Beyond the Microscope: Rethinking Microbial Diversity Measurement with the Model-Based Account

Aline Potiron[1,2]

[1]Institute of Philosophy and Scientific Method, Johannes Kepler University, Altenberger Straase, 69, Linz, 4040, Austria.
[2]Freudenthal Institute, Utrecht University, Princetonplein 5, Utrecht, 3584 CC, The Netherlands.

Contributing authors: potiron.aline@gmail.com;

Abstract

Measuring diversity in microbial ecology and microbiome studies is fraught with challenges, rendering the assessment of its "real-world" value nearly impossible. The instability of taxonomic classification, difficulty in isolating individuals, and reliance on DNA-based methods and statistical tools all contribute to the complexity of measuring diversity reliably. This manuscript explores the underlying philosophical issues, relating them to the measurement problem in philosophy. I argue that traditional philosophical accounts of measurement, including representational, operationalist, and realist approaches, are insufficient to address these issues. Instead, I examine these challenges through the lens of a model-based perspective on measurement, which can remain agnostic about entities and property ontologies, clarify the role of assumptions in diversity measurement, and provide solutions for justifying measurement procedures. This work emphasizes the importance of calibration and clearly defining measurement purposes, providing avenues for scientists to improve their measurement procedures. Ultimately, I contribute to a deeper understanding of the challenges and opportunities in measuring microbial diversity by bridging the gap between philosophy and scientific practice.

Keywords: Amplicon Sequencing, Diversity, Measurement Theories, Microbial Ecology, Microbiome, Philosophy in Practice

ORCID: 0000-0003-3521-882X

Acknowledgements I thank my PhD supervisors, Emanuele Ratti, and Julian Reiss, for carefully reading this manuscript. Earlier versions of this work were presented at the EPSA23, where it received many valuable comments. I thank Alexander Linsbischler, Margitta Mätzke, and Andreas Roser, who also gave me helpful comments and feedback. I am grateful to Robert Frühstückl and Jacqueline Wallis for exchanges on this work and follow-up comments. I thank Charlie Pauvert for helping compute the data in R and proofread the last version of the manuscript. Finally, I created the figures for this manuscript with Biorender.com. I also thank Abigail Nieves Delgado and Elian Schure for their reading and help on the revised version of the paper. I am grateful to three anonymous reviewers for their careful reading of the manuscript and their comments that have improved the current manuscript.

Statements and Declarations No specific funding was received to conduct this study. Most of it was done while employed at Johannes Kepler University, Linz, and the revisions were conducted while employed at Utrecht University, Utrecht. The author has no competing interests to declare relevant to this article's content.

1 Introduction

Microbial ecology studies microorganisms (bacteria, Archea, and fungi) and their critical role in environmental processes (Falkowski, Fenchel, & Delong, 2008). Additionally, microorganisms colonize all animals, including humans, and the human gut microbiome has been linked to various health and disease phenotypes (Berg et al., 2020). As a result, microbiologists have warned that microorganisms are closely linked to change at a global scale. Microbial diversity is supposed to be associated with a certain stability of the community, understood as the community's capacity to answer some internal or external disturbances, its stability over time, etc. (Normand, Duran, Le Roux, Morris, & Poggiale, 2015; Shade, 2017). Fluctuations in microbial diversity and activities could impact other organisms' resilience and ability to respond to climate change (Cavicchioli et al., 2019) and to health perturbations (Doré, Tap, & Ehrlich, 2017).

Microbial ecology and microbiome studies need to compare and rank the diversity of two or several communities. However, measuring diversity is challenging. The notion is fundamentally broad, as discussed in conservation biology (Maclaurin & Sterelny, 2008; Santana, 2014; Sarkar & Margules, 2002), and the delineation of a community and an area is arbitrary (Mittelbach & McGill, 2019). Consequentially, researchers have developed statistical indices that aggregate different facets of the notion of diversity. While it is already difficult in this situation to choose which index is the best to capture diversity, microbial ecology and microbiome studies display specific issues, making assessing diversity's 'real/world' value almost impossible. In section 2, I focus on these issues: the necessity of delineating local and global areas, microorganismal isolation, fluctuating classification, and the lack of a consensual species concept. Amplicon sequencing/metagenomics (AS/MGS) measuring practices worsen these issues.

In section 3, I propose that these specific issues are related to the fundamental problem of measurement (Reiss, 2008, 64) and examine several philosophical accounts

to solve these issues. I argue that all of the traditional accounts – representational, operationalist, and realist – have general limitations that found echoes in the measurement of microbial diversity. Diversity measurement in microbial ecology also presents specific challenges for these accounts.

Given these limitations, I argue that the model-based account is more adapted to the problem of measuring microbial diversity (section 4). First, it can remain agnostic about entities and property ontologies. Second, the appeal to models can help clarify the role of different assumptions in the retrieved value for the quantity diversity. This, in turn, can help scientists to reflect on their measures and identify their needs to develop reliable diversity measurement procedures.

In section 5, using this account, I ground in a conceptual framework solutions emerging (but not widely accepted and even less widely used) from the scientific community: the idea of calibration. The model-based account emphasizes the calibration procedure's role in increasing the measurements' reliability. Current practices (AS/MGS) are still under development (Parker, 2017) or at the pre-measurement stage (Frigerio, Giordani, & Mari, 2010). Moreover, I point to and develop a more original solution of integrating systematically the purpose of measurement in the measurement procedure model. The purpose constrains the choice of the diversity index.

2 Specific Issues About Measuring Diversity in Microbial Ecology[1]

The intuitive idea about microbial diversity at a spatial scale can be illustrated by the difference one perceives between an assemblage of microorganisms in the human gut and an assemblage of microorganisms in a yogurt. The intuition is that the first assemblage contains more microorganisms in number and kind (probably billions) than the second assemblage (often only two microorganisms). The former is more diverse

[1] From now on in this manuscript, I consider microbiome studies to be part of microbial ecology if not specified explicitly otherwise.

than the latter. Diversity also has a time component. The intuition is the one we have when we use antibiotics: there will be a change in the community after the treatment. Microbial ecologists want to assess these changes in terms of the number of individuals, the number of kinds of individuals, and also which kind is affected by the perturbation.

In this section, I summarize these intuitions in three diversity-related questions: (A) How many taxa compose this community? (B) How are these taxa distributed? and (C) How different are these taxa? I focus on specific issues of microbial ecology: the necessity of delineating local and global areas, microorganismal isolation, fluctuating classification, and the lack of a consensual species concept that deepens these issues. They are worsened by the amplicon sequencing/metagenomics (AS/MGS) measuring practices. All these issues force scientists to make choices that influence diversity measurement.

Local versus Global Area Issue Three measurements of diversity are traditionally used. The α-diversity is the diversity found at a very local scale. The γ-diversity is the diversity at a bigger scale than the one delineated for α-diversity. How big is left open, that is, the difference in spatial scale between α- and γ-diversity, can vary across studies and is somewhat arbitrary. Finally, the β-diversity represents how quickly species composition changes across regional sites. It is often expressed as a turnover or a rate of change between α- (local spatial scale) and γ- (bigger spatial scale) diversity (Mittelbach & McGill, 2019).

In microbial ecology and microbiome studies, there is no consensus on the definition of local versus global area. For this specific point, I need to separate microbial ecology in general and microbiome studies. On the one hand, the definition of "local" in microbial communities is ambiguous. Indeed, a lot of environments, such as the soil, are highly heterogeneous even at a tiny scale "so that environmental samples (e.g., a soil core) are often already a mixture of local communities." (Normand et al., 2015, 266). So, the relative spatial scale between α- and γ-diversity is unclear. Consequentially, the

calculation of β-diversity, i.e., how quickly species composition changes across sites, is also vague, and the distinction between α- and β-diversity is rarely used in microbial ecology.

On the other hand, this distinction is sometimes used in microbiome studies. For example, what the authors Daybog and Kolodny call "the microbiome β-diversity conundrum" concerns conflicting findings on the relationship between the diversity of a host's microbiome and its fitness. The conundrum bears on the fact that different healthy individuals in the same population have different microbiome compositions, while the evolutionary assumption is that if a trait is beneficial for a population, it should not vary so much and should be conserved (Daybog & Kolodny, 2023). So in this study, the local community for which α-diversity is measured is the individual host, and the β-diversity is the diversity difference between two hosts. However, other studies compute β-diversity for expressing differences between both individuals and human groups (Chong et al., 2015). In both types of studies, the γ-diversity is not used, creating confusion about what β-diversity actually measures. Overall, the ambiguity surrounding the use of these measurements hinders comparisons between studies.

Taxa Delineation Issue Species richness, S (but also Shannon index, H', and all other indices), is sensitive to the concept used to delineate taxa. This is one of the central issues in measuring diversity in microbial ecology, especially at the species level, because of the ambiguity of this concept[2]. In the case of microorganisms, the traditional definition of species, such as the biological species concept considering species as reproductively isolated groups, does not hold because bacteria is a group where sexuality is rare and atypical (Normand et al., 2015). Many bacterial species are distinguished, but it is difficult to link them to a particular species concept[3]. Therefore, answering diversity-related question (A) is difficult. The risk associated with errors

[2]It is not my aim here to discuss the species concept in microbiology in detail. I am only outlining the points related to the problem of measuring diversity.

[3]Microbial taxonomists tend to use a concept close to the phylogenetic species concept, considering species to be ancestral individuals sharing relevant ancestors. However, this classification has inconsistencies with several polyphyletic taxa (Hugenholtz, Chuvochina, Oren, Parks, & Soo, 2021).

in taxa delineation is an overestimation or underestimation of the richness of the community. Additionally, without a clear way of distinguishing taxa, it is also difficult to count how many individuals pertained to a given taxa with risks of misattributing a species to an individual. This situation impairs the possibility of reliably answering question (B). Finally, without a clear concept for taxa delineation, it is difficult to assess question (C), how far individuals in a community are from each other.

