




Review

Mining Synergistic Microbial Interactions: A Roadmap on How to Integrate Multi-Omics Data

Joao Pedro Saraiva¹, Anja Worrich¹, Canan Karakoç^{1,2}, Rene Kallies¹, Antonis Chatzinotas^{1,2,3}, Florian Centler¹  and Ulisses Nunes da Rocha^{1,*} 

¹ Department of Environmental Microbiology, Helmholtz Centre for Environmental Research-UFZ, 04318 Leipzig, Germany; joao.saraiva@ufz.de (J.P.S.); anja.worrich@ufz.de (A.W.); canan.karakoc@ufz.de (C.K.); rene.kallies@ufz.de (R.K.); antonis.chatzinotas@ufz.de (A.C.); florian.centler@ufz.de (F.C.)

² German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

³ Institute of Biology, Leipzig University, 04103 Leipzig, Germany

* Correspondence: ulisses.rocha@ufz.de

Abstract: Mining interspecies interactions remain a challenge due to the complex nature of microbial communities and the need for computational power to handle big data. Our meta-analysis indicates that genetic potential alone does not resolve all issues involving mining of microbial interactions. Nevertheless, it can be used as the starting point to infer synergistic interspecies interactions and to limit the search space (i.e., number of species and metabolic reactions) to a manageable size. A reduced search space decreases the number of additional experiments necessary to validate the inferred putative interactions. As validation experiments, we examine how multi-omics and state of the art imaging techniques may further improve our understanding of species interactions' role in ecosystem processes. Finally, we analyze pros and cons from the current methods to infer microbial interactions from genetic potential and propose a new theoretical framework based on: (i) genomic information of key members of a community; (ii) information of ecosystem processes involved with a specific hypothesis or research question; (iii) the ability to identify putative species' contributions to ecosystem processes of interest; and, (iv) validation of putative microbial interactions through integration of other data sources.

Keywords: microbial communities; synergistic interactions; ecosystem processes; multi-omics



Citation: Saraiva, J.P.; Worrich, A.; Karakoç, C.; Kallies, R.; Chatzinotas, A.; Centler, F.; Nunes da Rocha, U. Mining Synergistic Microbial Interactions: A Roadmap on How to Integrate Multi-Omics Data. *Microorganisms* **2021**, *9*, 840. <https://doi.org/10.3390/microorganisms9040840>

Academic Editor: Jean Armengaud

Received: 9 February 2021

Accepted: 8 April 2021

Published: 14 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this review, we discuss a roadmap to mine inter-species interactions in microbial communities. To define this roadmap, we will use an ecosystem process as the unity from which to mine them. Here, we define an ecosystem process as a specific set of metabolic functions (e.g., benzoate degradation or in nitrification). In this review we use the term ecosystem process to define a unit to explore microbial interactions in order to limit microbial community richness to a manageable scale [1]. Currently, a mechanistic understanding of microbial interactions lies on known connections among genes, protein, reactions and their participation in an ecosystem process. Hence, we focus on the synergistic inter-species interactions as these can be directly linked to ecosystem processes. Nevertheless, predicting antagonistic or competitive interactions may be achieved by the inclusion of information such as enzyme kinetics and/or metabolic fluxes [2] involved in these interactions.

2. Synergistic Interspecies Interactions Drive Ecosystem Processes

In nature, microbes do not exist alone but rather as members of complex communities [3]. Synergistic interspecies interactions play an essential role in ecosystems by either improving adaptation of microbial communities to their habitats or allowing microorganisms to survive in environments for which they lack the complete metabolic capacity [4].

The type and degree of changes in physicochemical conditions of ecosystems such as the addition of chemicals affect ecosystem processes but also determine how microbial communities respond to those alterations. Species-level functional profiling of (meta)genomes is already possible by providing coverage and abundance estimates on individual pathways across microbial communities as well as for individual species [5]. Expanding this approach to ecosystem processes would generate data allowing the definition of groups of species that cover a full ecosystem process. One first step to determine microbial interactions that are affected by environmental changes (e.g., the introduction of chemicals or temperature shifts) and their relationship to ecosystem functioning relies on the identification and characterization of the constituents of microbial communities as well as their functional potential [6].

The concept of “everything is everywhere, but the environment selects” proposed by Baas Becking in 1934 [7] for microbial community assembly has gathered much debate as reviewed by Peter Girguis [8]. Extrapolating this concept to the functional potential (i.e., the complete set of functions) of species, however, is not straightforward. While for core functions such as glucogenesis, the idea that they are genetically widespread in all microorganisms might seem plausible, this is not true for every ecosystem process. For example, methanotrophic bacteria [9] and certain filamentous fungi [6] possess genes that encode a key enzyme in methane degradation—methane monooxygenase—not present in other microorganisms.

Microbial communities will be able to address changing environmental conditions (e.g., introduction of chemicals/substrates, shift in temperature and change in pH) if their combined functional potential encompasses the set of metabolic functions required to handle these changes. Here, functional profile is defined as a subset of metabolic functions from the complete functional pool that are needed for a given ecosystem process (definition adapted from Oh and collaborators [10]). The functional profile can be achieved due to action of single microbes if they are able to solely perform the set of metabolic functions for the required ecosystem process. Alternatively, this functional profile can be achieved by the interaction between two or more species (e.g., acquisition of antibiotic resistance genes via horizontal gene transfer [11], production of secondary metabolites [12] or cross-feeding [13]). It is likely that microbial communities with higher species richness will have a higher number of unique functional traits due to the individual species’ metabolic potential or as the result of the combined metabolic capabilities of multiple species that arises from interspecies interactions. Additionally, microbial community functional potential and functional redundancy is positively correlated with species richness [14]. Fetzer and collaborators [14] also showed that environmental conditions influenced the type and number of microbial interactions (Figure 1). Therefore, more studies characterizing microbial interaction that drive ecosystem processes are necessary as they link microbial diversity and ecosystem function responses to a changing environment [9,14].

From Where Will One Extract the Genomes to Explore Microbial Interactions?

Currently, advances in single cell sequencing [15], recovery of metagenome assembled genomes [16,17] and advanced cultivation strategies [18] enormously increased the power to identify species and their genomes from natural environments such as soils or deep subsurface. Still, datasets generated from high-throughput sequencing do not provide absolute abundances of species in a microbial community [19] and extra experiments are necessary to generate this data (e.g., quantitative PCR or in situ fluorescence hybridization). Nevertheless, it has been speculated that, in the near future, it will be possible to obtain the genomic information from all species in complex microbial communities [20] thus overcoming one of the bottlenecks in predicting interspecies interactions. Moreover, these approaches have allowed us to gather information on species relative abundance, phylogeny and function [21]. The increase in computational power and advances in sequencing technologies also provides a favorable environment to produce tools and to generate and analyze data in a timewise manner leading to faster and more reliable data mining [22].

These improvements allow the prediction of species interactions and estimation of their functional contribution to ecosystem processes in natural environments, as demonstrated by Kirwan et al. [23]. In the next section, we will discuss advantages and limitations of current approaches to predict interspecies interactions and strategies to overcome them.

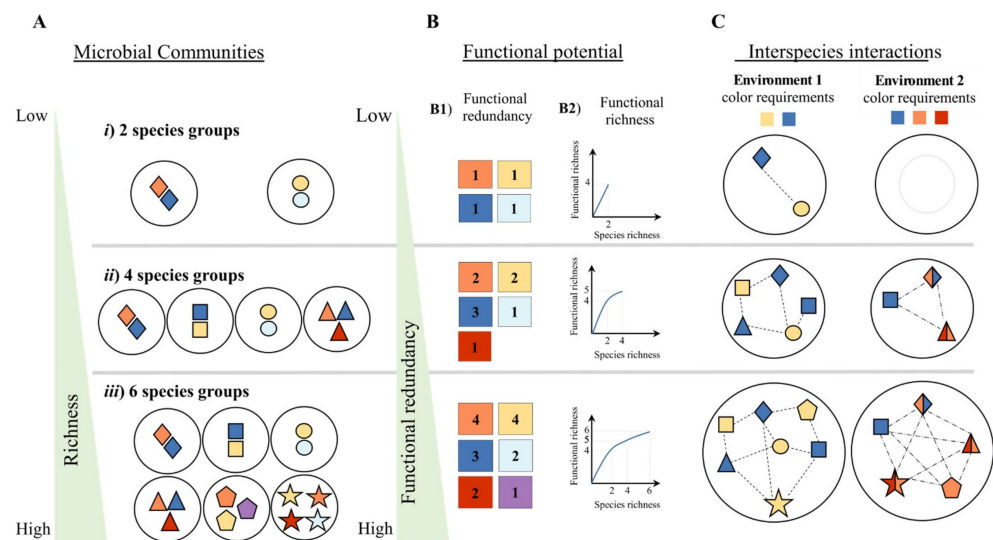


Figure 1. Interplay between microbial community size, functional potential and interspecies interactions. Functional richness and redundancy increase with microbial community richness. Three microbial communities (A) are represented with different levels of species richness (*i*, *ii* and *iii*), i.e. different number of unique species (represented by a geometric shape). Each species is capable of performing a number of functions that are represented by a specific color (B). For example, community *i* is composed by two species each capable of performing two different sets of functions. The number of unique functions illustrates the functional richness of each microbial community. Functional redundancy is determined by the number of microbes with the genetic potential to perform the same function. Thus, an increase in the number of unique species is more likely to result in an increase of the functional redundancy (B1) and richness (B2) of a microbial community. For example, from community *ii* to community *iii* there is an increase of two species and one unique function but the number of species capable of performing multiple functions (i.e., functional redundancy) also doubled for four functions: orange, yellow, blue and red. Furthermore, the combinations of interspecies interactions (C) is not only dependent on the individual microorganism's genetic potential but also determined by the environmental conditions. For example, growth in **Environment 1** requires a microbe's ability to perform two functions (blue and yellow) while for **Environment 2** three functions are required (orange, blue and red). Although not a linear relationship, the higher the number of species in a microbial community, the higher the probability of an increased number of interspecies interactions (as long as the genetic potential is present).

