

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

Haris Zafeiropoulos¹, Ermis Ioannis Michail Delopoulos¹,
Andi Erega², Annelies Geirnaert², John Morris³, 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

*Looks like Chris Quince is [our editor](#).

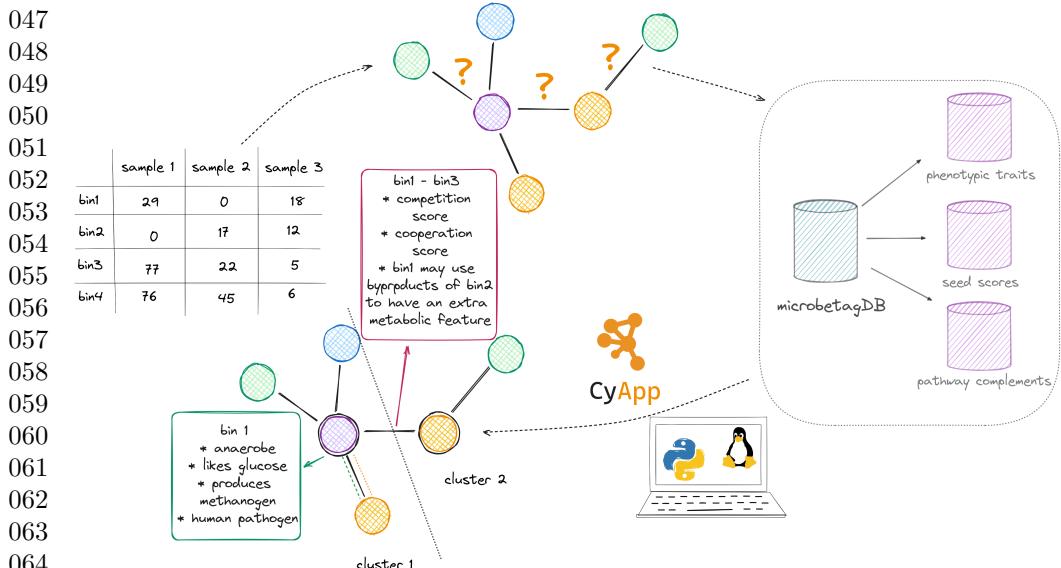


Figure abstract.

065
066
067
068 **Keywords:** microbial associations, enrichment analysis, data integration, pathway
069 complementarity, seed set

070 071 072 073 **Background** ¹ ²

074 Microbial ecology plays a fundamental role in the stability and resilience of ecosystems
075 and their processes; from soils, aquatic environments and biogeochemical cycles [1] to
076 host-associated environments and the human health [2, 3]. Most microbial species live
077 only in communities [4] and most natural microbial communities consist of hundreds
078 or even thousands of species [5]. Each species exhibits a unique repertoire of metabo-
079 lites and showcases adaptability across various metabolic niches, each with specific
080 nutrient and environmental requirements. To unravel microbial ecology involves dis-
081 entangling the principles that dictate the organization of a community, encompassing
082 both its composition and metabolic activities. This, in turn, entails understanding the
083 dynamics governing interactions among microbial species and their relationships with
084 the surrounding environment [6].

085 In recent years, there has been a compelling suggestion proposing an elegant
086 paradigm for microbial ecology: the correlation between community functional and
087 taxonomic composition hinges on the relative importance of metabolic niche effects

088
089 ¹We are to submit in the Microbiome journal as a "Software" manuscript, thus we follow [these rules](#)

090
091 ²The introduction should not include subheadings. The Background section should explain the relevant
092 context and the specific issue that the software described is intended to address.

relative to the processes inducing variability within functional groups [7]. Comparable environments should foster similar microbial community functions, even though there may be taxonomic variations within individual functional groups, while in more heterogeneous environments functional β diversity would be strongly correlated with taxonomic β diversity. In the latter, the decoupling between community composition and metabolic functioning is concealed by robust metabolic niche effect [7]. Thus, an interaction between two taxa may vary in different environments. Microbial interactions can be the result of multiple phenomena, such as exchange of metabolic products [8], biofilm formation [9], gene transferring [10] and signaling [11].

High-throughput sequencing (HTS) has provided great insight into the diversity and composition of microbial communities [12]. Uncultivated species can now be detected and their features can be inferred through their genomic information [13]. Moreover, the composition of thousands of microbiome samples is now accessible allowing for the inference of patterns among sets of samples. A widely used approach to extract such patterns, is the creation of co-occurrence networks based on metagenomic read data (amplicon and/or shotgun) [14]. A great number of approaches are available for co-occurrence network inference based on a range of statistical concepts such as: correlation (e.g., CoNet [15], SparCC [16]), linear regression (e.g., SpeicEasi [17]) and causal inference (FlashWeave [18]). Nevertheless, microbial co-occurrence networks continue to encounter various challenges [19], encompassing issues associated with data analysis and network construction, leading to tool-dependent analysis [4, 20, 21]. But also, challenges regarding the interpretation of the networks. Addressing the well-articulated question of *What can we learn from the hairballs* posed by Röttjers et al. [4] could provide essential insight on the mechanisms of the interactions.