Unstable Microbial Classification The consequence of the difficulties of delineating taxa is a fluctuating microbial classification with a regular changing of the terminology within the taxonomy. So far, in 2024, only 36,240 taxon names are validly published, and an additional 12,951 are published but not valid[4]. The validity depends on the *International Code of Nomenclature of Prokaryotes (ICPN)* "the document that contains the internationally accepted rules that regulate the naming of prokaryotic taxa." (Oren, 2024, 1). Moreover, the names changed, and in recent years, reclassifications have been proposed for several genera and species. For example, since 2018, more than 2,500 names have changed[5]. These reclassifications include microorganisms of industrial or medical importance (Oren, 2024). This is an issue for measuring diversity. First, even if a species might be isolated, it might be that the individual has no name or label. Without a name or label that attaches it to an organized classification, it is difficult to assess how different the isolated individual is from the others in the community. Hence, the information about the qualitative question (C) of how different these taxa are and what they are is unanswered. Second, this fluctuation can change the classification by adding or subtracting species. Two microorganisms considered different species at time t can be regarded as the same species at time $t+1$. This is an issue for the quantitative questions (A) and (B) and for comparing studies measuring diversity at different times.

[4] https://lpsn.dsmz.de/statistics/figure/130, consulted 14/03/2024
[5] https://lpsn.dsmz.de/statistics/figure/20, consulted 14/03/2024

Microorganism Isolation Issue Another related problem is the difficulties of isolating microorganisms.

> The rules of the ICNP only apply to cultivated prokaryotes. [...], as of 1 January 2001, the valid publication of the name of a new species must include the designation of a type strain, and a viable culture of that strain must be deposited in at least two publicly accessible culture collections in different countries from which subcultures must be available. (Oren, 2024, 3)

Most microorganisms are uncultivated. Thus, it is challenging to obtain an isolated specimen (= type strain) for each species discovered in a given microbiome and comply with the ICNP rule. It is then challenging to have a stable and encompassing classification. Additionally, the microorganism isolation issue also impacts the answer to question (B) about the distribution of each taxon. If individuals are not well delineated, the evenness information of diversity is inaccessible.

Illustration This situation is worsened by the practices used to obtain the equivalent of the observation tables that usually register diversity information (for each species encountered, the number of individuals is registered): meta-barcoding (now called amplicon-sequencing, AS) and meta-genomics (MGS). Diversity assessment in microbiology uses AS/MGS methods because cultivation-based methods underestimate microbial diversity (Hugenholtz, Goebel, & Pace, 1998). Amplicon sequencing relies on methods similar to DNA barcoding[6], with the significant difference being the pooling of the DNA of the individuals of the studied community. Once the DNA sequences have been computerized, they are analyzed and sorted according to their variations. Various mathematical and statistical tools are available, leading to different bioinformatic pipelines. The assumptions are: 1) the molecular marker tracks the patterns of variation that result from multiple evolutionary processes (Lowe & Ingram,

[6]DNA barcoding is the practice of selecting particular stretches of DNA and using specific sequences of nucleotides contained in them as indicators of the species identity of the organism from which the DNA was derived.

2023, 32), and 2) statistical methods allow scientists to distinguish between natural variation, i.e., from evolutionary processes and artifactual variation, i.e., caused by the AS method itself. Groups of DNA sequences emerge from this sorting. These groups correspond to the taxonomic level of the analysis, and the number of groups is the value associated with taxonomic diversity. A simplification of the steps of the AS analysis is available in Figure 1.

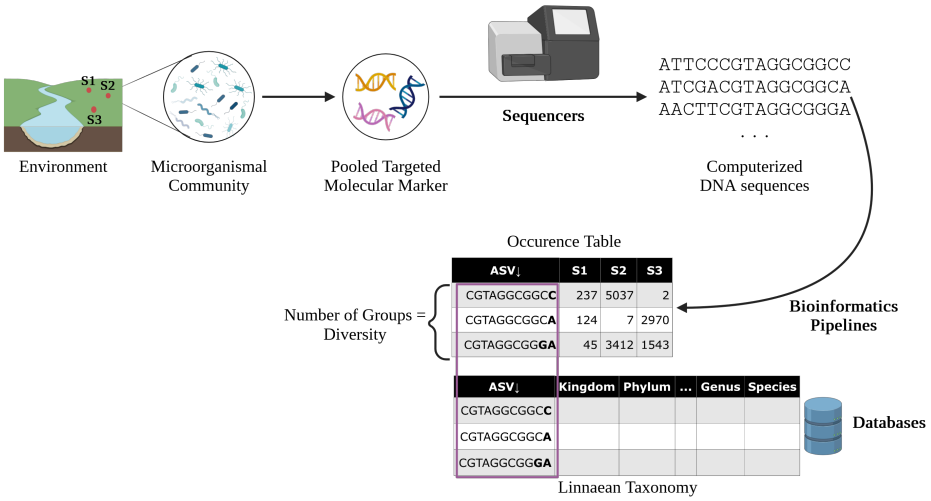

Fig. 1 Simplified representation of the steps of the amplicon sequencing method. Created in BioRender Created in BioRender. Potiron, A. (2025) https://BioRender.com/f61f354.

In AS and other MGS methods, there is no direct access to the individuals. Taxonomic identification by DNA barcoding associates one barcode or molecular marker with one individual. In AS, all individuals are mixed so that the DNA retrieved is not linked anymore to a physical, tangible, and observable individual. It is a methodological substitution for microbial observation, which disrupts the link between microorganisms and their DNA. This is related to the microorganism isolation issue because the AS method i) cannot always link a given DNA sequence to a known microorganism, and ii) it has been shown that sometimes it is unwise to equate the

sequence of one gene to one microorganism individual as this individual can possess several copies of the same gene and these copies might vary in their sequences (McLaren, Nearing, Willis, Lloyd, & Callahan, 2022).

There is a tension with these methods that deepens the issues of taxa delineation and unstable microbial classification. On the one hand, while traditional ecology relies on the descriptor "species" for measuring diversity, "species" is rarely used as a first descriptor in microbial ecology. "This problem of the taxonomic level to use is constant in the field of microbial ecology. [...], thus the widespread use of the term OTU (Operational Taxonomic Unit)" (Normand et al., 2015, 286). The concept of OTU results from different aggregation procedures used in different software for individuals or groups of microorganisms (Sneath, 2005). The definition of an OTU would be "a group of phylogenetically related organisms used in a study without specifying its taxonomic rank." (Normand et al., 2015, 285). OTUs are bioinformatic proxies for taxa. An essential factor to consider is then the resolution power of available techniques. Given the range of mathematical and statistical tools available, the method chosen and its assumptions impact the inquiry's conclusions and, here, the value assigned to the parameter diversity. For example, the threshold at which two DNA sequences are considered as belonging to the same OTU can vary. It has been traditionally accepted that a threshold of 97% of similarity for a given molecular marker is enough to distinguish between species. More recently, it has been argued that this threshold might not be enough and should be 100%. This is the case for Amplicon Sequence Variants (ASV). That is, two DNA sequences of the same molecular marker are considered to belong to the same group only when their sequence is exactly – 100% – the same. This situation adds another layer of difficulties to the taxa delineation issue. Indeed, researchers need to make choices not only on how to delineate species but also on how to delineate these bioinformatic proxies. Depending on this threshold, the value assigned to the diversity property will vary. Additionally, OTUs and ASVs may track

different taxonomic levels, e.g., genus, species, strains which adds assumptions about which taxonomic level is best suited for measuring diversity in relation to the stability of the community.

On the other hand, microbial ecology and, particularly, microbiome studies often try to link those bioinformatic proxies to the traditional Linnaean taxonomy through local or online databases. This is done for at least three reasons.

First, by analogy with the measurement of length using feet and meters, the measurement of diversity could keep OTUs, ASVs, species, or others as units. Contrary to feet and meters, there is no direct and consistent equivalence between OTUs and ASVs or between OTUs/ASVs and species. This means that studies using different units are not directly comparable. Hence, referring to the Linnean taxonomy is a way to use only one reference unit and compare studies between them.

However, the taxonomic level may be different. For example, one OTU/ASV can be identified at the genus level, while another is identifiable at the species level. Beyond the difficulty of identifying these DNA sequences, this situation signifies that the granularity of the measure is not consistent. It is as if the length of a set of objects was measured using both centimeters and meters, but here, one does not have a simple way of converting one into another, i.e., genus into species.

Second, traditional ecology is about species. One of the reasons is that beyond the number found for the measurement of diversity (e.g., five species), the measure is also supposed to give some ecological information about the community (see, as a minimum, questions (B) and (C) above). If the name *Escherichia coli* is well-documented, OTU1 (which the DNA sequence attached to it will differ between studies) is not. To do that, researchers rely on accumulated and organized knowledge: the Linnaean taxonomy stored in databases (Reimer et al., 2021; Schoch et al., 2020) and used in articles (Rosonovski et al., 2023).

Third, going a bit beyond the numerical value attributed to the parameter diversity, the possibility to link an OTU/ASV to known taxa (family, genus, or species depending on the resolution of the molecular marker chosen) opens another inferential layer, the possibility to have access to some function of the OTUs/ASVs. One hypothesis is that maybe it is at the level of the functions that diversity explains stability. Suppose a given OTU/ASV is identified as a known genus with known properties. In that case, the researcher can infer that the OTU/ASV in her community shares these properties, notably some metabolic functions. But this step encounters several problems. The association between a sequence of a molecular marker and a species depends on the completeness of the database used to make this association. Moreover, the molecular marker chosen might not be tuned enough to detect a difference at the species level or, on the contrary, can detect variations between strains depending on the species (McLaren & Callahan, 2018). That is, there is no reliable way of linking an organism to its functions through its belonging to a given OTU/ASV.

The last issue with MGS is the lack of reflexivity that arises from using bioinformatic pipelines. The lack of reflexivity is at two levels. The first is an absence of reflection on why it is necessary (or not) to measure diversity. The second level is on which indices to choose and what kind of corrections will be needed depending on the context (including the natural environment observed and the research question). In other words, those indices are often calculated without clear enunciation of a hypothesis to test, without purpose in microbial ecology in general (Shade, 2017) and microbiome studies (Reese & Dunn, 2018). This situation is saillant in MGS. DNA sequencing technologies increased the amount of data, and their digital format made them readily amenable to computer-helped statistical analysis. Bioinformatic pipelines make the calculation of various diversity indices easily accessible (it is doable by calling one function in a code). The problems are several depending on the background knowledge and skills of the researchers. First, these functions often present

too many choices of indices. For example, the function calculating the α-diversity in the phyloseq R package produces a graphic with seven indices[7]. There is no empirical constraint on how to choose among them, and the researchers need to appeal to some other justification. Second, the bioinformatic pipelines and even the functions calculating several diversity indices may appear as "black-boxed" for non-specialists[8]. Moreover, the question is specialist in what? Indeed, these functions embed statistical and ecological assumptions.