3. Current Approaches to Predict Microbial Community Functional Profiles and Interspecies Interactions

As stated in the previous section, the functional potential of microbial communities might shed light on potential synergistic interspecies interactions. Several modeling approaches have been proposed to predict microbial community functional profiles [24]. Here, we focus on approaches that take advantage of (meta)genomic data and represent microbial activity in the context of metabolic networks [25] as well as their suitability to infer synergistic interspecies interactions. Three main concepts have been proposed which differ in the level of detail in which they represent the metabolic activity in microbial communities: the supra-organism approach, the population-based approach and the guild approach.

3.1. The Supra-Organism Approach

In the supra-organism approach [2], a community is considered as a singular organism and the phylogenetic origin of individual enzymes detected to be present in the community is neglected. Based on metagenomic data, a global metabolic network of enzymatically catalyzed reactions is constructed, allowing for the prediction of shifts in pathway activity when comparing two samples. For example, pathway activity has been inferred in gut metagenomes [26] by calculating gene relative abundance in each individual sample. The supra-organism approach considers interactions between genes (or the respective reactions their enzymes catalyze) rather than between species. This allows the comparison of the functional profiles of microbial communities as whole. However, these models do not allow the prediction of interspecies interactions since the contribution of individual species to any metabolic function is not determined.

3.2. The Population-Based Approach

In contrast to the supra-organism approach, species boundaries are explicitly considered in population-based approaches [27]. Here, community members are represented by independent species in the model. These models are especially suited for the analysis of individual species' functional profile and have been extensively used in metabolic network reconstruction of single genomes [25,28,29]. Furthermore, this approach allows for the inclusion of direct metabolic interactions between community members. The inference of interspecies interactions is possible under this approach since the metabolic contribution of each species to ecosystem processes can be mapped to their genome. However, generating genome-scale metabolic models for all species in a community as well as estimating features such as biomass composition remain challenging [30,31].

3.3. The Guild-Based Approach

In the guild-based approach [32], microbial species performing the same metabolic function(s) are grouped together and they are represented by a unique entity in the model, reducing model complexity at the expense of individual species resolution. This approach has been used in ordinary differential equation-based modeling [33] and can be expanded to metabolic network-based approaches [34]. The guild approach is useful when microorganisms are known to possess similar functional traits (e.g. similar methods of organic matter decomposition [35]). However, similar to the supra-organism approach, predicting interspecies interactions in these models is hampered by their inability to identify the contribution of individual species to ecosystem processes [2].

Besides the supra-organism, population-based and guild-based approaches, statistical methods such as Pearson's and Spearman's correlation may assist in the identification of interspecies interactions. These methods identify significant relationships by correlating the taxon abundance within microbial communities [36]. However, correlating abundance does not provide information on the underlying mechanisms by which microbes potentially interact. Furthermore, the vast amount of putative interspecies interactions would also require intensive experimental validation.

3.4. Advantages and Limitation of Current Approaches Mining Microbial Interactions

As described in Sections 3.1–3.3, the current approaches used to mine microbial interactions rely on links among genes, enzymes and metabolic reactions. The main limitations and advantages of the different approaches can be found in Table 1. Further, the computational and manual curation efforts in model generation varies strongly as well as the level of detail at which predictions are generated. The supra-organism approach requires the least effort, but only allows for the prediction of general changes in pathways when comparing different samples. Population-based approaches deliver the most detailed quantitative predictions down to intracellular metabolic fluxes of individual species, but require substantial efforts in model generation. These include the generation of genome-scale models and definition of the biomass equation for each community member [37]. Although an ever-

increasing number of models become available and semi-automatic pipelines support the generation of novel genome-scale metabolic models, this step often remains difficult to perform. The determination of model parameters regarding cellular maintenance requirements and uptake kinetics are examples of these difficulties. Furthermore, genome scale-based approaches require additional types of data such as transcriptomics, metabolomics and proteomics to validate their results. The inclusion of additional data types would provide information on gene regulation, structure of microbial communities and possibly a link to ecosystem processes. In addition, computational challenges (e.g., random access memory requirements and time) arise in the prediction of putative interspecies interactions together with the increase of species and pathways studied (Equation (1)). For example, for a community of 35 species and a set of 3 reactions, a total of 6545 combinations are possible. A 10-fold increase in the number of species (350) will result in 7,084,700 possible combinations. Expanding the number of reactions to five will yield more than 42 billion possible combinations.

$$C_k(n) = \frac{n!}{k!(n-k)!} \quad (1)$$

where, n is the number of species, k is the number of reactions and C is the number of possible combinations.

Hence, the main limitations to mine microbial interactions can be summarized as:

- Identification of all species in a community;
- Incomplete functional annotation of genomes;
- Data integration and experimental validation; and
- Exponential increase of search space with relatively small increase of number of species or pathway size.

In the next section we discuss the knowns and unknowns in microbial interaction studies through a literature meta-analysis.

Table 1. Pros and cons of current modelling approaches to predict microbial interactions, environments where the selected approaches have been used and respective references.

Approach	Pros	Cons	Environments	References
Supra-organism	Global reaction network is possible and allows for prediction of shifts in pathway activity by measuring gene relative abundance.	Genetic potential of individual species not determined	Anaerobic mixed culture fermentations	[38]
		Contribution of individual species to shifts in pathway activity not determined since interactions are based on genes/reactions.	Agricultural soil and seep sea “whale fall” carcasses	[39]
Population-based	Species boundaries explicitly defined. Individual species functional potential can be determined. Allows determining direct metabolic interactions between species.	High computational and manual curation efforts since full genome-scale metabolic models for each species is required.	Corals	[40]
			Anoxic sediments	[41]
			Batch and Continuous cultures	[42]
			Synthetic microbial systems	[43]
Guild-based	Less complex models since grouping of species is based on their known functional traits.	Requires previous knowledge on functional traits. Individual contribution of species to ecosystem processes is unknown.	Soil	[44,45]

4. Beyond Genetic Potential: Drawing a Strategy to Mine and Validate Microbial Interactions

Based on the analysis discussed in Sections 3.3 and 3.4, a road to mine microbial interactions lies in the exploration of the genetic potential involved in specific ecosystem processes. Therefore, we suggest dividing mining of microbial interactions into smaller blocks of information generated from (meta)genomic data. Thus, inference of synergistic interspecies interactions would be determined based on the combined functional potential of all species in a microbial community similar to the study by Jiménez and collaborators [46]. Nevertheless, our meta-analysis indicates that to validate putative microbial interactions extra data is needed (see Section 4.1).

Further, we postulate that prior to identifying microbial interaction it is relevant to focus on specific ecosystem processes. Therefore, estimating the contribution of microbes to ecosystem processes could be determined by looking at its genomic content as long as the connections between genes, proteins and reactions involved in said ecosystem process are known. The identification of a subset of species in a microbial community with the potential, even partial, to participate in an ecosystem process reduces the search space of putative microbial interactions. As discussed in Section 3.4, the reduced search space will make mining relevant inter-species interactions feasible in microbial communities.

In summary, the study of microbial communities is challenging due to our limited capacity to identify all members of a community, to connect genes to proteins and reactions, and to determine species interactions. Hence, mining microbial interactions would include: (A) genomic information of all (or key) members of a community; (B) information of ecosystem processes involved with a specific hypothesis or research question; and (C,D) the ability to identify putative species' contributions to ecosystem processes of interest. A step forward would include standardizing the validation of putative microbial interactions

(G) through integration of other data sources (F). In Section 4.1, we discuss different strategies to validate putative microbial interactions. In Section 4.2, we describe a hypothetical workflow including the mapping of specific ecosystem processes in different members of a microbial community and the validation of putative microbial interactions. In Section 4.3, we provide an example of a study combining omics, data mining and experimental data to infer species contributions to specific ecosystem processes.

4.1. Validation of Putative Microbial Interaction through Integration of Different Data Sources

High throughput sequencing has illuminated the black box of microbial diversity. Metagenomics provided an insight into the functional potential of microbial communities without the necessity of cultivating and characterizing thousands of isolates [47,48]. Therefore, metagenomics might provide information on the links between the genetic differences within species and their effects on hosts or adaptability to novel environmental conditions [49]. However, current approaches focusing on predictions of gene functions based on (meta)genomes have four major drawbacks, extensively reviewed by Prosser [50]. First, genetic potential studies may assume that gene presence is directly linked to function [51]. Second, microbial communities are three-dimensional structures that play a crucial role in ecosystem functioning and are not directly assessed by gene presence [52]. Third, different levels of protein activity can be found in different species or among populations of the same species due to transcriptional or post-translational modifications [53]. Fourth, temporal and spatial variability of environmental conditions and community dynamics need to be accounted for when demonstrating microbial interactions [50]. Based on these four limitations, the next four paragraphs discuss strategies that can be added to genetic potential to validate microbial interactions.

4.1.1. Assumption that Gene Presence is Directly Linked to Function

The presence of genetic potential to perform an ecosystem process does not guarantee this process is active [54]. Doolittle and Zhaxybayeva [55] and later Jansson and Hofmockel [56] suggested the investigation of the metaphecome to better understand the functions that are carried out by the active microbial communities under given environmental conditions. The metaphecome considers both microbial functions encoded in the metagenome and biotic and abiotic factors influencing the activity of community members. Thus, it encompasses not only (meta)genomics but also (meta)transcriptomics, (meta)proteomics, metabolomics and factors such as gene silencing.

