Genomic, metabolic and literature oriented annotation of microbial co-occurrence networks enhances associations confidence level and hypothesis generation

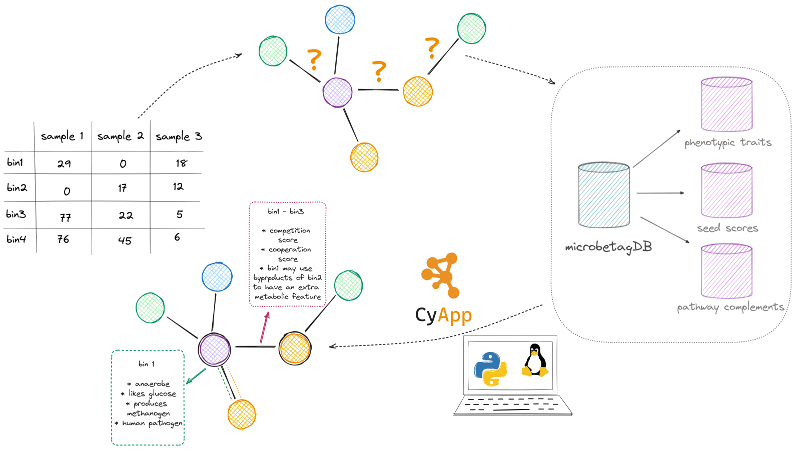
Haris Zafeiropoulos[1], Ermis Ioannis Michail Delopoulos[1], Andi Erega[2], Annelies Geirnaert[2], John Morris[3], Karoline Faust[1*]

[1*]Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, KU Leuven, Herestraat, Leuven, 3000, , Belgium.
[2]Institute of Food, Nutrition and Health, ETH Zurich, Street, Zurich, 8092, , Switzerland.
[3]Department of Pharmaceutical Chemistry, University of California San Francisco, Street, San Francisco, 94143, California, USA.

*Corresponding author(s). E-mail(s): karoline.faust@kuleuven.be;
Contributing authors: haris.zafeiropoulos@kuleuven.be;
ermisioannis.michaildelopoulos@student.kuleuven.be;
andi.erega@hest.ethz.ch; annelies.geirnaert@hest.ethz.ch;
scooter@cgl.ucsf.edu;

Abstract

Up to 350 words.
The abstract must include the following separate sections:
Background: the context and purpose of the study
Results: the main findings
Conclusions: a brief summary and potential implications



	sample 1	sample 2	sample 3
bin1	29	0	18
bin2	0	17	12
bin3	77	22	5
bin4	76	45	6

Figure abstract. From a count table (bins or OTUs/ASVs) one can come up with a co-occurrence network. To better understand and assess the confidence level of these associations, microbetag annotates both taxa (nodes) and
integrating genomic data

Keywords: microbial associations, enrichemnt analysis, data integration, pathway complementarity, seed set

1 Introduction

[1] [2]

A widely used approach is the creation of co-occurrence networks based on community data. To build such networks, there is a great number of approaches: Spearman and Pearson correlations, CoNet [1] SparCC [2] SpiecEasi [3], MAGMA [4] and FlashWeave [5] are just a few of them. However, the outcome is usually tool-dependent [6, 7].

microbeAnnotator [8]

Related literature: Karaoz U and Brodie EL (2022) microTrai [9] , a computational pipeline that infers and distills ecologically relevant traits from microbial genome sequences. It does not apply networks

2 Implementation

[3]

[1] We are to submit in the Microbiome journal as a ”Software” manuscript, thus we follow these rules

[2] The introduction should not include subheadings.

[3] This should include a description of the overall architecture of the software implementation, along with details of any critical issues and how they were addressed.

Genomes included

Using the GTDB v202 metadata files, we retrieved the NCBI genome accessions of the representative genomes of high quality, i.e. completeness $\geq 95\%$ and contamination $\leq 5\%$. That resulted a set of $26,778$ covering $22,009$ unique NCBI Taxonomy Ids. Using these accession numbers, we were able to download their corresponding `.faa` files when available (`get_gtdb_faa.py`) leading to a set of $16,900$ amino acid sequence files.

