Quantifying the Prokaryotic Resource Niche

By Dhylan Patel (Faculty of Natural Science, Imperial College London), Professor Samraat Pawar (Faculty of Natural Science, Imperial College London).

# Abstract

We propose a novel method for quantifying prokaryotic ecological niches based on the concept of a resource niche: the resource dependencies of an organism for optimal growth. Through this breakthrough study, we shed light on this important concept for enhancing culture methodologies, prebiotic design, and our understanding of ecosystem responses to climate change.

Through the DMSZ MediaDive database, we collected data on 11534 prokaryote-growth medium pairings. We obtained information on growth-medium resource composition, minimum pH, and maximum pH, before collecting prokaryote specific information on genome-size, gene counts and protein-coding genes. This enabled us to perform niche width distributional analysis. We condensed the dataset to 2364 species level comparisons and correlation tested against genome size, gene counts and protein-coding genes. Finally, we performed K means clustering and tSNE dimensional reduction to visualise the full dataset’s clustering morphology.

We observed that niche widths were distributed in a left skew pattern, whereby very few species had high resource niche widths – explaining the relative rarity of generalist species in comparison to specialists and opportunists.

From our correlation analyses, we determined no significant correlation – contrary to our hypotheses - between the log scaled fundamental niche (represented by genome size, gene counts and protein-coding genes) and the log scaled resource niche width. This finding indicated that a more complex measure of the fundamental niche is required – potentially to account for the high functional redundancy of niche genes.

Finally, we found that clear clustering patterns emerged based on phylogenetic differences, although these became more prominent at lower taxonomic classifications. This indicated the importance of niche conservatism acting on smaller timescales, thus demonstrating that prokaryote ecological niches diverge at higher taxonomic ranks.

Ultimately, we found that the resource niche concept was a rigorous measure of prokaryote niche classification. To further affirm its predictive capacity, correlation analyses should be completed with a more robust measure of the fundamental niche – such as genome-scale metabolic modelling. Based on this investigation, there is scope to further examine the utility of the resource niche in predicting prokaryotic growth media for cultivation.

# Acknowledgements

## Table of Contents

* *List of Abbreviations*
* *Introduction*
* *Materials and Methods*
* *Results*
* *Discussion*
* *References*

## List of Abbreviations

t-SNE = T Distributed Stochastic Neighbourhood Embedding.

DSMZ = Deutsche Sammlung von Mikroorganismen und Zellkulturen (German collection of microbes and cell cultures GmbH)

# Introduction

#### The fundamental and realised niche

The concept of an ecological niche can be defined as an organism’s function within its habitat. This can be further understood in prokaryotes as the fundamental and realised niche (Malard and Guisan, 2023). The fundamental niche defines the range of genes a prokaryote possesses, to facilitate its potential functions, thereby delineating its intrinsic metabolic capacity. The realised niche defines the actual occupied niche of an organism in its habitat. This is typically described through metatranscriptomics; however, we propose a novel measure: resource requirements.

#### Importance of this study

This is the first time that the prokaryotic realised niche has been quantified via resource dependencies shedding new insights on the impact of the fundamental niche. Previous efforts have included diffusion mapping to span the bacterial metabolic niche (Fahimipour and Gross, 2020), however, since these maps are based on annotated genomes, there is no way to distinguish the realised niche from the fundamental niche.

Culture media have been catalogued previously into the KOMODO database which utilised media recipes in the DSMZ repository and identified microbe-strain media combinations (Oberhardt et al., 2015). However, this database is no longer maintained and also only catalogued a small number of the currently used growth media. Therefore, the DSMZ developed the more comprehensive MediaDive database (Koblitz et al., 2023) which converts the complete DSMZ collection of growth media into standardized recipes. This database contains growth media catalogued by resource composition, alongside which prokaryotes grow optimally on each recipe.