4.1.2. Spatial (Three Dimensional) Structure of Microbial Communities

The three-dimensional structure plays an important role in interspecies interactions [57]. Species are not homogeneously distributed within microbial communities but rather structured based on their relationships with each other [58] and are shaped by their metabolic and physiological needs. For example, in biofilms found in suburban bath surfaces such as marble and plaster, growth and microbial community structure are influenced by their susceptibility to light [59]. Heterotrophic microorganisms adapted to dark were mostly found in plaster while low-light adapted microbes were found in mortar [59]. In biological soil crusts bacterial diversity was also shown to be heterogeneous across the different layers [52] and these might influence the number and degree and microbial interactions. Further, in wastewater treatment plant granular sludges presented different growth requirements, community structure and microbial relationships dependent on nitrogen and phosphorous availability [60]. Methods to determine the 3D structure of microbial communities include protein structures [61] and advanced microscopy [57].

4.1.3. Different Levels of Protein Activity within Species or Populations

For an ecosystem process to be active the required genes need to be translated, transcribed and encounter favorable environmental conditions. In cases where the focus is to understand the effect or fate of specific compounds on different ecosystem processes stable

isotope probing can be used to trace their transformation and turnover within the different members of a community. Methods such as stable isotope labelling metagenomics [62], metatranscriptomics [63], metaproteomics [64,65] or metabolomics [66] could serve as validation strategies by illustrating metabolite fluxes through groups of predicted interspecies interactions. Other methods such as co-culture experiments could also be employed to validate predicted inter-species interactions [36]. Stable Isotope Probing has also been coupled with NanoSIMS imaging and fluorescent in situ hybridization (FISH) to identify microbial interactions based on the flux of labelled isotopes between microbes [67].

4.1.4. Temporal Variability

It is widely known that microbial communities do not vary only in space but also in time; as deeply studied by the Earth Microbiome Project [68], the NIH Human Project [69] and Tara Oceans [70] among others. A major drawback from most, if not all, techniques to profile microbial communities is that they are based on snapshots of a given community for that specific technique [71]. The distribution of organisms and their functions are not homogeneous and species interactions are not constant in time due to abiotic forces and adaptive mechanisms [72–74]. Therefore, to validate patterns of microbial interactions one need not only to follow the community over time, but also to have appropriate number of samples in time and space [66,75]. Finally, the time frame for collection of samples and numbers of biological replicates will depend on the biodiversity of a given ecosystem and how dynamic the given process is [76].

When designing strategies to validate microbial interactions, it is necessary to keep in mind that different methods will encompass different limitations and potential levels of detail regarding inter-species microbial interactions (Table 2). For example, studies employing 16S rRNA gene sequencing provide an estimate of the phylogeny of the different species in a microbial community but not their functional potential. Predicting inter-species interactions based on the microbial community's combined functional potential is in this case not possible. Metabolic network reconstruction of individual genomes from environmental samples can provide substantial detail on the metabolic capabilities of individual species and potential metabolic exchanges. However, the high computational, physiological and chemical information requirements (e.g., biomass equation, reaction reversibility, ATP) to generate reliable genome-scale metabolic networks for such complex environments are seemingly difficult.

The influence of species on the growth of other community members can also be inferred from co- and mixed-culture experiments by comparing their growth to pure cultures. However, understanding the mechanisms by which species interact requires additional data from complex wet-lab experiments such as stable isotope labelling. A last bottleneck to the validation of interspecies interactions in complex environments is our inability to grow most species in the laboratory [77]. Hence, only a small number of species interactions can be experimentally validated without highly-complex wet-lab procedures.

4.2. From Mining to Validation: A Workflow to Identify Mechanisms Underlying Microbial Interactions

In this review, we indicate that defining putative interspecies interactions based solely on genomic potential is only the first piece of the puzzle in gaining a greater in-depth understanding of microbial interactions. Additional information from other data sources and experiments are required (Section 4.1). In this section, we discuss a workflow for iterative mining of microbial interactions by integrating *in silico* approaches and experimental validation using a thought experiment. In this thought experiment, we use a hypothetical microbial community composed of four species and a hypothetical ecosystem process composed of seven enzymatic reactions (Figure 2). We also assume these reactions require the presence of seven single protein-encoding genes and five protein-encoding genes that participate in the formation of two protein complexes (as you may have multiple proteins involved in the same reaction).

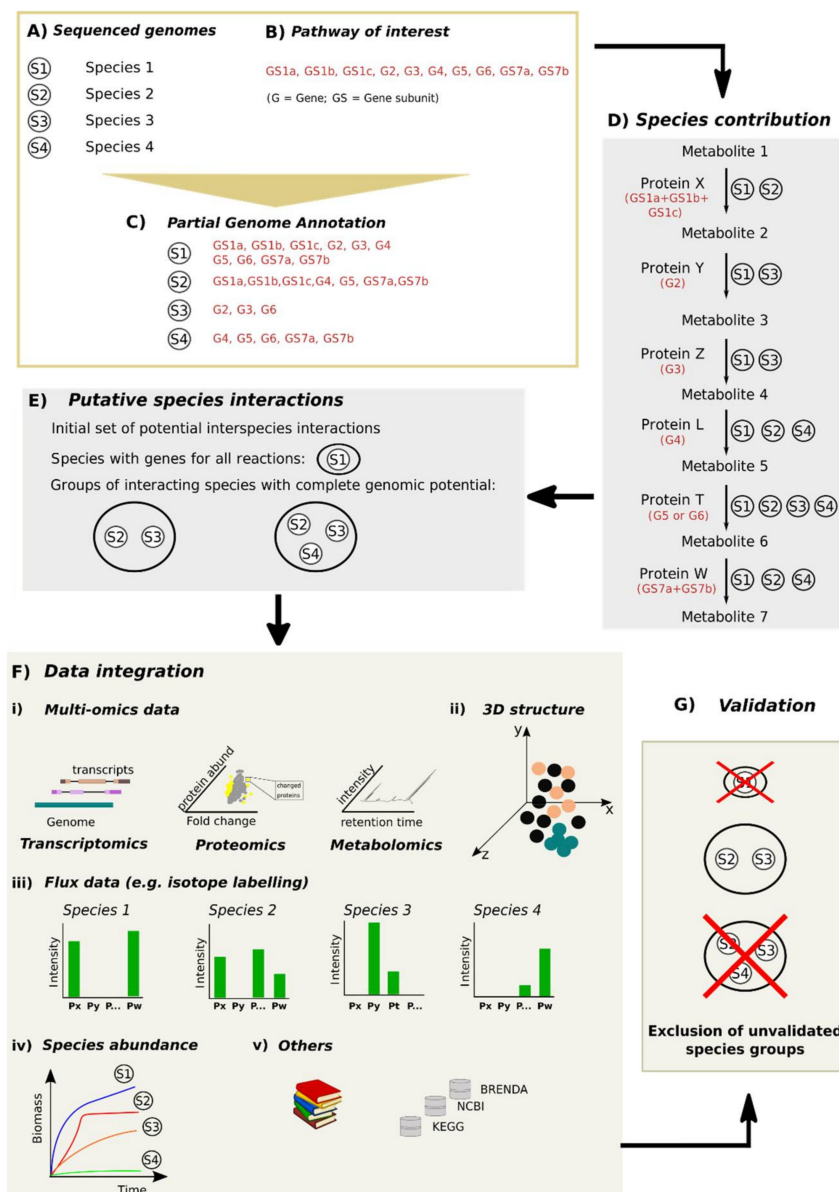


Figure 2. Theoretical work frame for prediction of interspecies interactions. Here, a microbial community is composed of four species (A) and a hypothetical ecosystem process/pathway of interest (B) requires the presence of five single protein-encoding genes and five protein-encoding genes that participate in the formation of two protein complexes. First, annotation of individual genomes from the microbial community is performed using only the set of target genes (C). From here, one can determine which species possess the complete functional potential to perform the target pathway (D). Additionally, one can also determine which species or groups of species possess a combined genomic potential to perform the complete ecosystem process (E) putative interacting species. Further refinement of the generated lists can be achieved by the inclusion of experimental data, species absolute abundances, literature searches, specialized databases and other omics data types (e.g., transcriptomics, metabolomics and proteomics) (F) leading to increased robustness of predictions and reduction of the number of interspecies interactions for experimental validation (G).

First, it is necessary to obtain the genomes of all or key members of the microbial community (Figure 2A). Second, a set of genes are selected that represent the reactions involved in the pathway of interest (Figure 2B). Third, genome annotation is performed using only the set of genes defined in the second step (Figure 2C). Fourth, the genetic potential of each species is mapped to the pathway of interest (Figure 2D). This mapping

allows to determine which species or groups of species possess the genetic potential for all reactions in the pathway of interest (Figure 2E). Next, additional types of data are integrated in the analysis to validate putative species interactions (Figure 2F). Among others, these can include multi-omics data, such as: (i) (meta)transcriptomics (gene expression), (meta)proteomics (protein expression) and metabolomics (measurement of metabolite production); (ii) three-dimensional structure of the microbial community; (iii) protein abundances measured through stable isotope labelling; (iv) species growth profiles; and, (v) literature and specialized databases. The integration of all these types of data confirms or excludes putative synergistic species interactions (Figure 2G). To note, this process should be iterative as new insights concerning microbial communities and the addition of data to repositories are obtained. For example, genome re-annotation might be necessary if novel species are identified or a better understanding of the reactions that are involved in a specific ecosystem process.