Taxonomy schemes

microbetag maps the taxonomy of each entry in the abundance table to its corresponding NCBI Taxonomy id and if available its closest GTDB representative genome(s). Two well established taxonomy schemes are supported. The Genome Taxonomy DataBase (GTDB) [10] that is being broadly used in bins and/or MAGs taxonomical classification and the Silva database [11] that has NCBI Taxonomy [12]. The primer links the representative genomes included to their corresponding NCBI Taxonomy ids too.

There is a great number of taxonomies that are being used in such studies, e.g. Silva [11], Ribosomal Database Project (RDP) [13], manually curated ones and more, As a consequence, there is not a standardised format of the taxonomies assigned, from bioinformatics pipelines used for the analysis of such data. microbetag makes use of the `fuzzywuzzy` library that implements the Levenshtein Distance Metric to get the closest NCBI taxon name and thus its corresponding NCBI Taxonomy id. ++ ncbi nodes dump A relatively high similarity score is used (90) to avoid false positives.

DADA2 formatted 16S rRNA gene sequences for both bacteria and archaea [14] were used to trained the TAXID classifier [15] of the DECIPHER package.

Network inference

FlashWeave [5]

a computational approach based on a flexible Probabilistic Graphical Model framework that integrates metadata and predicts direct microbial interactions from heterogeneous microbial abundance data sets with hundreds of thousands of samples.

A flexible Probabilistic Graphical Model framework is used in a computational approach that incorporates metadata and predicts direct microbial interactions. This is done using heterogeneous microbial abundance datasets consisting of hundreds of thousands of samples.

Literature oriented node annotation

Using a set of Tara Oceans samples [16] FAPROTAX [17] estimates the functional potential of the bacterial and archaeal communities, by classifying each taxonomic unit into functional group(s) based on current literature, announcements of cultured representatives and/or manuals of systematic microbiology. In this manually curated approach, a taxon is associated with a function if and only if all the cultured species within the taxon have been shown to exhibit that function. In its current version, FAPROTAX includes more than 80 functions based on 7600 functional annotations and covering more than 4600 taxa. Contrary to gene content based approaches, e.g. PICRUSt2, FAPROTAX estimates metabolic phenotypes based on experimental evidence.

microbetag invokes the accompanying script of FAPROTAX and converts the taxonomic microbial community profile of the samples included in the user's abundance table or of the taxa present in the provided network, into putative functional profiles. Then, it parses FAPROTAX's subtables to annotate each taxonomic unit present on the user's data with all the functions for which they had a hit. FAPROTAX annotations are not part of the microbetagDB but are computed on the fly.

Genomic oriented node annotation

phenDB [18] is a publicly available resource that supports the analysis of bacterial (meta)genomes to identify 47 distinct functional traits. It relies on support vector machines (SVM) trained with manually curated datasets based on gene presence/absence patterns for trait prediction. More specifically, the model for a particular trait is trained using a collection of EggNOG annotated genomes where the knowledge of whether that trait is present or absent among its members is available. The `compute-genotype` program of phenotrex supports the creation of such tabular *genotype* files. A *genotype* file can be used along with a *phenotype* one, i.e., a file containing true phenotypic trait values for each input genome on which to train the model, and the `train` program of phenotrex can then be performed. Last, the models can now be used to predict their corresponding traits; based on the completeness/contamination of the genomes, the accuracy varies.

In the frameowrk of microbetagDB, phenotrex classifiers were re-trained using the genomes provided by phenDB for each trait to sync with the latest version of eggNOG. Genomes were downloaded from NCBI using the Batch Entrez program. Then, *genotype* files were produced for all the high quality GTDB representative genomes. Each model was then used against all the GTDB *genotype* files to annotate each with the presence or the absence of the trait.