The use of microbial network inference as a means for predicting interactions has underscored its limited accuracy, and the fact that the biological implications of network properties remain unclear [22]. Theoretical principles derived from network studies might provide indications of emergent biological characteristics [4, 23]. For example, modules (highly interconnected nodes) within microbial co-occurrence networks could serve as indicators of ecological processes that govern community structure, including niche filtering and habitat preference[24]. Data integration and clustering have been suggested to address this challenge [19]. Clusters identified in microbial association networks have demonstrated their ability to mirror key drivers of community composition [25] and several algorithms and implementations are available [26]. However, data integration approaches in microbial co-occurrence networks are so-far limited. Here, we present **microbetag**, a microbial co-occurrence network annotator that exploits several channels of information to enhance/diminish the confidence of the associations suggested on the network and generate hypotheses for further investigation both at the paired-interaction and the community level

microbetag serves as a comprehensive platform that consolidates information on taxa along with their potential metabolic interactions from multiple channels (see Implementation 3). Key concept on the presented approach is the exploitation of the *reverse ecology* approach [27]. Reverse ecology leverages genomics to explore community ecology with no *a priori* assumptions about the taxa involved. Making the most of advancements in systems biology and genomic metabolic modeling, as well as

139 system-level analysis of intricate biological networks, the reverse ecology framework
140 enables the prediction of ecological traits for less-understood microorganisms, their
141 interactions with others, and the overall ecology of microbial communities [28]. In
142 this context, seed set analysis has been a major contribution in the study of both the
143 species and the community ecology based on their genetic information.

144 A metabolic network's "seed set" is the set of compounds that, based on the net-
145 work topology, need to be acquired exogenously [29] (see Figure 1). Such nodes might
146 be independent, i.e. they cannot be activated by any other node in the network, or they
147 can be interdependent forming groups of seed nodes. Seeds have been proven them-
148 selves a successful proxy for the habitat of the organism and an essential tool in the
149 frameowrk of reverse ecology [29, 30]. Based on the seed concept, several graph theory-
150 based metrics (indices) have been described to predict species interactions directly
151 from their networks' topologies [31–34]. Over the last years, the seed approach has
152 been implemented at the Genome-scale metabolic network reconstructions (GENREs)
153 level. GENREs encapsulate mathematical representations capturing the biochemical
154 reactions that could take place within an organism [35–37].

155 Metabolic complementarity among species, serving as a reflection of potential coop-
156 eration within communities, assesses the capacity for collaboration; cross-feeding or
157 syntrophy interactions are typical examples of such a collaboration. Contary, metabolic
158 competition refers to the metabolic overlap between two species leading to exploitative
159 competition, e.g. for nutrient resources. Seed and non-seed sets can be used to com-
160 pute such indices. Thorough examination of such complements can reveal metabolic
161 interactions leading to patterns observed on the co-occurrence network.

162 Considering complementarity as a range of alternatives and paired-wise microbial
163 interactions in the context of the community as a whole, microbial species may also
164 exchange metabolic compounds that may be not seeds at a certain time but may allow
165 them to perform functions that currently are not capable of [38, 39]. Such by-products
166 may be even metabolites not even necessary for themeselves but for the community as
167 a whole [40]. To explore the potenial of a species metabolism given they benefit from
168 a partner of theirs, genome annotations combined with collections of functional units
169 to highlight can provide a valid proxy. We present here a naive approach exporting all
170 possible complements between a pair of species based on their KEGG ORTHOLOGY
171 (KOs) annotations and the KEGG MODULES database [41].

172 **microbetag** integrates user's co-occurrence network integrating phenotypic traits
173 on the taxa present on the network (nodes) and potential metabolic interactions to
174 their suggested associations (edges). A Graphical User Interface (GUI) is supported
175 as a CytoscapeApp providing a user-friendly environment to investigate annotations
176 in a straightforward way. All annotations present on microbetagDB are also available
177 though an Application Programming Interface (API). **microbetag**'s source code is
178 under a GNU GPL v3 license and available on GitHub. Documentation and further
179 support on how to use **microbetag** is available at [documentation web-site](#). To the best
180 of our knowledge there is not a software with which **microbetag** could be compared
181 with directly. To validate our annotations we used a recently published network with
182 partially known interactions between some pairs of species found associated [42] (see
183 Results section, paragraph 3). To demonstrate **microbetag**'s potential, we present
184

the main features of its interface and we discuss a real-world use-case (see Discussion section, paragraph 3). 185
186
187

Implementation³

Genomes included

Using the Genome Taxonomy Database (GTDB) v207 [metadata files](#), we retrieved the NCBI genome accessions of the high quality representative genomes, i.e. completeness $\geq 95\%$ and contamination $\leq 5\%$. A set of 26,778 genomes was obtained, representing 22,009 unique NCBI Taxonomy Ids. Using these accession numbers, we were able to download their corresponding .faa files when available leading to a set of 16,900 amino acid sequence files. The latter were annotated and used to obtain potential pathway complementarities between pairs of genomes (see paragraph 3). Last, when available, their corresponding annotations on PATRIC database [43] were retrieved to reconstruct GENREs (see paragraph 3).

Taxonomy schemes

`microbetag` maps the taxonomy of each entry in the abundance table to their corresponding NCBI Taxonomy Id and, if available, their closest GTDB representative genome(s), since several GTDB representative genomes may map to the same NCBI Taxonomy Id. Two well established taxonomy schemes are supported: the GTDB [44] that is being broadly used for bins and/or MAGs taxonomical classification and the Silva database [45] that is widely used in amplicon studies. Both taxonomy schemes link their taxonomies to NCBI Taxonomy Ids [46]. In case none of those two taxonomies was used and the abundance table contains less than 1,000 taxa, `microbetag` maps the user provided taxonomies to NCBI Taxonomy. To this end, `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; a relatively high similarity score is used (90) to avoid false positives. Also, using the nodes dump file of NCBI Taxonomy, `microbetag` may retrieve the children taxa of a taxon in user's data, along with their corresponding NCBI Taxonomy Ids, if requested by the user. If the user provides their abundance table with taxonomies already mapped to the GTDB taxonomy, `microbetag` will report the best possible annotations in a time efficient manner.

