Imperial College London

Department of Life Sciences

Microbial Cell Size Distributions in Gut Communities: Deviations from Null Models and Functional Implications

Author:
Charmaine Wan

Supervisor:
Dr. Samraat Pawar

October 2025

Abstract

Microbial cell size is a fundamental trait shaping physiological function and ecosystem processes, yet the drivers of size structuring in host-associated gut microbiomes remain unresolved. This study quantified cell size distributions in gut microbiomes from 712 host species across diverse taxonomic and physiological groups. By integrating microbial cell size data with amplicon sequencing, we tested whether observed cell size distributions deviate from random assembly and evaluated the influence of host taxonomy and thermal physiology. Empirical randomization analyses showed that most gut microbiomes have mean cell sizes significantly larger than expected by chance, indicating deterministic assembly shaped by host traits rather than stochastic processes. Endothermic and vertebrate hosts favoured larger, more metabolically efficient microbes, while phylogeny and thermal physiology jointly structured community cell size. These patterns reveal strong ecological filtering and host selection in shaping microbiome function, with implications for carbon cycling and ecosystem stability under environmental change. Our findings advance trait-based understanding of microbial community assembly and highlights the importance of integrating empirical trait distributions into theoretical and predictive host–microbiome models.

Introduction

Microbial life, through its extraordinary diversity and metabolic activity, is a cornerstone of ecological and biogeochemical dynamics. Among prokaryotes, cell size is a foundational phenotypic trait that shapes physiological function, ecological strategy, and evolutionary adaptation (DeLong et al., 2010; Kempes et al., 2011). Prokaryotic cells can range more than four orders of magnitude in volume, from ultramicrobacteria at 0.2 μm diameter to giants exceeding 750 μm (Schulz and Jorgensen, 2001). This reflects adaptation to gradients in resource supply, environmental stress, and trophic interactions (Portillo et al., 2013). Cell size dictates surface-to-volume ratios, thereby determining nutrient uptake efficiency, metabolic rates, and competitive dynamics (Levin and Angert, 2015). As microbial communities drive essential biogeochemical cycles that regulate Earth's elemental fluxes, understanding how cell size structures these communities offers crucial insights into ecosystem functioning under global environmental change.

The ecological significance of microbial cell size extends beyond individual physiology to influence community-level carbon cycling and ecosystem stability. The metabolic theory of ecology (MTE) links physiological rates to organismal size using allometric scaling laws (West, 1997; DeLong et al., 2010). In contrast to animals, heterotrophic prokaryotes exhibit hyperallometric scaling where larger cells demonstrate disproportionately higher metabolic rates and carbon use efficiency compared to smaller cells (DeLong et al., 2010; Kempes et al., 2012). This biological scaling has direct consequences for ecosystem carbon cycling. Communities dominated by smaller cells tend to exhibit lower carbon use efficiency (CUE) and higher respiration, ultimately respiring a greater fraction of organic carbon back to the atmosphere (Dang and Morrissey, 2024). In contrast, size-diverse communities display more efficient and stable substrate conversion to biomass, with implications for long-term ecosystem carbon storage. Such linkages between the size-abundance spectrum and carbon fluxes are increasingly recognized as important modulators of feedback to climate change (Daufresne et al., 2009).

Beyond metabolic scaling, the distribution of cell sizes in microbial communities is a major axis of trait structure influencing resilience and dynamical stability. Recent advancements in consumer-resource theory extend classic models, showing that size structure mediates the range of feasible, stably coexisting communities (Posfai et al., 2017). Communities that maintain a broad span of cell sizes buffer resource variability, combine fast responders with slower stabilizers, and can resist collapse following perturbations (Kempes et al., 2016). Theoretical modeling predicts that shifts in cell size distribution, whether induced by environmental change or ecological filtering, can reverberate through food webs and ecosystem function (Jiang and Morin, 2007; Portillo et al., 2013).

Building on this theoretical foundation, host-associated gut microbiomes offer a powerful natural system for dissecting the ecological and evolutionary factors that structure microbial size organization. Microbial communities within animal guts are shaped not only by resource and physical conditions, but also by host taxonomy, thermal physiology, and evolutionary history (Smits et al., 2016; Groussin et al., 2017). Whether and how host traits drive size-structuring of resident microbes remains largely unresolved, especially across large phylogenetic and physiological gradients.

Despite growing recognition of cell size as a key microbial trait, systematic analyses of cell size patterns across diverse host-associated communities remain rare, and the processes structuring these distributions remain poorly understood. This study addresses these knowledge gaps by systematically analyzing microbial cell size distributions in gut microbiomes across 712 host species spanning diverse taxonomic and physiological groups. We explicitly test observed gut microbial cell size distributions deviate from null model expectations, providing a direct assessment of ecological filtering versus random assembly (Portillo et al., 2013). We also aim to quantify the contributions of host taxonomy and thermal physiology, based on established evidence for their strong impact on community structure through metabolic demands and selective environments (Xue et al., 2024). Understanding these drivers is critical for predicting how host-microbe interactions respond to environmental change and for explaining the evolutionary processes structuring microbial functional diversity.

Methods

Data collection and cell volume calculation

We collected microbial cell morphology data (cell length, width, and shape) and taxonomic information for bacterial and archaeal species from Bacterial Diversity (BacDive) Metadatabase using the BacDiveClient API. The dataset included taxonomic classifications from domain to species level for each organism. Minimum and maximum cell length and width data were retrieved in µm. Cell volume calculations were performed using shape-specific geometric formulas with the following approaches.

For rod-shaped, filamentous, vibrio-shaped, helical, spiral, curved, and spore-shaped cells, each was modeled as a capsule, using the formula:

$$V = \pi r^2 (\left(\frac{4}{3}\right) r + L)$$

where r is the radius (width/2) and L is the cylindrical length (total length - width). For ellipsoidal, coccus-shaped, ovoid-shaped, oval-shaped, sphere-shaped, and crescent-shaped cells, we approximated volumes using ellipsoidal geometry:

$$V = (\frac{4}{3})\pi r^2 (\frac{L}{2})$$

where r is the radius and L is the major axis length. For each species, minimum and maximum volumes were calculated in μm^3 based on reported size ranges, and the geometric mean volume was computed as the square root of the product of minimum and maximum volumes.