Pathway complementarity

For the subset of the 16,900 high quality GTDB representative genomes that a `.faa` was available, *kofamscan* [19] was performed to annotate them with KEGG ORTHOLOGY terms (KOs) [20]. Their KOs were then mapped to their corresponding KEGG modules. A KEGG module is defined as a functional unit within the KEGG framework, that represents a set of enzymes and reactions involved in a specific biological process or pathway [21]. A module's definition is a logical expression and consists of KOs and the following symbols: a. the space, representing a connection in the pathway b. plus sign, representing a molecular complex, c. comma, representing alternatives and d. minus sign, designates an optional item in the complex. Both (a) and (b) cases should be considered as ”AND” logical operators, while (c) would be the ”OR”. For example, the definition of the D-Galacturonate degradation in Bacteria (M00631) is:

K01812 K00041 (K01685,K16849+K16850) K00874 (K01625,K17463) that once breaking down, it leads to 4 alternative sets of KOs (pathways):

K01812 K00041 K01685 K00874 K01625

K01812 K00041 K16849+K16850 K00874 K01625

K01812 K00041 K01685 K00874 K17463

K01812 K00041 K16849+K16850 K00874 K17463

We define a genome as having a ”complete” module if and only if all of the KOs present in any of the module's alternatives are also found among the annotated KOs of the genome. All modules definitions were retrieved using the KEGG API and parsed (parse_module_definitions.py). A dictionary was built with all the alternatives, i.e. alternative sets of KOs, for a module to be complete (module_definition_map.json). Each pair of the KEGG annotated genomes was then investigated for potential pathway complementarities, i.e. whether a genome lacking a number of KOs ($genome_A$) to have a complete module ($module_x$) could benefit from another's species genome(s) ($genome_B$). In that case, $genome_B$ does not necessarily have a complete alternative of $module_x$; as long as it has the missing KOs that $genome_A$ needs to complete an alternative of it, $genome_B$ potentially complements $genome_A$ with respect to $module_x$. In total, $341,568$ unique complementarities were exported (`pathway_complementarity.py`). Thanks to the graphical user interface (GUI) of the KEGG pathway map viewer, each complementarity can be visualised as part of the closest KEGG metabolic map; where the KOs coming from the donor are shown with a blue-green colour, while those from the beneficiary's genome itself with rose.

As several GTDB representative genomes might map to the same NCBI Taxonomy Id, all the possible genomes' combinations are annotated in the edge of a pair of species level taxonomically annotated OTUs/ASVs/bins. On top of that, as co-occurrence networks are undirected, both nodes of a suggested association are considered as potential donors and beneficiary species.

Seed scores using genome scale metabolic reconstructions

A metabolic network's ”seed set” is the set of compounds that, based on the network topology, need to be acquired exogenously [22]. Such nodes might be independent, i.e. they cannot be activated by any other node in the network, or they can be interdependent forming groups of seed nodes. In that case, once a seed is assured, it activates all the rest of that group. Therefore, a confidence level (C) ranging from 0 to 1, has been previously described to quantify the relevance of each seed:

$$C_i = 1/seed's\ group\ with\ i\ size \qquad (1)$$

$C = 0$ corresponds to a non-seed node, while $C = 1$ represents an independent node.

Based on the seed concept, several graph theory-based metrics have been described to predict species interactions directly from their networks' topologies. The Metabolic

Complementarity Index ($MI_{Complementarity}$) measures the degree to which two microbial species can mutually assist each other by complementing each other's biosynthetic capabilities. As described in [23], it is defined as the proportion of seed compounds of a species that can be synthesized by the metabolic network of another, but are not included in the seed set of the latter.

$$MI_{Complementarity} = \frac{|SeedSet_A \cap \neg SeedSet_B|}{|SeedSet_A \cap (SeedSet_B \cup \neg SeedSet_B)|} \tag{2}$$

$MI_{Complementarity}$ offers an upper bound assessment of the potential for syntrophic interactions between two species.