Measurement of diversity in microbial ecology and microbiome studies is not as stable and consistent as length or temperature. I argue that this situation is similar to the "fundamental problem of measurement" and seek help in the main available philosophical accounts of measurement. After evaluating them and describing their problems in the case of microbial ecology and microbiome studies, I argue that the philosophical model-based account of measurement is the best to account for diversity measurement in microbial ecology. It clarifies some of the measurement issues encountered and grounds or develops solutions.

3 Philosophical Accounts of Measurement and Their Problems

There are numerous issues arising from the diversity measurement. First, even when the categories (taxa) are well-defined and individuals are accessible, it is a challenge to measure diversity. Several indices are used to aggregate different information about diversity. Second, in microbial ecology, the situation is worse. The specific issues of the taxa delineation, unstable classification, and microorganism isolation make independent access to the real-world value of diversity challenging. Thus, it is difficult to know how close the diversity value measured is to the actual value (or if

[7] https://joey711.github.io/phyloseq/plot_richness-examples.html, consulted 05/02/2025

[8] I am not saying that the information about the detailed calculus is never accessible, but it is not directly available while using those packages and thus it necessitates an extra effort from the researcher to have access to this information. Additionally, statistical and/or ecological skills are also needed to process and understand this information.

there even is an actual value). Third, without access to this real-world value, it is difficult to know whether the measuring procedure used in microbial ecology (MGS) is accurate and gives us the correct value for the parameter or even if this parameter exists. I suggest that it constitutes what Julian Reiss called the "fundamental problem of measurement":

> To know the value of a variable, we need to know that the measurement procedure associated with it is veridical (that is, that the procedure gives the correct result). However, we need to know the variable's value to check whether the procedure is veridical. Since we have no independent access to either the value of the variable or the accuracy of the procedure, we can never know whether the measurement procedure is veridical or what the value of the variable is. (Reiss, 2008, 64)

I will follow the historical development of measurement theory (Tal, 2020). I argue that none of the traditional accounts will solve the measurement problem, and the best framework in microbiology is the model-based account.

3.1 The Representational Account of Measurement

The central point in the representational account is the idea that relations between numbers can express relations among objects. The two main aims of this account are to understand the assumptions behind using mathematical structures to describe aspects of the empirical world and to analyze the adequacy and limits of using these mathematical structures. "The representational theory of measurement defines measurement as the construction of mappings from empirical relational structures into numerical relational structures." (Krantz, Luce, Suppes, & Tversky, 1971, 9). A measurement scale is "a many-to-one mapping — a homomorphism — from an empirical to a numerical relational structure, and measurement is the construction of scales." (Tal, 2020). Figure 2 gives an example of a homomorphism. The empirical relational structures are a set of empirical objects associated with the qualitative relations among them. The numerical relational structures are a set of numbers (e.g., real numbers)

associated with specific mathematical relations among them (e.g., "greater than", addition). A homomorphism is a structure-preserving map between the empirical relational structures and the numerical relational structure such that the relations between the elements in the empirical relational structure (e.g., "longer than") are preserved by relations between numbers in the numerical relational structure (e.g., "greater than"), see Figure 2 for an illustration.

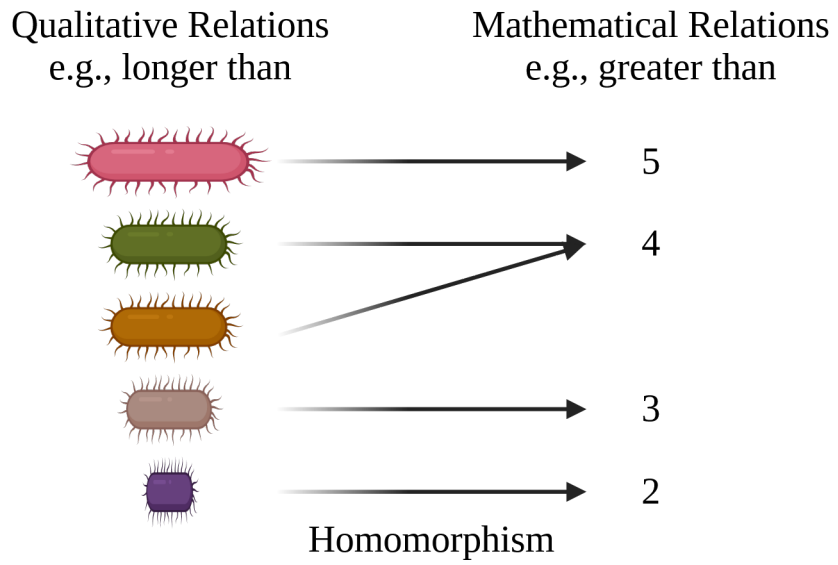

Fig. 2 Example of a homomorphism between empirical objects and a set of real numbers. Created in BioRender. Potiron, A. (2025) https://BioRender.com/c55i883

Measurement, then, is the assignment of numbers to attributes of the phenomenon. The problem is that relations between numbers, such as equality, do not always correspond to empirical information. The questions then include which assignments are adequate and under which conditions. The assignment is not random (Reiss, 2008, 65), and "numbers are adequate for expressing magnitudes insofar as algebraic operations among numbers mirror empirical relations among magnitudes." (Tal, 2020). There are two steps for the assignment. The first one is the formulation of axioms that

describe the qualitative empirical structures. These axioms describe qualitative laws the phenomenon of interest is supposed to obey. For example, for length, elements of the empirical structures are longer, smaller, or equal to one another. The second step uses those axioms to establish logical proof of the adequacy between assigning numbers to magnitudes possessing such structures (Tal, 2020). The axioms help derive two theorems, the representation theorem to "establish a formal similarity between an empirical and a numerical structure and the uniqueness theorem establishes what kinds of transformations are permissible on the mapping." (Reiss, 2008, 65). For example, suppose it is true that elements can be ordered by their length. In that case, it should be possible to derive a function such that, when applied to elements of the empirical structure, it assigns numbers to these elements that conserve the relation, e.g., "greater than" to represent "longer than" (see Figure 2).

As noticed elsewhere, the representational theories of measurement have the disadvantage of needing a lot of background knowledge. Moreover, the laws of nature governing the relations among empirical objects need to be stable for the measuring scale to be relevant (Reiss, 2008, 66). However, this abundance of knowledge and the stability of it are not always available. Even for the well-known and established length, mass, time, and others, it has not always been the case that we have much knowledge and stability. According to Reiss then, the problem with this account is that it: "tells us what structure an attribute of a phenomenon must have to be measurable *given we have a reliable measurement instrument* but it does not tell us where and how to look for a reliable instrument in the first place." (Reiss, 2008, 67, original emphasis). In the case of diversity measurement in microbial ecology and microbiome studies, taxonomic knowledge is lacking, and the instability of taxonomic classification is an impairment for the stability of the measure. The empirical relational structure is the relations among individuals within one or several communities. These relations are "similar to," "different from" within a community and "more diverse than," "less diverse than,"

or "equally diverse than" between communities. From this, we derived the possibility of assigning individuals to categories using a nominal scale (taxonomic assignation); we can also derive the possibility of mapping the relations attached to diversity to a numerical relational structure "greater than, smaller than, equal to." The problem I pinpoint here is with the first part of the measurement. The ordinal scale used is not stable. That is, the function that assigns a category to an individual is not stable, as is the homomorphism. In the case of the Linnaean taxonomy, it is evident that the scale is not stable – see above section 2. The scale is also unstable when using bioinformatic proxies – that is, "OTU1" in one study is different from "OTU1" in another one. For ASVs, the DNA sequence is unique, so, in principle, the ordinal scale could be stable; that is, the DNA sequence would be the scale, but it is not widely used and could be impractical (the units would be entire DNA sequences – or we could use labels, hence Linnaean taxonomy). The general problem is the comparison between studies.

Moreover, when comparing the diversity of two communities, there is no stable series of tests for which one can be sure that the results will always agree about the ranking between the two communities. That is, it is possible in one series of tests that a community is more diverse than the other and vice versa in another series of tests. Taking several samples of the same community is an attempt at stabilizing the series, but it is never guaranteed. This is so because, in AS, sample analysis is seldom reproducible. It is physically impossible to re-analyze the same set of samples in the same manner because it is a disruptive practice. The procedure itself destroys the initial mixes of microorganisms by extracting their DNA. What can be done at best is to repeat the series of tests on samples judged similar enough to the initial experiment to represent the same communities[9]. Yet, even when doing so, the protocols used to extract, amplify, and sequence the DNA introduce errors. The same can be said about

[9]In this sense, it is similar to other samples (e.g., water from the same spring) for which one needs to assume a certain similarity between them for the sake of the analysis. In the case of a water sample, for example, it is possible to re-analyze the same sample even if it may happen that the sample has changed over time (e.g., contamination). The case of microbial communities is peculiar because the targeted living system is destroyed to be studied and analyzed. So, even if one wanted to re-analyze the sample, it would be impossible. I thank an anonymous reviewer for pointing out this issue.

the bioinformatic analysis. Depending on the choices made in this analysis, even if only one index is calculated, the communities can be ordered differently. For example, a meta-analysis on nine datasets obtained from the gut microbiome of lean and obese rats (four datasets) and mice (five datasets) shows discordant results for α-diversity measures using the Shannon index (see table 1 for the formula). I follow the authors of this study, and I comment on these results by grouping both animals. In three of the nine datasets, the microbial community of the gut in obese rodents is ranked more diverse than the microbial community of the gut in lean rodents. In one of them, the inverse ranking is obtained, and in all the others (five out of nine), the communities are similarly diverse (Jiao et al., 2018, 247).

One minimal condition of a measurable attribute of a phenomenon is that the objects or phenomena that have the attribute can be ordered. For example, in the case of length, the elements of the empirical relational structure should be able to be ordered according to the attribute "length." Length uses a ratio scale that necessitates the fixation of a unit, e.g., the length of a standard rod = 1m. By measuring these elements using the predefined scale, one can say for every element of the empirical relational structure which element is longer than, smaller than, or equal to. The issue in microbial ecology is whether communities of microorganisms can be ordered. In theory, it should be possible to say one community is more diverse than another. However, without explicit assumptions and hypotheses, it is difficult to tell which of the two communities is more diverse. First, microbial ecology is subjected to the general issues of community and spatial area delineation and the problems it poses in comparing diversity indices between two (or more) communities. Second, microbial ecology is subjected to specific issues that impact the interpretation of the indices used to measure diversity. Table 1 illustrates some of these issues.