Network inference

When a co-occurrence network is not provided by the user, `microbetag` exploits FlashWeave [18] to build one on the fly. Yet, `microbetag` supports the annotation of networks built from any algorithm/software, in any format Cytoscape can load.

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

231 **microbetag pre-processing**

232 In order to aid the user to map their sequences to the GTDB taxonomy, DADA2-
233 formatted 16S rRNA gene sequences for both bacteria and archaea [47] were used to
234 trained the TAXID classifier of the DECIPHER package [48] and are available through
235 the [microbetag preprocess Docker image](#). Likewise, when the abundance table consists
236 of more than 1,000 taxa, providing a network as an input is mandatory. Again, in order
237 to facilitate the user, [microbetag](#) preprocess Docker image supports the inference of
238 a network using FlashWeave.
239

240 **Literature based nodes annotation**

241 Using a set of Tara Oceans samples [49] FAPROTAX [50] estimates the functional
242 potential of the bacterial and archaeal communities, by classifying each taxonomic
243 unit into functional group(s) based on current literature, announcements of cultured
244 representatives and/or manuals of systematic microbiology. In this manually curated
245 approach, a taxon is associated with a function if and only if all the cultured species
246 within the taxon have been shown to exhibit that function. In its current version,
247 FAPROTAX includes more than 80 functions based on 7600 functional annotations
248 and covering more than 4600 taxa. Contrary to gene content based approaches, e.g.
249 PICRUSt2 [51], FAPROTAX estimates metabolic phenotypes based on experimental
250 evidence.
251

252 [microbetag](#) invokes the accompanying script of FAPROTAX and converts the
253 taxonomic microbial community profile of the samples included in the user's abun-
254 dance table or of the taxa present in the provided network, into putative functional
255 profiles. Then, it parses FAPROTAX's subtables to annotate each taxonomic unit
256 present on the user's data with all the functions for which they had a hit. FAPROTAX
257 annotations are not part of the microbetagDB but are computed on the fly.
258

259 **Genomic based nodes annotation**

260 phenDB [52] is a publicly available resource that supports the analysis of bacterial
261 (meta)genomes to identify 47 distinct functional traits, e.g. whether a species is pro-
262 ducing butanol or it has an halophilic lifestyle. It relies on support vector machines
263 (SVM) trained with manually curated datasets based on gene presence/absence pat-
264 terns for trait prediction. More specifically, the model for a particular trait is trained
265 using a collection of EggNOG annotated genomes where the knowledge of whether
266 that trait is present or absent among its members is available. These models (classi-
267 fiers) are used to predict presence/absence of their corresponding traits in non-studied
268 species.
269

270 In the frameowrk of microbetagDB, classifiers were re-trained using the genomes
271 provided by phenDB for each trait to sync with the latest version of eggNOG [53]
272 and the phenotrex [52] software tool. Genomes were downloaded from NCBI using
273 the [Batch Entrez](#) program. Then, *genotype* files were produced for all the high quality
274 GTDB representative genomes. Each model was then used against all the GTDB
275 *genotype* files to annotate each with the presence or the absence of the trait. A list of all
276

the phenotypic traits available for the genomes present on microbetagDB is available
on [microbetag](#)'s documentation site. The updated models are also available

Pathway complementarity

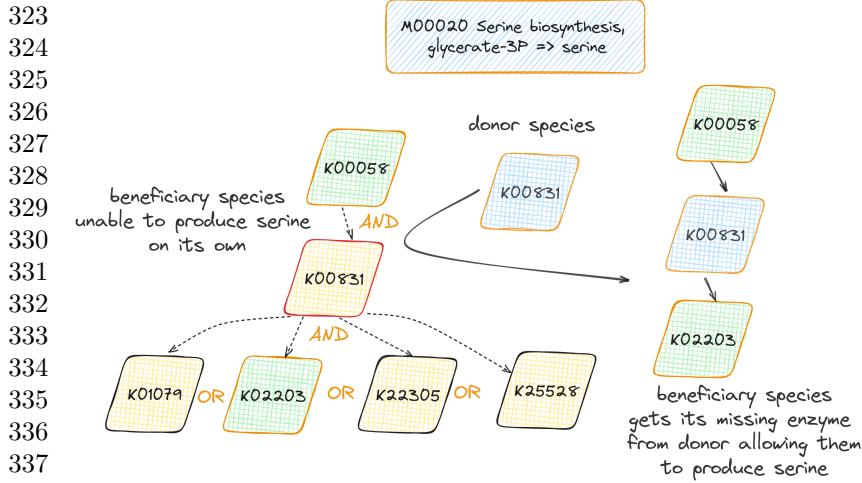
To infer potential pathway complementarities we consider the modules described in KEGG MODULES database [41]. A KEGG module is defined as a functional unit within the KEGG framework, that represents a set of enzymes and reactions involved in a specific biological process or pathway [54]. Such a unit consists of several *steps*, each of which may have more than one molecular ways to occur (Figure 1). A module's definition is a logical expression and consists of KOs that may be coupled with one another as: a. connected steps of the pathway b. parts of a molecular complex, c. alternatives of the same step, and d. optional entities of a complex. Both (a) and (b) cases should be considered as the AND logical operator, while (c) would be the OR (Figure 1). Given a module's definition, we will consider as an *alternative* any subset of the KO terms mentioned in the definition, that has exactly one way to perform each step, provided that all the steps of the module are covered. We define a genome as having a *complete* module, if and only if all of the KOs of at least one alternative are present on the genome.

Within this framework, `kofamscan` [55] was used to annotate with KEGG ORTHOLOGY terms (KOs) the 16,900 high quality GTDB representative genomes for which a `.faa` was available [56]. The KOs of each genome were then mapped to their corresponding KEGG modules; a KO may map to more than one modules (1 : n).