When cell shape data were missing, we imputed values using the dominant cell shape within each genus. For unknown shapes, rod-shaped morphology was assumed by default. This method ensured comprehensive coverage of cell size data while maintaining geometric accuracy across diverse bacterial and archaeal morphologies. A cell size database of 5455 bacterial species and 76 archaeal species were composed.

Amplicon data processing and taxonomic integration

Amplicon sequencing data targeting the 16S rRNA v3-4 regions were compiled from a large-scale comparative gut microbiome dataset. Processed amplicon sequencing data from

approximately 2700 gut microbiome samples representing 712 host species were integrated with a cell size database at both genus and family levels to enable weighted cell size calculations. Raw amplicon data were filtered to exclude contaminants including mitochondrial and chloroplast sequences. Taxonomic matches between amplicon sequencing and cell size databases were performed using flexible algorithms employing regex patterns to accommodate nomenclature differences. Amplicon read counts were pooled at the genus or family level for each sample, and relative abundances were calculated as proportions of total reads. These abundance data were merged with cell size records by matching taxonomic names, assigning values for cell geometric mean volume. Where multiple entries existed for a taxon, median values across all species within the genus or family provided representative cell geometric mean volume estimates. Merged datasets were produced with each row consisting of a taxon, its relative abundance, and corresponding cell size measurements, supporting downstream weighted analyses of microbiome cell size distributions.

Taxonomic classification quality in the amplicon sequencing data was systematically evaluated by quantifying the proportions of reads assigned versus unassigned across taxonomic levels for all host species. Reads were classified as unassigned if they contained ambiguous labels, were empty or null, or indicative of uncertain taxonomy. Assessment focused on both the family level, distinguishing reads confidently assigned to well-defined families from ambiguous assignments, and the genus level, identifying reads with reliable family classification but unresolved genus assignment. Analyses were conducted at the level of individual samples and aggregated by host species to characterize variation in taxonomic resolution and identify samples with substantial proportions of unclassified reads. This comprehensive evaluation of taxonomic assignment informs the interpretation of subsequent cell size analyses and highlights inherent limitations of the amplicon-based approach.

Cell volume distribution visualisation

Cell volume distributions were visualized at family and phylum taxonomic levels to assess prokaryote cell size diversity patterns. Log_{10}-transformed geometric mean volumes were calculated for all species with available data, and taxonomic groups with fewer than 30 species were excluded to ensure statistical robustness.

We created individual histogram subplots for each family with consistent bin edges spanning the global range of $\log_{10}$ volumes. Each histogram displayed the count of species within 35 bins, with mean values indicated by vertical dashed lines. A complementary normalized kernel density estimation (KDE) plot was created to visualize the probability density distributions across families, using Gaussian KDE with a bandwidth method of 0.5. The KDE plot employed normalized density values scaled to a maximum of 1.0.

Randomization analysis of cell size distribution using a global pool approach

To assess the statistical significance of observed cell size distributions in microbiome communities, a global pool randomization approach was implemented to test whether observed cell size patterns deviate from what would be expected under random assembly of cell sizes across all communities. The global pool was constructed by aggregating all cell geometric mean volumes across all samples, representing the full bacterial size range in the dataset. For each microbiome sample, the observed relative abundances of taxonomic groups were preserved while randomly assigning cell sizes without replacement from the global pool, repeated 1000 times per sample. We calculated the weighted mean, variance, and kurtosis of cell size using the original relative abundance as weights in each randomized iteration. These metrics capture the central tendency, spread, and distribution shape of cell sizes.

The null hypothesis tested by t-test and F-test is that the observed weighted mean and variance of cell sizes do not differ from those expected under random cell size assignment. To assess the statistical significance of observed metrics, we compared them to randomized distributions using t-test for means and F-tests for variances, generating 1000 p-values per sample. We calculated empirical p-values as the proportion of randomizations where $p<0.05$ to represent how often the observed pattern significantly differed from random assignments. High empirical p-values (approaching 1.0) indicate that the observed cell size distribution consistently deviates from randomness, suggesting ordered and non-random biological patterns. Meanwhile, low empirical p-values indicate that observed patterns are likely due to chance alone.

Phylogenetic analysis

To quantify the evolutionary component of variation in microbial cell size distributions, we assessed phylogenetic signals in traits mapped to host species. A dated phylogenetic tree was obtained using TimeTree, covering 507 metazoan host species present in both our dataset and the published supertree. For each host, we calculated the average difference between observed and randomized median microbial cell sizes at both family and genus taxonomic levels. These deviations were then assigned as continuous traits to the tips of the tree.

The phylogenetic signal was quantified by estimating Pagel's λ (Pagel, 1999) and Blomberg's K (Blomberg, Garland and Ives, 2003) using established comparative methods. These statistics were used to determine whether closely related hosts tended to harbor gut microbiomes with more similar cell size deviations than expected by chance.

Host group classification and comparative analysis

Host species were classified as vertebrates or invertebrates using taxonomic information. Vertebrates were defined as species belonging to the phylum Chordata, while invertebrates encompassed all other Metazoa species. Thermal physiology categories were assigned based on class-level taxonomy: endotherms included species within Mammalia and Aves, whereas ectotherms comprised all remaining Metazoa classes, excluding Pavlovophyceae and Magnoliopsida. Species lacking complete taxonomic information were removed from subsequent analyses.

To compare microbial cell size patterns across host groups, group-specific global pools of cell sizes were generated for vertebrates versus invertebrates and for endotherms versus ectotherms. For each microbiome sample, we performed 1000 randomizations by sampling cell sizes exclusively from the corresponding group-specific pool while preserving the original relative abundance distribution. Weighted medians for both observed and randomized distributions were calculated using fixed size-class bins, and the mean difference between observed and randomized medians was computed per sample.

We tested the significance of differences in median cell size between host groups using Wilcoxon rank-sum tests. Violin plots with integrated box plots were used to visualize these distributions, showing probability density and summary statistics for each host group comparison.

Results

Dataset overview and taxonomic integration

The curated cell morphology database comprised 5,457 bacterial and 78 archaeal species with complete cell dimension data, representing over 1,200 families and 2,500 genera. Given the limited number of archaeal species available, we focused exclusively on bacterial taxa for merging and subsequent analyses. We integrated amplicon sequencing data from 712 host species, totalling approximately 2,700 microbiome samples, with this database at both genus and family taxonomic levels.

