



# Microbiome structure and function: A new framework for interpreting data

Gregor P. Greslehner

## ► To cite this version:

Gregor P. Greslehner. Microbiome structure and function: A new framework for interpreting data. BioEssays, 2020, 42 (7), pp.1900255. 10.1002/bies.201900255 . hal-02558255

**HAL Id: hal-02558255**

**<https://hal.science/hal-02558255>**

Submitted on 29 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Microbiome structure and function: A new framework for interpreting data

Gregor P. Greslehner

This is the peer reviewed version of the following article: “Microbiome structure and function: A new framework for interpreting data”, *BioEssays* **2020**, 1900255, which has been published in final form at [doi:10.1002/bies.201900255](https://doi.org/10.1002/bies.201900255). This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

## Abstract

I suggest a distinction between different notions of ‘structure’ and ‘function’ for interpreting the overwhelming amount of data on microbiome structure and function: sequence data, biochemical agents, interaction networks, taxonomic communities, and their dynamics can be linked to potential or actual biochemical activities, causal roles, and selected effects, respectively. This conceptual clarification has important methodological consequences for how to interpret existing data and approach open questions in contemporary microbiome research practice. In particular, the field will have to start thinking about notions of function more directly.

**keywords:** microbiome; structure; function; research practice; sequence data; taxonomic composition

## 1. Introduction

Microbiome research has recently gained a lot of attention and promises to revolutionize the biomedical field. Major impacts are expected from cataloging and characterizing different microorganisms in different conditions and body sites, linking the microbial composition to particular health outcomes. Knowing “who is there” (i.e. the taxonomic composition of the

microbiome) is a perfectly reasonable starting point, but understanding “what is going on” (i.e. the biochemical activities and their phenotypic effects) is much more important, thus taking us from structure to function. Both notions have played a major role in microbiome research, as reflected in the title of one of the Human Microbiome Project Consortium’s first and most influential publications:<sup>[1]</sup> ‘structure, function and diversity of the healthy human microbiome’.

Despite their prominent conceptual place, however, the notions of ‘structure’ and ‘function’ are notoriously polysemic, i.e. refer to different things. This gives rise to some confusion in many areas of biology on all levels of organization. Here, I argue that in order to think properly about microbiome structure and function, we have to distinguish between different notions of ‘structure’ and ‘function’. Doing so is not only a matter of philosophical conceptual clarification, but also addresses some theoretical problems that the field of microbiome research is currently struggling with.

The framework in which we think about microbiome structure and function greatly impacts our understanding of, and experimental approaches to, the microbiome, particularly within the context of different methods and their interpretation.

## 2. What Does ‘Structure’ Mean?

The term ‘structure’ can refer to many different things that are ordered or composed in a particular way. Thinking of biological objects like a molecule, for example, ‘structure’ can denote either (1) its genetic or amino acid sequence (“primary structure”), (2) its three-dimensional shape (“tertiary structure”), or (3) its interaction network (“organizational structure”). In microbiome research, however, scientists usually have a fourth meaning in mind: (4) “community structure”, i.e. the taxonomic composition of species within a microbial population. These taxonomic profiles are usually specific for certain habitats, body sites, and different stages of development and aging. This leaves us with four different meanings of ‘structure’ (I will use the numbers to refer to these different meanings throughout the text):

1. *genetic sequence*
2. *biochemical agent*
3. *interaction network*
4. *taxonomic composition*

All four different meanings of ‘structure’ are illustrated in Figure 1. Frequently it is sequence data (1) from which the other notions are directly or indirectly inferred. With sequence data being a useful and easily available starting point for forming hypotheses, there is a danger of relying too much on just one notion of structure. It is customary to use (parts of) 16S rRNA sequences (1) for predicting the taxonomic composition (4) of samples. Similarly, the encoded biochemical agents (2) are inferred from whole genome sequencing. Post-translational protein modifications and other biomolecules represent biochemical agents that cannot be easily predicted on the basis of sequence data. Interaction network structures (3), on the other hand, require different techniques and data to be studied. Microbial networks and their dynamics are much more informative than snapshots in terms of taxonomic composition,<sup>[2,3]</sup> sequences, or biochemical agents.<sup>[4]</sup> Besides the static properties of these interaction networks, their dynamics promise to reveal insights into the behavioral changes of these interactions over time – and their functional consequences.<sup>[5–7]</sup> This way, also disruptions can be given a more precise meaning that is helpful in revealing the underlying mechanisms, rather than summarizing them under the vague label of “dysbiosis” – a concept that has received a lot of criticism.<sup>[8–10]</sup> Additionally, spatial organization or three-dimensional network structure has been shown to be of importance.<sup>[11–14]</sup> If we take the edges of a network as the distances between individual microbes, one can easily reconstruct the three-dimensional spatial structure as an important special case of a network structure.

Simply focusing on “who is there” in terms of taxonomic composition (4) only holds limited information,<sup>[15]</sup> especially when trying to understand the functional impacts of the microbiome on the health of a human host or their effects on a complex ecosystem. Although it is important to keep these different notions of structure separate, they rarely lead to confusion. It is only when trying to predict either one notion based on another, or when trying to link them to notions of function, that we run into trouble. Calls for putting more emphasis on the notion of function have become louder in the last few years.<sup>[16,17]</sup> What, exactly, does ‘function’ mean in the first place? In the following section, I offer four different meanings one should distinguish when thinking and talking about microbiome function.