Further, the Metabolic Competition Index ($MI_{Competition}$) represents the similarity in two species' nutritional profiles. This index establishes an upper limit on the level of competition that one species may face from another. As also described in [23], it is calculated as the proportion of compounds in a species' seed set that coincide with those in an other's, while also factoring in the confidence scores associated with seed compounds.

$$MI_{Competition} = \frac{\sum C(SeedSet_A \cap SeedSet_B)}{\sum C(SeedSet_A)} \tag{3}$$

Those indices have been thoroughly described and implemented in the NetCooperate [24] and NetCompt [25] tools correspondingly. We will be referring to those two indices as "seed scores". Most recently, the PhyloMint Python package [23] was released supporting the calculation of the seed scores of genome scale metabolic network reconstructions (GENREs) in SBML format.

In the framework of microbetag, seed scores were computed using PhyloMint and draft GENREs for all pair-wised combinations of GTDB representative genomes that have been RAST annotated in the framework of the PATRIC database [26]. GENREs were reconstructed using the Model SEED pipeline [27] through its Python interface ModelSEEDpy.

Clustering network

manta is a heuristic network clustering algorithm that clusters nodes within weighted networks effectively, leveraging the presence of negative edges and discerning between weak and microbetag invokes manta [28] to infer clusters from the microbial network. A taxonomically-informed layout is

strong cluster assignments. ++ taxonomy layout

Groups of annotations

Biologically meaningful groups were built using the micrO ontology [29].

Building the CytoscapeApp

The microbetag CytoscapeApp was build based on the source code of the scVizNet [30]. Java @Ermis to add

Enrichment analysis is supported. Hypergeometric distribution FDR +++

Dependencies, Web server and API

The microbetag web service is container - based and consists of three Docker [31] (v24.0.2) images: a. the MySQL database b. an nginx [32] web server and c. the app itself. The latter uses Gunicorn (20.1.0) to build an application server which communicates with the web server using the Web Server Gateway Interface (WSGI) protocol and handles incoming HTTP requests. microbetag is implemented as a Flask application (v2.3.2); Flask is a micro web framework for developing Python web applications and RESTful APIs. A thorough description of microbetag's API is available at the ReadTheDocs web site. The source code of the microbetag web service is available on GitHub.

python 3.11 slim docker image julia 1.7.1 for flashweave mysql.connector 8.0.27 python library pandas 2.1.1. numpy 1.26.0 multiprocessing