Taxonomic classification quality assessed the success of assigning amplicon sequencing reads to known bacterial taxa from reference databases. At the family level, a median of approximately 75–80% of reads were confidently assigned to well-defined bacterial families, while the remaining 20–25% remained unclassified due to ambiguous labels or insufficient taxonomic resolution in the reference database. Classification proportions varied across host species, ranging from 45% to 95%. At the genus level, approximately 30–35% of reads remained unclassified below the family level, meaning these reads were assigned to a family but not to a specific genus, with variation ranging from 10% to 60% across samples. Distribution histograms for both per-sample and per-host-species taxonomic assignments demonstrated that the majority of samples achieved high family-level classification success, with a right-skewed distribution indicating most samples had low proportions of unclassified reads.

Taxonomic integration efficiency measured the success of matching classified amplicon taxa to entries in the cell size database. Integration was more successful at the family level, where we successfully matched approximately 80–85% of classified reads and 60–70% of unique classified families to corresponding cell size records in the BacDive database. At the genus level, matching rates were lower, with 65–70% of classified reads and 50–55% of unique classified

genera successfully linked to cell size data. Distribution histograms of proportion hits across samples and host species revealed that most achieved high integration rates, particularly at the family level, though a minority of samples exhibited lower matching success due to limited representation of certain taxa in the cell size database. Overall, the majority of sequencing data were successfully assigned cell volume values, enabling robust downstream analyses of community-level cell size distributions. (Given the higher classification success and integration efficiency at the family level, we proceeded with family-level data for all subsequent cell size distribution analyses.)

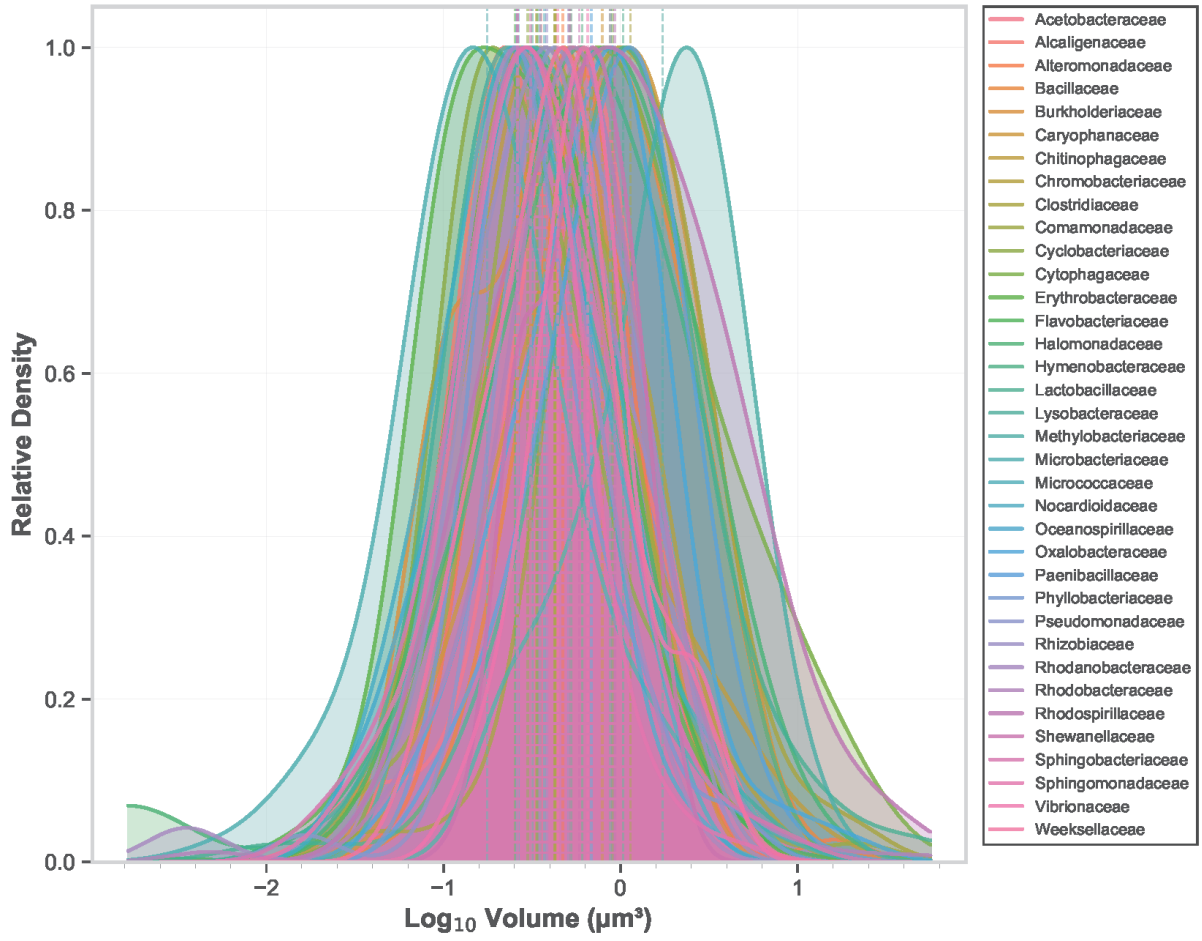

Figure 1. Normalized kernel density estimation (KDE) plot of family-level $\log_{10}$-transformed geometric cell volume distributions of 36 bacterial families ($n \geq 30$ species) with individual family densities color-coded. The color-coded dashed lines represent mean $\log_{10}$ cell volume of that specific family.