All modules definitions were retrieved using the KEGG API and parsed. A dictionary was built with all the alternatives of each module. 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.

Thanks to the graphical user interface (GUI) of the [KEGG pathway map viewer](#) [57, 58], 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.

`microbetag` annotates the edges of a co-occurrence network by isolating pairs where both taxa map to an annotated genome present on microbetagDB. Since co-occurrence networks are undirected, both nodes of a suggested association are considered as potential donors and beneficiary species. When more than one GTDB representative genomes map to the same NCBI Taxonomy Id all the possible genomes' combinations are considered. Finally, two edges are added in such pairs of taxa in the annotated network: one considering $species_A$ as the potential beneficiary and $species_B$ as the potential donor species, and one vice-versa.



323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338 **Fig. 1:** Pathway complementarity approach. The high quality GTDB genomes were
339 annotated with KEGG ORTHOLOGY (KO) terms. The various ways of getting a
340 KEGG module complete were enumerated and all the possible ways a donor species
341 could "fill" a beneficiary's non-complete module were calculated. In this case, there
342 are 4 unique ways for having the serine biosynthesis module complete; in all of them
343 K00831 is required. However, it is missing from the beneficiary species that supports
344 the 2 out of the 3 steps of the module's definition. A donor species having and poten-
345 tially sharing the corresponding enzyme of K00831 may enable the beneficiary species
346 to produce serine.
347

348 349 Seed scores using genome scale metabolic reconstructions

350 The Metabolic Complementarity Index ($MI_{Complementarity}$) measures the degree to
351 which two microbial species can mutually assist each other by complementing each
352 other's biosynthetic capabilities. As described in [59], it is defined as the proportion
353 of seed compounds of a species that can be synthesized by the metabolic network of
354 another, but are not included in the seed set of the latter. $MI_{Complementarity}$ offers
355 an upper bound assessment of the potential for syntrophic interactions between two
356 species. Further, the Metabolic Competition Index ($MI_{Competition}$) represents the sim-
357 ilarity in two species' nutritional profiles. This index establishes an upper limit on the
358 level of competition that one species may face from another. Those indices have been
359 thoroughly described and implemented in the NetCooperate [31] and NetCompt [32]
360 tools correspondingly. We will be referring to those two indices as "seed scores".
361

362 Most recently, the PhyloMinttool [59] was released supporting the calculation of
363 the seed scores of GENREs in SBML format.
364

In the framework of microbetag, seed scores were computed using PhyloMint and
365 draft GENREs for all pair-wised combinations of GTDB representative genomes that
366 have been RAST annotated in the framework of the PATRIC database [43]. GENREs
367 were reconstructed using the Model SEED pipeline [60] through its Python interface
368 [ModelSEEDPy](#).

Moreover, the computed seed and the non-seed (i.e., set of metabolic compounds a genome can build on its own) sets of each genome were used to get their overlap among all the pair-wised combinations of those genomes. More specifically, the overlap of <i>seed set_{speciesA}</i> with the <i>non seed set_{speciesB}</i> was retrieved. microbetag then annotates again the edges of the co-occurrence network where both taxa have been mapped to a at least one GTDB genome, mentioning all the KEGG maps for which there is at least one seed compound of the potentially beneficiary species	369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
Clustering network	377
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 [26] to infer clusters from the microbial network. A taxonomically-informed layout is	378 379 380 381 382 383 384 385 386 387
strong cluster assignments. ++ taxonomy layout	383
Groups of annotations	384
Biologically meaningful groups were built using the micrO ontology [61].	385 386 387
Building the CytoscapeApp	388
The microbetag CytoscapeApp was build based on the source code of the scVizNet [62]. Java @Ermis to add	389 390 391
Enrichment analysis is supported. Hypergeometric distribution FDR +++	392 393
Dependencies, Web server and API	394
The microbetag web service is container - based and consists of three Docker [63] (v24.0.2) images: a. the MySQL database b. an nginx [64] 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 .	395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
python 3.11 slim docker image julia 1.7.1 for flashweave mysql.connector 8.0.27	405
python library pandas 2.1.1. numpy 1.26.0 multiprocessing	406
text processing using awk	407
KEGG API	408

415 **Results** ⁴

416 **microbetag and microbetagDB**

418

419

420 **PRE-CALCULATIONS**

421

422 high quality GTDB
423 representative genomes

424

per genome

425 phenotrex

426

per pair of genomes

pathway comple-
427 mentarity

428

GEMS reconstruc-
429 tion with
430 ModelSEED

431

432 NetCmpt

433 Net Cooperate

434

435 PhyloMint

436

437 microbe-microbe cooperation
438 based on seed scores

439

440 microbe-microbe competition
441 based on seed scores

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

WORKFLOW

459 abundance table
460 (bins, ASVs, OTUs)

461 if network
462 not provided

463 FlashWeave

464 map sequence/
465 taxon to GTDB
466 representative genome

467 assign

468 phenotypical traits based
469 on local phenDB

470 FAPROTAX

471 co-occurrence network

472 get associations
473 where both taxa
474 at the species/
475 strain level

476 manta

477 potential pathway
478 complementarities

479 network clusters

480 co-operation and
481 competition seed scores

482 enrichment analysis
483 CyApp

459 **Fig. 2:** Diagram of the **microbetag** pre - calculations and the on the fly workflow.
460 GTDB v207 representative genomes were filtered and for those of high-quality 33
461 phenotypic traits were predicted using **phenotrex**. To this end, models were re-trained
462 to sync with recent version of eggNOG.