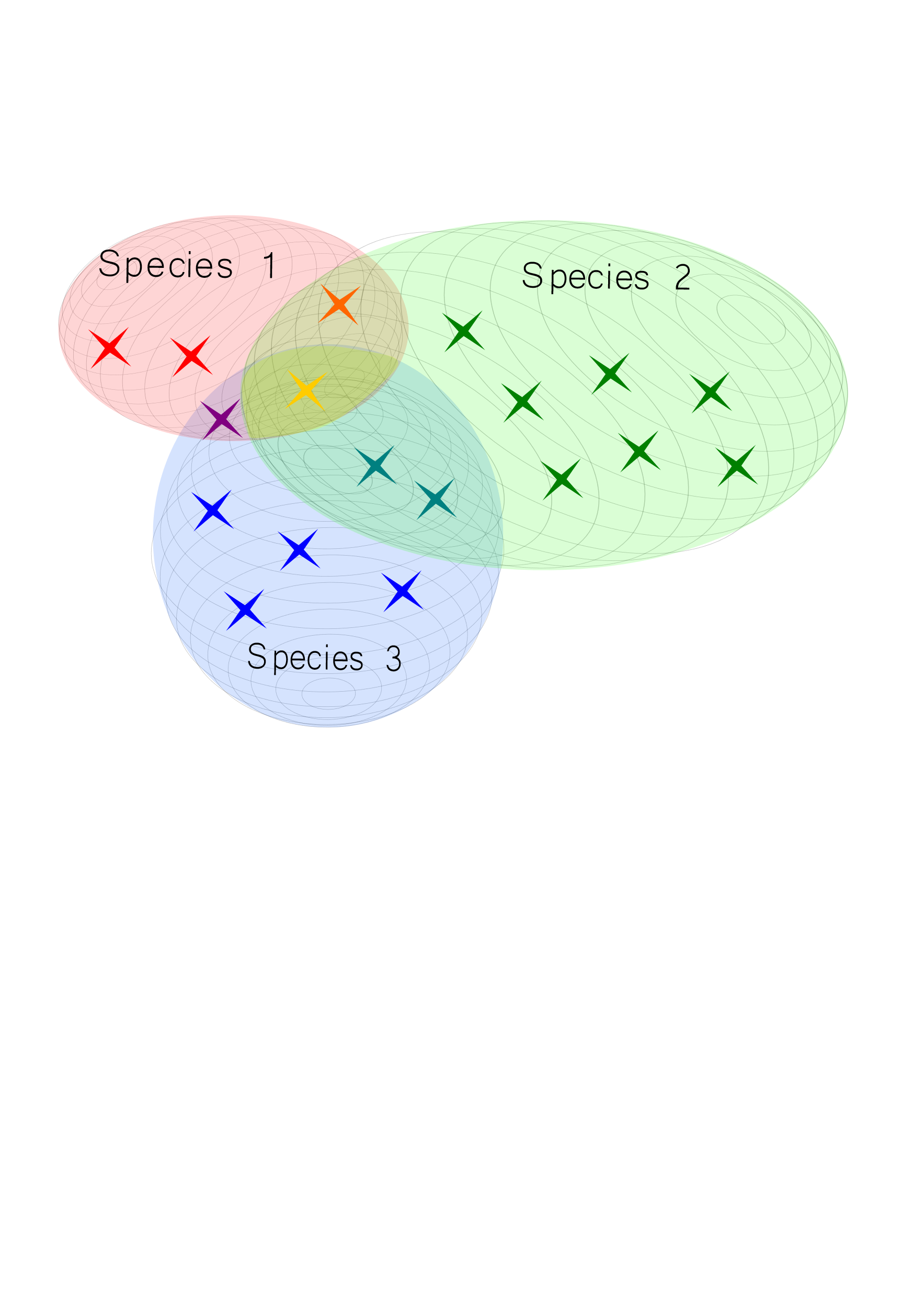
The prokaryotic resource niche has wide-ranging importance since it indicates the environmental makeup required for certain species to grow. This may have implications for the process of culturing newly discovered species which is often challenging without much time and experimentation. Cultures generally require water, mineral salts, a carbon source and a nitrogen source alongside growth factors (which the bacteria cannot synthesise itself) (Bonnet et al., 2020), however, the exact compositions are highly variable between species.

On a broader scale, there are also implications for whole ecosystem applications, since ecologists may be able to better conserve important ecologically relevant microbiomes through new prebiotic design. These supplements may adjust the resource balance in existing ecosystems to favour improved microbiome health based on our understanding of prokaryotic resource requirement niches.

Another important consideration is the relationship between broader climactic niche widths and reduced vulnerability to local extinctions (Grinder and Wiens, 2023). Since a prokaryote’s climactic niche is dictated by its fundamental niche (just as a prokaryote’s resource niche is), the distribution of niche widths may have important implications for conservation too.

#### Purpose of this investigation

This investigation aims to instigate a new measure of the prokaryotic niche – as illustrated in figure 1- understood via the resource dependencies between different species. We are seeking to observe whether the newly constructed resource niche follows expected patterns of niche width distribution and niche conservatism. Niche conservatism is defined as the retention of traits across time (Wiens et al., 2010) and it can be observed through phylogenetically similar organism’s occupying similar resource niche spaces. On top of this, we are interested in understanding whether an organism’s fundamental niche maps onto its realised resource niche. This will demonstrate whether the realised niche width can be characterized by limited genomic information.



*Figure 1: Illustration of the resource niche. The resources are marked with ‘x’ and circle area represents niche width. Niche conservatism would imply that closely related species occupy a similar geospatial area – although the niche*

1. Observe how niche width is distributed.
2. Determine whether niche width is correlated with the fundamental niche.
3. Identify whether niche conservatism creates clustering patterns.

#### Hypotheses:

1. Larger cell sizes and larger genomes scale with protein coding genes (DeLong et al., 2010) and hence more protein coding genes result in more complex gene networks. Taken together, larger genomes may occupy a broader realised resource niche. The large number of genes may provide the metabolic plasticity required for a habitat generalist to adapt to a greater variety of environmental conditions (Malard and Guisan, 2023).
2. Phylogenetic distances between co-occurring species was observed to be greatest at intermediate resource abundances and lowest at extreme resource abundances (Lin et al., 2021).

## Materials and Methods

We obtained reference Json files from the DSMZ MediaDive database detailing ingredients (catalogued by their unique identifier) and growth media recipes (catalogued with unique identifiers).