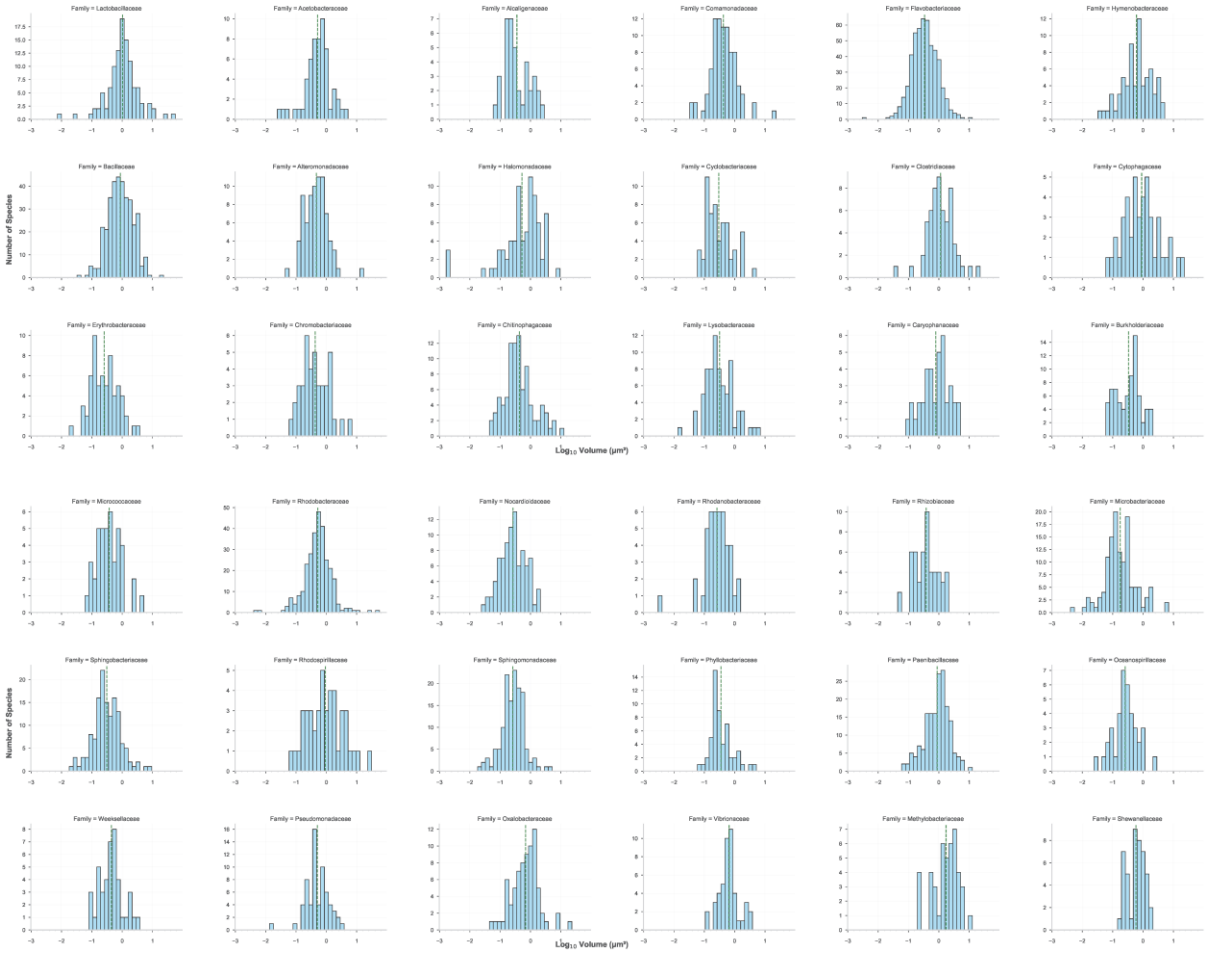

Figure 2. Histograms of family-specific distributions of $\log_{10}$-transformed geometric cell volumes in gut bacteria measured by the number of species of 36 bacterial families ($n \geq 30$ species). The green dashed lines in each family represent the mean $\log_{10}$ cell volume of that specific family.

Microbial cell volume distribution across bacterial families

To characterize the diversity of microbial cell sizes in our database, we visualized cell volume distributions at the family taxonomic level. Across 36 bacterial families with at least 30 species, $\log_{10}$-transformed geometric mean cell volumes spanned approximately -2.8 to 1.8 $\log_{10}$ µm^3, representing over 4.5 orders of magnitude in cell volume. This substantial range reflects the broad diversity of prokaryotic cell sizes in the dataset.

Individual family-level histograms revealed that most families exhibited unimodal, approximately log-normal distributions centered around distinct mean values, indicated by vertical dashed lines in each subplot (Figure 2). Family-level mean volumes typically ranged from -1.5 to 1.2 $\log_{10}$ µm^3, with a median of approximately 0.1 $\log_{10}$ µm^3 (equivalent to 1.26 µm^3). Families such as Micrococcaceae and Halomonadaceae displayed means shifted toward

larger volumes, while Acetobacteraceae and Sphingomonadaceae were centered at smaller sizes, illustrating considerable taxonomic variation in typical cell sizes.

The normalized kernel density estimation (KDE) plot aggregated density distributions across all families, revealing a clear peak at approximately 0.0 $\log_{10}$ µm^3 (1.0 µm^3), representing the modal cell volume across bacterial families (Figure 1). While most families contributed to this central mode, some extended into the distribution tails, contributing to the overall right-skewed shape. This pattern indicates that bacterial cell sizes are broadly conserved around a central value, with family-specific adaptations resulting in notable deviations in mean cell volume.

Together, these visualizations demonstrate that bacterial cell size diversity is substantial at the family level, with most families clustering around a common size range but displaying family-specific shifts in central tendency.

