies – a technical review. *J. Physiol.* **601**, 1611–1623 (2023).
710 57. Mercier, C., Elleouet, J., Garrett, L. & Wakelin, S. A. genomesizeR: An R package for
711 genome size prediction. Preprint at https://doi.org/10.1101/2024.09.08.611926 (2024).

58. Donhauser, J. *et al.* Modelling soil prokaryotic traits across environments with the trait sequence database ampliconTraits and the R package MicEnvMod. *Ecol. Inform.* **83**, 102817 (2024).
59. Pollak, S. *et al.* Public good exploitation in natural bacterioplankton communities. *Sci. Adv.* **7**, eabi4717 (2021).
60. Klimek, D., Herold, M. & Calusinska, M. Comparative genomic analysis of Planctomycetota potential for polysaccharide degradation identifies biotechnologically relevant microbes. *BMC Genomics* **25**, 523 (2024).
61. Riley, W. *et al.* A framework for integrating genomics, microbial traits, and ecosystem biogeochemistry. Preprint at https://doi.org/10.21203/rs.3.rs-4966902/v1 (2024).
62. Lu, D. & Ricciuto, D. Efficient surrogate modeling methods for large-scale Earth system models based on machine-learning techniques. *Geosci. Model Dev.* **12**, 1791–1807 (2019).
63. Abs, E., Chase, A. B., Manzoni, S., Ciais, P. & Allison, S. D. Microbial evolution—An under-appreciated driver of soil carbon cycling. *Glob. Change Biol.* **30**, e17268 (2024).
64. Allison, S. D. A trait-based approach for modelling microbial litter decomposition. *Ecol. Lett.* **15**, 1058–1070 (2012).
65. Nelson, A. R. *et al.* Wildfire-dependent changes in soil microbiome diversity and function. *Nat. Microbiol.* **7**, 1419–1430 (2022).
66. Chen, I.-M. A. *et al.* The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* **51**, D723–D732 (2023).
67. Mukherjee, S. *et al.* Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9. *Nucleic Acids Res.* **51**, D957–D963 (2023).
68. The Genome Standards Consortium *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
69. Nelson, A. R. *et al.* Wildfire-dependent changes in soil microbiome diversity and function.

Nat. Microbiol. **7**, 1419–1430 (2022).

70. Blas Benito. BlasBenito/collinear: CRAN release v1.0.1. Zenodo https://doi.org/10.5281/ZENODO.10039489 (2023).
71. Hebbali, A. olsrr: Tools for Building OLS Regression Models. (2024).
72. Eckert, A. parallelDist package - RDocumentation. (2022).
73. R Core Team. R: The R Project for Statistical Computing. R Foundation for Statistical Computing (2023).
74. Martinez Arbizu, P. pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. (2017).
75. Fong, C. R., Ryznar, E. R., Smith, L. L. & Fong, P. Towards a trait-based framework for marine macroalgae: Using categorical data to explore the nature of emergent functional groups. *J. Ecol.* **111**, 1848–1865 (2023).
76. Schloerke, B. *et al.* GGally: Extension to 'ggplot2'. (2024).

Supplementary Files

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

- YASisthenewYASSI.pdf
- SupplementaryTables.zip