In the next section we provide an example of a study combining the use of multi-omics, data mining and experimental data. We also provide examples of how their study could be supplemented with additional tools and methods to improve prediction and validation of interspecies interactions.

4.3. Assembling a Workflow to Determine Microbial Interactions

Generating and analyzing the data needed to predict microbial interactions focused on specific ecosystem processes can come from a variety of sources such as axenic, enrichment or co-cultures and environmental samples. Additionally, several methods and tools exist to assemble genomes and to assess their quality, to perform functional annotation and to determine the contributions of microbes to ecosystem processes.

In the next four paragraphs, we will discuss different strategies and tools involved in mining and validating microbial interactions. To foster this discussion, we will analyze the work from Tláskal and collaborators [91]. In this study the authors hypothesize that the decomposition of deadwood required the combined efforts of fungal and bacterial species.

4.3.1. Identifying Microbial Species and Their Genetic Potential

In order to test their hypothesis, Tláskal and collaborators [91] first identified which species were present in deadwood samples. As most soil microorganisms remain uncultured [92], the authors decided to explore the microbial diversity in their system using amplicon metagenomics and metatranscriptomics [91]. Regarding the identification of microbial species, metagenomics was used to recover metagenome-assembled genomes (MAGs) and Metatranscriptomics was used to identify the transcription levels of the small subunit ribosomal RNA. The analysis of small subunit ribosomal RNA transcriptional levels indicate that an organism may be active but does not make a direct link with genetic potential [18]. Although working well for dominant Prokaryotes and DNA viruses, the recovery of MAGs (extensively reviewed by Chen and collaborators [93]) has several limitations for low abundance taxa and Eukaryotes. Other techniques that can be used to identify simultaneously the genetic potential and phylogeny of microbial species encompass single-cell genomics [94] and advanced culturing techniques [95,96].

Table 2. Outcomes and limitations of different methods to study microbial interactions. We assigned four validation strategies to confirm microbial interaction as following: (1) Expression or activity assays (e.g., transcriptomics, proteomics, metabolomics, RT-PCR, FBA); (2) 3D structure and spatial variability; (3) Substrate specificity; and, (4) Temporal variability.

Outcome	Limitations	Methods	Environment	Validation				Ref. ^a
				1	2	3	4	
Improvement in the identification of microbial community species.	Lack of mechanistic understanding of species interactions.	Combination of MALDI-TOF MS ^b analysis and high-throughput sequencing 16S rRNA ^c .	Kimchi	✓	○	○	✓	[78]
		16S rRNA gene sequencing.	Human oral environments	○	✓	○	○	[79]
Demonstration of the influence of abiotic factors on microbial community dynamics.	High computational and data requirements for reconstruction of individual metabolic models.	Metagenomics, metabolic network reconstruction and FBA ^d .	Anaerobic digestion microbiomes	✓	○	✓	○	[80]
		PLS-PM ^e	Rice soil rhizosphere	○	✓	○	✓	[81]
	Lack of mechanistic understanding of species interactions.	16S rRNA gene sequencing.	Urban and forest park soil litter layers	○	✓	○	✓	[82]
		In vivo experiment of meadow steppe soil under different precipitation regimes.	Topsoil	✓	✓	○	✓	[83]
High computational and data requirements for reconstruction of individual metabolic models and complex wet-lab experiments required for validation.		Metabolic network reconstruction, EFM ^f and FBA.	Acid-sulfate-chloride springs	✓	○	✓	○	[84]

Table 2. Outcomes and limitations of different methods to study microbial interactions. We assigned four validation strategies to confirm microbial interaction as following: (1) Expression or activity assays (e.g., transcriptomics, proteomics, metabolomics, RT-PCR, FBA); (2) 3D structure and spatial variability; (3) Substrate specificity; and, (4) Temporal variability.

Outcome	Limitations	Methods	Environment	Validation				Ref. ^a
				1	2	3	4	
Demonstration of the influence of interspecies interactions on microbial community dynamics.	Lack of mechanistic understanding of species interactions.	Co-culture of isolates, RNA-Seq ^g and RT-qPCR ^h .	Wine fermentation	✓	○	○	✓	[85]
		qPCR ⁱ and 16S rRNA gene sequencing.	Mixed bacterial consortia	✓	○	○	✓	[86]
Improved mechanistic understanding of interspecies interactions.	Complex wet-lab experiments required for validation.	SIP ^j and Metagenomics.	Continuous up-flow anaerobic sludge blanket reactors	✓	○	✓	✓	[87]
		Pure and co-cultures and cyclic voltammetry analysis.	Palm oil mill effluent	○	○	✓	✓	[88]
	High computational and data requirements for reconstruction of individual metabolic models.	Mono- and co-culture, metabolic network reconstruction, bipartite graphs, HPLC ^k , CGQ ^l , GC-MS ^m ; SIP.	In silicon experiments with pure and co-culture	✓	○	✓	✓	[89]
		Metabolic network reconstruction and cFBA ⁿ .	In silicon experiments pure cultures	✓	○	✓	○	[27]
		Metabolic network reconstruction, evolutionary game theory and FBA.	In silicon experiments pure cultures	✓	○	○	○	[90]
	Metagenomics, Metatranscriptomics.	Synthetic human gut	✓	✓	○	○	[5]	

^a Ref., numbers in between brackets represent references for the different studies; ^b MALDI-TOF: matrix-assisted laser desorption/ionization; ^c rRNA: Ribosomal ribonucleic acid; ^d FBA: Flux Balance Analysis; ^e PLS-PM: Partial least squares - path model; ^f EFM: elementary flux mode; ^g RNA-Seq: Ribonucleic acid sequencing; ^h RT-qPCR: Real Time quantitative polymerase chain reaction; ⁱ qPCR: Quantitative polymerase chain reaction; ^j SIP: stable isotope probing; ^k HPLC: High-performance liquid chromatography; ^l CGQ: cell growth quantifier; ^m GC-MS: Gas chromatography mass spectrometry; ⁿ cFBA: Community Flux Balance Analysis.

4.3.2. Defining an Ecosystem Process and Links between Genes, Enzymes and Reactions for a Given Ecosystem Process

Assessing the contribution of species to any given ecosystem process requires functional annotation of the predicted coding sequences to the reactions involved in said process. The National Center for Biotechnology Information (NCBI) [97] is one of the largest repositories for functional annotation [98]. However, mining sequencing data from a large non-specialized database is time-consuming and requires substantial manual curation, as users need to search for all links between genes, enzymes and reactions and to identify and to correct misannotated entries [99]. On the other hand of the spectrum, subject-specific databases are often built upon experimentally validated and/or manually curated data. For example, the Comprehensive Enzyme Information System (BRENDA) [100] collects information on functional enzymes and metabolism. BRENDA's data includes manual curation obtained from text mining and linked to other curated databases of protein sequences such as the Swiss-Prot [101] from the Uniprot Knowledgebase [102]. In the study by Tláskal and collaborators [91], the authors used one general and two specialized databases for functional annotation. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [103], a non-specialized database, was used to identify all genes in deadwood present in the KEGG database. Using its output, Tláskal and collaborators [91] identified species with the genetic potential to perform methylotrophy and nitrogen cycling. Further, two specialized databases were used to annotate genes involved in carbohydrate utilization in bacteria; respectively, Functional Ontology Assignments for Metagenomes (FOAM) [104] and database of carbohydrate-active enzyme (CAZyme) sequence and annotation (dbCAN HMM) database [105].

4.3.3. Mining Putative Species Interactions

After functional annotation, microbial interactions can be mined by mapping the genetic potential of the different species to each reaction in the selected ecosystem process. In the study by Tláskal and collaborators [91], the contributions of each species to deadwood decomposition was manually assessed by integrating species abundance profiles and relative abundance of genes involved in deadwood decomposition. As a semi-automated alternative, the HMP Unified Metabolic Analysis Network (HumanN2) [5] processes metagenomic and metatranscriptomic data allowing the identification of microbial interactions in complex microbial communities. In the near future, studies involving high-throughput mining of microbial interactions may become broadly used by microbiologists. Therefore, the scientific community is developing automated pipelines such as the Network-Based Tool for Predicting Metabolic Capacities of Microbial Species and their Interactions (NetMet) [106]. NetMet performs automated mining of microbial interactions by linking species-specific enzymatic reactions to metabolites in user-defined environments. In addition, the number of potential combinations can create computational hurdles for mining microbial interactions (Equation (1)). Therefore, limiting the search space is an important step to predict microbial interactions and may be done in silico or in vivo. In silico approaches may consist of correlation analysis or machine learning identification of key microbial species, as previously demonstrated in different environments; such as, coral reefs [107], bioreactors [108] and batch cultures [108]. Current in vivo approaches to deal with the search space involve co-cultures and the use of stable isotopes. Co-cultures limit the search space by controlling the number of species used in the experiments [109]. In addition, stable isotope labeling may limit the search space for the active metabolites [110], proteins [111] or species [112] directly involved in transformations of the labeled compounds.