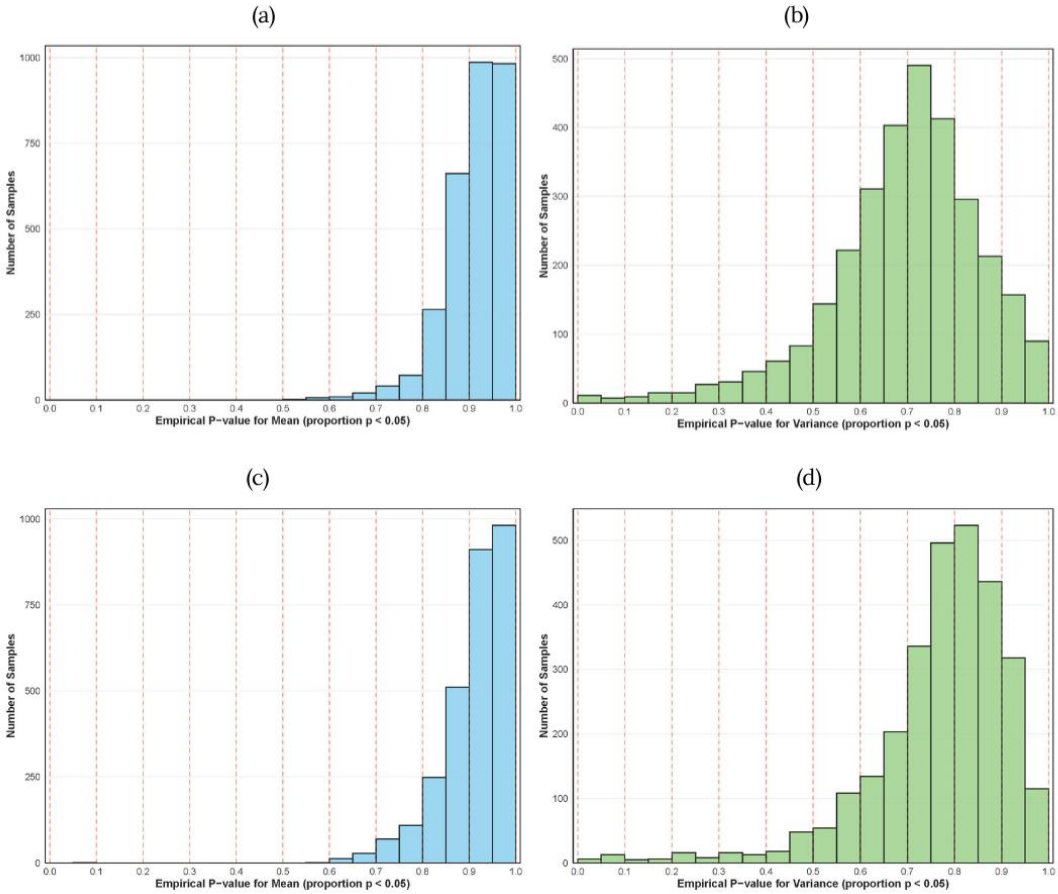

Figure 3. Histograms of empirical p-value distributions for cell size randomization tests on mean and variances at family and genus levels. Where (a) shows the empirical p-values for the weighted mean cell volume at family level (n = 3044 samples), (b) represents the variance in cell volume at the family level, (c) represents the weighted mean cell volume at genus level (n = 2872 samples), (d) represents the variance in cell volume at genus level. Vertical dashed lines indicate p-value bin cutoffs in all panels.

Table 1. Summary of empirical p-value-based randomization tests for weighted mean and variance of microbial cell size distributions on family and genus level.

Taxonomic level	Metric	$p > 0.9$	$p = 0.8$-0.9	$p < 0.05$
Family	Mean	64.7%	30.5%	0.0%
Family	Variance	8.1%	16.9%	0.4%
Genus	Mean	65.9%	26.6%	0.0%
Genus	Variance	15.1%	33.7%	0.2%

Empirical p-value analysis of cell size randomization

To assess the statistical significance of observed microbial cell size patterns, we determined empirical p-values for the weighted mean and variance of cell volume in each microbiome sample at both family and genus levels (Table 1). This analysis measures how frequently observed community cell size statistics deviate from null distributions generated by randomizing cell sizes while preserving abundance structure.

At the family level (n = 3,044 samples), 64.7% of samples exhibited empirical p-values greater than 0.9 for the mean (Figure 3a), indicating that observed mean cell volumes were substantially higher than would be expected by chance. For the genus level (n = 2,872 samples), this proportion was similar at 65.9% (Figure 3b), affirming the robustness of this pattern across taxonomic resolutions. Most samples fell into the p-value bins of 0.9–0.995 or 0.8–0.9 for both family and genus.

In contrast, variance results were less extreme. Only 8.1% of family-level samples (Figure 3c) and 15.1% of genus-level samples (Figure 3d) fell into the highest p-value bin (>0.9). Instead, variance p-values were more evenly distributed across intermediate categories, suggesting that cell size variation within communities frequently resembled that observed in randomized assemblages.

The consistently high empirical p-values for mean cell volume demonstrate that observed microbial communities are non-randomly assembled with respect to cell size, while the more moderate p-values for variance suggest that size variability is less strongly constrained by ecological or selective processes.

Phylogenetic signal in host-associated microbial cell size traits

At both family and genus levels, Blomberg's K (Blomberg, Garland and Ives, 2003) was much less than 0.1 with nonsignificant p-values ($p > 0.5$), indicating weak phylogenetic signal weaker than expected under a Brownian motion model. Pagel's λ (Pagel, 1999) was approximately 0.3 with highly significant p-values ($p < 0.01$), suggesting moderate but not strong phylogenetic dependence. These results indicate a weak to moderate phylogenetic signal in host microbial cell size traits, suggesting that while evolutionary relatedness plays a role, ecological and physiological factors likely have a greater influence.

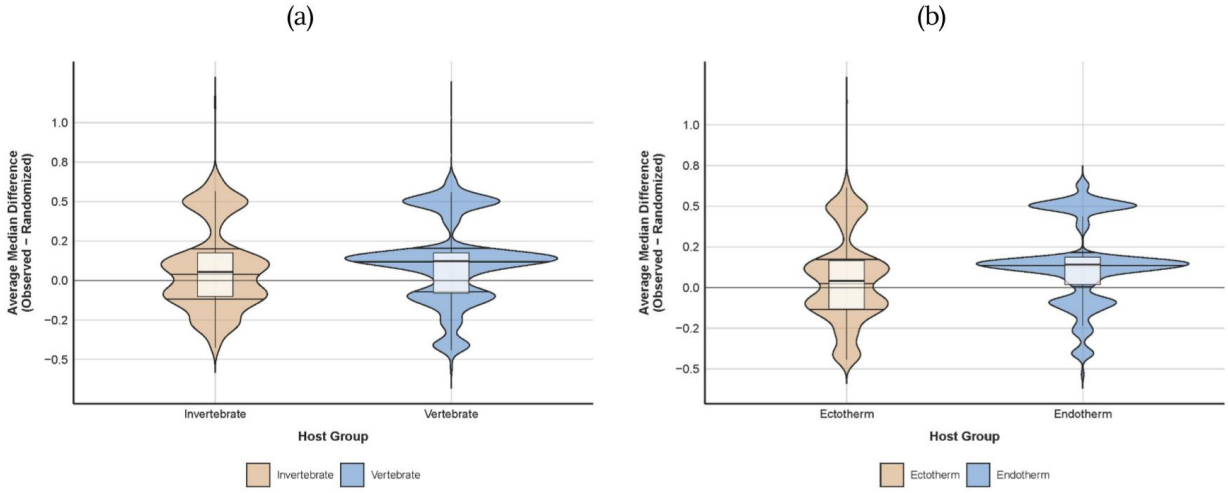

Figure 4. Violin plots of observed-randomized median cell size by host taxonomy and thermal physiology. Where (a) shows the distribution of average median cell size differences between invertebrate ($n = 518$) and vertebrate ($n = 2194$) hosts, and (b) shows the distribution of average median cell size differences between ectothermic ($n = 1115$) and endothermic ($n = 1629$) hosts.