### 3. What Does ‘Function’ Mean?

Microbes show a great functional diversity, as reflected in their genetic repertoire. While the number of microbial cells and human host cells are roughly the same, according to the latest estimations,<sup>[18,19]</sup> the number of genes encoded by the microbes still outperforms that of the human host by orders of magnitude.<sup>[20]</sup> The potential biochemical activities encoded in these

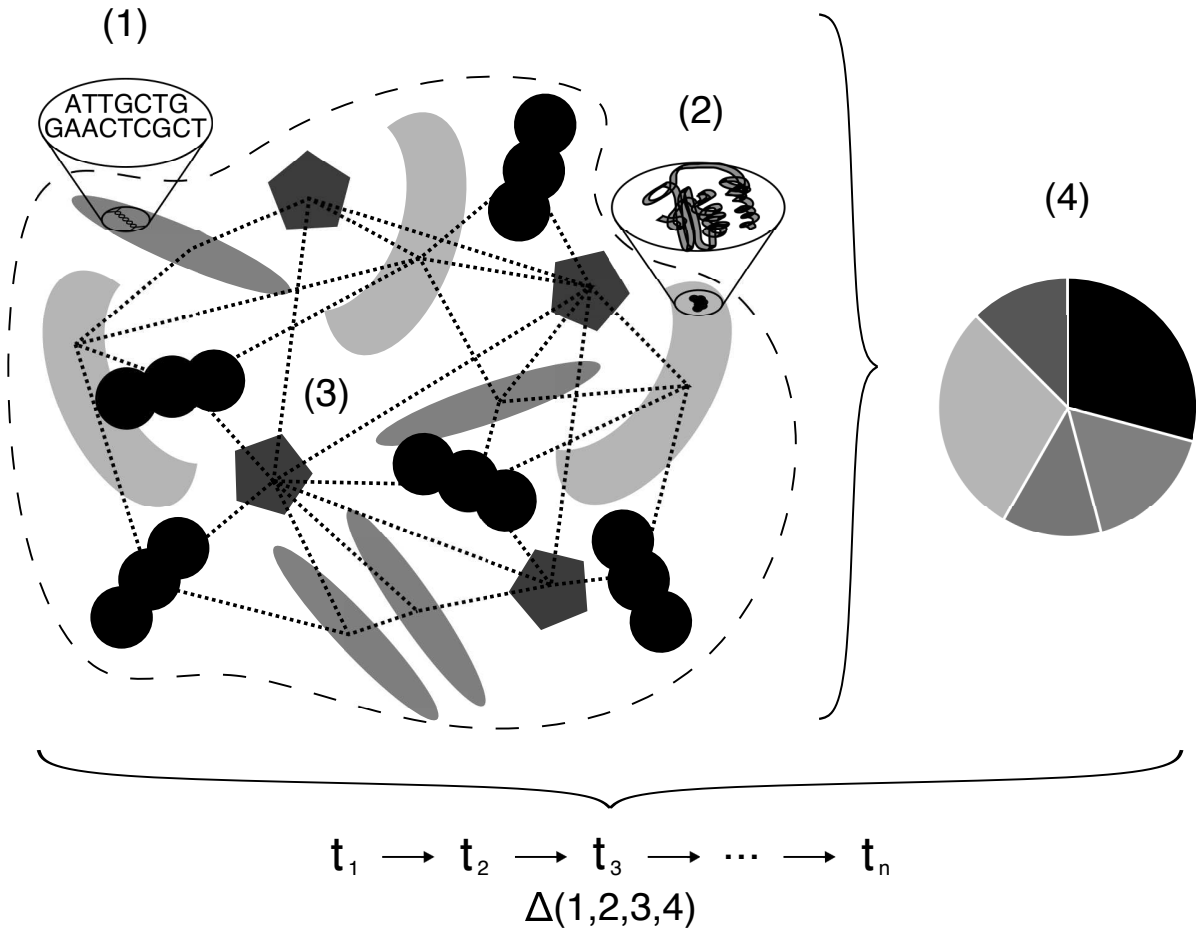


Figure 1: Different notions of structure: (1) genetic sequence; (2) biochemical agents, e.g. proteins; (3) interaction networks, shown with dotted lines, and the surrounding dashed line for the spatial structure as a special case; and (4) taxonomic community structure. Finally,  $t_1, t_2, t_3, \dots, t_n$  stand for different points in time, where variation and dynamics of all these structures –  $\Delta(1,2,3,4)$  – can be observed by longitudinal studies.

genes might be involved in physiological processes that – directly or indirectly – influence the physiology and health of the host organism.

Philosophical debates about different notions of function have been going on for decades,<sup>[21,22]</sup> with the main positions being function understood as *selected effect* on the one hand and *causal role* on the other hand (Box 1 summarizes the main philosophical discussion of function). However, having clear definitions and concepts about function are also highly relevant in scientific practice.<sup>[23]</sup> Microbiologist Jonathan L. Klassen recently argued for the importance of using clear definitions when using the term ‘function’ in microbiome research.<sup>[16]</sup> Besides discussing classical notions from philosophy, this raises important conceptual ques-

tions about how we should think about microbiome function. Given the different notions

#### Box 1: Philosophical debates on the notions of function

The two major accounts discussed in philosophy are the “selected effect”<sup>[24]</sup> and “causal role”<sup>[25]</sup> notions of function. In a nutshell, the selected effect function is why something is there to do what it does, what it has been selected for. The causal role account, on the other hand, only addresses how the components contribute to the capacity of the system of which they are part. Other refined notions and their relations have been intensively discussed over the last decades. Suffice it to say that there are many different notions that have been discussed in philosophy in order to explain the apparent goal-directedness of many biological systems and their operations.<sup>[22]</sup>

Arno G. Wouters distinguishes (in addition to other, evolutionary considerations) between activities of single entities and their roles as part of larger interacting systems.<sup>[26]</sup> Different evolutionary notions of function are summarized under the term ‘selected effect’ in the present paper, despite the fact that there are different evolutionary aspects to be distinguished here as well (fitness advantages, targets of selection, etc.). Since we are not addressing the intricacies of evolutionary considerations here, we will just stick with the potential and actual biochemical activities, biological roles, and selected effect notions of function in this paper.