We created a web scraping script in the python environment which utilized calls to REST APIs produced by the DSMZ MediaDive. This provided information on the growth media including information on complexity, minimum pH, maximum pH, and which microbes grew optimally on the media. Growth media were standardized by filtering out recipes which were not composed of 1000ml of distilled water.

Utilizing this data, the microbes were filtered for only prokaryotes with complete information on lineage, gene counts, genome size, protein coding genes and genome GC content. To achieve this, we used the NCBI REST API v2 and produced a data frame which listed 11534 prokaryote-media pairings.

Niche width could be represented by summing up the total number of unique resources required in each growth medium across rows in the data frame. This information was plotted as a histogram to demonstrate the distribution of niche widths.

This dataset also catalogued organisms with their individual resource dependencies – allowing for correlation tests between the resource abundances in the dataset against aspects of the fundamental niche – genome size, gene counts and protein coding genes. Spearman’s Ranked Correlation Coefficients were measured and plotted to illustrate which resource abundances were most significantly correlated (both positively and negatively) with increases in prokaryote complexity.

Since certain species were observed to grow optimally on multiple growth media, there were some prokaryotes in the data frame which were over-represented – potentially by sampling biases in the MediaDive database. Therefore, to build a representative measure of niche width, the total resource requirements were taken from all the growth media attributed to an organism. These total niche widths per organism could then be correlated against the fundamental niche of each organism (determined by the genome size, gene counts and protein coding genes). These plots were based of 2364 fundamental niche-realised niche pairings.

In order to investigate niche conservatism, lineage information was utilized to determine any clustering within the prokaryote-media pairings dataset. To visualise this, we performed K-means cluster analysis to cluster data into distinct clusters, with random centroids which are adjusted iteratively. This method has been used before for phylogenetic classification of large datasets containing metagenomic information (Choudhury et al., 2023). T-SNE was utilized for dimensional reduction here, owing to its strength in handling large multidimensional data reduction in microbiome datasets (Xu et al., 2020). This clustered, dimensionally reduced dataset was then visualized utilizing the python seaborn package (Waskom, 2021) to produce cluster analysis plots. The datapoints were classified based on phylum and class level differences for the most abundant groups in the dataset.

# Results

#### Distribution of Niche Widths

As depicted in figure 2, there is a left skewed bimodal distribution of niche width, with most prokaryotes growing optimally on 4 and 10 total resources across all their known optimal growth media. The left skew indicates that prokaryotes occupying a broader niche width are far less frequent, prokaryotes with narrower niche widths are more common and most prokaryotes occupy intermediate niche widths. However, this may be influenced by sampling bias whereby certain major prokaryotes are cultured on a variety of growth media and, therefore, numerous optimal resources are known for these heavily studied species. 2) Realised Niche Width Correlated with Fundamental Niche

*Figure 2: Histogram showing the frequency distribution of log scaled prokaryotic niche widths.*



To reduce the effects of sampling bias, a global measure of niche width per prokaryote is required. Therefore, by aggregating resource requirements on a species level, we could plot relationships between the fundamental and realised niche on an individual organism level. This is demonstrated in figure 3.

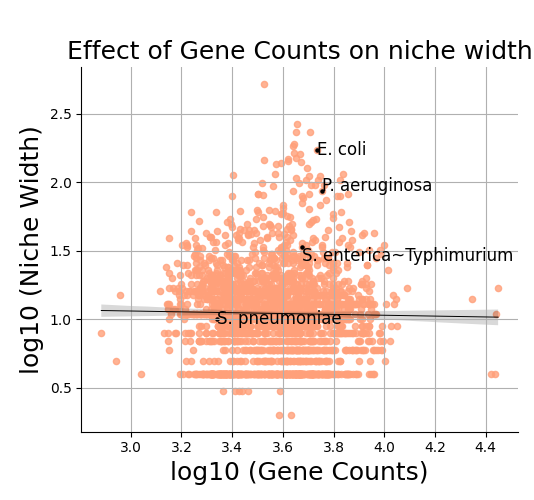
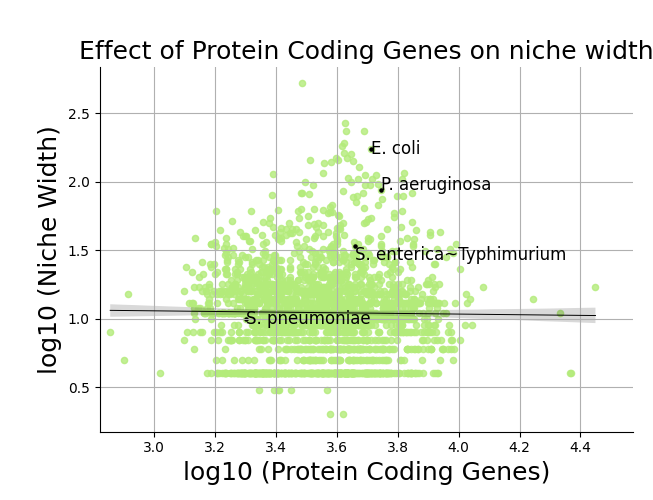
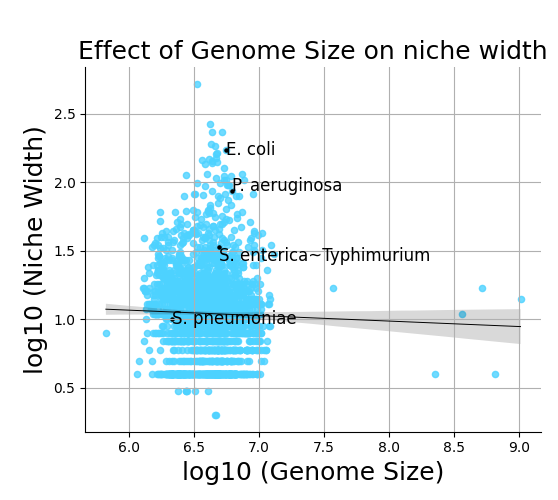
|  |  |  |
| --- | --- | --- |
| Regression | Y intercept | Slope |
| log10(Gene Counts) ~ log10 (Niche Width) | 3.58  (P<0.05) | -0.01  (P>0.05) |
| log10(Protein Coding Genes) ~ log10 (Niche Width) | 3.56  (P<0.05) | -0.01  (P>0.05) |
| log10(Genome Size) ~ log10 (Niche Width) | 6.62  (P<0.05) | -0.02  (P>0.05) |