4.3.4. Validating Microbial Interactions

As discussed in Section 4.2, validation of microbial interaction may consist of the integration of other data beyond genetic potential. For example, Tláskal and collaborators [91] integrated gene expression patterns obtained from metatranscriptomics and experimentally

determined C and N gas fluxes with the metagenomic data to validate the microbial contributions to deadwood decomposition. Microbial interactions were validated when genes mapped to each species and their expression patterns correlated with C and N production rates. However, these links are based on statistical methods (such as correlation analysis) which do not guarantee causality. Furthermore, correlation analysis does not necessarily identify indirect interactions between species [113]. Although integration of metagenomics, metatranscriptomics, CO₂ production and nitrogen fixation measurements increased substantially the robustness of the predictions, further validation may expand the conclusions achieved by Tláskal and collaborators [91]. For example, studies have shown that spatial structure of microbial communities influence ecosystem processes [114–116] and should be taken into account. The 3D structure of the microbial communities could be determined using a molecular ecology network approach as previously demonstrated in biological soil crusts [117]. Tracking the flow of ¹³C-labelled substrates through the metabolism of microbes has also been shown as a valuable method to assess microbial community activity and function [118] even in highly complex and diverse environments [112,119]. Nevertheless, such an approach cannot be carried out in natural forest studies such as the one by Tláskal and collaborators [91].

In summary, each of the different steps from mining to validation of microbial interactions is not limited to a single method or strategy. The selection of approaches to determine microbial interactions will depend on the research questions posed by a study, its source material, the acquisition of the genetic potential for species in a given community and complexity of ecosystem processes of interest.

5. Conclusions

Understanding how interspecies interactions contribute to ecosystem functioning is a central issue not only in microbiology but the large field of ecology. Rather than solely depending on diversity measures or correlations, future research should be directed into searching mechanisms underlying causal relationships between community components and their abiotic environments. To optimize the road to uncover interspecies interactions, it is of utmost importance to first identify an ecosystem process of interest. It is relevant to highlight that this ecosystem process should have known connections between genes, proteins and reactions. The next step is to collect genomes of all (or key) species in the microbial community of interest. Mapping of the different reactions to the different genomes is a crucial step as it determines the functional potential of each individual species. This step should be carefully performed since the databases being used to link gene, proteins and reactions are constantly being updated. One must be aware that the size of the pathways involved with the ecosystem process of interest together with the number of species in the respective microbial community has a high-impact on the computational demands to determine microbial interactions and the interpretation of the analysis. In silico correlation analysis and machine learning or in vivo experiments involving co-cultures and stable isotope labeling may circumvent this hurdle. In addition, genetic potential alone is not indicative of microbial interactions. Therefore, the methods needed to validate putative microbial interactions should be chosen based on the complexity of the ecosystem process, the diversity of the microbial community and other biotic and abiotic factors in the ecosystem.

Author Contributions: Conceptualization, U.N.d.R.; Writing—Original Draft Preparation, J.P.S. and U.N.d.R.; writing—review and editing, U.N.d.R., A.W., C.K., R.K., F.C. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Helmholtz Young Investigator grant VH-NG-1248 Micro “Big Data”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Currie, W.S. Units of Nature or Processes across Scales? The Ecosystem Concept at Age 75. *New Phytol.* **2011**, *190*, 21–34. [CrossRef]
- Song, H.-S.; Cannon, W.; Beliaev, A.; Konopka, A. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. *Processes* **2014**, *2*, 711–752. [CrossRef]
- Gentry, T.J.; Pepper, I.L.; Pierson, L.S. Chapter 19—Microbial Diversity and Interactions in Natural Ecosystems. In *Environmental Microbiology*, 3rd ed.; Pepper, I.L., Gerba, C.P., Gentry, T.J., Eds.; Academic Press: San Diego, CA, USA, 2015; pp. 441–460. ISBN 978-0-12-394626-3.
- Stams, A.J.M.; Bok, F.A.M.D.; Plugge, C.M.; Eekert, M.H.A.V.; Dolfig, J.; Schraa, G. Exocellular Electron Transfer in Anaerobic Microbial Communities. *Environ. Microbiol.* **2006**, *8*, 371–382. [CrossRef] [PubMed]
- Franzosa, E.A.; McIver, L.J.; Rahnard, G.; Thompson, L.R.; Schirmer, M.; Weingart, G.; Lipson, K.S.; Knight, R.; Caporaso, J.G.; Segata, N.; et al. Species-Level Functional Profiling of Metagenomes and Metatranscriptomes. *Nat. Methods* **2018**, *15*, 962–968. [CrossRef]
- Hall, E.K.; Bernhardt, E.S.; Bier, R.L.; Bradford, M.A.; Boot, C.M.; Cotner, J.B.; del Giorgio, P.A.; Evans, S.E.; Graham, E.B.; Jones, S.E.; et al. Understanding How Microbiomes Influence the Systems They Inhabit. *Nat. Microbiol.* **2018**, *3*, 977–982. [CrossRef] [PubMed]
- Baas Becking, L.G.M.; Nicolai, E. On the Ecology of a Sphagnum Bog. *Blumea Biodivers. Evol. Biogeogr. Plants* **1934**, *1*, 10–45.
- Girguis, P. Microbial Ecology: Here, There and Everywhere. *Nat. Microbiol.* **2016**, *1*, 16123. [CrossRef]
- Bell, T.; Newman, J.A.; Silverman, B.W.; Turner, S.L.; Lilley, A.K. The Contribution of Species Richness and Composition to Bacterial Services. *Nature* **2005**, *436*, 1157–1160. [CrossRef]
- Oh, Y.K.; Palsson, B.O.; Park, S.M.; Schilling, C.H.; Mahadevan, R. Genome-Scale Reconstruction of Metabolic Network in *Bacillus Subtilis* Based on High-Throughput Phenotyping and Gene Essentiality Data. *J. Biol. Chem.* **2007**, *282*, 28791–28799. [CrossRef] [PubMed]
- Huddlestone, J.R. Horizontal Gene Transfer in the Human Gastrointestinal Tract: Potential Spread of Antibiotic Resistance Genes. *Infect. Drug Resist.* **2014**, *7*, 167–176. [CrossRef]
- Tyc, O.; Song, C.; Dickschat, J.S.; Vos, M.; Garbeva, P. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends Microbiol.* **2017**, *25*, 280–292. [CrossRef] [PubMed]
- Seth, E.C.; Taga, M.E. Nutrient Cross-Feeding in the Microbial World. *Front. Microbiol.* **2014**, *5*, 350. [CrossRef] [PubMed]
- Fetzer, I.; Johst, K.; Schäwe, R.; Banitz, T.; Harms, H.; Chatzinotas, A. The Extent of Functional Redundancy Changes as Species' Roles Shift in Different Environments. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14888–14893. [CrossRef]
- Gawad, C.; Koh, W.; Quake, S.R. Single-Cell Genome Sequencing: Current State of the Science. *Nat. Rev. Genet.* **2016**, *17*, 175–188. [CrossRef] [PubMed]
- Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.-A.; Woodcroft, B.J.; Evans, P.N.; Hugenholtz, P.; Tyson, G.W. Recovery of Nearly 8000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life. *Nat. Microbiol.* **2017**, *2*, 1533–1542. [CrossRef] [PubMed]
- Nayfach, S.; Roux, S.; Seshadri, R.; Udwy, D.; Varghese, N.; Schulz, F.; Wu, D.; Paez-Espino, D.; Chen, I.-M.; Huntemann, M.; et al. A Genomic Catalog of Earth's Microbiomes. *Nat. Biotechnol.* **2020**. [CrossRef]
- Da Rocha, U.N.; Cadillo-Quiroz, H.; Karaoz, U.; Rajeev, L.; Klitgord, N.; Dunn, S.; Truong, V.; Buenrostro, M.; Bowen, B.P.; Garcia-Pichel, F.; et al. Isolation of a Significant Fraction of Non-Phototroph Diversity from a Desert Biological Soil Crust. *Front. Microbiol.* **2015**, *6*, 277. [CrossRef]
- Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **2017**, *8*, 2224. [CrossRef]
- Clingenpeel, S.; Clum, A.; Schwientek, P.; Rinke, C.; Woyke, T. Reconstructing Each Cell's Genome within Complex Microbial Communities—Dream or Reality? *Front. Microbiol.* **2014**, *5*, 771. [CrossRef]
- Velsko, I.M.; Frantz, L.A.F.; Herbig, A.; Larson, G.; Warinner, C. Selection of Appropriate Metagenome Taxonomic Classifiers for Ancient Microbiome Research. *mSystems* **2018**, *3*. [CrossRef]
- Frigg, R.; Hartmann, S. Models in Science. 2006. Available online: <https://seop.illc.uva.nl/entries/models-science/> (accessed on 26 January 2021).
- Kirwan, L.; Connolly, J.; Finn, J.A.; Brophy, C.; Lüscher, A.; Nyfeler, D.; Sebastià, M.-T. Diversity–Interaction Modeling: Estimating Contributions of Species Identities and Interactions to Ecosystem Function. *Ecology* **2009**, *90*, 2032–2038. [CrossRef] [PubMed]
- Kreft, J.-U.; Plugge, C.M.; Prats, C.; Leveau, J.H.J.; Zhang, W.; Hellweger, F.L. From Genes to Ecosystems in Microbiology: Modeling Approaches and the Importance of Individuality. *Front. Microbiol.* **2017**, *8*, 2299. [CrossRef]
- Biggs, M.B.; Medlock, G.L.; Kolling, G.L.; Papin, J.A. Metabolic Network Modeling of Microbial Communities. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2015**, *7*, 317–334. [CrossRef]
- Yarygin, K.; Tyakht, A.; Larin, A.; Kostyukova, E.; Kolchenko, S.; Bitner, V.; Alexeev, D. Abundance Profiling of Specific Gene Groups Using Precomputed Gut Metagenomes Yields Novel Biological Hypotheses. *PLoS ONE* **2017**, *12*, e0176154. [CrossRef]