Table 2. Results of Wilcoxon rank-sum test on the comparisons of observed versus randomized median microbial cell size across host groups.

Host group	Averaged observed median ± SE	Averaged observed – randomized median difference ± SE	Wilcoxon W	p-value
Vertebrate	0.628 ± 0.006	0.114 ± 0.006	645217	< 0.0001
Invertebrate	0.590 ± 0.012	0.078 ± 0.012		
Endotherm	0.685 ±0.006	0.146 ± 0.250	1151099	< 0.0001
Ectotherm	0.562 ± 0.008	0.045 ± 0.268		

Influence of host taxonomy and thermal physiology on microbial cell size organization

To evaluate the extent to which microbial cell size distributions deviate from random assembly and how this varies with host biological traits, we compared median cell size differences between observed communities and randomized counterparts across major host taxonomic and physiological groups (Table 2).

Significant differences in microbial cell size distributions were observed between major host taxonomic and physiological groups. Vertebrates exhibited higher median differences from randomized expectations than invertebrates (Figure 4a). Across vertebrate hosts (n = 2,194), the mean observed–randomized median difference was 0.114 ± 0.006, compared with 0.078 ± 0.012 for invertebrates (n = 518). The Wilcoxon rank-sum test confirmed this difference as highly significant (W = 645,217, $p < 0.0001$), indicating that vertebrate-associated microbiomes display cell size distributions deviating more strongly from random assembly than those of invertebrates.

Thermal physiology further revealed a clear separation between endothermic and ectothermic hosts (Figure 4b). Endotherms (n = 1,629) showed substantially greater observed–randomized median differences of 0.146 ± 0.250 than ectotherms (n = 1,115) with the value of 0.045 ± 0.268. This contrast was highly significant with the Wilcoxon rank-sum test (W = 1,151,099, $p < 0.0001$), demonstrating that endothermic hosts maintain microbial communities with consistently larger cell sizes relative to random expectations. Together, these results suggest that both host taxonomy and thermal physiology strongly influence microbial cell size organization across gut microbiomes.

Discussion

This study aims to systematically characterize how microbial cell size distributions are structured within gut microbiomes across diverse animal hosts, and to test whether these distributions deviate from random assembly. The results demonstrate that microbial cell size distributions are significantly non-random in assembly (Figure 3 and Table 1) and are systematically structured by host traits, particularly phylogeny and thermal physiology (Figure 4 and Table 2).

Empirical p-value analyses show that around 65% of samples at family level have observed mean cell sizes significantly larger than expected by chance (Figure 3a). This is consistent with previous studies evaluating the relationship between deterministic trait-environment and microbial community size assembly in both terrestrial and aquatic ecosystems (Luan et al., 2020; Li et al., 2023). While size variability within communities aligns more closely with randomness, the consistent non-random assembly of mean cell size implies strong ecological filtering and host selection shaping community structure.

Notably, endotherms and vertebrates showed the largest deviations from random assemblies (Figure 4), supporting that host species with higher metabolic rates and more stable internal environments result in preferential selection of larger and metabolically efficient microbial cells (Liu et al., 2016; Youngblut et al., 2019; Yang et al., 2020). The more pronounced separation of cell size distribution between endotherms and ectotherms highlights the pivotal role of thermal physiology in shaping gut microbiome structure and function (Kokou et al., 2018; Huus and Ley, 2021). Endotherms maintain stable, high body temperatures and metabolic rates that create a warm, high resource flux, and deeply anaerobic gut environment, This environment favours the persistence and dominance of larger, metabolically efficient, and obligate anaerobic microbes, which are well-adapted to stable conditions to support complex host metabolic needs (Li et al., 2019; Khakisahneh et al., 2020; Huus and Ley, 2021). In contrast, ectotherms experience fluctuating body temperature with less stable gut environments, leading to less specialized and generally smaller microbial communities (Huus and Ley, 2021). While phylogenetic effects contribute to microbiome variation, the influence of the thermal environment appears to be primary and hierarchical. In previous studies, it is shown that thermal physiology overrides phylogenetic signals and drives convergent shifts in microbiome structure across distantly related taxa (Amato et al., 2019; Song et al., 2020; Sepulveda and Moeller, 2020).

These results highlight a critical link between host-associated microbiomes and carbon cycling, with broad ecosystem implications. In particular, larger microbial cells in endotherm and vertebrate guts are associated with higher carbon use efficiency (CUE), meaning more carbon is retained in microbial biomass and less is lost as carbon dioxide through respiration (Saifuddin et al., 2019). Changes in host distribution, diet, and thermal physiology can be driven by climate change. As a result, these changes can lead to shifts in gut microbiome structure,

altering microbial CUE and the fate of carbon in terrestrial systems (Williams, Williams and Logan, 2023). For instance, climate-driven changes in gut microbiota can affect the balance between carbon storage and greenhouse gas emissions, causing cascading effects on global carbon cycling and climate feedback (Abs et al., 2024; Marie Le Geay et al., 2024). Therefore, the Integration of microbial cell size and CUE into carbon-cycle models is essential for accurately predicting ecosystem responses to climate change (Abs et al., 2024).

Furthermore, consumer-resource theory offers a rigorous framework for interpreting the non-random and positively skewed cell size distributions in gut microbiomes. Our empirical data suggest that stable environments and host resource filtering select for larger microbial cells, supporting theoretical predictions that efficient and larger-celled taxa are preferred under stable or high-resource supply regimes (Yoshiyama and Klausmeier, 2007; Serbanescu, Nikola Ojkic and Banerjee, 2021). Integrating empirical trait distributions into consumer-resource models will allow quantitative testing of whether the feasibility, stability, and shape of predicted community size distribution truly match observed microbiomes under varied environmental conditions.