There are no significant relationships found via linear regression (figure 3D) between log scaled fundamental niche characteristics and the realised niche to indicate a specific correlation. The realised niche width for all metrics (figure 3A, 3B and 3C) forms a slight normal distribution, with a slight left skew for log10(Genome Size) in figure 3B.

*Figure 3: Scatterplot of log scaled fundamental niche characteristics. Measuring realized niche width as number of total resources which the organism can be grow on (across all optimal media recipes).*

*Four medically relevant prokaryotes (E. coli, P. aeruginosa, S. enterica (serovar Typhimurium) and S. pneumoniae, are labelled across all scatterplots.*

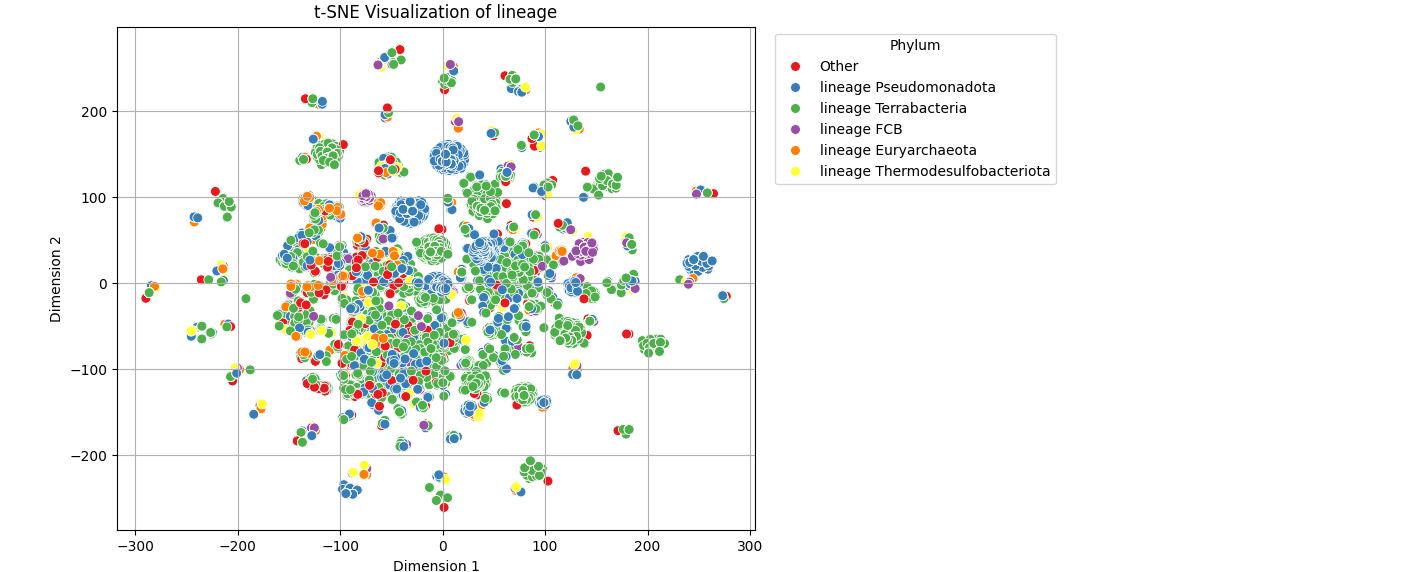
*(A) Gene Counts, (B) Genome Size, (C) Protein Coding Genes plotted against log scaled realized niche width. (D) Table of linear regression models and parameter significance.*



#### Niche Conservatism

To visualise the effects of niche conservatism, clustering patterns were visualised using t-SNE for each of the 11534 organism-nutrient media combinations (including multiple media for the same prokaryote). The dimensionally reduced data, had a trustworthiness score of 0.9915, indicating a stable dimensional reduction process. The data was then colour coded based on the major (most frequent) phyla groups for each prokaryotic lineage in the data frame. The visualisation utilized 6 cluster centroids as initial inputs to visualise the major phyla.