For the same size area and the same total number of individuals $N = 50$, in community C1, the richness S is five taxa, the relative abundance (i.e., the number

Table 1 Two imagined microbial communities and their associated diversity indices

	Number of Individuals (N)	Taxa Richness (S)	Relative Abundance	Chao1 Index[1]	Shannon Index[2] (H')
Community 1 (C1)	50	5	uneven (see Table 2)	6.50	0.46
Community 2 (C2)	50	5	even = 10/50	5.00	1.61

[1]Chao1 index calculated using R package 'vegan' (Oksanen et al., 2024). Chao1 Formula = $S_{obs} + N_1^{(N_1-1)}/(2 \times (N_2+1))$ where N_1 and N_2 are count of singletons and doubletons respectively.

[2]Shannon index calculated using R package 'vegan' (Oksanen et al., 2024). Shannon formula = $H' = -\sum_{i=1}^{S} p_i \ln p_i$ where i ranges between 1 and S and p_i is the proportion of individuals belonging to the i-th species or taxon.

Table 2 Uneven distribution in the microbial community C1

Taxa (s_i)	s_1	s_2	s_3	s_4	s_5
Relative Abundance	45/50	2/50	1/50	1/50	1/50

of individuals pertaining to each taxon) is uneven (Table 2), the Chao1 index gives a diversity of 6.50, and the Shannon index gives a diversity of $H' = 0.46$. In community C2, the richness S is five taxa, the relative abundance is even (10 individuals among the 50 individuals (10/50) belong to each of the five taxa), the Chao1 index is equal to 5.00, and the Shannon index is equal to 1.61. Now, which community is "more diverse"?

We can imagine principles connecting diversity to more readily observable attributes. For example, there could be a principle that prescribes that C1 is more diverse than C2 if and only if $S_{C1} > S_{C2}$, that is, the taxonomic richness in C1 is greater than in C2. In this case, C1 is as diverse for the taxa observed as C2 (this assumes that the taxa are clearly and reliably delineated and that attributing an individual to a given taxon is reliably possible). Alternatively, we can imagine that the information given in Table 1 is only a sample for each community; thus, the taxonomic richness S is an incomplete estimation of the taxa richness of both communities. A similar principle will prescribe that C1 is more diverse than C2 if the richness estimate for an infinite number of samples is higher for C1 than for C2. The Chao1 index

is this sort of index. In this case, because $Chao1_{C1} > Chao1_{C2}$, then community C1 is more diverse than community C2. In addition to the assumptions made for S, this principle assumes that the relative abundance (needed to compute the Chao1 index) is known to be reliably measurable. Another principle that can be used to compare the diversity of two areas of the same size and with relatively similar qualities is to imagine a law that prescribes that the community C2 is more diverse relative to the community C1 if and only if the relative abundance for each taxon is more even for C2 than C1. As the evenness is considered in the Shannon index H', it means that because $H'_{C2} > H'_{C1}$, then community C2 is more diverse than community C1. Again, this assumes that the relative abundance is known to be reliably measurable.

The three solutions sketched here need additional assumptions to settle the question of which community is the more diverse. The first solution assumes a stable taxonomic classification. However, as mentioned in section 2, classifying microorganisms according to some principles is far from easy. The concept of species is even more controversial than the notion of diversity because of its widespread significance in biology. In microbiology, OTUs and ASVs are used. Let's suppose that S is expressed in "OTUs." This means that both communities have a richness of five OTUs. If these OTUs are the same for both communities, i.e., the DNA sequences of the five OTUs in C1 are exactly the same as the DNA sequences of the five OTUs in C2, then the diversity of C1 and C2 is equal, at least according to S. If they aren't, some OTUs may be representative enough of the genus level of a category of microorganisms, while others are representative enough for the species level. Thus, the measure uses a heterogeneous scale, one at the species level and the other at the genus level. Traditional taxonomy is a hierarchical scale: one can infer the genus level from the species level but not vice versa. Thus, when one wants to convert to traditional taxonomy, one has to take the common level, the genus level, and cannot access diversity at the species

level. This is problematic because one can ask whether the ordering of the communities C1 and C2 holds when changing the taxa level. It could be that C2 is more diverse than C1 at the genus level but not at the species level. For example, the five species of C2 could be from five different genera, while the five species of C1 could be from only three distinct genera. Therefore, the ordering of communities might differ according to the level of the taxa chosen. It becomes more challenging to compare the diversity between two communities.

The two other solutions use part or total information about the relative abundance (evenness). Yet, this principle assumes that it is possible to delineate individuals reliably. Again, this is difficult to obtain because of the issue of microorganism isolation. In AS and MGS, the number of DNA sequences per sample is used as a proxy for the number of individuals per sample. However, it has been shown that it is not always possible to equate one DNA molecule with one individual microorganism (see section 2).

Finally, the first solution answers the question (A) How many taxa compose this community? But not question (B) How are these taxa distributed? Nor (C) How different are these taxa? The two other solutions answer (A) and (B), but not (C). None of those solutions assess the functions of the communities, while it might be that C2 is more *functionally* diverse than C1.

Similar to the case encountered in the price index in economics (Reiss, 2008), the representational account of measurement cannot settle the question of which community is the more diverse. We need an index to measure diversity, but different indices give different results (Table 1), and the choice of the index matters for the conclusion.

3.2 The Operationalist Account of Measurement

Operationalism is about the meaning and use of quantity terms such as "length" (Tal, 2020). It is more a theory of concept legitimization than a theory of measurement

(Reiss, 2008, 69). More specifically, it asserts that the meaning of these quantity terms is determined by the set of operations used for their measurement, "In general, we mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations.*" (Bridgman, 1927, 5, original emphasis). In the length example, "length" means precisely the set of operations needed to obtain the value, e.g., the result of the concatenating rigid rods. In the extreme version of operationalism, different sets of operations measure different quantities. However, if the results of these different operations agree with experimental error, Bridgman conceded that it remains meaningful to label with the same name the corresponding quantities (Bridgman, 1927, 16).

Strong operationalism, as presented here, has several issues. One is that it is difficult to understand what is part of the set of operations. In the diversity measurement procedure described above (Section 2, Figure 1), different operations exist to extract the DNA before it is computerized. For example, scientists can use a mechanical lysis of the microorganismal membrane (usually performed using tiny rigid rods) or a biological lysis (using specific enzymes to degrade the microorganismal membrane). The issue with operationalism is whether these operations are relevant to determining the meaning of the quantity term' diversity.' In an extreme version of it, it does. That is, two slightly different operations of DNA extraction will lead to two different quantity terms. Then, the question is, how similar do these two operations need to be considered to play a part in the same set of operations?

A related issue is what an operation is. Or, more precisely, which scientific action is relevant to be part of the set of operations considered as defining the use of the quantity term? For example, is it relevant whether the DNA extraction scientist is wearing white or blue gloves? If we argue that neither of these examples should change the meaning and use of the quantity term 'diversity' because they should not influence

the measurement result, "we already depart from strict operationalism because we presuppose that these operations measure the same concept." (Reiss, 2008, 70).

Another issue with operationalism is known as 'concept proliferation.' If each set of operations corresponds to a different quantity term, science will be full of these non-related quantity terms. This is not the case in microbial ecology and microbiome studies. For example, if a mathematical formula counts as an operation, there is a difference in the operations needed to compute the Shannon index and the Chao1 index (see Table 1 for the equations). However, most researchers consider these two (and other indices) to measure α-diversity. The quantity term is the same, but not the methods. Yet, science is full of practices that use various methods to measure the same quantity term.

A way out of the issue of concept legitimization and toward a theory of measurement is the retention of some aspect of operationalism, e.g., the relevance of the operations used to measure a quantity in the meaning of that quantity but releasing the reductive pressure of the absolute equality between these operations and the meaning of the term. Conventionalism is an in-between approach that accepts that the application of a quantity term is somewhat chosen by convention yet still under some empirical property either of the attribute to be measured or of the operations used to measure the quantity (e.g., the physical property of a fluid which variation of its volume is used to assess the temperature) (Tal, 2020). An example using Table 1 is to decide by convention which of the indices presented here gives the correct quantity for diversity. Without going as far as that, Reese and colleagues have noticed that Shannon's index is the most commonly reported diversity index in the studies on gut microbiomes (36 studies on human and non-human gut microbiomes) they have considered in their paper (Reese & Dunn, 2018, 4). Moreover, after a comparison with other indices used in those studies, they suggest that "the data from comparisons of

Shannon diversity may be the most robust and informative, *particularly given the frequency with which this metric is already reported in the field.*" (Reese & Dunn, 2018, 7, emphasis added). This suggests that we are not far from a convention emerging from a natural consensus in the field. Yet, as pointed out elsewhere, "no matter which procedure is used, it is important that the decision about which index is correct is socially, and only socially, validated." (Reiss, 2008, 70).

The problem of the conventionalist solution is a certain fixity that stops the empirical investigation of the justification for measurement procedures. It demands too little empirical background knowledge for the justification of a specific procedure (Reiss, 2008, 70). How do you provide reasons other than convention and preferably empirically grounded to justify using specific measurement procedures compared to others? For example, the Shannon index is often used in microbial ecology and microbiome studies. However, there is no empirical justification to do so; why not, for example, use the Simpson index.

Despite its pitfalls, operationalism reminds us that each measurement procedure has its assumptions and idiosyncrasies and that "it is a fallacy to think different methods measure the same concept just because we attach the same name to them." (Reiss, 2008, 71). Although most micro ecologists would agree that diversity as measured by a richness index, whatever the taxa level, is a slightly different concept from diversity as measured by the Shannon index (which adds the distribution information). Other metrics are often compared between themselves as if this difference does not exist or at least as if they measure the same quantity at the end (see, e.g., Dubois, Girard, Lapointe, and Shapiro (2017); Schnorr et al. (2014)). Because "there is no universally accepted, absolute value of diversity for a given community" (Reese & Dunn, 2018, 2), the method used to measure it matters in the meaning of the quantity retrieved.

3.3 The Realist Account of Measurement

Realist accounts commit themselves to a particular ontology of the measurement procedure. Realist accounts consider measurement as the empirical examination of an objective or mind-independent property and/or relation. Objectivity here means independent from the measurer beliefs and conventions. The property is also independent of the methods and procedures used for measuring it (Tal, 2020). Moreover, the measurement procedure can never lead to the property's real value. It is only an estimation that gives approximations of true value. Realist accounts do not consider that the main aim of measurement procedures is the assignment of values to individual properties of an object but obtaining knowledge about properties and relations. Most of the defenders of realism about measurement have also defended some sort of realism about properties. For example, the property 'length' to which relations such as "longer than" and "sum of" can be attributed exist independently of whether the object to which we attach this property exists and independently of whether a human is evaluating and ordering objects.