Despite the robust findings, several methodological limitations must be acknowledged. Only bacterial analysis was performed in this study due to the incomplete taxonomic and trait coverage. Many abundant or ecologically important microbial lineages, particularly within archaea and uncultured taxa, lack cell size measurements, potentially biasing trait distributions. Only family-level aggregation was possible for trait-taxonomy mapping as amplicon sequencing that targets 16S rRNA v3-4 regions in this study. This approach is unable to resolve microbial identities to the genus or species level due to both insufficient sequence variations and confounding effects of horizontal gene transfer (Earl et al., 2018; Johnson et al., 2019; Paul, 2022). As a result, collapsing analyses to family level increases match rates between sequencing and cell size databases but it also obscures potentially important intra-family functional and ecological variation at a finer resolution (Poretsky et al., 2014; Joos et al., 2020). Furthermore, randomization-based null models are valuable for distinguishing stochastic from deterministic microbial community assembly. However, they do not account for non-random trait correlations and ecological dependencies among taxa (Zhou and Ning, 2017).

Future work should prioritize expansion and refinement of morphometric trait databases, potentially using genome-based predictions of microbial cell sizes (Vieira-Silva and Rocha,

2010). It would enable more comprehensive mapping of microbiome diversity, especially for archaeal and uncultured lineages currently underrepresented in analyses. Longitudinal sampling of microbial cell sizes and their relative abundances would benefit looking into the dynamics of trait assemble, plasticity, and eco-evolutionary adaptation within host-associated communities. This can facilitate the detection of temporal responses to environmental or physiological changes under a changing climate. Moreover, integrating empirical trait distributions with consumer-resource models, especially incorporating non-random assembly cell size abundance, will further clarify the mechanisms governing assembly, stability, and functional outcomes in microbiome ecology.

Conclusion

This study demonstrates that gut microbial communities are not random assemblages but are strongly structured by host taxonomy and thermal physiology. Deterministic ecological processes, particularly those linked to host thermal physiology, select for larger, more metabolically efficient microbes in endothermic and vertebrate hosts. These community patterns have direct implications for carbon use efficiency and ecosystem-scale carbon cycling. Moreover, our null model and phylogenetic analyses reveal that thermal physiology often exerts a stronger influence than evolutionary history in determining community cell size and structure. Together, these findings highlight the importance of integrating microbial trait-based models to better predict host-microbe and trophic interactions under global change scenarios.

References

Abs, E., Chase, A.B., Manzoni, S., Ciais, P. and Allison, S.D. (2024). Microbial evolution—An under-appreciated driver of soil carbon cycling. Global Change Biology, 30(4). doi:https://doi.org/10.1111/gcb.17268

Amato, K.R., Mallott, E.K., McDonald, D., Dominy, N.J., Goldberg, T., Lambert, J.E., Swedell, L., Metcalf, J.L., Gomez, A., Britton, G.A.O., Stumpf, R.M., Leigh, S.R. and Knight, R. (2019). Convergence of human and Old World monkey gut microbiomes demonstrates the importance of human ecology over phylogeny. Genome Biology, [online] 20(1). doi:https://doi.org/10.1186/s13059-019-1807-z

Blomberg, S.P., Garland, T. and Ives, A.R. (2003). TESTING FOR PHYLOGENETIC SIGNAL IN COMPARATIVE DATA: BEHAVIORAL TRAITS ARE MORE LABILE. Evolution, 57(4), p.717. doi:https://doi.org/10.1554/0014-3820(2003)057[0717:tfpsic]2.0.co;2

Dang, C. and Morrissey, E.M. (2024). The size and diversity of microbes determine carbon use efficiency in soil. Environmental Microbiology, 26(5). doi:https://doi.org/10.1111/1462-2920.16633

Daufresne, M., Lengfellner, K. and Sommer, U. (2009). Global warming benefits the small in aquatic ecosystems. Proceedings of the National Academy of Sciences, 106(31), pp.12788–12793. doi:https://doi.org/10.1073/pnas.0902080106

DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M. and Brown, J.H. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. Proceedings of the National Academy of Sciences, 107(29), pp.12941–12945. doi:https://doi.org/10.1073/pnas.1007783107

Earl, J.P., Adappa, N.D., Krol, J., Bhat, A.S., Balashov, S., Ehrlich, R.L., Palmer, J.N., Workman, A.D., Blasetti, M., Sen, B., Hammond, J., Cohen, N.A., Ehrlich, G.D. and Mell, J.C. (2018). Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. Microbiome, 6(1). doi:https://doi.org/10.1186/s40168-018-0569-2

Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W. and Alm, E.J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nature Communications, 8(1). doi:https://doi.org/10.1038/ncomms14319

Huus, K.E. and Ley, R.E. (2021). Blowing Hot and Cold: Body Temperature and the Microbiome. mSystems, 6(5). doi:https://doi.org/10.1128/msystems.00707-21

Jiang, L. and Morin, P.J. (2007). Temperature fluctuation facilitates coexistence of competing species in experimental microbial communities. Journal of Animal Ecology, 76(4), pp.660–668. doi:https://doi.org/10.1111/j.1365-2656.2007.01252.x

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E. and Weinstock, G.M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature Communications, 10(1), pp.1–11. doi:https://doi.org/10.1038/s41467-019-13036-1

Joos, L., Beirinckx, S., Haegeman, A., Debode, J., Vandecasteele, B., Baeyen, S., Goormachtig, S., Clement, L. and De Tender, C. (2020). Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. BMC Genomics, 21(1). doi:https://doi.org/10.1186/s12864-020-07126-4

Kempes, C.P., Dutkiewicz, S. and Follows, M.J. (2011). Growth, metabolic partitioning, and the size of microorganisms. Proceedings of the National Academy of Sciences, 109(2), pp.495–500. doi:https://doi.org/10.1073/pnas.1115585109

Kempes, C.P., Wang, L., Amend, J.P., Doyle, J. and Hoehler, T. (2016). Evolutionary tradeoffs in cellular composition across diverse bacteria. The ISME Journal, 10(9), pp.2145–2157. doi:https://doi.org/10.1038/ismej.2016.21

Khakisahneh, S., Zhang, X.-Y., Nouri, Z. and Wang, D.-H. (2020). Gut Microbiota and Host Thermoregulation in Response to Ambient Temperature Fluctuations. mSystems, 5(5). doi:https://doi.org/10.1128/msystems.00514-20