10

⁴Significant 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 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	461		
annotated network returned in .cyjs format	462		
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 [48] of the DECIPHER R package [65]. A co-occurrence network is also built using FlashWeave [18], as microbetag also does.	463		
	464		
	465		
	466		
	467		
	468		
	469		
	470		
	471		
	472		
	473		
	474		
	475		
	476		
microbetag CytoscapeApp	477		
Overall comment, the CytoscapeApp returns averages and s.d. for example in seed scores. If you want the exact values, go through the API.	478		
	479		
	480		
	481		
	482		
	483		
A. GTDB-tk: 480 bins	484		
B. GTDB 16S: 3000 ASVs	485		
C. Silva:	486		
D. fuzzywuzzy:	487		
Step	Time(sec)	Notes	
Taxonomy mapping	Cell 1,2	on the fly	
Network inference	Cell 2,2	on the fly	
microbetag annotations	Cell 3,2	on the fly	
manta clustering	Cell 4,2	on the fly	
Step	Time(sec)	Notes	
Taxonomy mapping	Cell 1,2	Cell 1,3	
Network inference	Cell 2,2	Cell 2,3	
microbetag annotations	Cell 3,2	Cell 3,3	
manta clustering	Cell 4,2	Cell 4,3	
Step	Time(sec)	Notes	
Taxonomy mapping	Cell 1,2	Cell 1,3	
Network inference	Cell 2,2	Cell 2,3	
microbetag annotations	Cell 3,2	Cell 3,3	
manta clustering	Cell 4,2	Cell 4,3	
Table 1: Computing times per step using an abundance table of 400 taxa with taxonomy: A. taxonomy scheme B. C. D. ⁵ specs of the laptop used.	490		
The app was based on the StringApp and supported by the NRNB group.	491		
	492		
	493		
	494		
	495		
Validation of microbetag potential	496		
vitamin dataset [42]	497		
Metagenomic or metabarcoding data are often used to predict microbial interactions in complex communities, but these predictions are rarely explored experimentally. Here, we use an organism abundance correlation network to investigate factors that control community organization in mine tailings-derived laboratory microbial consortia grown under dozens of conditions.	498		
The network is overlaid with metagenomic information about functional capacities to generate testable hypotheses.	499		
Thiamine alternative pathway [66, 67]	500		
	501		
	502		
	503		
	504		
	505		
	506		

Study
those
2 to
under-
stand
our
find-
ings

507 **Discussion** ⁶

508

509 **Interpetating a real-world network with microbetag**

510

511 Annelies' dataset.

512

513 **microbetag as a resource**

514

515 **Limitations**

516

517 As shown in [68] (see Figure 6b), the original version of CheckM [69] that is still used on
518 GTDB returns lower completeness scores to genomes that correspond to phyla known
519 for having shorter genomes in general, e.g. Patescibacteria representative genomes on
520 GTDB have an average completeness ~65%. **microbetag** inherits this in the filtering
521 process for getting only high quality genomes and thus, only few representatives from
522 these taxonomic groups are present on microbetagDB.

523

524 It is well known that higher-order interactions, i.e. interactions involving more
525 than two species [33] Pairwise relationships do not capture more complex forms of
526 ecological interactions, in which one species depends on (or is influenced by) multiple
527 other species. [3]

528

529 **Future work**

530

531 Further indices using the seed concept have been also presented such as the metabolic
532 interaction potential (*MIP*) and the metabolic resource overlap (*MRO*). *MIP* is
533 defined as the difference between the minimal number of components required for the
534 growth of all members in a noninteracting community and an interacting community,
535 i.e. the maximum number of essential nutritional components that a community can
536 provide for itself through interspecies metabolic exchanges [33]. Similarly, *MRO* is
537 defined as the maximum possible overlap between the minimal nutritional require-
538 ments of all member species [33]. Regression and association rule mining [70] can be
539 applied to address this challenge.

540

- 541 • pathway and seed complementarities for higher-order interactions
- 542 • spatial dimension
- 543 • transcriptomics data integration: compare potential complementarities with what
- 544 is going on
- 545 •

546

547 **Conclusions**

548

549 ⁷

550 Data integration

551

552 ⁶The user interface should be described and a discussion of the intended uses of the software, and the
553 benefits that are envisioned, should be included, together with data on how its performance and functionality
554 compare with, and improve, on functionally similar existing software. A case study of the use of the software
555 may be presented. The planned future development of new features, if any, should be mentioned.

556

557 ⁷This should state clearly the main conclusions and provide an explanation of the importance and

558 relevance of the case, data, opinion, database or software reported.

Supplementary information.	⁸	553
		554
Declarations		555
		556
• Availability of data and materials		557
– Raw sequences for the use case:		558
– Raw data for the validations case:		559
		560
• Funding		561
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 E.I.M.D.		562
		563
		564
		565
• Conflict of interest/Competing interests		566
The authors declare that they have no other competing interests.		567
• Authors' contributions	⁹	568
Conceptualization: K.F. Methodology: K.F. and H.Z. Software: H.Z., E.I.M.D. and J.M Validation: H.Z. and K.F. Formal analysis: H.Z. and K.F. Investigation: H.Z. Resources: K.F., A.E. and A.G. Data Curation: H.Z. Writing - Original Draft: H.Z. and K.F. Writing - Review & Editing: all Visualization: H.Z. Supervision: K.F., H.Z. and S.M. Project administration: K.F. Funding acquisition: K.F., A.E.		569
		570
		571
		572
		573
• Acknowledgements		574
We would like to thank Dr Christina Pavloudi and ++ for the insight on how to organise the trait groups.		575
		576
• Ethics approval		577
Not applicable		578
• Consent to participate		579
Not applicable.		580
• Code availability:		581
– microbetagDB related scripts: https://github.com/hariszaf/microbetag		582
– microbetagApp and webserver: https://github.com/msysbio/microbetagApp .		583
– CytoscapeApp: https://github.com/ermismd/MGG/		584
– Validation and use case: <i>jthink of having that under the 3D'omics organization;</i>		585
– Documentation web-site: https://hariszaf.github.io/microbetag/		586
		587
Appendix A Mappings		588
<i>n : 1 n : n etc</i>		589
		590
		591
		592
		593
		594
		595
		596
		597
		598

⁸If your article has accompanying supplementary file(s) please state so here. E.g. supplementary figures and tables captions.