Realist accounts have been developed against robust versions of operationalism and conventionalism. The main argument of realists is based on the observation of the practice of science. They claim that their accounts make sense of several scientific practices, such as the observation that usually, "quantities are ontologically prior to the procedure that measures them" (Tal, 2020), or that the scientific discourse is full of discourses about "measurement accuracy" and "measurement error." Realists also claim to better account for the idea that new measurement procedures are better in the sense of being more accurate than the older ones because the measured values are closer to the real ones.

If pressed upon, despite the lack of an unambiguous definition of diversity, microbial ecologists could argue that diversity is an objective property in a sense developed by

the realists: mind-independent, independent from their beliefs and conventions, independent from the methods and procedures used for measuring. At least some authors have argued that diversity should be considered along those lines, e.g., "diversity is not good or bad, it simply 'is'" (Shade, 2017, 1) or the paper entitled "In Nature, There Is Only Diversity" (McLaren & Callahan, 2018). However, being a realist about diversity does not solve the issue of choosing between different indices, especially when those indices give different values and do not correlate. The answer may depend on what one is a realist about: Is diversity believed to be a simple phenomenon that all the indices can capture, but these indices are noisy? (as if one was measuring the length of an object, but all our procedures were inaccurate and prone to error and uncertainties) or is diversity believed to be a complex phenomenon, with each index capturing a facet of this phenomenon? (Arnillas & Carscadden, 2024). In the first case, combining in some ways the values of different indices may help. In the second case (closer to microbial ecology), one needs to choose which facet is the most interesting to answer the research question and which index (among several) that captures this facet is the most suitable. Being a realist does not help to answer this situation.

Moreover, these realist positions often forget that diversity is, by definition, dependent on categories. A set of objects can be as diverse as we have categories to classify them. As mentioned several times in this paper, the idea of diversity depends on the concept of species or any concepts that delineate taxa (e.g., OTUs/ASVs). Even if we consider functional diversity, a community can be at maximum as diverse as we have categories to classify the functions it contains. In theory, we can devise as many functions as we want or as we see fit for our investigation, but this is partly decided by convention and not empirically. So, advocating a certain objectivity of the quantity diversity assumes that, in theory, there is an objective and natural way of separating microorganisms. It is a form of species (or at least taxa) realism. For example,

McLaren and Callahan (2018) consider that functional variations might be revealed at the species and/or strain level. But this is already assuming that such levels exist.

3.4 The Model-based Account of Measurement

Model-based account has been developed by studying contemporary work on metrology (Parker, 2017; Tal, 2020). It follows the International Vocabulary of Metrology (VIM):

> measurement assigns a symbol—the measurement result—to the object under measurement so that the symbol intends to provide descriptive information on the current state of the object with respect to one of its properties, called the *measurand*, i.e., the property intended to be measured (Mari, Carbone, & Petri, 2012, p. 2109, original emphasis).

The model-based account is a development and extension of the information-theoretic account proposed by van Fraassen. The critical point in both these accounts is the consideration of two levels in measurement procedures. The first one is the physical level. The measuring apparatus interacts with an object in a given state. This interaction is the input of the measuring system. It is encoded in a signal and then converted to a reading, constituting the system's output. This reading should reliably reflect the state (at least in some respect) of the object at the beginning of the interaction (van Fraassen, 2008, 150). The second one is an abstract level. Background theory represents the object's possible states on a parameter space. The "measurement outcome" is a location in this space. It is an incomplete representation of the item measured obtained "by displaying values of some physical parameters that - according to the theory governing this context - characterize that object." (van Fraassen, 2008, 179-80). A measurement outcome can be complex, and measuring procedures can include calculations and input from a model or theory (van Fraassen, 2008, 177).

The model-based account adds a modeling stage of the measurement procedure. In the modeling stage, the quantity to be measured (measurand) is idealized. The model of the measurand

> specifies the relevant structure of the object under measurement, the definition of the measurand itself in reference conditions (i.e., as if it were measurable by an ideal measuring system), and the set of quantity values which could be assigned to the measurand Giordani and Mari (2012).

Moreover, in the modeling stage, the entire measurement procedure is modeled. The quantities involved in the procedure, their interactions, the conditions of their interactions, and the whole experimental environment—including the measuring system, the measuring subject, and the empirical surrounding environment—in which the experiment is expected to be performed are idealized. The output of the measurement procedure then is the measurement result based on the information from the modeling stage and the information obtained in calibration. It is associated with a related uncertainty.

Assigning a specific value to the measurand and a related uncertainty must satisfy specific epistemic criteria. The model-based account emphasizes calibration, which is a methodological interpretation of the coherence criterion employed by van Fraassen. The coherence criterion emphasizes that the assumptions made to pass from the physical state of the object under measurement to the measurement outcome should be coherent with the background knowledge or other substantive presuppositions about the measured quantity (Tal, 2020). Calibration must guarantee consistency by ensuring the traceability of the measurement results to given references. Consistency is understood as different measurements of the same quantity should be statistically consistent with one another once uncertainties are considered (Parker, 2017, 6). Moreover, calibration also permits the estimation of uncertainties in a measurement procedure. It can be part of the measurement procedure itself, so in the modeling stage, scientists will attribute precisely and exhaustively the sources of error and uncertainty in the measurement procedure or assume in the uses of measuring apparatus or system already calibrated with available uncertainty estimate (Parker, 2017).

The model-based account of measurement and the account offered by van Fraassen share similar ideas, notably the one according to which measurement is relative to a theory or a model and the notion that coherence and consistency play a role in measurement. The main difference is that the modeling stage permits the distinction between relevant aspects of the object under measurement from artifacts produced during the measurement procedure (Parker, 2017, 6). This is so through the appeal to a goal of the measurement procedure, which will drive choices made in the modeling stage.

In summary, "a model of a measurement [procedure] is a representation of how an instrument or apparatus can be used to learn about the value of a parameter." (Parker, 2017, p. 4, 'process' changed for 'procedure' to remain consistent within the paper). It should consider the physical interactions that will happen and the data's subsequent treatment. It is idealized and does not account for all the tiny details of an actual measurement procedure. The assumption remains that these idealizations do not significantly impact the measurement results or are in a range acceptable for the accuracy demanded by the measurement goal, or these idealizations will be corrected at some point in the measurement procedure. Depending on the accuracy level requested, the model of the measurement procedure should also include steps to correct the deviations between the assumptions of the model and the actual state of the measurement procedure, complexifying the model.

The representational view of measurement has the drawback of needing a lot of background knowledge that is lacking in current microbiology investigations. Moreover, the ranking of the communities according to diversity necessitates some stability of the results, while so far, it is unclear that this is the case in microbial ecology. Finally, this account can not make sense of the aggregation of different quantities in one index for measuring diversity (as shown by the analysis of Table 1). The problem

of the operationalist account is the proliferation of various diversity concepts; this situation is not observed in the literature. It is also difficult to define exactly what an operation is and what it contains. One problem with the realist account is that it does not help choose between different diversity indices. Moreover, many of the measures used in microbial ecology are a function of the number of species or use the concept of species in one way or another. This means that being realists about the diversity property implies being realists about the concept of species. This is an uneasy move in the microbial world.

The model-based account provides a valuable tool for analyzing the issues in diversity measurement in microbial ecology and microbiome studies. First, it can remain agnostic about entities and property ontologies. Second, as shown below in section 4, the appeal to models can help clarify the role of different assumptions in the retrieved value for the quantity diversity. This can help solve the issue of the lack of reflexivity in MGS. Finally, I show in section 5 that it grounds scientific solutions in a broader conceptual framework (i.e., calibration) and develops and extends new solutions to justify measurement procedures (i.e., the use of the measurement's purpose).

4 Application of the Model-Based Account to the Measurement of Diversity in Microbial Ecology

To illustrate the adequation between model-based account and diversity measurement practices, I will now apply the model-based account of measurement to the assessment of the diversity of the microorganismal community and schematized in Figure 3.

The modeling stage first idealized the system under measurement. The system under measurement is the microorganismal community C delineated in a given environment. It can be modeled as a collection of DNA sequences obeying the assumption that one sequence is one individual, C^*. Second, it idealized the measurand, the property of interest associated with the system under measurement. The measurand is the

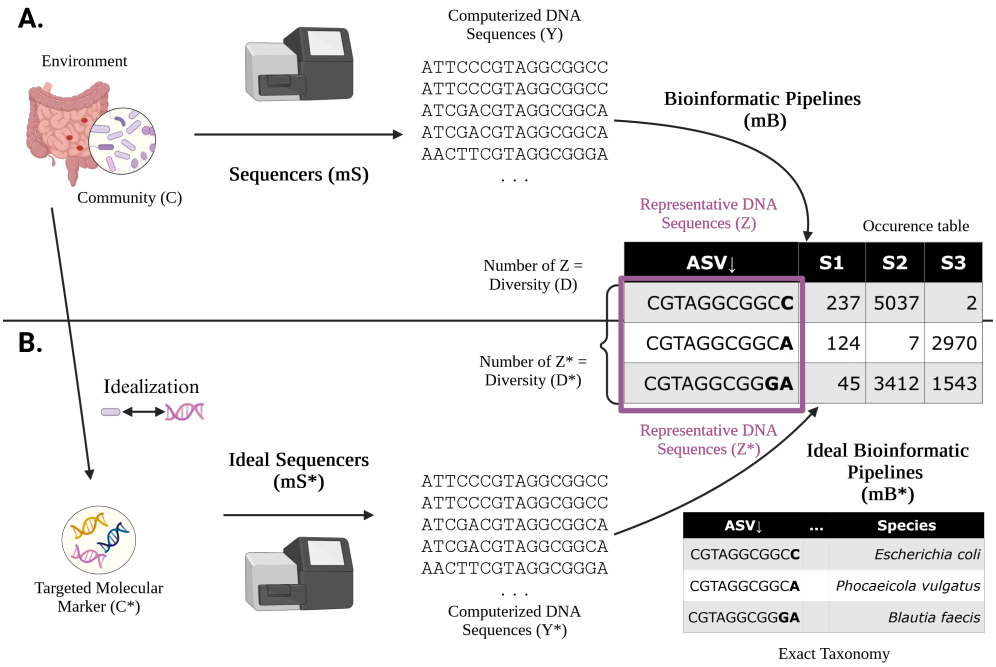

Fig. 3 Application of the model-based account to the amplicon sequencing method. **A.** Operative stage of the measurement procedure. **B.** Example of a model of the measurement procedure and its idealizations. Created in BioRender. Potiron, A. (2025) https://BioRender.com/k41m293

diversity, D. The modeler can model the measurand differently, e.g., as the taxonomic α-diversity understood as the species richness, D^*. This choice will also impact the measuring scale, e.g., positive integers.