The Terrabacteria group are the most frequently occurring phylum in the dataset and appeared to aggregate centrally, whilst also forming smaller, tight clusters elsewhere (figure 4). For these smaller clusters, there is distinct separation, however the large central aggregation of Terrabacteria indicates that a lower taxonomic classification is required for clearer separation.



*Figure 4: tSNE plot demonstrating the clustering pattern of prokaryotes, color coded by the most frequent taxonomic phyla. Dimensional reduction was performed on a database of 11532 prokaryotes across 723 dimensions (including the grouping factor, minimum pH, and maximum pH).*

Pseudomonadota, on the other hand, formed four clearly distinct clusters (figure 4). This demonstrates that differences at a lower taxonomic classification level may characterize the formation of distinct clusters, instead of the phyla level differences.

This pattern is clearly visualised in the FCB group, whereby two clusters form in spatially distinct parts of the plot. This highlights the effectivity of resource requirements for clustering phylogenetically similar prokaryotes together, however, it also emphasizes the insufficiencies of utilizing phylum level classifications.

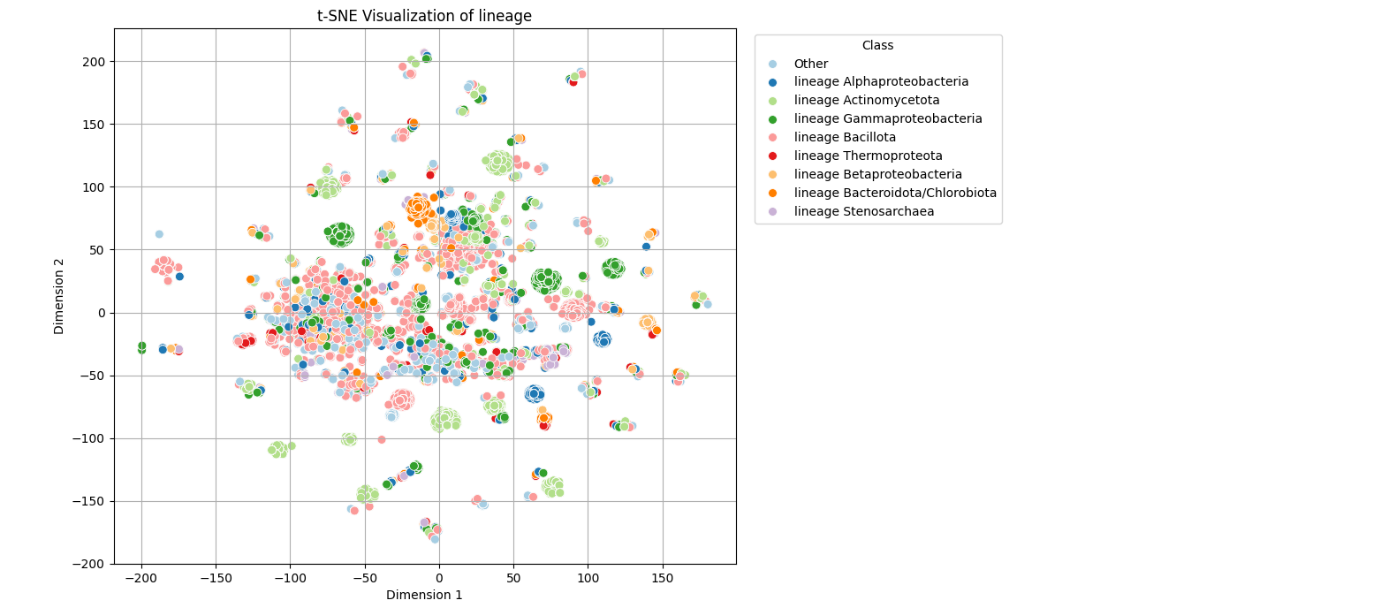
Intriguingly, there does seem to be a loose aggregation of Euryarchaeota towards the upper left quadrant of the plot with another loose aggregation of Thermodesulfobacteriota toward the lower left quadrant.

This indicates that phyla level classifications may be sufficient to loosely characterize the distributions of members of Euryarchaeota and Thermodesulfobacteriota, but not the larger phyla groups. For groups such as Terrabacteria and Pseudomonadota, are lower classification level is required to distinguish differences between clusters.

As a result of this, figure 5 was produced to highlight the clustering patterns of prokaryotes at the class level – depicting the major taxonomic classes and achieving a trustworthiness score of 0.9933 based on 9 initial cluster centroids. Whilst the classification system in figure 5 does not exactly form direct class categories for the phyla in figure 4, they do subdivide the major phyla listed.

Critically, the formerly indistinguishable Terrabacteria group becomes split between the groups of Actinomycetota and Bacillota, forming spatially distinct patterns in figure 5. Bacillota occupies a loosely clustered, central aggregate in the plot, whereas Actinomycetota forms compact, distinct clusters on the peripheries of the plot. To distinguish between these clusters, a lower taxonomic subdivision would be required. Evidently, these clusters are well-separated demonstrating that even groups within the same class may differ greatly with regards to their realised niche.

