Quantifying the prokaryotic ecological niche using computational methods

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

(5) **Introduction:** Must be **accessible to the non-specialist** and normally explains the ‘big picture’ before providing an in-depth, critical assessment of the relevant research literature that will identify the gaps in knowledge and identify the questions to be answered or hypotheses to be tested. Figures (preferably self-made) should be included. These may include ones used on the poster.

(6) **Materials and** **Methods**: should describe fully the materials and methods used (including databases where appropriate) at a level that will allow the work to be repeated. No need to describe commonly used methods; just provide relevant references. Do not include results in this section.

(7) **Results:** should **provide a clear narrative** of the relevant experiments and analyses and will include relevant data to support the conclusions drawn. Do not show all your experimental data, just the relevant data. All figures should have legend and be fully annotated. Where appropriate, suitable statistical analyses should be performed to confirm significance.

(8) **Discussion**: will relate the results to current ideas in the literature and suggest future areas of study

* 6,000 word limit (examiners reward conciseness)

This excludes references and figures/figure legends but not in-text citations.

## Abstract

## Acknowledgements

## Table of Contents

List of Abbreviations

Introduction

Materials and Methods

Results

Discussion

Methods

## List of Abbreviations

OTU = Operational Taxonomic Unit

NTI = Nearest Taxon Index. Large positive values indicate phylogenetic clustering. Large negative values indicate phylogenetic overdispersion.

## Introduction

#### What is 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. 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.

#### How can the fundamental niche be quantified?

The fundamental niche represents an organisms’ unique genetic repertoire which thus define its metabolic capabilities. These features may be studied by examining protein coding genes, overall gene counts and genome size. These intrinsic characteristics provide a simple measure for organism complexity and therefore metabolic capacity. A more detailed and representative procedure, utilises diffusion maps which creates a coordinate system to span the bacterial metabolic niche (Fahimipour and Gross, 2020) where presence or absence of metabolic traits are inferred from annotated genomes.

#### How can the realised niche be quantified?

The realised niche defines the actual occupied niche of a prokaryote – which includes metabolising environmental resources for maintenance processes and biomass synthesis. Maintenance processes may include extracellular enzyme production (for resource degradation), repair, maintaining osmotic balance, signalling and motility (Malik et al., 2020). Prokaryotes differ in their maintenance requirements and therefore require different quantities and types of environmental resources. To investigate which resources are required, experimental culturing approaches may highlight which growth media specific prokaryotes are best cultured on.

#### How can we gather information on optimal growth media?

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, because the database is no longer maintained and catalogued only a small number of the currently used growth media, 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.

#### What do we know already about the metabolic fundamental niche?

The metabolic niche space can be visualised as filamentous with branches from a common origin ending in clusters (Fahimipour and Gross, 2020). This demonstrates that there are large areas of niche space which are unoccupied - either due to being suboptimal and redundant, or yet to be identified. Through 2D mapping of diffusion variables it is also evident that there is a general correlation with metabolic clusters and taxonomic class, indicating that phylogenetic similarities correlate roughly to metabolic similarities. It would therefore be expected that more closely related prokaryotes would have closely overlapping resource niches.

#### What do we know already about resource requirements defining niches?

The effects of resource abundances on microbial communities are documented through studying community structure and niche differentiation. Niche differentiation is the process by which organisms occupy different niches to subdivide resources: organisms which share trait niches will undergo competitive exclusion. The degree to which niche differentiation occurs can be measured by species dissimilarity in a community.

No correlation was determined between nutrient availability and niche differentiation (measured by NTI) (Lin et al., 2021). This could be due to environmental selection (whereby shared environmental preferences cause species co-occurrence) playing a greater role than taxa competitions (which causes exclusion).

For gradients of increasing resource abundance, it was observed that community network complexity decreased (Lin et al., 2021) due to strong environmental selection at high resource abundances. However, as resource abundance increased, gene network complexity increased. This may be due to more resources facilitating more biomass production requiring more diverse genes and pathways.

However, phylogenetic distance between co-occurring species was observed to be greatest at intermediate resource abundances and lowest at extreme resource abundances (Lin et al., 2021).

Overall, resource abundance does not affect niche differentiation (due to environmental selection) but may decrease community complexity and increase genetic network complexity (Lin et al., 2021). Ultimately, these effects are influenced by interspecific interactions when observed on a community scale.

On an individual species level, it would be expected that fewer prokaryotes are adapted to optimal growth at high resource abundances and that they would be associated with a higher genetic repertoire in their fundamental niche.

#### What are the goals of this investigation?

1. Determine any links between the prokaryotic fundamental niche (complexity) and the realised niche.
2. Determine whether phylogenetic distance between prokaryotes:
   1. Effects niche breadth
   2. Varies according to resource abundances.
3. Determine whether life history strategies play a role in categorizing prokaryotic niche space.