⁹Based on the CRedit system. Current list is indicative.

599 **Appendix B Background on seed scores and
600 complementarities**
601

602 **B.1 Background on seed scores**
603

604 In that case, once a seed is assured, it activates all the rest of that group. Therefore,
605 a confidence level (C) ranging from 0 to 1, has been previously described to quantify
606 the relevance of each seed:

607
$$C_i = 1/\text{seed}'s\ group\ with\ i\ size \quad (\text{B1})$$

608 $C = 0$ corresponds to a non-seed node, while $C = 1$ represents an independent
609 node.

610
$$MI_{Complementarity} = \frac{|SeedSet_A \cap \neg SeedSet_B|}{|SeedSet_A \cap (SeedSet_B \cup \neg SeedSet_B)|} \quad (\text{B2})$$

611 As also described in [59], it is calculated as the proportion of compounds in a
612 species' seed set that coincide with those in an other's, while also factoring in the
613 confidence scores associated with seed compounds.

614
$$MI_{Competition} = \frac{\sum C(SeedSet_A \cap SeedSet_B)}{\sum C(SeedSet_A)} \quad (\text{B3})$$

615 **B.2 Background on pathway complementarity**
616

617 For example, the definition of the D-Galacturonate degradation in Bacteria (M00631)
618 is:

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

621 K01812 K00041 K01685 K00874 K01625
622 K01812 K00041 K16849+K16850 K00874 K01625
623 K01812 K00041 K01685 K00874 K17463
624 K01812 K00041 K16849+K16850 K00874 K17463

625 **B.3 Complementarities**
626

627 KEGG compound ModelSEED compounds ModelSEED compounds mapped to
628 KEGG compounds and kept only those related to KEGG modules.

629 **References**
630

- 631 [1] Yuan, M.M., Guo, X., Wu, L., Zhang, Y., Xiao, N., Ning, D., Shi, Z., Zhou, X.,
632 Wu, L., Yang, Y., *et al.*: Climate warming enhances microbial network complexity
633 and stability. *Nature Climate Change* **11**(4), 343–348 (2021)

- [2] Raes, J., Bork, P.: Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology* **6**(9), 693–699 (2008) 645
646
647
- [3] Faust, K., Raes, J.: Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**(8), 538–550 (2012) 648
649
650
- [4] Röttjers, L., Faust, K.: From hairballs to hypotheses—biological insights from microbial networks. *FEMS microbiology reviews* **42**(6), 761–780 (2018) 651
652
653
- [5] Bálint, M., Bahram, M., Eren, A.M., Faust, K., Fuhrman, J.A., Lindahl, B., O’Hara, R.B., Öpik, M., Sogin, M.L., Unterseher, M., *et al.*: Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS microbiology reviews* **40**(5), 686–700 (2016) 654
655
656
657
- [6] Robinson, C.J., Bohannan, B.J., Young, V.B.: From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews* **74**(3), 453–476 (2010) 658
659
660
661
- [7] Louca, S., Jacques, S.M., Pires, A.P., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., Doebeli, M.: High taxonomic variability despite stable functional structure across microbial communities. *Nature ecology & evolution* **1**(1), 0015 (2016) 662
663
664
665
666
- [8] Kost, C., Patil, K.R., Friedman, J., Garcia, S.L., Ralser, M.: Metabolic exchanges are ubiquitous in natural microbial communities. *Nature Microbiology*, 1–9 (2023) 667
668
669
- [9] Arnaouteli, S., Bamford, N.C., Stanley-Wall, N.R., Kovács, Á.T.: *Bacillus subtilis* biofilm formation and social interactions. *Nature Reviews Microbiology* **19**(9), 600–614 (2021) 670
671
672
673
- [10] Sousa, J.M., Lourenço, M., Gordo, I.: Horizontal gene transfer among host-associated microbes. *Cell Host & Microbe* **31**(4), 513–527 (2023) 674
675
- [11] Keller, L., Surette, M.G.: Communication in bacteria: an ecological and evolutionary perspective. *Nature Reviews Microbiology* **4**(4), 249–258 (2006) 676
677
678
- [12] Finn, R., Balech, B., Burgin, J., Chua, P., Corre, E., Cox, C., Donati, C., Santos, V., Fosso, B., Hancock, J., Heil, K., Ishaque, N., Kale, V., Kunath, B., Médigue, C., Pafilis, E., Pesole, G., Richardson, L., Santamaria, M., Van Den Bossche, T., Vizcaíno, J., Zafeiropoulos, H., Willassen, N., Pelletier, E., Batut, B.: Establishing the elixir microbiome community [version 1; peer review: awaiting peer review]. *F1000Research* **13**(50) (2024) <https://doi.org/10.12688/f1000research.144515.1> 679
680
681
682
683
684
685
- [13] Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., *et al.*: A new view of the tree of life. *Nature microbiology* **1**(5), 1–6 (2016) 686
687
688
689
690

