

Opinion piece



Cite this article: Prosser JI. 2020 Putting science back into microbial ecology: a question of approach. *Phil. Trans. R. Soc. B* **375**: 20190240.
<http://dx.doi.org/10.1098/rstb.2019.0240>

Accepted: 14 January 2020

One contribution of 19 to a theme issue 'Conceptual challenges in microbial community ecology'.

Subject Areas:

ecology

Keywords:

scientific method, hypothesis, induction, community ecology, diversity, microbiome

Author for correspondence:

James I. Prosser
e-mail: j.prosser@abdn.ac.uk

Putting science back into microbial ecology: a question of approach

James I. Prosser

School of Biological Sciences, University of Aberdeen, Cruickshank Building, St Machar Drive, Aberdeen AB24 3UU, UK

JIP, 0000-0003-1757-5102

Microbial ecology, the scientific study of interactions between natural microbial communities and their environments, has been facilitated by the application of molecular and 'omics'-based techniques that overcome some of the limitations of cultivation-based studies. This has increased emphasis on community ecology and 'microbiome' studies, but the majority address technical, rather than scientific challenges. Most are descriptive, do not address scientific aims or questions and are not designed to increase understanding or test hypotheses. The term 'hypothesis' is increasingly misused and critical testing of ideas or theory is restricted to a small minority of studies. This article discusses current microbial ecology research within the context of four approaches: description, induction, inference to the best explanation and deduction. The first three of these do not follow the established scientific method and are not based on scientific ecological questions. Observations are made and sometimes compared with published data, sometimes with attempts to explain findings in the context of existing ideas or hypotheses, but all lack objectivity and are biased by the observations made. By contrast, deductive studies address ecological questions and attempt to explain currently unexplained phenomena through the construction of hypotheses, from mechanism-based assumptions, that generate predictions that are then tested experimentally. Identification of key scientific questions, research driven by meaningful hypotheses and adoption of scientific method are essential for progress in microbial ecology, rather than the current emphasis on descriptive approaches that address only technical challenges. It is, therefore, imperative that we carefully consider and define the fundamental scientific questions that drive our own research and focus on ideas, concepts and hypotheses that can increase understanding, and only then consider which techniques are required for experimental testing.

This article is part of the theme issue 'Conceptual challenges in microbial community ecology'.

1. Introduction

Microbial ecology is arguably the most important and least developed area of ecology. Microbes are ubiquitous, occupy the broadest range of environments, with the broadest range of environmental conditions, and are essential for all biogeochemical processes and for the existence of all animals and plants. Despite historical lack of awareness of their importance, we are now considered by many to be in a golden age of microbial ecology. The number of microbial ecology research papers has certainly increased significantly in the past three decades, although at a similar rate to those on plant and animal ecology. Increased research activity is, in part, due to the development of cheaper and faster sequencing methodologies and their use in characterizing microbial communities. This has led to the discovery of unexpectedly high diversity, previously uncultured microbes, indications of their potential function and popularization of the 'microbiome' concept. Microbial ecologists have, of course, been studying microbial communities for decades, but increased research has highlighted potential issues regarding the motivation and aims of microbial ecology research in general.

Microbial ecologists aim to gain an understanding of the relationships and interactions between microorganisms and their environments. Through the scientific method, attempts are made to explain observations and phenomena that cannot currently be explained, to find general principles or theories that operate across organisms and environments and to test these by experimentation. The scientific method has advanced over the past four centuries and remains an area of active study within the philosophy of science. Many aspects are still subject to debate and the details of scientific method should not be considered as fixed or final; nor is there necessarily a perfect approach for each ecological study. There are, however, imperfect approaches that are not designed to, and are incapable of, increasing understanding and which lead to confusion, misunderstanding, propagation of wrong ideas and wasted resources. Here, I will consider the approaches used to study microbial ecology and their benefits and limitations.

I will discuss four approaches: description, induction, inference to best explanation and deduction. This 'classification' is not perfect, but it provides a framework. A detailed discussion of these approaches is beyond the scope of this article and readers are referred to textbooks, reviews and online resources on scientific philosophy (e.g. [1–6]). (Note that I am concerned with those aspects of scientific philosophy that provide analysis of, and guidance on how research should be performed, rather than those that describe how scientists behave.) I will consider fundamental, rather than applied studies, and will illustrate ideas through examples from microbial community ecology, particularly my own area of research (soil ammonia oxidizers). The issues raised, however, apply to all areas of microbial ecology and are not unique to microbial ecology, or even ecology.