#### What are the hypotheses for these investigation objectives?

-------------------------------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------------------------

## Materials and Methods

Accessed the DSMZ database to obtain two reference json files: 1) data on the ingredients listed in the database, catalogued with a unique identifier and 2) media listed in the database catalogued with their own unique identifier. All API calls, data handling, analysis and visualization were performed in python.

Used REST API calls to the DSMZ database to access recipe information to understand the exact makeup of each growth medium as well as the strains which grow optimally upon it.

Cleaned the data to create a new data framework where microbes were catalogued alongside fields containing information about which growth media they grew optimally on.

Utilized REST API calls to the NCBI database to access information on each taxon to filter for prokaryotes only – as per the scope of this investigation. Taxa which were missing information on genome size, GC content, protein coding genes and gene counts were excluded from the resultant dataset.

Produced scatterplots of the three intrinsic characteristics for organism complexity: protein coding genes, gene counts and genome size with respect to niche width. Niche width was represented by the number of growth media the organism could be cultured optimally on.

Constructed a new dataset which leveraged the DSMZ ingredients json file to detail the media recipes by ingredient ratios and therefore quantify the resource dependencies for each taxa.

Produced a correlation analysis script which determined the Spearman’s ranked correlation coefficient for each resource (ingredient) by determining the relationship between resource quantity and intrinsic characteristic.

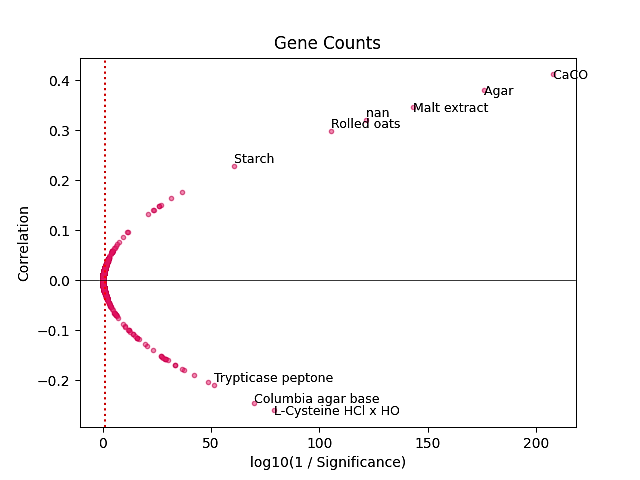
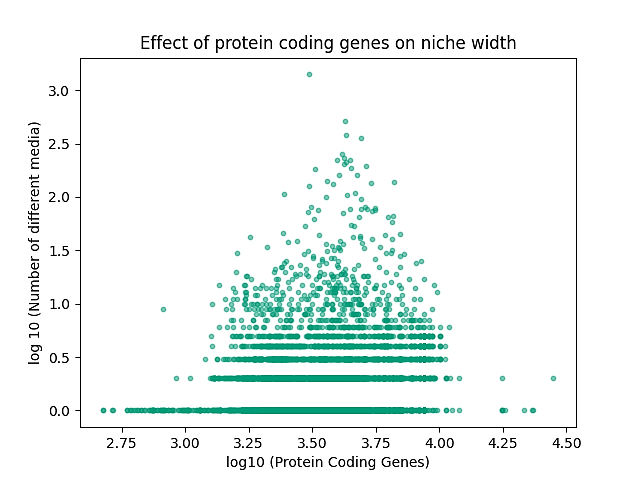
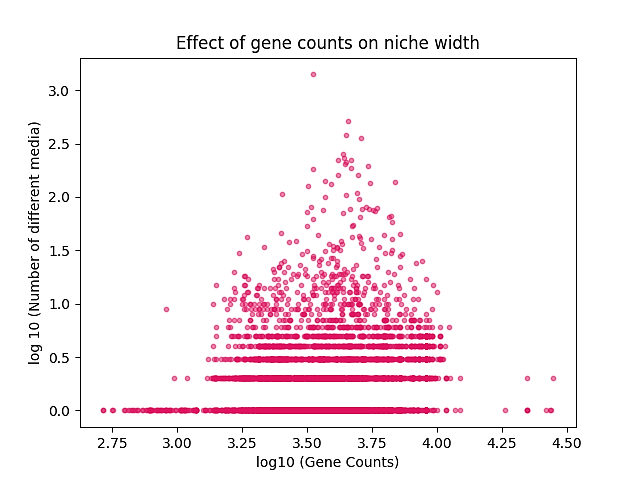
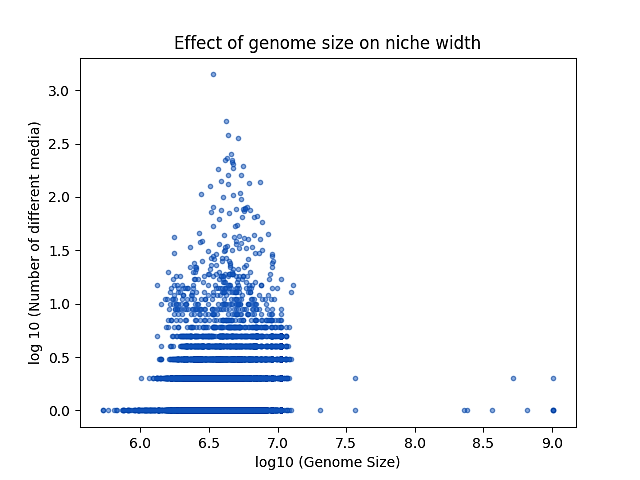
Scatterplots depicting the significance of these relationships were produced to show which ingredients were significantly correlated with organism complexity.