27. Khandelwal, R.A.; Olivier, B.G.; Röling, W.F.M.; Teusink, B.; Bruggeman, F.J. Community Flux Balance Analysis for Microbial Consortia at Balanced Growth. *PLoS ONE* **2013**, *8*, e64567. [[CrossRef](#)]
28. Rocha, I.; Förster, J.; Nielsen, J. Design and Application of Genome-Scale Reconstructed Metabolic Models. In *Microbial Gene Essentiality: Protocols and Bioinformatics*; Methods in Molecular Biology™ Book Series; Humana Press: Totowa, NJ, USA, 2008; Volume 416, pp. 409–431.
29. Dias, O.; Rocha, M.; Ferreira, E.C.; Rocha, I. Reconstructing Genome-Scale Metabolic Models with Merlin. *Nucleic Acids Res.* **2015**, *43*, 3899–3910. [[CrossRef](#)]
30. Gottstein, W.; Olivier, B.G.; Bruggeman, F.J.; Teusink, B. Constraint-Based Stoichiometric Modelling from Single Organisms to Microbial Communities. *J. R. Soc. Interface* **2016**, *13*, 20160627. [[CrossRef](#)]
31. Gu, C.; Kim, G.B.; Kim, W.J.; Kim, H.U.; Lee, S.Y. Current Status and Applications of Genome-Scale Metabolic Models. *Genome Biol.* **2019**, *20*, 121. [[CrossRef](#)]
32. Blondel, J. Guilds or Functional Groups: Does It Matter? *Oikos* **2003**, *100*, 223–231. [[CrossRef](#)]
33. Kettle, H.; Louis, P.; Holtrop, G.; Duncan, S.H.; Flint, H.J. Modelling the Emergent Dynamics and Major Metabolites of the Human Colonic Microbiota. *Environ. Microbiol.* **2015**, *17*, 1615–1630. [[CrossRef](#)] [[PubMed](#)]
34. Hanemaaijer, M.; Röling, W.F.M.; Olivier, B.G.; Khandelwal, R.A.; Teusink, B.; Bruggeman, F.J. Systems Modeling Approaches for Microbial Community Studies: From Metagenomics to Inference of the Community Structure. *Front. Microbiol.* **2015**, *6*, 213. [[CrossRef](#)] [[PubMed](#)]
35. Talbot, J.M.; Martin, F.; Kohler, A.; Henrissat, B.; Peay, K.G. Functional Guild Classification Predicts the Enzymatic Role of Fungi in Litter and Soil Biogeochemistry. *Soil Biol. Biochem.* **2015**, *88*, 441–456. [[CrossRef](#)]
36. Faust, K.; Raes, J. Microbial Interactions: From Networks to Models. *Nat. Rev. Microbiol.* **2012**, *10*, 538–550. [[CrossRef](#)] [[PubMed](#)]
37. Thiele, I.; Palsson, B.Ø. A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121. [[CrossRef](#)]
38. Rodríguez, J.; Kleerebezem, R.; Lema, J.M.; van Loosdrecht, M.C.M. Modeling Product Formation in Anaerobic Mixed Culture Fermentations. *Biotechnol. Bioeng.* **2006**, *93*, 592–606. [[CrossRef](#)]
39. Tringe, S.G.; von Mering, C.; Kobayashi, A.; Salamov, A.A.; Chen, K.; Chang, H.W.; Podar, M.; Short, J.M.; Mathur, E.J.; Detter, J.C.; et al. Comparative Metagenomics of Microbial Communities. *Science* **2005**, *308*, 554–557. [[CrossRef](#)]
40. Hay, M.E.; Parker, J.D.; Burkepille, D.E.; Caudill, C.C.; Wilson, A.E.; Hallinan, Z.P.; Chequer, A.D. Mutualisms and Aquatic Community Structure: The Enemy of My Enemy Is My Friend. *Annu. Rev. Ecol. Evol. Syst.* **2004**, *35*, 175–197. [[CrossRef](#)]
41. Boetius, A.; Ravenschlag, K.; Schubert, C.J.; Rickert, D.; Widdel, F.; Gieseke, A.; Amann, R.; Jørgensen, B.B.; Witte, U.; Pfannkuche, O. A Marine Microbial Consortium Apparently Mediating Anaerobic Oxidation of Methane. *Nature* **2000**, *407*, 623–626. [[CrossRef](#)]
42. Traore, A.S.; Fardeau, M.-L.; Hatchikian, C.E.; Le Gall, J.; Belaich, J.-P. Energetics of Growth of a Defined Mixed Culture of *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: Interspecies Hydrogen Transfer in Batch and Continuous Cultures. *Appl. Environ. Microbiol.* **1983**, *46*, 1152–1156. [[CrossRef](#)]
43. Chuang, J.S.; Rivoire, O.; Leibler, S. Simpson’s Paradox in a Synthetic Microbial System. *Science* **2009**, *323*, 272–275. [[CrossRef](#)]
44. Wertz, S.; Degrange, V.; Prosser, J.I.; Poly, F.; Commeaux, C.; Guillaumaud, N.; Roux, X.L. Decline of Soil Microbial Diversity Does Not Influence the Resistance and Resilience of Key Soil Microbial Functional Groups Following a Model Disturbance. *Environ. Microbiol.* **2007**, *9*, 2211–2219. [[CrossRef](#)]
45. Gill, A.S.; Purnell, K.; Palmer, M.I.; Stein, J.; McGuire, K.L. Microbial Composition and Functional Diversity Differ across Urban Green Infrastructure Types. *Front. Microbiol.* **2020**, *11*, 912. [[CrossRef](#)] [[PubMed](#)]
46. Jiménez, D.J.; Chaves-Moreno, D.; van Elsas, J.D. Unveiling the Metabolic Potential of Two Soil-Derived Microbial Consortia Selected on Wheat Straw. *Sci. Rep.* **2015**, *5*, 13845. [[CrossRef](#)] [[PubMed](#)]
47. Jeffries, T.C.; Rayu, S.; Nielsen, U.N.; Lai, K.; Ijaz, A.; Nazaries, L.; Singh, B.K. Metagenomic Functional Potential Predicts Degradation Rates of a Model Organophosphorus Xenobiotic in Pesticide Contaminated Soils. *Front. Microbiol.* **2018**, *9*, 147. [[CrossRef](#)] [[PubMed](#)]
48. Rondon, M.R.; August, P.R.; Bettermann, A.D.; Brady, S.F.; Grossman, T.H.; Liles, M.R.; Loiacono, K.A.; Lynch, B.A.; MacNeil, I.A.; Minor, C.; et al. Cloning the Soil Metagenome: A Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Appl. Environ. Microbiol.* **2000**, *66*, 2541–2547. [[CrossRef](#)] [[PubMed](#)]
49. Van Rossum, T.; Ferretti, P.; Maistrenko, O.M.; Bork, P. Diversity within Species: Interpreting Strains in Microbiomes. *Nat. Rev. Microbiol.* **2020**, *18*, 491–506. [[CrossRef](#)]
50. Prosser, J.I. Dispersing Misconceptions and Identifying Opportunities for the Use of “omics” in Soil Microbial Ecology. *Nat. Rev. Microbiol.* **2015**, *13*, 439–446. [[CrossRef](#)]
51. Ance Meyers, L.; Ance, F.D.; Lachmann, M. Evolution of Genetic Potential. *PLoS Comput. Biol.* **2005**, *1*, e32. [[CrossRef](#)]
52. Garcia-Pichel, E.; Johnson, S.L.; Youngkin, D.; Belnap, J. Small-Scale Vertical Distribution of Bacterial Biomass and Diversity in Biological Soil Crusts from Arid Lands in the Colorado Plateau. *Microb. Ecol.* **2003**, *46*, 312–321. [[CrossRef](#)]
53. Querton, T.; Kang, T.; Papakis, V.; Nguyen, K.; Nowak, C.; Li, Y.; Bleris, L. Uncoupling Gene Expression Noise along the Central Dogma Using Genome Engineered Human Cell Lines. *Nucleic Acids Res.* **2020**, *48*, 9406–9413. [[CrossRef](#)]
54. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. Studying Gene Expression and Function. In *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, NY, USA, 2002.
55. Doolittle, W.F.; Zhaxybayeva, O. Metagenomics and the Units of Biological Organization. *BioScience* **2010**, *60*, 102–112. [[CrossRef](#)]