discussed above, I suggest that ‘function’ can mean the following four different things (again, I will use the letters to refer to these different meanings throughout the text):

- A. *potential biochemical activity*
- B. *actual biochemical activity*
- C. *causal role of these activities in biological processes*
- D. *selected effect*

Let’s walk through these notions individually. Not all encoded genes are expressed, their presence as genetic sequences (1) does not guarantee that they are actually carrying out their functions; it is an indication for potential biochemical activity (A), but not necessarily for actual biochemical activity (B).<sup>[27–29]</sup> Inferring causal role of these activities in biological processes (C) is an even bigger challenge. Finally, selected effects (D) are a matter of evolutionary dynamics. (D) and (C) have received a lot of philosophical attention. In particular, philosophical perspectives have focused on the teleological (goal-directed) notions of function. However,

from the perspective of biological research practice, where structure and function of the microbiome are being characterized mostly based on sequence data (1), biochemical activities (A, B) and biological roles (C) outweigh evolutionary considerations of selected effects.

The confusion caused by mixing up different notions of function can be quite significant. Think, for example, of the debate surrounding the percentage of DNA that has a “function” following the ENCODE estimation.<sup>[30–32]</sup> While the ENCODE consortium took any kind of biochemical activity to be a function, thus ending up with a functional percentage of 80%, conservative notions of only considering sequences that have been selected for protein-coding to be functional would attribute a function to only less than 2% of human DNA. These conflicting numbers are simply a result of using different notions of function.

Similarly, lack of clarity and inconsistency can cause confusion in microbiome research.<sup>[5]</sup> When we are trying to understand biological phenomena at higher levels of organization, the one-dimensional information from genetic sequences provides valuable and easily accessible data; their interpretation in terms of function, however, will often depend on additional data that cannot be inferred with certainty solely from sequences. Although the tools and methods used to make these inferences are far from the accusation of genetic reductionism, they still make some simplistic assumptions that inherently limit the scope of (functional) understanding that can be obtained from the focus on genetic sequence data.

Different methods are required to study different notions of structure and function (Figure 2). Studying biochemical activities and their place in metabolic networks requires metabolomic strategies.<sup>[33]</sup> Neither network interactions, actual biochemical activities, nor biological roles can straightforwardly be attributed to whole genome shotgun sequence data, let alone to taxonomic profiles based on 16S rRNA. This is not to deny the importance of these large high-throughput studies that investigate exactly that, thus providing excellent preliminary proxies for directions in which to look further. However, without having other kinds of data, which are often difficult to obtain, these might be the only informative starting points for pursuing further investigations. It is important to highlight that these associations are just the starting point for a mechanistic understanding, not only of “who is there” but also of “what can they do”<sup>[34]</sup> – and “what are the *actually* doing”. “What effect do these activities have on host phenotypes” and other questions in terms of targets of selection, evolutionary dynamics, and fitness advantages, are entirely different explanatory projects altogether, where all these different kinds of data can be used in evolutionary frameworks.<sup>[35,36]</sup>

Being aware of these different concepts and their respective methodological approaches is important for how we should think about microbiome structure and function.

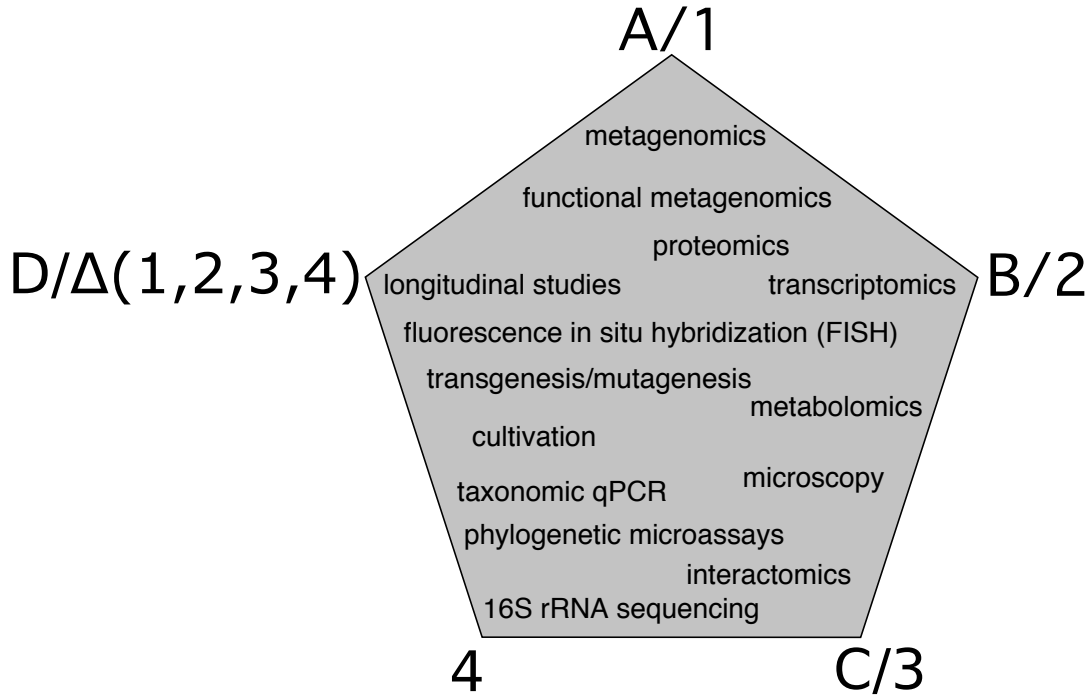


Figure 2: Studying different notions of microbiome structure and function with different methods. See text and Table 1 for details on the different notions of structure and function: (A) potential biochemical activity; (B) actual biochemical activity; (C) causal role (in biological processes); (D) selected effect; (1) genetic sequence; (2) biochemical agent; (3) interaction network (including spatial structure); (4) taxonomic community; ( $\Delta(1,2,3,4)$ ) variation and dynamics of changes in 1, 2, 3, or 4.