2. Description

Descriptive or 'look-see' studies involve observations and measurements of microbes and their environments, but with no intention of explaining these observations or increasing understanding.

(a) Who is there?

The most obvious examples of descriptive studies are surveys that catalogue the microbes present in an environment. This approach is termed nature study, when performed by amateurs with basic techniques, or natural history, if more advanced techniques and analysis are used by professional scientists; it also applies to mining of large quantities of data. Until the 1990s, this required laboratory cultivation, from an environmental sample, of as many microbes as feasible and phenotypic classification of each (mainly through determination of numerous physiological characteristics during laboratory growth). The current approach is high-throughput sequencing of 16S rRNA genes, with identification and phylogenetic analysis following comparison with database sequences. Sequence analysis of functional genes provides a similar description of organisms with the potential for a particular function, e.g. nitrogen fixation, denitrification. These molecular techniques provide limited information on phenotype but are relatively cheap and rapid; crucially, they do not involve laboratory cultivation and may provide better phylogenetic information.

The major limitation of descriptive studies is that they are not driven by scientific questions or theories. They are aimless and cannot, in themselves, answer scientific questions. Sequence surveys may, and probably will lead to 'discovery' of new phylotypes, but they cannot increase understanding of their ecology or ecosystem function. In the absence of an aim, there is no basis for determining or justifying study design, sampling protocols, choice of gene(s) or analysis methods. There is no way of determining, before or after, when or whether enough sequences have been obtained. Indeed, there are no criteria for assessing or justifying the need for the study, the resources (time and money) required or the value of the data obtained.

Descriptions are also, necessarily, limited by and wholly reliant on the techniques that are available, affordable, feasible and sufficiently rapid, and on their accuracy. A criticism of cultivation-based surveys was selectivity of laboratory growth media and conditions, a major concern acknowledged by microbial ecologists prior to 1990. Although the remarkable findings of early molecular surveys demanded an assessment of biases and limitations of molecular techniques, familiarity with these techniques and use of standardized methods and analytical software have reduced awareness and consideration of their limitations. Examples include cell lysis bias, extraction efficiency, extracellular nucleic acids, primer bias, variation in gene copy number with growth rate, and other intrinsic and unavoidable biases. Molecular techniques, therefore, enable characterization of uncultivated organisms, but have introduced new biases and limitations and are far from perfect. In 2050, current molecular techniques will probably be considered as primitive as we currently view those used in the 1980s.

(b) What is meant by 'there'?

The above descriptions relate to organisms and which phylotype is (potentially) doing what, but ecology is the interaction between microbes and their environment, implying that physical, chemical and biological characteristics of the environment and biogeochemical process rates should also be surveyed. Techniques for measuring environmental characteristics have also advanced, and those that are cheap and easy are frequently measured, but surveys tend to focus on the organisms (or at least gene sequences), rather than their environment. Again, though, the lack of scientific aims means that there are no criteria for assessing how many and which environmental characteristics should be measured.

(c) Who is doing what?

Descriptive studies may also seek information on which phylotype is doing what. Cultivation-based approaches automatically generated information on function, but 16S rRNA surveys provide only limited information. Asexual reproduction and horizontal gene transfer at all phylogenetic levels prevent consistent or meaningful definitions of species, and other taxonomic units, allowing only subjective arbitrary operational definitions (hence my use of 'phylotype' throughout this article). They also severely limit prediction of phenotypic characteristics from phylogeny. The links between function and phylogeny are not completely destroyed; if they were destroyed, our task as microbial ecologists would be impossible. They are, however, limited and a major challenge is to understand these limitations, their extent and the consequences (see, for example, [7,8] and other articles in this theme issue). In addition, large proportions of natural

communities, as determined by 16S rRNA gene studies, have not been cultivated and their potential functions are not known.

Current attempts to describe microbial function involve metagenomics and other 'omics' approaches or, for cultivated organisms, analysis of genomes. Omics approaches also have major limitations [9] and, at best, provide information only on potential activity. Although avoiding cultivation, descriptions of potential function again rely on the efficiency and accuracy of techniques used. Many genes will be transcribed and translated only under specific conditions, many will be in dormant or dying cells, the predicted function of many (often most) will be unknown or inaccurate, quantitative functional information is lacking and many important ecophysiological characteristics have no obvious genetic determinant.