56. Jansson, J.K.; Hofmockel, K.S. The Soil Microbiome—from Metagenomics to Metaphenomics. *Curr. Opin. Microbiol.* **2018**, *43*, 162–168. [[CrossRef](#)] [[PubMed](#)]
57. Harrison, J.J.; Ceri, H.; Yerly, J.; Stremick, C.A.; Hu, Y.; Martinuzzi, R.; Turner, R.J. The Use of Microscopy and Three-Dimensional Visualization to Evaluate the Structure of Microbial Biofilms Cultivated in the Calgary Biofilm Device. *Biol. Proced. Online* **2006**, *8*, 194–215. [[CrossRef](#)] [[PubMed](#)]
58. Stubbendieck, R.M.; Vargas-Bautista, C.; Straight, P.D. Bacterial Communities: Interactions to Scale. *Front. Microbiol.* **2016**, *7*, 1234. [[CrossRef](#)] [[PubMed](#)]
59. Natale, A.D.; Mele, B.H.; Cennamo, P.; Mondo, A.D.; Petraretti, M.; Pollio, A. Microbial Biofilm Community Structure and Composition on the Lithic Substrates of Herculaneum Suburban Baths. *PLoS ONE* **2020**, *15*, e0232512. [[CrossRef](#)]
60. Gonzalez-Gil, G.; Thomas, L.; Emwas, A.-H.; Lens, P.N.L.; Saikaly, P.E. NMR and MALDI-TOF MS Based Characterization of Exopolysaccharides in Anaerobic Microbial Aggregates from Full-Scale Reactors. *Sci. Rep.* **2015**, *5*, 14316. [[CrossRef](#)]
61. Ruppé, E.; Ghazlane, A.; Tap, J.; Pons, N.; Alvarez, A.-S.; Maziers, N.; Cuesta, T.; Hernando-Amado, S.; Clares, I.; Martínez, J.L.; et al. Prediction of the Intestinal Resistome by a Three-Dimensional Structure-Based Method. *Nat. Microbiol.* **2019**, *4*, 112–123. [[CrossRef](#)]
62. Starr, E.P.; Shi, S.; Blazewicz, S.J.; Probst, A.J.; Herman, D.J.; Firestone, M.K.; Banfield, J.F. Stable Isotope Informed Genome-Resolved Metagenomics Reveals That Saccharibacteria Utilize Microbially-Processed Plant-Derived Carbon. *Microbiome* **2018**, *6*, 122. [[CrossRef](#)] [[PubMed](#)]
63. Fortunato, C.S.; Huber, J.A. Coupled RNA-SIP and Metatranscriptomics of Active Chemolithoautotrophic Communities at a Deep-Sea Hydrothermal Vent. *ISME J.* **2016**, *10*, 1925–1938. [[CrossRef](#)]
64. Kong, Y.; Kuzyakov, Y.; Ruan, Y.; Zhang, J.; Wang, T.; Wang, M.; Guo, S.; Shen, Q.; Ling, N. DNA Stable-Isotope Probing Delineates Carbon Flows from Rice Residues into Soil Microbial Communities Depending on Fertilization. *Appl. Environ. Microbiol.* **2020**, *86*, e02151-19. [[CrossRef](#)]
65. Seifert, J.; Taubert, M.; Jehmlich, N.; Schmidt, F.; Völker, U.; Vogt, C.; Richnow, H.-H.; von Bergen, M. Protein-Based Stable Isotope Probing (Protein-SIP) in Functional Metaproteomics. *Mass Spectrom. Rev.* **2012**, *31*, 683–697. [[CrossRef](#)] [[PubMed](#)]
66. Chokkathukalam, A.; Kim, D.-H.; Barrett, M.P.; Breitling, R.; Creek, D.J. Stable Isotope-Labeling Studies in Metabolomics: New Insights into Structure and Dynamics of Metabolic Networks. *Bioanalysis* **2014**, *6*, 511–524. [[CrossRef](#)] [[PubMed](#)]
67. Musat, N.; Musat, F.; Weber, P.K.; Pett-Ridge, J. Tracking Microbial Interactions with NanoSIMS. *Curr. Opin. Biotechnol.* **2016**, *41*, 114–121. [[CrossRef](#)] [[PubMed](#)]
68. Thompson, L.R.; Sanders, J.G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K.J.; Prill, R.J.; Tripathi, A.; Gibbons, S.M.; Ackermann, G.; et al. A Communal Catalogue Reveals Earth’s Multiscale Microbial Diversity. *Nature* **2017**, *551*, 457–463. [[CrossRef](#)] [[PubMed](#)]
69. Turnbaugh, P.J.; Ley, R.E.; Hamady, M.; Fraser-Liggett, C.M.; Knight, R.; Gordon, J.I. The Human Microbiome Project. *Nature* **2007**, *449*, 804–810. [[CrossRef](#)] [[PubMed](#)]
70. Sunagawa, S.; Acinas, S.G.; Bork, P.; Bowler, C.; Eveillard, D.; Gorsky, G.; Guidi, L.; Iudicone, D.; Karsenti, E.; Lombard, F.; et al. Tara Oceans: Towards Global Ocean Ecosystems Biology. *Nat. Rev. Microbiol.* **2020**, *18*, 428–445. [[CrossRef](#)] [[PubMed](#)]
71. Manzoni, C.; Kia, D.A.; Vandrovцова, J.; Hardy, J.; Wood, N.W.; Lewis, P.A.; Ferrari, R. Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences. *Brief. Bioinform.* **2018**, *19*, 286–302. [[CrossRef](#)]
72. Lauber, C.L.; Ramirez, K.S.; Aanderud, Z.; Lennon, J.; Fierer, N. Temporal Variability in Soil Microbial Communities across Land-Use Types. *ISME J.* **2013**, *7*, 1641–1650. [[CrossRef](#)]
73. Hannula, S.E.; Kielak, A.M.; Steinauer, K.; Huberty, M.; Jongen, R.; Long, J.R.D.; Heinen, R.; Bezemer, T.M. Time after Time: Temporal Variation in the Effects of Grass and Forb Species on Soil Bacterial and Fungal Communities. *mBio* **2019**, *10*, e02635-19. [[CrossRef](#)]
74. Ladau, J.; Eloë-Fadrosch, E.A. Spatial, Temporal, and Phylogenetic Scales of Microbial Ecology. *Trends Microbiol.* **2019**, *27*, 662–669. [[CrossRef](#)]
75. Karakoç, C.; Clark, A.T.; Chatzinotas, A. Diversity and Coexistence Are Influenced by Time-Dependent Species Interactions in a Predator–Prey System. *Ecol. Lett.* **2020**, *23*, 983–993. [[CrossRef](#)]
76. Bar-Joseph, Z.; Gitter, A.; Simon, I. Studying and Modelling Dynamic Biological Processes Using Time-Series Gene Expression Data. *Nat. Rev. Genet.* **2012**, *13*, 552–564. [[CrossRef](#)]
77. Da Rocha, U.N.; van Elsas, J.D.; van Overbeek, L.S. Real-Time PCR Detection of *Holophagae* (*Acidobacteria*) and *Verrucomicrobia* Subdivision 1 Groups in Bulk and Leek (*Allium porrum*) Rhizosphere Soils. *J. Microbiol. Methods* **2010**, *83*, 141–148. [[CrossRef](#)]
78. Kim, E.; Cho, E.-J.; Yang, S.-M.; Kim, M.-J.; Kim, H.-Y. Novel Approaches for the Identification of Microbial Communities in Kimchi: MALDI-TOF MS Analysis and High-Throughput Sequencing. *Food Microbiol.* **2021**, *94*, 103641. [[CrossRef](#)]
79. Moraes, L.C.; Lang, P.M.; Arcanjo, R.A.; Rampelotto, P.H.; Fatturi-Parolo, C.C.; Ferreira, M.B.C.; Montagner, F. Microbial Ecology and Predicted Metabolic Pathways in Various Oral Environments from Patients with Acute Endodontic Infections. *Int. Endod. J.* **2020**, *53*, 1603–1617. [[CrossRef](#)]
80. Basile, A.; Campanaro, S.; Kovalovszki, A.; Zampieri, G.; Rossi, A.; Angelidaki, I.; Valle, G.; Treu, L. Revealing Metabolic Mechanisms of Interaction in the Anaerobic Digestion Microbiome by Flux Balance Analysis. *Metab. Eng.* **2020**, *62*, 138–149. [[CrossRef](#)] [[PubMed](#)]