## 4. What Is the Relationship Between Structure and Function?

How do all these different notions of structure (1–4) and function (A–D) relate to each other? Based on genetic sequencing of 16S rRNA (1), the composition of the taxonomic community (4) can be obtained, and overall genomic profiling (1) gives information about potential biochemical activity (A). While genetic sequences might not actually be expressed into biochemical agents (2), their actual biochemical activity (B) can only be studied with omics approaches when they are expressed. Which biological roles (C) these biochemical activities (B) play in the interplay among themselves, the role of other microbes, and the host, are pressing relevant questions that depend on the topology and dynamics of interaction networks (3). If one wants to draw a connection between these different notions of structure and function, I argue that scientifically meaningful connections can only be drawn between the pairs shown in Table 1.

There are correlations between the taxonomic composition (4) and the genetic sequences



Table 1: How different notions of structure and function relate to each other. The predictive strength of these relationships differs between different notions. While taxonomic community (4) does not guarantee any inferences about function with certainty, it can serve as a proxy to make predictions – within certain limits. The other relationships are more reliable in their predictive strength.

<b>structure</b>	<b>function</b>
genetic sequence (1)	potential biochemical activity (A)
biochemical agent (2)	actual biochemical activity (B)
interaction network (3), including spatial structure	causal role of these activities in biological processes (C)
taxonomic community (4)	(limited) proxy for (A, B, and C); obtained from 16S rRNA sequence (1)
variation and dynamics of (1, 2, 3, or 4)	selected effect (D), selection for (B or C)

(1) present, together with their functional implications (at least, potentially). This makes taxonomic information a limited proxy and preliminary predictor for the notions of function (A, B, C). However, there are limitations when making inferences from sequence data.<sup>[37,38]</sup> Predicting functional profiles from genomic and taxonomic data is made possible by powerful bioinformatical tools.<sup>[39,40]</sup> Data science tools like BURRITO allow the visualization and analysis of the relationship between taxa and function.<sup>[41]</sup> To see how strong the links between genomic/taxonomic data and different notions of function actually are, and which scientific inferences and predictions can meaningfully be made, are a matter of current debate. While these data mining tools provide sophisticated and partially accurate predictions about the potential biochemical activities (A) of the encoded biochemical agents (2), their actual biochemical activities (B), biological roles (C), let alone selected effects (D) can hardly be inferred from genomic/taxonomic data (1, 4) alone. One cannot just assume functional similarity from some kind of taxonomic association. At the same time, functional dissimilarity does not necessarily follow from differences in taxonomy. Being aware of these facets of microbiome structure and functions allows to look for the correlations and causal connections in the right places. By confusing different notions for each other, or just using what is the most convenient to obtain technically, we might fail to notice the causally important difference makers.

What many studies understand by ‘functional composition’ is actually still sequence data (1) with functional annotation, i.e. inferred potential biochemical activities (A). There are many caveats with predicting functions based on genes. When studying the microbial world, homologous reference genomes can be scarce. The prediction and annotations of function based on genetic homology, although a powerful prediction tool, are far from being perfect. The explanatory power and predictive strength depend to a large extent on the models being used for

making these predictions and explanations. Their accuracy can be improved by keeping the different notions of structure and function and their relationships clearly apart, as summarized in Table 1.

Genetic homology, which is being used to predict the biochemical activities of the corresponding genes, can also fail in cases of high genetic homology, by a few changes that drastically change the three-dimensional structure and its actual biochemical activity of the corresponding protein. Even if closely related genomes can be used for reference, microbes with a similar genetic sequence – especially when only based on similarity in their 16S rRNA sequence – can still differ substantially in what they contribute functionally. Successful predictions on genetic data depend to a large extent on the so-called “central dogma of molecular biology”.<sup>[42]</sup> Even in cases where a biochemical function can be successfully predicted, i.e. the potential biochemical activity is also an actual one of the corresponding biochemical agent, the biological role it plays in a metabolic pathway or physiological process can hardly ever be attributed to such a biochemical agent (2), or its corresponding genetic sequence (1), in isolation. What is needed to annotate the biological role is the interaction network structure (3). We also need to understand the dynamics of changes in taxonomic composition, the resulting biochemical activities and roles, and their effects on each other. To do so, we need to study the “degree to which a shift in a community’s taxonomic composition will impact its functional capacities (a property that we refer to here as taxa-function robustness)”.<sup>[43]</sup> That there it is no one-to-one mapping, but a many-to-many mapping, between the different notions of structure and function is shown by the fact “that functional changes can occur in the absence of changes in the taxonomic profile”.<sup>[44]</sup>

Similarly, the functional vocabulary and concepts need to be improved. So far, and contrary to certain claims and expectations, including functional information did not improve the accuracy of microbial classification.<sup>[34]</sup> This might be less surprising, given the inference steps involved in obtaining functional data based on sequence data. Without a better understanding of microbiome structure and function, these error-prone inferences will remain rather murky and coarse-grained – in particular, as long as different notions of function are not clearly distinguished.

Going so far as to attempt to predicting the functional impact on health outcomes of the host based on microbial sequence data remains an even bigger explanatory gap, skipping several factors that also need to be taken into account when trying to predict and explain these health outcomes and their causes.<sup>[1,45,46]</sup> Attempts to use taxonomic biomarkers of “who is there” as short-cuts to understanding “what they are doing”<sup>[47]</sup> omit all these other notions and levels of microbiome structure and function.