text processing using awk

KEGG API

2.1 Running large datasets

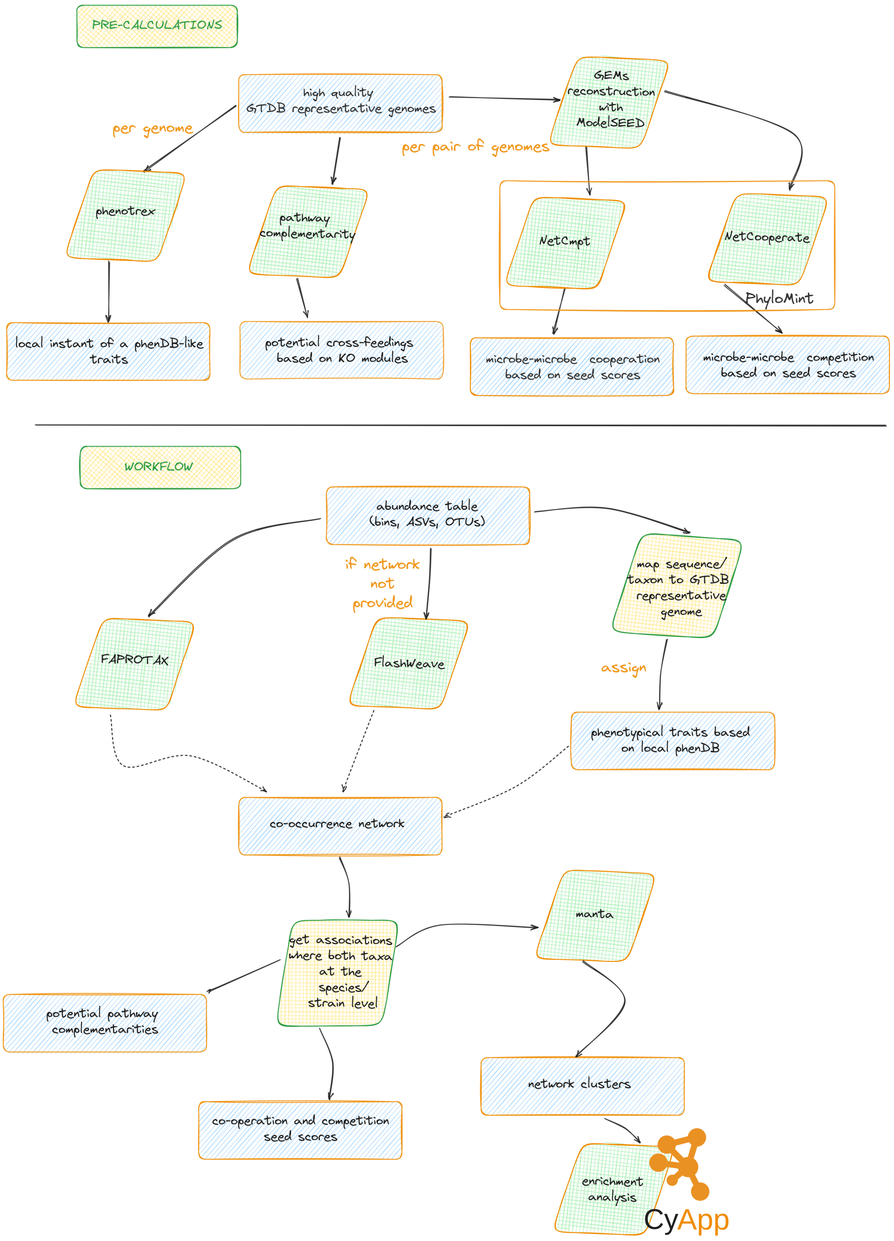
3 Results

[4]

[4] Significant advance over previously published software (usually demonstrated by direct comparison with available related software) This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures. This section may be combined with the Discussion section for Software articles.

microbetag and microbetagDB

Fig. 1: Diagram of the microbetag pre - calculations and the on the fly workflow. 33 phenotypic traits were predicted for a total of 34,607 high quality GTDB genomes using `phenotrex`



PRE-CALCULATIONS
high quality GTDB representative genomes
per genome
phenotrex
pathway complementarity
GEMs reconstruction with ModelSEED
per pair of genomes
NetCmpt
NetCooperate
PhyloMint
local instant of a phenDB-like traits
potential cross-feedings based on KO modules
microbe-microbe cooperation based on seed scores
microbe-microbe competition based on seed scores
WORKFLOW
abundance table (bins, ASVs, OTUs)
FAPROTAX
if network not provided
FlashWeave
map sequence/ taxon to GTDB representative genome
assign
phenotypical traits based on local phenDB
co-occurrence network
get associations where both taxa at the species/ strain level
manta
potential pathway complementarities
co-operation and competition seed scores
network clusters
enrichment analysis
CyApp

microbetag in numbers: 34,608 GTDB representative genomes 32 phen-model-oriented metabolic functions 92 FAPROTAX functions 341,568 unique complements involved in $>$ 184 million beneficiary - donor pairs' complementarities 30,755 GENREs leading to 1 billion competition and complementarity scores

annotated network returned in `.cyjs` format

For a computationally efficient way to annotate large networks, a Docker image is provided so the user runs a taxonomy assignment using the IDTAXA algorithm [15] of the DECIPHER R package [33]. A co-occurrence network is also built using FlashWeave [5], as microbetag also does.

microbetag CytoscapeApp

Overall comment, the CytoscapeApp returns averages and s.d. for example in seed scores. If you want the exact values, go through the API.

Validation of microbetag potential

vitamin dataset [34]

Interpetating a real-world network with microbetag

Annelies' dataset.