Caveats:

Some prokaryotes were excluded from the data-frame being analysed since the NCBI database lacked sufficient information to cover all fields required. This may make the data less representative of all cultured organisms in the DSMZ catalogue.

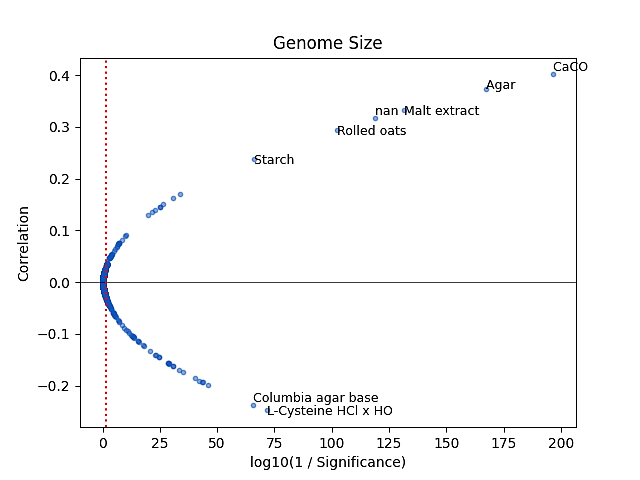
There are certain outlier results which have been cultured across many more media than others (which usually range from 0-25 media). This may be a result of sampling bias in the DSMZ catalogue.

## Results



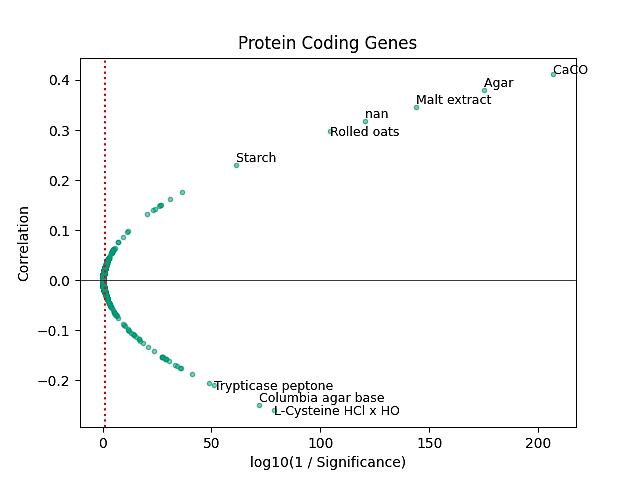
3

2



3

2



3

2

As demonstrated in fig 1, niche width is normally distributed for ranges in organism complexity

Notably, the ingredients which displayed the most significant positive correlations between abundance and organism complexity were CaCO3, starch, rolled oats, malt extract, agar and an unnamed ingredient. The ingredients which had the most negative correlations for abundance with organism complexity were trypticase peptone, Columbia agar base and L-cysteine HCL x H2O.

## Discussion

The normal distribution of niche width for prokaryotes with organism complexity is unexpected. We expected more complex organisms to contain a larger enzymatic repertoire to thus allow the metabolism of a wider range of growth resources. The result observed may occur due to the log scaling transformation which distinguishes lower values for niche width into more distantly spaced discrete values. Alternatively, the normal distribution may form because there is a constraint imposed on more complex organisms. One probability is that more complex organisms require more resources to replicate their larger genomes and , therefore, there is only a limited subset of suitable resources which can support this increased demand for phosphorus and nitrogen.

## References

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

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

Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., Allison, S.D., 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. ISME J. 14, 1–9. https://doi.org/10.1038/s41396-019-0510-0

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

### Appendix