Similarly, the large spatially separated clusters of Pseudomonadota in figure 4 are subdivided amongst Alphaproteobacteria, Betaproteobacteria and Gammaproteobacteria. Similarly to Actinomycetota, all three of these classes form compact, spatially separate clusters. The FCB group from figure 4 becomes the Bacteroidota/Chlorobiota group in figure 5 which still clusters distinctly.



*Figure 5: tSNE plot demonstrating the clustering pattern of prokaryotes, color coded by the most frequent taxonomic classes. Dimensional reduction was performed on a database of 11532 prokaryotes across 723 dimensions (including the grouping factor, minimum pH and maximum pH).*

Overall, clusters within the class may be distinctly spaced from each other, emphasizing again that phylogenetic similarities may govern niche similarity. However, for complete niche differentiation a lower classification level will be required.

## Discussion

#### Niche Widths

The left skewed distribution of log scaled niche width frequency is indicative of prokaryotes most commonly occupying narrower niche widths – filling the roles of environmental specialists. By contrast, far fewer prokaryotes occupy broader niche widths and act as environmental generalists (Malard and Guisan, 2023). Whilst there may be more specialists than generalists in our findings, it is evident that most samples fell in between these two categories as opportunists. Our findings support previous classifications of specialists, opportunists and generalists in cross-study bacterial analyses (Xu et al., 2021) where 27% and 9% of species studied were classed as specialists and generalists respectively, with the remainder being opportunists.

#### Fundamental and Realised Niche Relationship

Contrary to our hypotheses, there was no significant relationship between the fundamental niche and the realised resource niche from our findings. This may be explained by the high degree of functional redundancy in niche genes compared to the functional redundancy of niche proteins (Wang et al., 2024). Therefore, although genomes, gene counts and protein coding genes determine which niche proteins are produced, functional redundancy in niche genes can override their importance. Ultimately, it is the less redundant, niche proteins, which may delineate the fundamental niche best. A better measure of the fundamental ecological niche may potentially have involved using genome-scale metabolic reconstructions. Genome scale metabolic modelling was able to compute the metabolic niche distances between organisms and correlate them with COBRA-based distances (Heirendt et al., 2019; Régimbeau et al., 2022). These models have also been used to accurately generate phylogenetic trees (Schulz and Almaas, 2020)- making them suitable for niche conservatism analyses.

#### Niche Conservatism

In the context of resource niches, niche conservatism is the increased similarity between resource requirements of more closely related organisms than distantly related organisms. This is depicted visually in our tSNE analysis producing figures 4 and 5, where it is evident that clustering patterns emerge due to phylogenetic similarity. However, the level of phylogenetic classification required to spatially separate resource-sharing clusters is wide-ranging. As visualised in figure 4, the highly diverse phylum of the Terrabacteria group cannot be separated into its own distinct space-occupying resource niche. However, by splitting it into the classes of Actinomycetota and Bacillota, spatially distinct groups begin to emerge. There is further evidence of spatial distinction through members of the Pseudomonadota phylum forming more distinct clusters in figure 5. This pattern of increased spatial separation when the taxonomic classifier is reduced closer to the species level, indicates the presence of niche conservatism acting.

Evidently, the time scale on which niche conservatism acts is far less than the phylum or class level, since taxa within these groups still form spatially distinct groupings. Instead, it seems that niche conservatism plays an important role in shaping more recent evolutionary behaviour. Lower classifications are required to group individual clusters by resource dependencies.

This has important implications for our understanding of culture methodologies. Since resource dependencies are most similar at lower taxonomic levels, we cannot expect to find suitable growth media by merely observing members of the same phylum or class, instead we must seek the lowest taxonomic level for comparisons. With the speed and precision offered by next-generation sequencing, newly discovered organisms may be cultured at an accelerated rate by comparisons to the nearest phylogenetic relative of known resource dependencies.

#### Overall Utility of the Prokaryotic Resource Niche

The resource niche definition created by our study has indeed displayed utility as a viable measure of prokaryotic niche space – thus validating our web scraping, culture-based methodology. As demonstrated, the resource niche has distributional behaviour affirming our current understanding of niche distributions on the generalist-specialist spectrum. Further to this, it also displays the foundations of niche conservatism shown through the clustering patterns of figures 4 and 5.

To gain further insight into the factors which determine and shape the prokaryotic resource niche, a more comprehensive measure of the fundamental niche is required. Our approach, which used measurements of intrinsic cellular characteristics, was insufficient at predicting significant correlations with resource niche width. There is scope to expand our framework using genome-scale metabolic models as representations of fundamental niche space which may then be correlated with prokaryote resource requirements.