There is no one-to-one mapping between taxonomic composition and function (in any of its different notions).<sup>[48]</sup> Different microbial communities can behave very similarly, and the same biological roles can be fulfilled by many different taxonomic groups. Also, conversely, similar taxonomic composition can differ significantly in their ecological and other functional properties.<sup>[49,50]</sup> Focusing almost exclusively on the taxonomic composition of microbial communities is not satisfactory to study and understand the microbiome. As Rob Knight and colleagues recently put it: “Knowing the composition of a microbial community is no longer a sufficient research goal; we want to know the function of the community”.<sup>[42]</sup> Achieving this research goal requires the integration of omics data from different methods, targeting different notions of microbiome structure and function.<sup>[51,52]</sup> Interpretation of the data, as well as building models and hypotheses based on them, will all depend on a proper theoretical understanding.<sup>[53,54]</sup> As I argue here, it will be crucial to think about microbiome structure and function, with a clear distinction between different notions and their relations, in order to advance the field.

## 5. Does Structure Determine Function?

With the relationships between these different notions of structure and function at hand (Table 1), which notions can be inferred or predicted from the other ones? Which methods can be used to determine these notions directly or indirectly? The predominant methodology at the moment is 16S rRNA sequencing. In many ways, this sequencing methodology has its benefits: next-generation sequencing has become affordable and fast, as well as the computational power that is required to handle all the sequence data being fed into the data bases every day: The fact that the prokaryotic small ribosomal 16S subunit is highly conserved enables taxonomic profiling. However, there are a couple of limitations and caveats coming with this approach.

The first and most important limitation is the omission of all the other non-bacterial microorganisms (viruses, fungi, helminths, protozoa) that have no ribosomes, or eukaryotic ribosomes, and that receive much less attention in the current research practice. It is not out of the question to use other genomic markers, including eukaryotic ribosomal sequences. A critical limitation, however, still persists in any approach that focuses on genetic markers in order to identify “who is there” (Box 2).

So far, the dominant approach has been the attempt to predict the functional profile from 16S taxonomic profiles. Although there are correlations and predictions can be based on taxonomic proxies, inherent caveats in this approach make clear that it can only supplement, but

## Box 2: Beyond Taxa and “Who Is There”

Microbiome researchers and philosophers have argued in the last couple of years that function might be more stable, conserved, and important than taxonomic composition.<sup>[55]</sup> There is no good (causal) mapping between taxonomic composition and function<sup>[56]</sup> – let alone, taxonomic composition to host phenotype: “One of its [the Human Microbiome Project’s] biggest initial revelations was that the taxonomic composition of the microbiota in the human body was not a reliable predictor of host phenotypes, such as disease susceptibility”.<sup>[57]</sup> Others have also argued for the importance of moving beyond the focus on species towards a functional perspective<sup>[58–61]</sup> – sometimes with an emphasis on genes.<sup>[62]</sup> I agree that characterizations of the microbiome should focus on functional aspects. However, as discussed above, the functional information that can be obtained from genetic sequences (1) has its limitations, too.

The multiple realizability of functional features by different biochemical agents, interaction networks, and bacterial communities (2, 3, 4) can explain why there is such a diversity of species within the same “functional guild”. It might be function rather than taxonomic composition that matters – or, as the microbiologist W. Ford Doolittle and the philosopher Austin Booth suggested, what matters is the “song, not the singer”.<sup>[63]</sup> Other clever wordplays express a similar idea: “the players may change but the game remains”,<sup>[64]</sup> or simply “function first, taxa second”.<sup>[17]</sup>

If true, the aims to identify a “core microbiome”<sup>[65]</sup> should then also be more oriented towards functional aspects rather than taxonomy.<sup>[66]</sup> There is evidence for this in both biomedical and environmental microbiome contexts.<sup>[67]</sup> In order to study the additional notions of structure and function, different methods than already existing tools might be necessary, and those might be provided by different disciplines. This underscores the need for “multidisciplinary approaches to exploring human–microbiome interactions”.<sup>[68,69]</sup> Generally speaking, “what is going on” in terms of function might be more important than “who is there” structurally – presupposing a proper understanding of different notions of function.

not replace, studying these functional components directly. Besides the fact that taxonomic classification is always somewhat arbitrary, horizontal gene transfer is abundant in the microbial world<sup>[70,71]</sup> – which allows potential and actual biochemical activities and the biological roles that come with it to be transferred easily between different kinds of microbes. If the aim is to understand the mechanistic processes underlying the microbial contributions to host phe-

notypes and health outcomes, it is indispensable to move beyond this rudimentary taxonomical sequencing approach.

Take Proteobacteria, for example, which are named after the Greek god Proteus for the very reason that they show large versatility in physiology and function.<sup>[72]</sup> The relative abundance of taxa like these makes it difficult to hypothesize about their functional contribution to the ecological niche they inhabit. Interventions that aim at restoring or modifying microbial functions should then also not just focus on introducing certain percentages of taxa. The functions in question can be realized by multiple different taxa, or depend on a particular interaction network rather than the mere presence/absence of certain taxa. Different methods and machine learning approaches to linking taxonomic data and functional aspects will be useful for making more precise annotations and predictions, yet they do not provide any mechanistic understanding (Box 3).

It is well-known that correlation is not the same thing as causation. Assigning causation is a particularly tricky enterprise in microbiome research.<sup>[76]</sup> Acquiring more data on community structure (4) will probably not be enough to move beyond mere associations to the underlying biological mechanisms. Obtaining more metagenomic sequence data (1) and attempts to predict function, on the other hand, has its limits, too. While potential biochemical activity (A) can be quite reliably predicted (at least on a coarse-grained resolution, and given that similar sequences with annotated biochemical activities already exist in the data base), the actual biochemical activities (B) and biological roles (C) – which depend on interaction networks (3), not sequences (1) – have to be studied by different approaches (Figure 2).