- 691 [14] Matchado, M.S., Lauber, M., Reitmeier, S., Kacprowski, T., Baumbach, J., Haller,
692 D., List, M.: Network analysis methods for studying microbial communities: A
693 mini review. Computational and structural biotechnology journal **19**, 2687–2698
694 (2021)
- 695
- 696 [15] Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J.,
697 Huttenhower, C.: Microbial co-occurrence relationships in the human microbiome.
698 PLoS computational biology **8**(7), 1002606 (2012)
- 699
- 700 [16] Friedman, J., Alm, E.J.: Inferring correlation networks from genomic survey data.
701 PLoS computational biology **8**(9), 1002687 (2012)
- 702
- 703 [17] Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bon-
704 neau, R.A.: Sparse and compositionally robust inference of microbial ecological
705 networks. PLoS computational biology **11**(5), 1004226 (2015)
- 706
- 707 [18] Tackmann, J., Rodrigues, J.F.M., Mering, C.: Rapid inference of direct interac-
708 tions in large-scale ecological networks from heterogeneous microbial sequencing
709 data. Cell systems **9**(3), 286–296 (2019)
- 710
- 711 [19] Faust, K.: Open challenges for microbial network construction and analysis. The
712 ISME Journal **15**(11), 3111–3118 (2021)
- 713
- 714 [20] Kishore, D., Birzu, G., Hu, Z., DeLisi, C., Korolev, K.S., Segrè, D.: Inferring
715 microbial co-occurrence networks from amplicon data: a systematic evaluation.
716 Msystems, 00961–22 (2023)
- 717
- 718 [21] Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y.,
719 Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., *et al.*: Correlation detection strategies
720 in microbial data sets vary widely in sensitivity and precision. The ISME journal
721 **10**(7), 1669–1681 (2016)
- 722
- 723 [22] Berry, D., Widder, S.: Deciphering microbial interactions and detecting keystone
724 species with co-occurrence networks. Frontiers in microbiology **5**, 219 (2014)
- 725
- 726 [23] Guo, B., Zhang, L., Sun, H., Gao, M., Yu, N., Zhang, Q., Mou, A., Liu, Y.: Micro-
727 bial co-occurrence network topological properties link with reactor parameters
728 and reveal importance of low-abundance genera. npj Biofilms and Microbiomes
729 **8**(1), 3 (2022)
- 730
- 731 [24] Ma, B., Wang, Y., Ye, S., Liu, S., Stirling, E., Gilbert, J.A., Faust, K., Knight, R.,
732 Jansson, J.K., Cardona, C., *et al.*: Earth microbial co-occurrence network reveals
733 interconnection pattern across microbiomes. Microbiome **8**, 1–12 (2020)
- 734
- 735 [25] Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi,
736 Y., Audic, S., Berline, L., Brum, J.R., *et al.*: Plankton networks driving carbon
export in the oligotrophic ocean. Nature **532**(7600), 465–470 (2016)

- [26] Röttjers, L., Faust, K.: Manta: A clustering algorithm for weighted ecological networks. *Msystems* **5**(1), 10–1128 (2020) 737
738
739
- [27] Levy, R., Borenstein, E.: Reverse ecology: from systems to environments and back. In: *Evolutionary Systems Biology*, pp. 329–345. Springer, ??? (2012) 740
741
742
- [28] Levy, R., Borenstein, E.: Metagenomic systems biology and metabolic modeling of the human microbiome: From species composition to community assembly rules. *Gut Microbes* **5**(2), 265–270 (2014) 743
744
745
- [29] 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) 746
747
748
749
- [30] Parter, M., Kashtan, N., Alon, U.: Environmental variability and modularity of bacterial metabolic networks. *BMC evolutionary biology* **7**, 1–8 (2007) 750
751
752
- [31] 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) 753
754
755
756
- [32] 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) 757
758
759
760
- [33] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., Patil, K.R.: Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences* **112**(20), 6449–6454 (2015) 761
762
763
764
- [34] Belcour, A., Frioux, C., Aite, M., Bretaudieu, A., Hildebrand, F., Siegel, A.: Metage2metabo, microbiota-scale metabolic complementarity for the identification of key species. *Elife* **9**, 61968 (2020) 765
766
767
768
- [35] Thiele, I., Palsson, B.Ø.: A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols* **5**(1), 93–121 (2010) 769
770
771
- [36] Durot, M., Bourguignon, P.-Y., Schachter, V.: Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS microbiology reviews* **33**(1), 164–190 (2008) 772
773
774
775
- [37] Cerk, K., Ugalde-Salas, P., Nedjad, C.G., Lecomte, M., Muller, C., Sherman, D.J., Hildebrand, F., Labarthe, S., Frioux, C.: Community-scale models of microbiomes: Articulating metabolic modelling and metagenome sequencing. *Microbial Biotechnology* **n/a(n/a)**, 14396 <https://doi.org/10.1111/1751-7915.14396> <https://ami-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/1751-7915.14396>. e14396 MICROBIO-2023-392.R1 776
777
778
779
780
781
782