## References

Bonnet, M., Lagier, J.C., Raoult, D., Khelaifia, S., 2020. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. New Microbes New Infect. 34, 100622. https://doi.org/10.1016/j.nmni.2019.100622

Choudhury, N., Sahu, T.K., Rao, A.R., Rout, A.K., Behera, B.K., 2023. An Improved Machine Learning-Based Approach to Assess the Microbial Diversity in Major North Indian River Ecosystems. Genes 14, 1082. https://doi.org/10.3390/genes14051082

DeLong, J.P., Okie, J.G., Moses, M.E., Sibly, R.M., Brown, J.H., 2010. Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. Proc. Natl. Acad. Sci. U. S. A. 107, 12941–12945. https://doi.org/10.1073/pnas.1007783107

Fahimipour, A.K., Gross, T., 2020. Mapping the bacterial metabolic niche space. Nat. Commun. 11, 4887. https://doi.org/10.1038/s41467-020-18695-z

Grinder, R.M., Wiens, J.J., 2023. Niche width predicts extinction from climate change and vulnerability of tropical species. Glob. Change Biol. 29, 618–630. https://doi.org/10.1111/gcb.16486

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Žagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L.V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Guebila, M.B., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H.S., Bryant, W.A., Artacho, F.J.A., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B.Ø., Thiele, I., Fleming, R.M.T., 2019. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. Nat. Protoc. 14, 639–702. https://doi.org/10.1038/s41596-018-0098-2

Koblitz, J., Halama, P., Spring, S., Thiel, V., Baschien, C., Hahnke, R.L., Pester, M., Overmann, J., Reimer, L.C., 2023. MediaDive: the expert-curated cultivation media database. Nucleic Acids Res. 51, D1531–D1538. https://doi.org/10.1093/nar/gkac803

Lin, Q., Li, L., Adams, J.M., Heděnec, P., Tu, B., Li, C., Li, T., Li, X., 2021. Nutrient resource availability mediates niche differentiation and temporal co-occurrence of soil bacterial communities. Appl. Soil Ecol. 163, 103965. https://doi.org/10.1016/j.apsoil.2021.103965

Malard, L.A., Guisan, A., 2023. Into the microbial niche. Trends Ecol. Evol. 38, 936–945. https://doi.org/10.1016/j.tree.2023.04.015

Oberhardt, M.A., Zarecki, R., Gronow, S., Lang, E., Klenk, H.-P., Gophna, U., Ruppin, E., 2015. Harnessing the landscape of microbial culture media to predict new organism–media pairings. Nat. Commun. 6, 8493. https://doi.org/10.1038/ncomms9493

Régimbeau, A., Budinich, M., Larhlimi, A., Pierella Karlusich, J.J., Aumont, O., Memery, L., Bowler, C., Eveillard, D., 2022. Contribution of genome‐scale metabolic modelling to niche theory. Ecol. Lett. 25, 1352–1364. https://doi.org/10.1111/ele.13954

Schulz, C., Almaas, E., 2020. Genome-scale reconstructions to assess metabolic phylogeny and organism clustering. PloS One 15, e0240953. https://doi.org/10.1371/journal.pone.0240953

Wang, T., Li, L., Figeys, D., Liu, Y.-Y., 2024. Pairing metagenomics and Metaproteomics to characterize ecological niches and metabolic essentiality of gut microbiomes. ISME Commun. ycae063. https://doi.org/10.1093/ismeco/ycae063

Waskom, M.L., 2021. seaborn: statistical data visualization. J. Open Source Softw. 6, 3021. https://doi.org/10.21105/joss.03021

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Jonathan Davies, T., Grytnes, J.-A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M., Stephens, P.R., 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecol. Lett. 13, 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

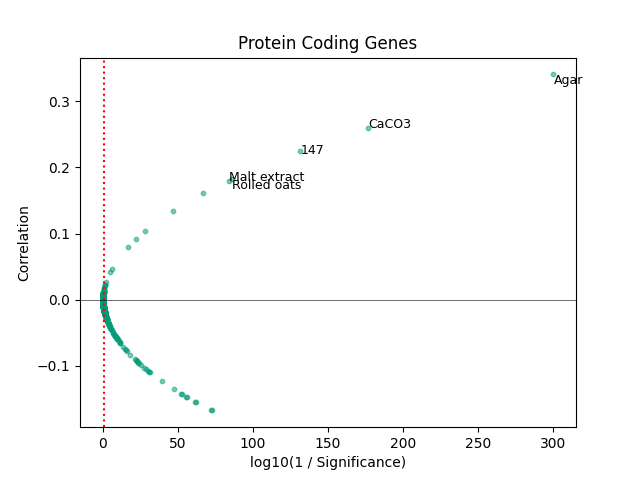
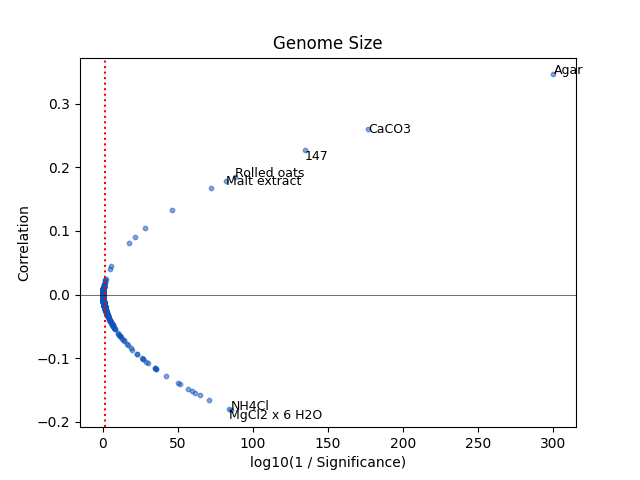
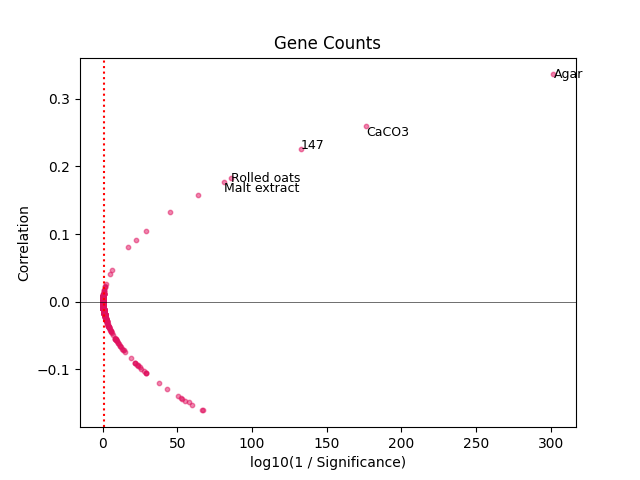
Xu, Q., Vandenkoornhuyse, P., Li, L., Guo, J., Zhu, C., Guo, S., Ling, N., Shen, Q., 2021. Microbial generalists and specialists differently contribute to the community diversity in farmland soils. J. Adv. Res. 40, 17–27. https://doi.org/10.1016/j.jare.2021.12.003