Being aware of these different notions of structure and function and the relations between them will be critical for moving from mere correlations to an understanding of mechanisms and cause-effect relationships. Taxonomic profiles usually provide only a snapshot of the current composition. Many functionally important aspects might only be perceptible in the interactions and dynamics within the community structure, which require ecological approaches<sup>[77]</sup> that might not yet be established within the classical canon or toolbox of microbiome researchers.<sup>[78]</sup> Other important approaches include meta-omics approaches, transcriptomics, proteomics, metabolomics, interactomics, etc. – with an emphasis on their potential and actual biochemical activities and biological roles (A, B, C).

Establishing causality in microbiome research will remain a major challenge and a tricky enterprise, for which philosophy can provide some valuable tools.<sup>[79]</sup> One important step, I think, is to distinguish clearly between different notions of microbiome structure and function when looking for causes and effects. Another crucial step is to make function a priority for microbiome research by implementing methods that target physiological functions directly,

### Box 3: From Associations to Mechanisms

Methods like principal components analysis remain powerful tools when interpreting sequence and taxonomy data, highlighting valuable correlations and even targets for intervention, monitoring shifts towards and away from a more or less “healthy” microbiome.<sup>[1,73]</sup> However, understanding of the underlying mechanisms and causal connections are unlikely to emerge from pushing forward with the same approach. More data of the same kind is not the solution either. What is needed are particular models (theoretical, computational, biochemical, model organisms) in which the biochemical agents, their activities, interactions networks, and biological roles can be investigated directly.

Are the “over-hyped” causal claims<sup>[74]</sup> due to overlooked limitations in current research practices? Linking structure and function in an unreflecting way can certainly lead to many misconceptions. The framework I offer for thinking about microbiome structure and function should help avoiding some of the underlying misunderstandings.

While some issues might be improved by technology (multi-omics and data-interpretation), the conceptual problem that I want to emphasize concerns the main focus on structure, i.e. sequences (1) and taxonomic composition (4). What does a taxonomic profile actually tell us about physiological states or health outcomes? A major problem being: how can we distinguish cause from effect? As Neeraj K. Surana and Dennis L. Kasper recently put it in a letter to *Nature*: “If the field is to move beyond correlations and begin to address causation, an effective system is needed for refining this catalogue of differentially abundant microbes and to allow subsequent mechanistic studies [...] The general inability to move beyond correlations and address causation has been the Achilles heel of microbiome research”.<sup>[75]</sup> By proceeding with the current framework of thinking about microbiome structure and function in an unreflected way, it is to be expected that these kinds of problems will persist.

instead of only inferring them indirectly from the sequence data. The common slogan that “structure determines function” should not lead us into believing that 16S rRNA sequence determines taxonomic composition, whose constituents’ genomic sequences, in turn, determine their biological functions, which would determine particular health outcomes. Basically, what do we mean when we talk about “the microbiome”? Which notions of structure and function are at work when thinking about the microbiome? The totality of genomic sequences, taxa, or the functional repertoire that comes with it?

Lita Proctor, coordinator of the Human Microbiome Project, recently argued in *Nature* that “most of the research so far has placed too much emphasis on cataloguing species names. We’ve been characterizing the human microbiome as if it were a relatively fixed property to be mapped and manipulated”.<sup>[80]</sup> Instead, she argues for ecological and evolutionary approaches to study host-microbe interactions in order to understand the underlying mechanisms. Different tools need to be applied when investigating all the notions of microbiome structure and function. Clearly distinguishing between the different notions outlined above when thinking about microbiome structure and function is an essential prerequisite.

## 6. Conclusions and Outlook

With these distinctions in mind, how should we think about microbiome structure and function? In addition to disambiguating conceptually different notions of these terms, I pointed out scientific and philosophical consequences. By distinguishing different notions of structure and function in their respective contexts, some scientific statements and questions become clearer and more accessible for empirical testing. At the same time, this conceptual clarification should have implications for the research and intervention practices in biology and medicine. Thus, I believe that thinking about microbiome structure and function in this way will benefit research practice by avoiding the confusion between different notions and their relations – and by putting more emphasis on function.

By shifting the focus from taxonomic composition and genetic sequences to notions of function, research and biomedical implications will lead to new conclusions and results based on a better understanding of microbiome structure and function. Given that ‘function’ is a notoriously ambiguous concept, the suggested framework of different notions should help clarify some common misunderstandings and dangers that microbiome research might run into when thinking about microbiome structure and function. The same framework will be helpful for moving from being descriptive to being explanatory, for developing mechanisms and models of the underlying physiological processes that influence host physiology and health. Restricting study of microorganisms to their genome sequences and community structure would be tantamount to ignoring the other rich notions of microbiome structure and function.

In order to study and understand the microbiome in all its different facets, we need to combine different aspects of structural and functional data – being aware of the conceptual differences and targeting the right entities and concepts for different explanatory projects.

## Acknowledgments

Many thanks to Thomas Pradeu and all members of the ERC IDEM ImmunoConcept group. This paper is based on a talk I gave at the ISHPSSB conference in Oslo, Norway on July 8, 2019; I am grateful for helpful feedback, especially from Rob Knight. Many thanks also to two anonymous reviewers and Andrew Moore, editor-in-chief of *BioEssays*, for helpful comments and suggestions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme – grant agreement no. 637647 – IDEM (P.I.: T. Pradeu).

## Conflict of Interest

The author declares no conflict of interest.