Importantly, these idealizations are not imposed on the inquirer. The modeler has to choose which idealizations (of the system under measurement and of the measurand) to use depending on the goal of the measurement, and these choices will lead to various models for the measurement problem (Giordani & Mari, 2012). For example, if the goal is to describe global and local diversity patterns of the soil microbiome, one can measure the variation in the number of species across time (e.g., species loss) or space (e.g., the latitudinal gradient in biodiversity is the pattern according to which species richness increases closer to the equator). An index of species richness is suitable

because changes in abundance and distribution may not be essential information for these purposes. This is not so if the goal is to assess the impact of local disturbances on diversity, such as antibiotics. In that case, one wants to estimate how the targeted treatment has modified the structure of the community. One, therefore, also has to look at the evenness of the biological community. It is then justified to choose an index of "heterogeneity," i.e., an index that combines the information of richness and evenness. In both cases, various indices exist that measure richness (observed richness, Chao1, Chao2, etc.) or a mix of richness and evenness (Shannon, Simpson, inverse Shannon, etc.). The goals can be refined so as to narrow down the number of possible indices. Moreover, as explained in section 5.2, other criteria, such as available resources and background knowledge, will constrain the choice of the indices.

Additionally, the measuring instrument can already contain idealizations of the system under measurement and the measurand. Those idealizations could be more or less numerous depending on what is integrated into the measuring instrument.

The next step in modeling the measurement procedure is to find an ideal solution. That is, the empirical and mathematical operations needed to obtain a solution to the measurement problem are specified. To do so, it is possible to decompose the ideal problem (C^*, D^*) into a set of more tractable problems. This is an important step because, again, the modeler can choose how many tractable problems she needs to solve. For example, several decompositions are possible in AS (Figure 3.B.). The modeler can choose to decompose the problem into two tractable problems by assuming that a quantity value for D^* of C^* can be obtained by first computerizing a collection of DNA sequences Y^* for each sample of C^*, subsequently comparing Y^* following a similarity threshold given by a theory or background assumptions to obtain DNA sequences representative of one taxon Z^*. The ideal problem (C^*, D^*) is thus decomposed into two ideal problems (C^*, Y^*) and (Y^*, Z^*). Note that I have omitted all the operations needed before obtaining computerized DNA sequences from modeling.

An ideal solution is then sought attainable by ideal measuring instruments, i.e., their results are perfectly traceable to appropriate references and by an ideal measurement procedure specifying how these ideal measuring instruments are to be applied to the system under measurement. Following the decomposition above:

1. An ideal DNA sequencer m_S* is applied to C^*. The ideal procedure using m_S* makes no errors, does not contaminate the samples, and respects the original sample's abundance and relative distribution.
2. An ideal bioinformatic pipeline m_B* is applied to Y^*. m_B* suggests a perfect way of grouping elements of Y^*. The sequences deemed similar are genuinely similar, and sequences deemed dissimilar are truly dissimilar. Moreover, for each Z^*, a known reference is associated, e.g., the name of the species to which this DNA sequence belongs.

The solution to the ideal problem (C^*, D^*) is thus: i) apply m_S* to $C*$ according to the procedure and find a quantity value y^* of Y^*, ii) apply m_B* to y^* according to the procedure and find a quantity value z^* of Z^*, finally iii) add the number of z^* obtained following such procedures and find the result, i.e., a quantity value d^* of D^*.

Then, the operative stage can take place. The operative stage uses the knowledge provided in the modeling stage to perform the measurement procedure. The first level is experimental. First, some instances m_S and m_B of the ideal instruments m_S* and m_B* are obtained. The sequencers used in the measurement procedure are fallible and thus give sequences that might contain errors or contaminations. The bioinformatics pipelines have no perfect way of grouping elements of y^*. It groups molecular DNA according to a similarity threshold at which two DNA sequences of the same molecular marker are considered similar. The choice of the threshold at which two DNA sequences are sorted similarly by the bioinformatic pipeline has a considerable impact on the number of z retrieved and thus on the measure d of C. The threshold might not reflect the delineation required to identify the taxa. For instance, the threshold might

be too restrictive; thus, the taxa found do not correspond to any known references. Moreover, the references might not be complete enough to track all the sequences z^* representative of taxa.

Once instances m_S and m_B of the ideal instruments m_{S*} and m_{B*} are obtained, the system under measurement, C is coupled with the measuring instruments m_S and a quantity value y of Y is found. Then, m_B is coupled with y, and a quantity value z of Z is found.

The second level of the operative stage is abstract, theoretical, and/or statistical. It assigns a value to the measurand and is generally accompanied by a related uncertainty. This way, a solution to the measurement problem is found according to the procedure specified in the ideal solution. The number of z is summed up to obtain the quantity value d of D. For instance, the lines in the table in Figure 3 are an example of z. The number of lines gives the taxa richness or α-diversity of the community, C. In this example, $d = 3$ is expressed in ASVs with an uncertainty that needs to be described in terms of the assumptions made throughout the measurement procedure.

I have illustrated in this section that the ontological status of property diversity is not central. The important thing is how it is idealized when stating the measurement problem. I have also incorporated a set of assumptions and idealizations needed to model the entire measurement procedure. This is only a subset because the procedure includes numerous other steps that are not detailed here. These idealizations need to be explicitly established if one wants to evaluate the performance of the measurement procedure. By doing so, the model-based account forces scientists to be aware of these idealizations and to reflect systematically on them and their justifications in a given context. The epistemic gain is to pinpoint these idealizations to make clear the choices scientists need to make. The model-based account can thus help solve the lack of reflexivity present in MGS in a more systematic way. Finally, I have tiptoed toward solutions to help scientists choose and construct their measurement models. In the next

section, I developed in more detail the solutions for calibration of the measurement procedure and the need for the purpose of the measurement.

5 Solving the Measurement Problem

5.1 Measurement Procedures in Microbial Ecology Need Calibration

5.1.1 Calibration is the Hallmark of a Reliable Measurement Procedure

Calibration is needed to achieve consistency (see section 3.4). It is achieved by guaranteeing the intersubjectivity of the measurement outcomes. That is, different subjects in different contexts should be able to interpret the measurement outcomes in the same way (Frigerio et al., 2010, 142). Calibration standardizes the measurement outcomes obtained by the different measurement procedures of the same system under measurement. Using calibration procedures permits that even if the measurement procedures are not exact replications of each other, they produce, in principle, the same (or at least statistically consistent) results for the same system under measurement.

To achieve intersubjectivity, objectivity must first be met. In the model-based account, the measurement procedure's objectivity is understood as the procedure is independent of the surrounding environment and the evaluating subject. The result of such a procedure depends on the system under measurement and on the procedure (Frigerio et al., 2010, 136). The objectivity of the measurement result depends also on the quality of the interaction between the system under measurement and the measuring instrument used in the procedure.

Three criteria define how close to objectivity a measurement procedure is: stability, selectivity, and non-invasiveness. The measuring instrument used in the measurement procedure should be stable in its interaction with the system under measurement. If

stability is achieved, the measurement indications are only dependent on the state of the system under measurement and not also on some internal interference of the measuring instrument (Frigerio et al., 2010, 136).

Stability is important because it ensures that the measuring instrument(s) are reliable in two senses: i) the same measuring instrument applied to the same system under measurement will give the same or similar enough results (statistically consistent), and ii) different measuring instruments applied to the same system should be statistically consistent with one another once uncertainties are considered. Stability is part of the reason researcher can rely on their results and the results of others. In the context of microbial ecology and microbiome studies, this is important because the extent to which the measurement procedure is stable gives an idea of the extent to which the results can be relied upon, interpreted by different researchers, and especially compared to one another. For example, if two studies compare the diversity of the soil microbiome but the measurement procedure is not stable enough, those two studies cannot be compared, or only vaguely.

Stability is achieved through the structural possibility of calibrating the measuring instrument(s) using stable measurement standards (Frigerio et al., 2010, 142). Measurement standards in turn "are instantiations of a given property to which given property values are conventionally assigned to be used as reference values." (Frigerio et al., 2010, 142). Recall that the model-based account of measurement was developed by studying metrologic systems. Measurement standards are important components of these systems.

Selectivity has a similar role to stability in ensuring the reliability of the measure. Selectivity is the idea that the measurement procedure should be able to distinguish between the effect of the state of the system under measurement and the effect of the environment when analyzing the measurement indication. So, the measurement results should only depend on the system under measurement and not also on its

environment (Frigerio et al., 2010, 137). For example, in AS/MGS, if the environment or the measurement procedure itself introduces DNA contaminations, the procedure should be able to distinguish that from the natural DNA variation of the sample so that the researcher can rely on the result of the measurement.

Finally, the criterion of non-invasiveness tries to answer uncertainties and errors produced by the potential side-effect of the physical interaction between the measuring instrument and the system under measurement. The measurement procedure should be able to interact with the system under measurement without modifying the state of the system (Frigerio et al., 2010, 137).

5.1.2 Measurement Procedures in Microbial Ecology Are Complex

I use the taxonomy created by Parker (2017). She uses a synthesis between the information account presented by van Fraassen (2008) and a model-based account from Tal (2013) to distinguish three types of measurement: direct, derived, and complex (Parker, 2017).

A complex measurement procedure involves using different results together from multiple direct and/or derived measurements. In a direct measurement procedure (e.g., rain gauge to measure rainfall depth), there is no explicit symbolic calculation to obtain a value for the parameter of interest. The value can be corrected (e.g., for the wind), but the subject matter of the value remains unchanged (e.g., rainfall depth) (Parker, 2017). By contrast, in a derived measurement procedure, the value assigned for the parameter of interest is derived from or is calculated from the values measured for other parameters. The derivation or calculation needs to follow scientific principles or definitions (Parker, 2017, 9).