81. Kumar, A.; Padhy, S.R.; Das, R.R.; Shahid, M.; Dash, P.K.; Senapati, A.; Panneerselvam, P.; Kumar, U.; Chatterjee, D.; Adak, T.; et al. Elucidating Relationship between Nitrous Oxide Emission and Functional Soil Microbes from Tropical Lowland Rice Soil Exposed to Elevated CO₂: A Path Modelling Approach. *Agric. Ecosyst. Environ.* **2021**, *308*, 107268. [[CrossRef](#)]
82. Yan, B.; Lu, Q.; He, J.; Qi, Y.; Fu, G.; Xiao, N.; Li, J. Composition and Interaction Frequencies in Soil Bacterial Communities Change in Association with Urban Park Age in Beijing. *Pedobiologia* **2021**, *84*, 150699. [[CrossRef](#)]
83. Yang, X.; Zhu, K.; Loik, M.E.; Sun, W. Differential Responses of Soil Bacteria and Fungi to Altered Precipitation in a Meadow Steppe. *Geoderma* **2021**, *384*, 114812. [[CrossRef](#)]
84. Hunt, K.A.; Jennings, R.M.; Inskeep, W.P.; Carlson, R.P. Multiscale Analysis of Autotroph-Heterotroph Interactions in a High-Temperature Microbial Community. *PLoS Comput. Biol.* **2018**, *14*, e1006431. [[CrossRef](#)] [[PubMed](#)]
85. Mencher, A.; Morales, P.; Curiel, J.A.; Gonzalez, R.; Tronchoni, J. *Metschnikowia pulcherrima* Represses Aerobic Respiration in *Saccharomyces cerevisiae* Suggesting a Direct Response to Co-Cultivation. *Food Microbiol.* **2021**, *94*, 103670. [[CrossRef](#)]
86. Yuan, H.; Huang, S.; Yuan, J.; You, Y.; Zhang, Y. Characteristics of Microbial Denitrification under Different Aeration Intensities: Performance, Mechanism, and Co-Occurrence Network. *Sci. Total Environ.* **2021**, *754*, 141965. [[CrossRef](#)] [[PubMed](#)]
87. Zhang, T.; Yin, Q.; Shi, Y.; Wu, G. Microbial Physiology and Interactions in Anammox Systems with the Intermittent Addition of Organic Carbons. *Bioresour. Technol.* **2021**, *319*, 124226. [[CrossRef](#)] [[PubMed](#)]
88. Islam, M.A.; Karim, A.; Mishra, P.; Dubowski, J.J.; Yousuf, A.; Sarmin, S.; Khan, M.M.R. Microbial Synergistic Interactions Enhanced Power Generation in Co-Culture Driven Microbial Fuel Cell. *Sci. Total Environ.* **2020**, *738*, 140138. [[CrossRef](#)]
89. Ravikrishnan, A.; Blank, L.M.; Srivastava, S.; Raman, K. Investigating Metabolic Interactions in a Microbial Co-Culture through Integrated Modelling and Experiments. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1249–1258. [[CrossRef](#)]
90. Zomorodi, A.R.; Segrè, D. Genome-Driven Evolutionary Game Theory Helps Understand the Rise of Metabolic Interdependencies in Microbial Communities. *Nat. Commun.* **2017**, *8*, 1563. [[CrossRef](#)] [[PubMed](#)]
91. Tláskal, V.; Brabcová, V.; Větrovský, T.; Jomura, M.; López-Mondéjar, R.; Monteiro, L.M.O.; Saraiva, J.P.; Human, Z.R.; Cajthaml, T.; da Rocha, U.N.; et al. Complementary Roles of Wood-Inhabiting Fungi and Bacteria Facilitate Deadwood Decomposition. *mSystems* **2021**, *6*, e01078-20. [[CrossRef](#)]
92. Da Rocha, U.N.; Van Overbeek, L.; Elsas, V.; Dirk, J. Exploration of Hitherto-Uncultured Bacteria from the Rhizosphere. *FEMS Microbiol. Ecol.* **2009**, *69*, 313–328. [[CrossRef](#)] [[PubMed](#)]
93. Chen, L.-X.; Anantharaman, K.; Shaiber, A.; Eren, A.M.; Banfield, J.F. Accurate and Complete Genomes from Metagenomes. *Genome Res.* **2020**, *30*, 315–333. [[CrossRef](#)]
94. Kaster, A.-K.; Sobol, M.S. Microbial Single-Cell Omics: The Crux of the Matter. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 8209–8220. [[CrossRef](#)]
95. Lagier, J.-C.; Hugon, P.; Khelaifia, S.; Fournier, P.-E.; Scola, B.L.; Raoult, D. The Rebirth of Culture in Microbiology through the Example of Culturomics To Study Human Gut Microbiota. *Clin. Microbiol. Rev.* **2015**, *28*, 237–264. [[CrossRef](#)]
96. Karimi, E.; Keller-Costa, T.; Slaby, B.M.; Cox, C.J.; da Rocha, U.N.; Hentschel, U.; Costa, R. Genomic Blueprints of Sponge-Prokaryote Symbiosis Are Shared by Low Abundant and Cultivable Alphaproteobacteria. *Sci. Rep.* **2019**, *9*, 1999. [[CrossRef](#)]
97. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [[CrossRef](#)]
98. Attwood, T.K.; Agit, B.; Ellis, L.B.M. Longevity of Biological Databases. *EMBnet. J.* **2015**, *21*, e803. [[CrossRef](#)]
99. Schnoes, A.M.; Brown, S.D.; Dodevski, I.; Babbitt, P.C. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* **2009**, *5*, e1000605. [[CrossRef](#)] [[PubMed](#)]
100. Chang, A.; Jeske, L.; Ulbrich, S.; Hofmann, J.; Koblitiz, J.; Schomburg, I.; Neumann-Schaal, M.; Jahn, D.; Schomburg, D. BRENDA, the ELIXIR Core Data Resource in 2021: New Developments and Updates. *Nucleic Acids Res.* **2021**, *49*, D498–D508. [[CrossRef](#)] [[PubMed](#)]
101. Bairoch, A.; Apweiler, R. The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TrEMBL. *Nucleic Acids Res.* **1996**, *24*, 21–25. [[CrossRef](#)] [[PubMed](#)]
102. The UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)] [[PubMed](#)]
103. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [[CrossRef](#)]
104. Prestat, E.; David, M.M.; Hultman, J.; Taş, N.; Lamendella, R.; Dvornik, J.; Mackelprang, R.; Myrold, D.D.; Jumpponen, A.; Tringe, S.G.; et al. FOAM (Functional Ontology Assignments for Metagenomes): A Hidden Markov Model (HMM) Database with Environmental Focus. *Nucleic Acids Res.* **2014**, *42*, e145. [[CrossRef](#)]
105. Huang, L.; Zhang, H.; Wu, P.; Entwistle, S.; Li, X.; Yohe, T.; Yi, H.; Yang, Z.; Yin, Y. DbCAN-Seq: A Database of Carbohydrate-Active Enzyme (CAZyme) Sequence and Annotation. *Nucleic Acids Res.* **2018**, *46*, D516–D521. [[CrossRef](#)]
106. Tal, O.; Selvaraj, G.; Medina, S.; Ofaim, S.; Freilich, S. NetMet: A Network-Based Tool for Predicting Metabolic Capacities of Microbial Species and Their Interactions. *Microorganisms* **2020**, *8*, 840. [[CrossRef](#)] [[PubMed](#)]
107. Rosado, P.M.; Leite, D.C.A.; Duarte, G.A.S.; Chaloub, R.M.; Jospin, G.; da Rocha, U.N.; Saraiva, J.P.; Dini-Andreote, F.; Eisen, J.A.; Bourne, D.G.; et al. Marine Probiotics: Increasing Coral Resistance to Bleaching through Microbiome Manipulation. *ISME J.* **2019**, *13*, 921–936. [[CrossRef](#)]

108. Lian, S.; Nikolausz, M.; Nijenhuis, I.; da Rocha, U.N.; Liu, B.; Corrêa, F.B.; Saraiva, J.P.; Richnow, H.H. Biotransformation of Hexachlorocyclohexanes Contaminated Biomass for Energetic Utilization Demonstrated in Continuous Anaerobic Digestion System. *J. Hazard. Mater.* **2020**, *384*, 121448. [[CrossRef](#)]
109. Aranda-Díaz, A.; Obadia, B.; Dodge, R.; Thomsen, T.; Hallberg, Z.F.; Güvener, Z.T.; Ludington, W.B.; Huang, K.C. Bacterial Interspecies Interactions Modulate PH-Mediated Antibiotic Tolerance. *eLife* **2020**, *9*, e51493. [[CrossRef](#)] [[PubMed](#)]
110. Yamazawa, A.; Iikura, T.; Morioka, Y.; Shino, A.; Ogata, Y.; Date, Y.; Kikuchi, J. Cellulose Digestion and Metabolism Induced Biocatalytic Transitions in Anaerobic Microbial Ecosystems. *Metabolites* **2014**, *4*, 36–52. [[CrossRef](#)] [[PubMed](#)]
111. Herbst, F.-A.; Bahr, A.; Duarte, M.; Pieper, D.H.; Richnow, H.-H.; von Bergen, M.; Seifert, J.; Bombach, P. Elucidation of In Situ Polycyclic Aromatic Hydrocarbon Degradation by Functional Metaproteomics (Protein-SIP). *Proteomics* **2013**, *13*, 2910–2920. [[CrossRef](#)] [[PubMed](#)]
112. Drigo, B.; Pijl, A.S.; Duyts, H.; Kielak, A.M.; Gamper, H.A.; Houtekamer, M.J.; Boschker, H.T.S.; Bodelier, P.L.E.; Whiteley, A.S.; van Veen, J.A.; et al. Shifting Carbon Flow from Roots into Associated Microbial Communities in Response to Elevated Atmospheric CO₂. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10938–10942. [[CrossRef](#)]
113. Biswas, S.; McDonald, M.; Lundberg, D.S.; Dangl, J.L.; Jojic, V. Learning Microbial Interaction Networks from Metagenomic Count Data. *J. Comput. Biol.* **2016**, *23*, 526–535. [[CrossRef](#)]
114. Khammar, N.; Malhautier, L.; Degrange, V.; Lensi, R.; Godon, J.-J.; Fanlo, J.-L. Link between Spatial Structure of Microbial Communities and Degradation of a Complex Mixture of Volatile Organic Compounds in Peat Biofilters. *J. Appl. Microbiol.* **2005**, *98*, 476–490. [[CrossRef](#)]
115. Van der Waals, M.J.; Atashgahi, S.; da Rocha, U.N.; van der Zaan, B.M.; Smidt, H.; Gerritse, J. Benzene Degradation in a Denitrifying Biofilm Reactor: Activity and Microbial Community Composition. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 5175–5188. [[CrossRef](#)] [[PubMed](#)]
116. Karaoz, U.; Couradeau, E.; da Rocha, U.N.; Lim, H.-C.; Northen, T.; Garcia-Pichel, F.; Brodie, E.L. Large Blooms of *Bacillales* (*Firmicutes*) Underlie the Response to Wetting of Cyanobacterial Biocrusts at Various Stages of Maturity. *mBio* **2018**, *9*, e01366-16. [[CrossRef](#)] [[PubMed](#)]
117. Zhou, H.; Gao, Y.; Jia, X.; Wang, M.; Ding, J.; Cheng, L.; Bao, F.; Wu, B. Network Analysis Reveals the Strengthening of Microbial Interaction in Biological Soil Crust Development in the Mu Us Sandy Land, Northwestern China. *Soil Biol. Biochem.* **2020**, *144*, 107782. [[CrossRef](#)]
118. Wegener, G.; Kellermann, M.Y.; Elvert, M. Tracking Activity and Function of Microorganisms by Stable Isotope Probing of Membrane Lipids. *Curr. Opin. Biotechnol.* **2016**, *41*, 43–52. [[CrossRef](#)]
119. Shi, S.; Nuccio, E.; Herman, D.J.; Rijkers, R.; Estera, K.; Li, J.; da Rocha, U.N.; He, Z.; Pett-Ridge, J.; Brodie, E.L.; et al. Successional Trajectories of Rhizosphere Bacterial Communities over Consecutive Seasons. *mBio* **2015**, *6*, e00746-15. [[CrossRef](#)]