## References

- [1] The Human Microbiome Project Consortium. *Nature* **2012**, 486, 207.
- [2] E. E. L. Muller, K. Faust, S. Widder, M. Herold, S. M. Arbas, P. Wilmes. *Curr. Opin. Systems Biol.* **2018**, 8, 73.
- [3] K. Faust, J. Raes. *Nat. Rev. Microbiol.* **2012**, 10, 538.
- [4] D. McDonald, Y. Vázquez-Baeza, W. A. Walters, J. G. Caporaso, R. Knight. *Biol. Philos.* **2013**, 28, 241.
- [5] D. R. Nemergut, S. K. Schmidt, T. Fukami, S. P. O'Neill, T. M. Bilinski, L. F. Stanish, J. E. Knelman, J. L. Darcy, R. C. Lynch, P. Wickey, S. Ferrenberg. *Microbiol. Mol. Biol. R.* **2013**, 77, 342.
- [6] K. Faust, D. G. Leo Lahti and, W. M. de Vos, J. Raes. *Curr. Opin. in Microbiol.* **2015**, 25, 56.
- [7] E. Guven-Maiorov, C.-J. Tsai, R. Nussinov. *PLoS Comput. Biol.* **2017**, 13, e1005579.
- [8] S. W. Olesen, E. J. Alm. *Nat. Microbiol.* **2016**, 1, 16228.
- [9] K. B. Hooks, M. A. O'Malley. *mBio* **2017**, 8, e01492.



- [10] C. R. Tiffany, A. J. Bäumler. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2019**, 317, G602.
- [11] K. A. Earle, G. Billings, M. Sigal, J. S. Lichtman, G. C. Hansson, J. E. Elias, M. R. Amieva, K. C. Huang, J. L. Sonnenburg. *Cell Host Microbe* **2015**, 18, 478.
- [12] G. P. Donaldson, S. M. Lee, S. K. Mazmanian. *Nat. Rev. Microbiol.* **2016**, 14, 20.
- [13] J. L. M. Welch, B. J. Rossetti, C. W. Rieken, F. E. Dewhirst, G. G. Borisy. *P. Natl. Acad. Sci. USA* **2016**, 113, E791.
- [14] C. Tropini, K. A. Earle, K. C. Huang, J. L. Sonnenburg. *Cell Host Microbe* **2017**, 21, 433.
- [15] T. A. C. Reydon. *Hist. Phil. Life Sci.* **2019**, 41, 40.
- [16] J. L. Klassen. *Nat. Microbiol.* **2018**, 3, 864.
- [17] A. Heintz-Buschart, P. Wilmes. *Trends Microbiol.* **2018**, 26, 563.
- [18] R. Sender, S. Fuchs, R. Milo. *Cell* **2016**, 164, 337.
- [19] R. Sender, S. Fuchs, R. Milo. *PLoS Biol.* **2016**, 14, e1002533.
- [20] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, M. Consortium, M. Antolin, F. Artiguenave, H. Blottiere, N. Borruel, T. Bruls, F. Casellas, C. Chervaux, A. Cultrone, C. Delorme, G. Denariáz, R. Dervyn, M. Forte, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, A. Jamet, C. Juste, G. Kaci, M. Kleerebezem, J. Knol, M. Kristensen, S. Layec, K. Le Roux, M. Leclerc, E. Maguin, R. Melo Minardi, R. Oozeer, M. Rescigno, N. Sanchez, S. Tims, T. Torrejon, E. Varela, W. de Vos, Y. Winogradsky, E. Zoetendal, P. Bork, S. D. Ehrlich, J. Wang. *Nature* **2010**, 464, 59.
- [21] A. G. Wouters. *Acta Biotheor.* **2005**, 53, 123.
- [22] J. Garson. *A Critical Overview of Biological Functions*. Springer, Cham, **2016**.

- [23] D. M. Keeling, P. Garza, C. M. Nartey, A.-R. Carvunis. *eLife* **2019**, 8, e47014.
- [24] L. Wright. *Philos. Rev.* **1973**, 82, 139.
- [25] R. Cummins. *J. Philos.* **1975**, 72, 741.
- [26] A. G. Wouters. *Stud. Hist. Phil. Biol. Biomed. Sci.* **2003**, 34, 633.
- [27] A. Tanca, M. Abbondio, A. Palomba, C. Fraumene, V. Manghina, F. Cucca, E. Fiorillo, S. Uzzau. *Microbiome* **2017**, 5, 79.
- [28] J. He, J. Lange, G. Marinos, J. Bathia, D. Harris, R. Soluch, V. Vaibhvi, P. Deines, M. A. Hassani, K.-S. Wagner, R. Zapien-Campos, C. Jaspers, F. Sommer. *BioEssays* **2019**, 42, 1900211.
- [29] A. T. Reese, S. M. Kearney. *Curr. Opin. Microbiol.* **2019**, 50, 20.
- [30] The ENCODE Project Consortium. *Nature* **2012**, 489, 57.
- [31] P.-L. Germain, E. Ratti, F. Boem. *Biol. Philos.* **2014**, 29, 807.
- [32] W. F. Doolittle, T. D. Brunet, S. Linquist, T. R. Gregory. *Genome Biol. Evol.* **2014**, 6, 1234.
- [33] G. A. Prosser, G. Larrouy-Maumus, L. P. S. de Carvalho. *EMBO Rep.* **2014**, 15, 657.
- [34] Z. Xu, D. Malmer, M. G. Langille, S. F. Way, R. Knight. *ISME J.* **2014**, 8, 2357.
- [35] K. R. Foster, J. Schluter, K. Z. Coyte, S. Rakoff-Nahoum. *Nature* **2017**, 548, 43.
- [36] A. L. Gould, V. Zhang, L. Lamberti, E. W. Jones, B. Obadia, N. Korasidis, A. Gavryushkin, J. M. Carlson, N. Beerenwinkel, W. B. Ludington. *P. Natl. Acad. Sci. USA* **2018**, 115, E11951.
- [37] J. A. Gilbert, C. L. Dupont. *Annu. Rev. Mar. Sci.* **2011**, 3, 347.
- [38] I. Zarraonaindia, D. P. Smith, J. A. Gilbert. *Biol. Philos.* **2013**, 28, 261.
- [39] K. P. Abhauer, B. Wemheuer, R. Daniel, P. Meinicke. *Bioinformatics* **2015**, 31, 2882.
- [40] M. G. I. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. V. Thurber, R. Knight, R. G. Beiko, C. Huttenhower. *Nat. Biotechnol.* **2013**, 31, 814.