In a complex measurement procedure, the measurement outcome is more informative than any outcome obtained using only a subset of the results obtained by direct and/or derived measurements. A complex procedure involves additional assumptions

about how to combine these results in a meaningful way (Parker, 2017, 11). An example of a complex measurement procedure is averaging instrument indications from different calibrated instruments measuring the same parameter. Often, the assumption is that the individual measurement procedures are independent.

The measurement procedures described in the previous section are complex. It is the result of the aggregation of several derived measurements. For example, the value assigned to d in one sample is derived from at least one value z measured for another parameter Z. The parameter Z measured is the taxa of the individuals; it is qualitative (to which taxa a given "observed" DNA sequence is assigned), while the targeted parameter is the parameter diversity D, the number of taxa in a given area. Thus, D is a composite parameter. The additional scientific principle is the one according to which the sum of the number of categories in Z reflects D. However, at the end of the procedure, the quantity value d of C results from aggregating the derived measurement procedures performed in multiple samples (Figure 3). It is complex because d is more informative than the outcome obtained using the measurement procedure performed only on one sample.

Complex measurement procedures are not more difficult in principle to calibrate than direct procedures (e.g., averaging the weight scale). Nevertheless, the additional layers of inference embedded in complex measurement procedures imply that more steps are susceptible to need calibration. It hinges on whether potential errors are likely to affect the outcome significantly. I will show below that calibration procedures are indeed challenging in the case of microbial ecology.

5.1.3 Calibration is Underway

As described below, scientists are attempting to calibrate the microbial diversity measurement procedure. The model-based account makes coherent these attempts by emphasizing the need for calibration. Additionally, by explaining the reasons (reliability of the measurement procedure), the model-base account can ground the calibration

solution in a broader philosophical and conceptual framework that helps address other issues of measuring microbial diversity (e.g., the lack of reflexivity).

I start by evaluating the diversity measurement procedure regarding the selectivity and non-invasiveness criteria. In AS and MGS practices, the most significant environmental problem that can impair the diversity measure is the presence of contamination. The sample was contaminated when DNA that was not initially present in the system under measurement, the microorganismal community, was retrieved from the sample. Current practices, though, avoid this quite efficiently, so the procedure is pretty selective. However, the measurement procedure is highly invasive regarding the system under measurement (the microorganismal community)[10]. MGS methods destroy this system to obtain the DNA of its individual microorganisms. That is, the membranes of the microorganisms are ruptured to have access to their DNA, which is degraded by successive DNA amplifications (the microorganisms are long dead at this stage). Therefore, the criterion of non-invasiveness regarding the system under measurement is never met. Uncertainties and errors are created at this stage of the procedure (Alteio et al., 2021; McLaren & Callahan, 2018; Pollock, Glendinning, Wisedchanwet, & Watson, 2018). Thus, the quantity value d refers to the modified state in which the system is after sampling and DNA extraction instead of the initial, unaffected state of the system under measurement. The issue is that, from this point, there is no easy way to connect this measure to the 'real-world' measure of the system itself.

Moreover, total stability is difficult to achieve in MGS practices. Different samples of the same community can have different diversity values. For example, in the material part of the procedure, i.e., DNA extraction, amplification, and sequencing, the membrane of some microorganismal species is easier to break than others, making the

[10]It might not be invasive regarding the environment to be sampled if the sample is tiny (some grams of soil) or in the case of fecal samples, it is noninvasive regarding the host. However, in both cases, the procedure is invasive regarding the microbial community. I thank an anonymous reviewer for helping me precise this point.

former microorganisms' DNA easier to access than the latter. It creates a modification of the real-world proportion of the species. Similarly, the choice of the molecular marker used can privilege the representation of specific taxa compared to others. The measurement procedures for diversity in microorganismal communities using amplicon sequencing and, more generally, MGS are full of biases and assumptions; for a thorough review of these biases and references, see, e.g., McLaren and Callahan (2018); McLaren et al. (2022); McLaren, Willis, and Callahan (2019); Pollock et al. (2018).

Several strategies have been attempted to achieve more stability in the physical part of the measurement procedure. The first one concerns the standardization of the material part of the procedure (see, e.g., Pollock et al., 2018). This attempt tries to increase the stability of diversity measurement by minimizing the range of possible DNA extraction, amplification, and sequencing methods, increasing the intersubjectivity and reliability of the procedure. One issue is that standardization might be optimized (made stable and reliable) for one particular purpose and not others. The risk is that scientists might become overconfident. They will use this specific method for every purpose, thinking it is reliable per se and not only for the particular purpose for which it has been optimized. A second issue is that even if different standardized protocols are optimized for other purposes, scientists will still need to judge how close their research question (the purpose of their measurement) is to the standardization available in the literature (in the optimistic case where such standardization will ever exist).

The second strategy uses measurement standards to calibrate the bioinformatic part of the measurement procedure, m_B. Attempts to calibrate this part of the procedure are uncommon, but some can be found in the literature. For example, one study sampled a community of known composition and abundance (Pauvert et al., 2019). That is, strains from different species of fungi have been cultured using strains available in publicly accessible culture collection (= measurement standards). The first steps of

the procedure are standardized; that is, DNA extraction, amplification, and sequencing of different samples of this constructed community have been using the same DNA extraction/amplification kit and protocol and the same location and technology of sequencers (m_S). Then, different available bioinformatic pipelines were tested. None of the pipelines found the same richness and composition as the original. When quantifying the relative abundance of each species present, finding accurate values became even more challenging (Pauvert et al., 2019). A similar attempt uses simulated mock communities instead of material, artificial communities and achieves similar results (Mathon et al., 2021). In addition, extending their results to different environments is difficult, even if these procedures could lead to a good approximation of the original community. It is difficult to evaluate the uncertainty of the measures precisely. Outside guessing that the procedure is overestimating diversity or underestimating it, there is no precise way of knowing how much.

So far, MGS practices to evaluate the diversity of a microbial community are non-calibrated or pre-measurement practices (Frigerio et al., 2010, p. 132, footnote 7). Equivalently, they are measuring practice under development, i.e., lacking a reliable and rigorous process of calibration (Parker, 2017, 22). Although invasive, MGS practices have a certain degree of objectivity by being mostly selective and partially stable. However, so far, calibration attempts have failed, and the disciplines don't yet have good models of the measurement procedures, so the results obtained are not inter-subjective and not directly comparable (Reese & Dunn, 2018; Shade, 2017), see also McLaren and Callahan (2018); McLaren et al. (2019); Pollock et al. (2018). Thus, the model-based account grounds these attempts in a more encompassing conceptual framework. It explains the situation in microbial ecology and why scientists are looking for calibrated procedures even if they are difficult to obtain.

5.2 Measurement Procedures in Microbial Ecology Need a Purpose

The model-based account is a goal-oriented framework. As sketched in section 3.4, the goal of the measurement allows the distinction between artifacts of the measurement procedure and the relevant aspects of the object under measurement. Notably, the goal of the measurement procedure is crucial because it gives the specifications of the measurement procedure, e.g., the types of idealization authorized in the model. Moreover, achieving this goal justifies the resources employed in the procedure. Thus, the measurement problem has to be interpreted given the aim of the measurement (Giordani & Mari, 2012).

Considering the aim of the measurement might seem uncommon in measurement theory (although see Reiss, 2008) because of the worry of shaping your measurement procedure according to the result one wants to see. However, I propose here that the purpose of the inquiry, not a specific result of this inquiry, constrains but does not entirely determine the choice of the index. The appeal to the aim of measurement has been made by scientists in general ecology (Magurran, 2004; Vellend, Cornwell, Magnuson-Ford, & Mooers, 2011) but takes its time to take roots in microbial ecology (McLaren & Callahan, 2018; Normand et al., 2015; Reese & Dunn, 2018; Shade, 2017).

The goal of the measurement constrains the choice of the diversity index (as illustrated briefly in section 4). The model-based account clarifies where this choice happens: in modeling the measurement problem, when the researcher needs to idealize the system under measurement and the measurand. For example, the aim of the inquiry can guide scientists toward which kind of indices to choose — simple, heterogeneity, more complex, etc. The idea broadly is that all (the few ones I have described but also many more) these indices are available in a space from which the inquirer can choose. The purpose of the inquiry narrows down the possibilities by selecting subset(s) within these indices. The more the purpose is well-defined and precise, the

smaller the subset is. The purpose of the inquiry acts as an eliminative process, selecting a category of indices that are adequate for the purpose by eliminating those that are not or less adequate.

A minimal criterion for adequacy is that an index's assumptions and understanding of diversity do not contradict the purpose and are relevant to the context. Another criterion will be whether the index helps fulfill the purpose of the measurement, e.g., ranking unambiguously two or several communities. Different indices may be adequate for different purposes, as I have briefly illustrated in section 4.

While the purpose of an inquiry constrains the choice of diversity index, it does not entirely determine it. Other components will be considered, such as available resources, background knowledge, and empirical facts about the phenomenon. The availability and accessibility of material and digital resources, including human resources, equipment, statistical tools, software, and databases, can influence a scientist's choice of index, with familiarity and ease of use being essential considerations. For example, scientists may opt for an index they're familiar with and can compute using available software rather than one they're unsure about (circling back to the issue of lack of reflexivity). Background knowledge, including the inquirer's training and understanding of different indices, also narrows down the pool of options, while empirical constraints, such as the specificities of the community under study, further shape the choice of index. For example, facts about how the natural phylogenetic diversity has been obtained (e.g., macroevolutionary processes, environmental constraints on fitness, competition for resources, etc.) will influence the form of the cladogram that can be constructed for a given community. In turn, the form of the cladogram influences the choice of the index within available phylogenetic indices (Vellend et al., 2011, 195, 200).

6 Conclusion

In this manuscript, I show how issues specific to microbial ecology and microbiome studies render assessing the 'real/world' value of diversity (if such a thing exists) almost impossible. The instability of the taxonomic classification and the difficulty of isolating individuals increase the difficulty of measuring diversity reliably in such disciplines. In addition, the use of DNA-based methods and statistical tools renders the assessment of diversity easy to get, with little reflection on the reasons for this assessment and the problematic interpretation of the value obtained by such principles. I relate these issues to the measurement problem in philosophy, so I explore the central positions available in philosophy. After showing that the traditional positions encounter issues in the case of measuring diversity, I have argued that the more recent one, the model-based account of measurement, is the best framework to analyze these issues. The main advantage is that it points towards some solutions for scientists to justify their measurement procedures. The modelization of the measurement procedures can help solve the issue of the lack of reflexivity in using MGS methods. The model-based account also grounds solutions developed by scientists (i.e., calibration) in a comprehensive theoretical framework. Moreover, it points toward the development of clearly defining their purposes more systematically.