- 783 [38] Mori, M., Ponce-de-León, M., Peretó, J., Montero, F.: Metabolic complementation
784 in bacterial communities: necessary conditions and optimality. *Frontiers in*
785 *Microbiology* **7**, 1553 (2016)
- 786
- 787 [39] Zientz, E., Dandekar, T., Gross, R.: Metabolic interdependence of obligate intra-
788 cellular bacteria and their insect hosts. *Microbiology and Molecular Biology*
789 *Reviews* **68**(4), 745–770 (2004)
- 790
- 791 [40] Kallus, Y., Miller, J.H., Libby, E.: Paradoxes in leaky microbial trade. *Nature*
792 *communications* **8**(1), 1361 (2017)
- 793
- 794 [41] Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., Kanehisa, M.:
795 Modular architecture of metabolic pathways revealed by conserved sequences of
796 reactions. *Journal of Chemical Information and Modeling* **53**(3), 613–622 (2013)
797 <https://doi.org/10.1021/ci3005379> <https://doi.org/10.1021/ci3005379>. PMID:
798 23384306
- 799
- 800 [42] Hessler, T., Huddy, R.J., Sachdeva, R., Lei, S., Harrison, S.T., Diamond, S.,
801 Banfield, J.F.: Vitamin interdependencies predicted by metagenomics-informed
802 network analyses and validated in microbial community microcosms. *Nature*
803 *Communications* **14**(1), 4768 (2023)
- 804
- 805 [43] Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N.,
806 Dietrich, E.M., Disz, T., Gabard, J.L., *et al.*: Improvements to patric, the
807 all-bacterial bioinformatics database and analysis resource center. *Nucleic acids*
808 *research* **45**(D1), 535–542 (2017)
- 809
- 810 [44] Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugen-
811 holtz, P.: Gtdb: an ongoing census of bacterial and archaeal diversity through
812 a phylogenetically consistent, rank normalized and complete genome-based
813 taxonomy. *Nucleic acids research* **50**(D1), 785–794 (2022)
- 814
- 815 [45] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies,
816 J., Glöckner, F.O.: The silva ribosomal rna gene database project: improved data
817 processing and web-based tools. *Nucleic acids research* **41**(D1), 590–596 (2012)
- 818
- 819 [46] Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya,
820 R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., *et al.*: Ncbi taxonomy:
821 a comprehensive update on curation, resources and tools. *Database* **2020**, 062
822 (2020)
- 823
- 824 [47] Alishum, A.: DADA2 Formatted 16S rRNA Gene Sequences for Both Bacteria
825 & Archaea. <https://doi.org/10.5281/zenodo.6655692> . <https://doi.org/10.5281/zenodo.6655692>
- 826
- 827 [48] Murali, A., Bhargava, A., Wright, E.S.: Idtaxa: a novel approach for accurate
828 taxonomic classification of microbiome sequences. *Microbiome* **6**(1), 1–14 (2018)

- [49] 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) 829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
- [50] Louca, S., Parfrey, L.W., Doeblei, M.: Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**(6305), 1272–1277 (2016)
- [51] Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., Langille, M.G.: Picrust2 for prediction of metagenome functions. *Nature biotechnology* **38**(6), 685–688 (2020)
- [52] Feldbauer, R., Schulz, F., Horn, M., Rattei, T.: Prediction of microbial phenotypes based on comparative genomics. *BMC bioinformatics* **16**(14), 1–8 (2015)
- [53] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., *et al.*: eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research* **47**(D1), 309–314 (2019)
- [54] 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)
- [55] 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)
- [56] 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)
- [57] Kanehisa, M., Sato, Y.: Kegg mapper for inferring cellular functions from protein sequences. *Protein Science* **29**(1), 28–35 (2020)
- [58] Kanehisa, M., Sato, Y., Kawashima, M.: Kegg mapping tools for uncovering hidden features in biological data. *Protein Science* **31**(1), 47–53 (2022)
- [59] 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)
- [60] Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Lindsay, B., Stevens, R.L.: High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology* **28**(9), 977–982 (2010)

- 875 [61] Blank, C.E., Cui, H., Moore, L.R., Walls, R.L.: Micro: an ontology of pheno-
876 typic and metabolic characters, assays, and culture media found in prokaryotic
877 taxonomic descriptions. *Journal of biomedical semantics* **7**(1), 1–10 (2016)
- 878
- 879 [62] Choudhary, K., Meng, E.C., Diaz-Mejia, J.J., Bader, G.D., Pico, A.R., Morris,
880 J.H.: scnetviz: from single cells to networks using cytoscape. *F1000Research* **10**
881 (2021)
- 882
- 883 [63] Merkel, D., *et al.*: Docker: lightweight linux containers for consistent development
884 and deployment. *Linux j* **239**(2), 2 (2014)
- 885
- 886 [64] Reese, W.: Nginx: The high-performance web server and reverse proxy. *Linux J.*
887 **2008**(173) (2008)
- 888
- 889 [65] Wright, E.S.: Using decipher v2. 0 to analyze big biological sequence data in r. *R*
890 *Journal* **8**(1) (2016)
- 891
- 892 [66] Llavero-Pasquina, M., Geisler, K., Holzer, A., Mehrshahi, P., Mendoza-Ochoa,
893 G.I., Newsad, S.A., Davey, M.P., Smith, A.G.: Thiamine metabolism genes
894 in diatoms are not regulated by thiamine despite the presence of predicted
895 riboswitches. *New Phytologist* **235**(5), 1853–1867 (2022)
- 896
- 897 [67] Romine, M.F., Rodionov, D.A., Maezato, Y., Osterman, A.L., Nelson, W.C.: Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors
898 in microbial communities. *The ISME journal* **11**(6), 1434–1446 (2017)
- 899
- 900 [68] Chklovski, A., Parks, D.H., Woodcroft, B.J., Tyson, G.W.: Checkm2: a rapid,
901 scalable and accurate tool for assessing microbial genome quality using machine
902 learning. *Nature Methods* **20**(8), 1203–1212 (2023)
- 903
- 904 [69] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W.: Checkm:
905 assessing the quality of microbial genomes recovered from isolates, single
906 cells, and metagenomes. *Genome research* **25**(7), 1043–1055 (2015)
- 907
- 908 [70] Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of
909 items in large databases. In: *Proceedings of the 1993 ACM SIGMOD International
910 Conference on Management of Data. SIGMOD '93*, pp. 207–216. Association for
911 Computing Machinery, New York, NY, USA (1993). <https://doi.org/10.1145/170035.170072> . <https://doi-org.kuleuven.e-bronnen.be/10.1145/170035.170072>
- 913
- 914
- 915
- 916
- 917
- 918
- 919
- 920