- [41] C. P. McNally, A. Eng, C. Noecker, W. C. Gagne-Maynard, E. Borenstein. *Front. Microbiol.* **2018**, 9, 365.
- [42] R. Knight, B. C. T. Alison Vrbanc and, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciulek, L.-I. McCall, D. McDonald, A. V. Melnik, J. T. Morton, J. Navas, R. A. Quinn, J. G. Sanders, A. D. Swafford, L. R. Thompson, A. Tripathi, Z. Z. Xu, J. R. Zaneveld, Q. Zhu, J. G. Caporaso, P. C. Dorrestein. *Nat. Rev. Microbiol.* **2018**, 16, 410.
- [43] A. Eng, E. Borenstein. *Microbiome* **2018**, 6, 45.
- [44] K. M. Ellegaard, P. Engel. *Front. Microbiol.* **2016**, 7, 1475.
- [45] J. A. Gilbert, R. A. Quinn, J. Debelius, Z. Z. Xu, J. Morton, N. Garg, J. K. Jansson, P. C. Dorrestein, R. Knight. *Nature* **2016**, 535, 94.
- [46] V. B. Young. *BMJ* **2017**, 356, j831.
- [47] N. Ottman, H. Smidt, W. de Vos, C. Belzer. *Front. Cell. Infect. Mi.* **2012**, 2, 104.
- [48] M. G. I. Langille. *mSystems* **2018**, 3, e00163.
- [49] M. Hamady, R. Knight. *Genome Res.* **2009**, 19, 1141.
- [50] C. J. Robinson, B. J. M. Bohannan, V. B. Young. *Microbiol. Mol. Biol. R.* **2010**, 74, 453.
- [51] S. Mondot, P. Lepage. *Ann. NY Acad. Sci.* **2016**, 1372, 9.
- [52] D. Jiang, C. R. Armour, C. Hu, M. Mei, C. Tian, T. J. Sharpton, Y. Jiang. *Front. Genet.* **2019**, 10, 995.
- [53] T. S. Schmidt, J. Raes, P. Bork. *Cell* **2018**, 172, 1198.
- [54] J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, R. Knight. *Nat. Med.* **2018**, 24, 392.
- [55] S. Louca, S. M. S. Jacques, A. P. F. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, M. Doebeli. *Nat. Ecol. Evol.* **2016**, 1, 0015.
- [56] S. A. Inkpen, G. M. Douglas, T. D. P. Brunet, K. Leuschen, W. F. Doolittle, M. G. I. Langille. *Biol. Philos.* **2017**, 32, 1225.
- [57] Nature Editorial. *Nature* **2019**, 569, 599.

- [58] J. L. Green, B. J. M. Bohannan, R. J. Whitaker. *Science* **2008**, 320, 1039.
- [59] C. E. Rosen, N. W. Palm. *BioEssays* **2017**, 39, 700032.
- [60] S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, L. W. Parfrey. *Nat. Ecol. Evol.* **2018**, 2, 936.
- [61] T. S. Stappenbeck. *Nat. Microbiol.* **2018**, 3, 6.
- [62] C. Burke, P. Steinberg, D. Rusche, S. Kjelleberg, T. Thomas. *P. Natl. Acad. Sci. USA* **2011**, 108, 14288.
- [63] W. F. Doolittle, A. Booth. *Biol. Philos.* **2017**, 32, 5.
- [64] T. M. Taxis, S. Wolff, S. J. Gregg, N. O. Minton, C. Zhang, J. Dai, R. D. Schnabel, J. F. Taylor, M. S. Kerley, J. C. Pires, W. R. Lamberson, G. C. Conant. *Nucleic Acids Res.* **2015**, 43, 9600.
- [65] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, J. I. Gordon. *Nature* **2007**, 449, 804.
- [66] P. Lemanceau, M. Blouin, D. Muller, Y. Moënné-Loccoz. *Trends Plant Sci.* **2017**, 22, 583.
- [67] S. Louca, L. W. Parfrey, M. Doebeli. *Science* **2016**, 353, 1272.
- [68] T. C. G. Bosch. *BioEssays* **2019**, 41, 1900130.
- [69] K. R. Amato, C. F. Maurice, K. Guillemin, T. Giles-Vernick. *BioEssays* **2019**, 41, 1900007.
- [70] H. Jeong, B. Arif, G. Caetano-Anollés, K. M. Kim, A. Nasir. *Sci. Rep.* **2019**, 9, 5953.
- [71] R. Sitaraman. *Microbiome* **2018**, 6, 163.
- [72] P. H. Bradley, K. S. Pollard. *Microbiome* **2017**, 5, 36.
- [73] J. Lloyd-Price, G. Abu-Ali, C. Huttenhower. *Genome Med.* **2016**, 8, 51.
- [74] P. Bourrat. *BioEssays* **2018**, 40, 1800178.
- [75] N. K. Surana, D. L. Kasper. *Nature* **2017**, 552, 244.

- [76] M. A. O'Malley, D. J. Skillings. *Perspect. Sci.* **2018**, 26, 239.
- [77] J. Raes, P. Bork. *Nat. Rev. Microbiol.* **2008**, 6, 693.
- [78] J. A. Gilbert, S. V. Lynch. *Nat. Med.* **2019**, 25, 884.
- [79] K. E. Lynch, E. C. Parke, M. A. O'Malley. *Biol. Philos.* **2019**, 34, 62.
- [80] L. Proctor. *Nature* **2019**, 569, 623.