4 Discussion

[5]

Discussions should be brief and focused. In some disciplines use of Discussion or 'Conclusion' is interchangeable. It is not mandatory to use both. Some journals prefer a section 'Results and Discussion' followed by a section 'Conclusion'. Please refer to Journal-level guidance for any specific requirements.

The user interface should be described and a discussion of the intended uses of the software, and the benefits that are envisioned, should be included, together with data on how its performance and functionality compare with, and improve, on functionally similar existing software. A case study of the use of the software may be presented. The planned future development of new features, if any, should be mentioned.

5 Conclusion

[6]

Conclusions may be used to restate your hypothesis or research question, restate your major findings, explain the relevance and the added value of your work,

[5] The user interface should be described and a discussion of the intended uses of the software, and the benefits that are envisioned, should be included, together with data on how its performance and functionality compare with, and improve, on functionally similar existing software. A case study of the use of the software may be presented. The planned future development of new features, if any, should be mentioned.

[6] This should state clearly the main conclusions and provide an explanation of the importance and relevance of the case, data, opinion, database or software reported.

highlight any limitations of your study, describe future directions for research and recommendations.

In some disciplines use of Discussion or 'Conclusion' is interchangeable. It is not mandatory to use both. Please refer to Journal-level guidance for any specific requirements.

Supplementary information. If your article has accompanying supplementary file/s please state so here.

Authors reporting data from electrophoretic gels and blots should supply the full unprocessed scans for key as part of their Supplementary information. This may be requested by the editorial team/s if it is missing.

Please refer to Journal-level guidance for any specific requirements.

Acknowledgments. [1]

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 was initiated thanks to an EMBO Scientific Exchange Grant to HZ. It was then supported by the 3D'omics Horizon project (101000309). We would also like to thank the National Resource for Network Biology (NRNB) and the Google Summer of Code 2023 for the support of EIMD.
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors' contributions

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Editorial Policies for:

Springer journals and proceedings: https://www.springer.com/gp/editorial-policies

Nature Portfolio journals: https://www.nature.com/nature-research/editorial-policies

Scientific Reports: https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals: https://www.biomedcentral.com/getpublished/editorial-policies

[1] Acknowledgments are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged. Please refer to Journal-level guidance for any specific requirements.

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

References

[1] Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., Huttenhower, C.: Microbial co-occurrence relationships in the human microbiome. PLoS computational biology **8**(7), 1002606 (2012)

[2] Friedman, J., Alm, E.J.: Inferring correlation networks from genomic survey data. PLoS computational biology **8**(9), 1002687 (2012)

[3] Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A.: Sparse and compositionally robust inference of microbial ecological networks. PLoS computational biology **11**(5), 1004226 (2015)

[4] Cougoul, A., Bailly, X., Wit, E.C.: Magma: inference of sparse microbial association networks. BioRxiv, 538579 (2019)

[5] Tackmann, J., Rodrigues, J.F.M., Mering, C.: Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. Cell systems **9**(3), 286–296 (2019)

[6] Kishore, D., Birzu, G., Hu, Z., DeLisi, C., Korolev, K.S., Segrè, D.: Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation. Msystems, 00961–22 (2023)

[7] Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., *et al.*: Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME journal **10**(7), 1669–1681 (2016)

[8] Ruiz-Perez, C.A., Conrad, R.E., Konstantinidis, K.T.: Microbeannotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. BMC bioinformatics **22**, 1–16 (2021)

[9] Karaoz, U., Brodie, E.L.: microtrait: a toolset for a trait-based representation of microbial genomes. Frontiers in Bioinformatics **2**, 918853 (2022)

[10] Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P.: Gtdb: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic acids research **50**(D1), 785–794 (2022)

[11] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O.: The silva ribosomal rna gene database project: improved data processing and web-based tools. Nucleic acids research **41**(D1), 590–596 (2012)

[12] Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., *et al.*: Ncbi taxonomy: a comprehensive update on curation, resources and tools. Database **2020**, 062 (2020)