Xu, Xueli, Xie, Z., Yang, Z., Li, D., Xu, Ximing, 2020. A t-SNE Based Classification Approach to Compositional Microbiome Data. Front. Genet. 11, 620143. https://doi.org/10.3389/fgene.2020.620143

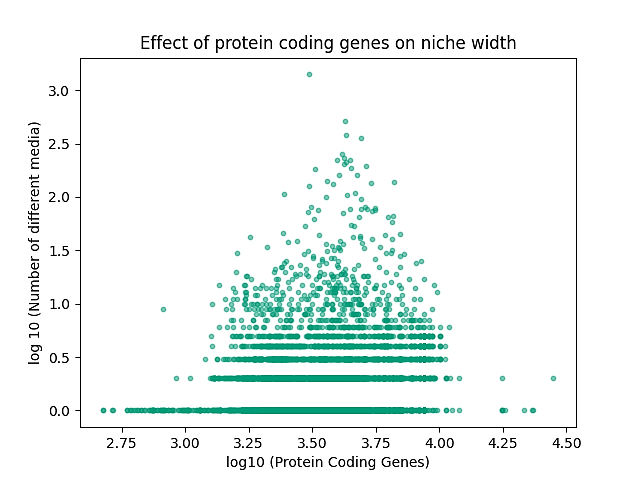
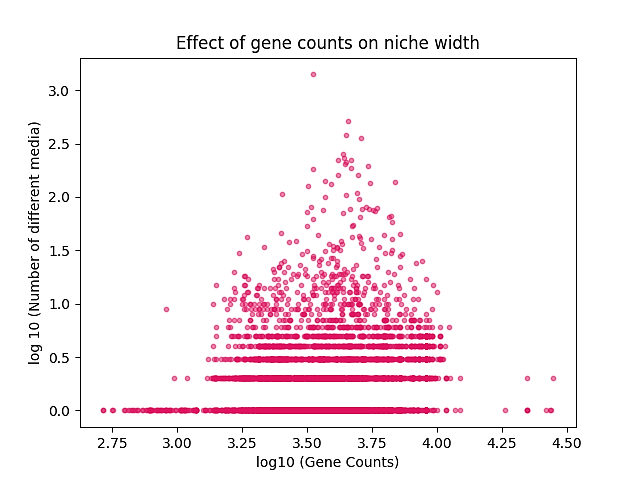
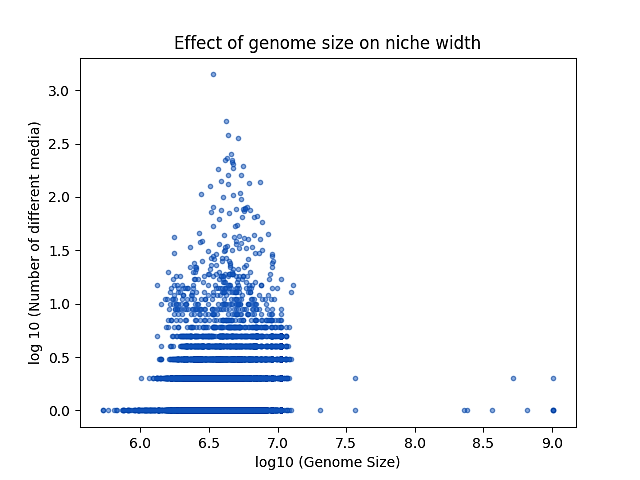
### 

# Appendix

# 



*Supplementary Figure 1: Scatterplots depicting the ingredients which displayed the most significant correlations between abundance and factors of the fundamental niche (gene counts, protein coding genes and genome size). The red dashed line indicates the 5% significance level. The most significant positively correlated ingredients were CaCO3, rolled oats, malt extract, agar and an unnamed ingredient. The ingredients which had the most negative correlations were NH4Cl and MgCl2 x 6H2O.*



*Supplementary Figure 2: Scatterplots measuring log scaled niche width (measuring niche width as number of total growth media per organism) across the log scaled characteristics of the fundamental niche.*