Kokou, F., Sasson, G., Nitzan, T., Doron-Faigenboim, A., Harpaz, S., Cnaani, A. and Mizrahi, I. (2018). Host genetic selection for cold tolerance shapes microbiome composition and modulates its response to temperature. eLife, [online] 7, p.e36398. doi:https://doi.org/10.7554/eLife.36398

Levin, P.A. and Angert, E.R. (2015). Small but Mighty: Cell Size and Bacteria. Cold Spring Harbor Perspectives in Biology, [online] 7(7), p.a019216. doi:https://doi.org/10.1101/cshperspect.a019216

Li, B., Li, L., Li, M., Lam, S.M., Wang, G., Wu, Y., Zhang, H., Niu, C., Zhang, X., Liu, X., Hambly, C., Jin, W., Shui, G. and Speakman, J.R. (2019). Microbiota Depletion Impairs Thermogenesis of Brown Adipose Tissue and Browning of White Adipose Tissue. Cell Reports, 26(10), pp.2720-2737.e5. doi:https://doi.org/10.1016/j.celrep.2019.02.015

Li, X., Stegen, J.C., Yu, Y. and Huang, J. (2023). Coordination and divergence in community assembly processes across co-occurring microbial groups separated by cell size. Frontiers in Microbiology, 14. doi:https://doi.org/10.3389/fmicb.2023.1166322.

Liu, H., Guo, X., Gooneratne, R., Lai, R., Zeng, C., Zhan, F. and Wang, W. (2016). The gut microbiome and degradation enzyme activity of wild freshwater fishes influenced by their trophic levels. Scientific Reports, [online] 6, p.24340. doi:https://doi.org/10.1038/srep24340

Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., Geisen, S. and Sun, B. (2020). Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. Nature Communications, 11(1). doi:https://doi.org/10.1038/s41467-020-20271-4

Marie Le Geay, Béatrice Lauga, Walcker, R. and Vincent E.J. Jassey (2024). A meta-analysis of peatland microbial diversity and function responses to climate change. Soil Biology and Biochemistry, 189, pp.109287–109287. doi:https://doi.org/10.1016/j.soilbio.2023.109287

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature, [online] 401(6756), pp.877–884. doi:https://doi.org/10.1038/44766

Paul, B. (2022). Concatenated 16S rRNA sequence analysis improves bacterial taxonomy. F1000Research, 11, p.1530. doi:https://doi.org/10.12688/f1000research.128320.1

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D. and Konstantinidis, K.T. (2014). Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. PLoS ONE, [online] 9(4). doi:https://doi.org/10.1371/journal.pone.0093827

Portillo, M.C., Leff, J.W., Lauber, C.L. and Fierer, N. (2013). Cell Size Distributions of Soil Bacterial and Archaeal Taxa. Applied and Environmental Microbiology, 79(24), pp.7610–7617. doi:https://doi.org/10.1128/aem.02710-13

Posfai, A., Thibaud Taillefumier and Wingreen, N.S. (2017). Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. Physical Review Letters, 118(2). doi:https://doi.org/10.1103/physrevlett.118.028103

Saifuddin, M., Bhatnagar, J.M., Segrè, D. and Finzi, A.C. (2019). Microbial carbon use efficiency predicted from genome-scale metabolic models. Nature Communications, 10(1). doi:https://doi.org/10.1038/s41467-019-11488-z

Schulz, H.N. and Jørgensen, B.B. (2001). Big Bacteria. Annual Review of Microbiology, 55(1), pp.105–137. doi:https://doi.org/10.1146/annurev.micro.55.1.105.

Sepulveda, J. and Moeller, A.H. (2020). The Effects of Temperature on Animal Gut Microbiomes. Frontiers in Microbiology, 11(384). doi:https://doi.org/10.3389/fmicb.2020.00384

Serbanescu, D., Nikola Ojkic and Banerjee, S. (2021). Cellular resource allocation strategies for cell size and shape control in bacteria. FEBS Journal, 289(24), pp.7891–7906. doi:https://doi.org/10.1111/febs.16234

Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., Dominguez-Bello, M.G. and Sonnenburg, J.L. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science, [online] 357(6353), pp.802–806. doi:https://doi.org/10.1126/science.aan4834

Song, S.J., Sanders, J.G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M.W., Mazel, F., Lutz, H.L., Winker, K., Graves, G.R., Humphrey, G., Gilbert, J.A., Hackett, S.J., White, K.P., Skeen, H.R., Kurtis, S.M., Withrow, J., Braile, T., Miller, M. and McCracken, K.G. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. mBio, 11(1). doi:https://doi.org/10.1128/mbio.02901-19

Vieira-Silva, S. and Rocha, E.P.C. (2010). The Systemic Imprint of Growth and Its Uses in Ecological (Meta)Genomics. PLoS Genetics, 6(1), p.e1000808. doi:https://doi.org/10.1371/journal.pgen.1000808

West, G.B. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. Science, [online] 276(5309), pp.122–126. doi:https://doi.org/10.1126/science.276.5309.122

Williams, C.E., Williams, C.L. and Logan, M.L. (2023). Climate change is not just global warming: Multidimensional impacts on animal gut microbiota. Microbial Biotechnology, [online] 16(9), pp.1736–1744. doi:https://doi.org/10.1111/1751-7915.14276

Xue, P., Budiman Minasny, Alexandre, Dobarco, M.R., McBratney, A., Bissett, A. and Caritat, P. de (2024). Drivers and human impacts on topsoil bacterial and fungal community biogeography across Australia. Global Change Biology, 30(3). doi:https://doi.org/10.1111/gcb.17216

Yang, L., Su, X., Gao, Y., Chenxiao Lv, Gao, Z., Liu, Y., Wang, Y., Li, S. and Wang, Z. (2020). The potential role of the gut microbiota in modulating renal function in experimental diabetic nephropathy murine models established in same environment. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1866(6), pp.165764–165764. doi:https://doi.org/10.1016/j.bbadis.2020.165764

Yoshiyama, K. and Klausmeier, C. (2007). Optimal Cell Size for Resource Uptake in Fluids: A New Facet of Resource Competition. The American Naturalist, 171(1), pp.59–70. doi:https://doi.org/10.1086/523950

Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R.E. and Farnleitner, A.H. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nature Communications, [online] 10(1), p.2200. doi:https://doi.org/10.1038/s41467-019-10191-3

Zhou, J. and Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiology and Molecular Biology Reviews, 81(4). doi:https://doi.org/10.1128/mmbr.00002-17