[13] Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M.: Ribosomal database project: data and tools for high throughput rrna analysis. Nucleic acids research **42**(D1), 633–642 (2014)

[14] Alishum, A.: DADA2 Formatted 16S rRNA Gene Sequences for Both Bacteria & Archaea. https://doi.org/10.5281/zenodo.6655692 . https://doi.org/10.5281/zenodo.6655692

[15] Murali, A., Bhargava, A., Wright, E.S.: Idtaxa: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome **6**(1), 1–14 (2018)

[16] Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., *et al.*: Structure and function of the global ocean microbiome. Science **348**(6237), 1261359 (2015)

[17] Louca, S., Parfrey, L.W., Doebeli, M.: Decoupling function and taxonomy in the global ocean microbiome. Science **353**(6305), 1272–1277 (2016)

[18] Feldbauer, R., Schulz, F., Horn, M., Rattei, T.: Prediction of microbial phenotypes based on comparative genomics. BMC bioinformatics **16**(14), 1–8 (2015)

[19] Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., Ogata, H.: Kofamkoala: Kegg ortholog assignment based on profile hmm and adaptive score threshold. Bioinformatics **36**(7), 2251–2252 (2020)

[20] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M.: Kegg for integration and interpretation of large-scale molecular data sets. Nucleic acids research **40**(D1), 109–114 (2012)

[21] Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., Kanehisa, M.: Modular architecture of metabolic pathways revealed by conserved sequences of reactions. Journal of chemical information and modeling **53**(3), 613–622 (2013)

[22] Borenstein, E., Kupiec, M., Feldman, M.W., Ruppin, E.: Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proceedings of the National Academy of Sciences **105**(38), 14482–14487 (2008)

[23] Lam, T.J., Stamboulian, M., Han, W., Ye, Y.: Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. PLoS computational biology **16**(10), 1007951 (2020)

[24] Levy, R., Carr, R., Kreimer, A., Freilich, S., Borenstein, E.: Netcooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. BMC bioinformatics **16**(1), 1–6 (2015)

[25] Kreimer, A., Doron-Faigenboim, A., Borenstein, E., Freilich, S.: Netcmpt: a network-based tool for calculating the metabolic competition between bacterial species. Bioinformatics **28**(16), 2195–2197 (2012)

[26] Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., *et al.*: Improvements to patric, the all-bacterial bioinformatics database and analysis resource center. Nucleic acids research **45**(D1), 535–542 (2017)

[27] Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L.: High-throughput generation, optimization and analysis of genome-scale metabolic models. Nature biotechnology **28**(9), 977–982 (2010)

[28] Röttjers, L., Faust, K.: Manta: A clustering algorithm for weighted ecological networks. Msystems **5**(1), 10–1128 (2020)

[29] Blank, C.E., Cui, H., Moore, L.R., Walls, R.L.: Micro: an ontology of phenotypic and metabolic characters, assays, and culture media found in prokaryotic taxonomic descriptions. Journal of biomedical semantics **7**(1), 1–10 (2016)

[30] Choudhary, K., Meng, E.C., Diaz-Mejia, J.J., Bader, G.D., Pico, A.R., Morris, J.H.: scnetviz: from single cells to networks using cytoscape. F1000Research **10** (2021)

[31] Merkel, D., *et al.*: Docker: lightweight linux containers for consistent development and deployment. Linux j **239**(2), 2 (2014)

[32] Reese, W.: Nginx: The high-performance web server and reverse proxy. Linux J. **2008**(173) (2008)

[33] Wright, E.S.: Using decipher v2. 0 to analyze big biological sequence data in r. R Journal **8**(1) (2016)

[34] Hessler, T., Huddy, R.J., Sachdeva, R., Lei, S., Harrison, S.T., Diamond, S., Banfield, J.F.: Vitamin interdependencies predicted by metagenomics-informed network analyses and validated in microbial community microcosms. Nature Communications **14**(1), 4768 (2023)