The implications of this analysis are double: it gives some avenues to scientists to improve their measurement procedures in these disciplines. As already mentioned, the calibration path is underway, and the extent of its achievements will remain to be seen in future years. Additionally, it is a successful use of the philosophical model-based account. Further studies beyond the scope of the one presented here could assess its usefulness to deal with measurement problems encountered in other disciplines, like biology (e.g., cancer cell population delineation). It would be interesting to see how similar or dissimilar these cases would be from measuring diversity in microbial ecology.

References

Alteio, L.V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., ... Schmidt, H. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry*, *160*, , https://doi.org/10.1016/j.soilbio.2021.108357

Arnillas, C.A., & Carscadden, K. (2024). When indices disagree: facing conceptual and practical challenges. *Trends in Ecology and Evolution*, *39*(7), 634-643, https://doi.org/https://doi.org/10.1016/j.tree.2024.02.001 Retrieved from https://www.sciencedirect.com/science/article/pii/S0169534724000387

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.C.C., Charles, T., ... Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, *8*(1), 1–22, https://doi.org/10.1186/s40168-020-00875-0

Bridgman, P.W. (1927). *The logic of modern physics.* Oxford, England: Macmillan.

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., ... Webster, N.S. (2019, 9). Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, *17*(9), 569–586, https://doi.org/10.1038/s41579-019-0222-5 Retrieved from https://www.nature.com/articles/s41579-019-0222-5

Chong, C.W., Ahmad, A.F., Lim, Y.A.L., Teh, C.S.J., Yap, I.K.S., Lee, S.C., ... Chua, K.H. (2015). Effect of ethnicity and socioeconomic variation to the gut microbiota composition among pre-adolescent in malaysia. *Scientific Reports*, *5*(1), , https://doi.org/10.1038/srep13338

Daybog, I., & Kolodny, O. (2023, 12). A computational framework for resolving the microbiome diversity conundrum. *Nature Communications*, *14*(1), , https://doi.org/10.1038/s41467-023-42768-4

Doré, J., Tap, J., Ehrlich, D. (2017). Composition du microbiote intestinal. *Le microbiote intestinal : Un organe à part entière* (John Libbe ed., pp. 13–23). Montrouge.

Dubois, G., Girard, C., Lapointe, F.J., Shapiro, B.J. (2017, 11). The Inuit gut microbiome is dynamic over time and shaped by traditional foods. *Microbiome*, *5*(1), 151, https://doi.org/10.1186/s40168-017-0370-7

Falkowski, P.G., Fenchel, T., Delong, E.F. (2008, 5). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*(5879), 1034–1039, https://doi.org/10.1126/science.1153213 Retrieved from https://www.science.org/doi/10.1126/science.1153213

Frigerio, A., Giordani, A., Mari, L. (2010). Outline of a general model of measurement. *Synthese*, *175*(2), 123–149, https://doi.org/10.1007/s11229-009-9466-3

Giordani, A., & Mari, L. (2012). Measurement, models, and uncertainty. *IEEE Transactions on Instrumentation and Measurement*, *61*(8), 2144–2152, https://doi.org/10.1109/TIM.2012.2193695

Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D.H., Soo, R.M. (2021, 7). *Prokaryotic taxonomy and nomenclature in the age of big sequence data* (Vol. 15) (No. 7). Springer Nature.

Hugenholtz, P., Goebel, B.M., Pace, N.R. (1998). *Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity* (Vol. 180; Tech. Rep. No. 18). Retrieved from http://crab2.berkeley.edu/pacelab/176.htm.

Jiao, N., Baker, S.S., Nugent, C.A., Tsompana, M., Cai, L., Wang, Y., ... Zhu, L. (2018). Gut microbiome may contribute to insulin resistance and systemic inflammation in obese rodents: a meta-analysis. *Physiol Genomics*, *50*, 244–254, https://doi.org/10.1152/physiolgenomics.00114 Retrieved from https://gordonlab.wustl.edu/

Krantz, D., Luce, R., Suppes, P., Tversky, A. (1971). *Foundations of Measurement Vol 1: Additive and Polynomial Representations.* San Diego and London: Academic Press.

Lowe, J.W., & Ingram, D.S. (2023, 8). DNA barcoding and the changing ontological commitments of taxonomy. *Biology and Philosophy*, *38*(4), , https://doi.org/10.1007/s10539-023-09916-y

Maclaurin, J., & Sterelny, K. (2008). *What Is Biodiversity?* The University of Chicago Press.

Magurran, A.E. (2004). *Measuring Biological Diversity*. Oxford: Blackwell Science.

Mari, L., Carbone, P., Petri, D. (2012). Measurement fundamentals: A pragmatic view. *IEEE Transactions on Instrumentation and Measurement*, *61*(8), 2107–2115, https://doi.org/10.1109/TIM.2012.2193693

Mathon, L., Valentini, A., Guérin, P.E., Normandeau, E., Noel, C., Lionnet, C., ... Manel, S. (2021, 10). Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources*, *21*(7), 2565–2579, https://doi.org/10.1111/1755-0998.13430

McLaren, M.R., & Callahan, B.J. (2018). In Nature, There Is Only Diversity. *mBio*, *9*(1), 02149–17, https://doi.org/https://doi.org/10.1128/mBio.02149-17

McLaren, M.R., Nearing, J.T., Willis, A.D., Lloyd, K.G., Callahan, B.J. (2022, 1). Implications of taxonomic bias for microbial differential-abundance analysis. *bioRxiv*, 2022.08.19.504330, https://doi.org/10.1101/2022.08.19.504330 Retrieved from http://biorxiv.org/content/early/2022/10/08/2022.08.19.504330.abstract

McLaren, M.R., Willis, A.D., Callahan, B.J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *eLife*, , https://doi.org/10.7554/eLife.46923.001 Retrieved from https://doi.org/10.7554/eLife.46923.001

Mittelbach, G.G., & McGill, B.J. (Eds.). (2019). *Community Ecology* (Second Edition ed.). Oxford: Oxford University Press.

Normand, P., Duran, R., Le Roux, X., Morris, C., Poggiale, J.-C. (2015). Biodiversity and Microbial Ecosystems Functioning. *Environmental microbiology: Fundamentals and applications* (pp. 261–291). Dordrecht: Springer Netherlands.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., ... Weedon, J. (2024). *vegan: Community Ecology Package.* Retrieved from https://github.com/vegandevs/vegan

Oren, A. (2024, 3). On validly published names, correct names, and changes in the nomenclature of phyla and genera of prokaryotes: a guide for the perplexed. *npj Biofilms and Microbiomes*, *10*(1), 20, https://doi.org/10.1038/s41522-024-00494-9 Retrieved from https://www.nature.com/articles/s41522-024-00494-9

Parker, W.S. (2017, 3). Computer simulation, measurement, and data assimilation. *British Journal for the Philosophy of Science*, *68*(1), 273–304, https://doi.org/10.1093/bjps/axv037

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., ... Vacher, C. (2019, 10). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology*, *41*, 23–33, https://doi.org/10.1016/j.funeco.2019.03.005

Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M. (2018). The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome

Studies. *Applied and Environmental Microbiology.*, *84*(7), 1–12,

Reese, A.T., & Dunn, R.R. (2018, 7). Drivers of microbiome biodiversity: A review of general rules, feces, and ignorance. *mBio*, *9*(4), , https://doi.org/10.1128/mBio.01294-18

Reimer, L.C., Sardà Carbasse, J., Koblitz, J., Ebeling, C., Podstawka, A., Overmann, J. (2021, 10). Bacdive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Research*, *50*(D1), D741-D746, https://doi.org/10.1093/nar/gkab961 Retrieved from https://doi.org/10.1093/nar/gkab961 https://academic.oup.com/nar/article-pdf/50/D1/D741/42058564/gkab961.pdf

Reiss, J. (2008). Measurement between the absolute and the arbitrary. *Error in economics* (pp. 60–81).

Rosonovski, S., Levchenko, M., Bhatnagar, R., Chandrasekaran, U., Faulk, L., Hassan, I., ... Harrison, M. (2023, 11). Europe pmc in 2023. *Nucleic Acids Research*, *52*(D1), D1668-D1676, https://doi.org/10.1093/nar/gkad1085 Retrieved from https://doi.org/10.1093/nar/gkad1085 https://academic.oup.com/nar/article-pdf/52/D1/D1668/55040834/gkad1085.pdf

Santana, C. (2014, 10). Save the planet: eliminate biodiversity. *Biology and Philosophy*, *29*(6), 761–780, https://doi.org/10.1007/s10539-014-9426-2

Sarkar, S., & Margules, C. (2002). Operationalizing biodiversity for conservation planning. *Journal of Biosciences*, *27*(4), 299–308, https://doi.org/10.1007/

BF02704961 Retrieved from https://doi.org/10.1007/BF02704961

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., ... Crittenden, A.N. (2014, 4). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, *5*, , https://doi.org/10.1038/ncomms4654

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., ... Karsch-Mizrachi, I. (2020, 08). Ncbi taxonomy: a comprehensive update on curation, resources and tools. *Database*, *2020*, baaa062, https://doi.org/10.1093/database/baaa062 Retrieved from https://doi.org/10.1093/database/baaa062 https://academic.oup.com/database/article-pdf/doi/10.1093/database/baaa062/33570620/baaa062.pdf

Shade, A. (2017, 1). *Diversity is the question, not the answer* (Vol. 11) (No. 1). Nature Publishing Group.

Sneath, P.H. (2005). Numerical Taxonomy. D. Brenner, N. Krieg, J. Staley, & G. Garrity (Eds.), *Bergey's manual® of systematic bacteriology* (pp. 39–42). Boston, MA: Springer US.

Tal, E. (2013). *The Epistemology of Measurement: A Model-based Account* (Unpublished doctoral dissertation). University of Toronto.

Tal, E. (2020). *Measurement in Science* (Fall 2020 ed.). Edward N. Zalta. Retrieved from https://plato.stanford.edu/archives/fall2020/entries/measurement-science/

van Fraassen, B.C. (2008). *Scientific Representation*. Oxford University PressOxford.

Vellend, M., Cornwell, W.K., Magnuson-Ford, K., Mooers, A. (2011). Measuring Phylogenetic Biodiversity. A. Magurran & B.J. McGill (Eds.), *Biological diversity: Frontiers in measurement and assessment* (pp. 194–207). Oxford: Oxford University Press.