Importantly, however, even if we knew the function of every gene in every microbial cell in an environmental sample, and knew which genes were transcribed and which proteins were present, mere description of genes, transcripts and proteins would not increase understanding. Metagenomic or genomic surveys that lack a question or aim are, therefore, unbounded and lack criteria with which to determine endpoints, relevance or success. In the absence of such criteria, we must question the scientific reasons for their existence.

(d) What is the effect of ...?

Descriptive studies are, of course, not restricted to community surveys. For example, many investigate the effect of a particular environmental factor on microbes or their activities. Although these studies may have an objective and a question, e.g. 'does temperature influence soil microbial community composition or activity?', observing and describing the effect does not increase understanding or provide explanations that may be more broadly relevant. Again, the objectives and questions are technical and not scientific. Unfortunately, these studies are sometimes even presented as testing hypotheses (e.g. 'we hypothesized that temperature affects communities'), but these are not meaningful or scientific hypotheses and this represents misuse of the term 'hypothesis' (see below), often in an attempt to make a study seem more 'scientific'.

Descriptive studies, therefore, may address technical questions and challenges, but not scientific challenges. The ease with which molecular, genomic or metagenomic data can be obtained invites their collection in the hope that something interesting may 'fall out' of the data.¹ This may happen but the probability of answering an important ecological question without first asking or having a question is low. Similarly, the chances of finding an explanation for an environmental phenomenon without first observing the phenomenon is low. As a consequence, these studies lead to desperate attempts to find a question to which the data provide an answer, to provide belated justification for the study and to 'make a story out of the data', with the unavoidable bias and subjectivity that this entails.

(e) Are there purely descriptive studies?

In practice, descriptive studies may not always be performed 'in the dark'. Purely descriptive studies are sometimes justified as providing baseline data, e.g. prior to monitoring following environmental change, or when using new techniques or exploring previously unstudied environments. However, it would be surprising, and worrying, if there was no reason

for expecting environmental change to influence a community or for having an interest in a new site, and how it might differ from others. These reasons should provide the basis for scientific questions or hypotheses. Similarly, every environmental characteristic cannot be measured and every organism cannot be characterized. Decisions must, therefore, be made on which microbial groups (bacteria, archaea, fungi, protozoa, functional groups) and characteristics are measured. Again, it would be surprising and worrying if these decisions were not made on the basis of underlying views of what is important and interesting. There is, therefore, an implicit, if not stated reason for expecting something different or unusual that provides the basis for a meaningful study that, if defined and discussed prior to measurements, would provide the basis for a scientific question and rational experimental design. Indeed, results arising from descriptions are often described as surprising or unexpected and 'effect of' studies are often described as interesting or even 'successful'. All of these terms indicate that there was prior information that could and should have been formulated as a hypothesis. Another perceived benefit of descriptive studies is in extending databases that others might find useful in answering ecological questions. However, critical testing of ecological questions generally requires well considered experimental design, rather than analysis of data that have been collected randomly.

3. Induction

In its simplest form (enumerative) inductive reasoning involves creation of a general rule from a number of observations for a particular class and inference that all members of that class will follow the same rule. To employ a common example of induction, we examine 50 swans, observe that all are white and then infer that all swans are white. This approach was criticized by the eighteenth century philosopher David Hume, who questioned whether 'instances of which we have had no experience resemble those of which we have had experience' [10, part III, §VII], i.e. he questioned whether the past could tell us about the future. He concluded that this process was not based on reason and was interested in why it was adopted (the problem of induction), suggesting that it was based on imagination, custom and habit. It is not difficult to find support for this view. There are many cases in which the past has not predicted the future and previous success in predicting the future does not logically mean that future predictions will be successful.² Induction can be defined more broadly, e.g. relaxing the need for generalizations and considering the number of 'supportive' observations and the lack of contradictory observations (naive induction). However, the inductive approach is based solely on observations and does not provide information on causes or mechanisms. We might be correct in inferring that all swans are white (although black swans do exist), but we have no information on why this might be the case, i.e. we may have knowledge, but we have no understanding and no way of confirming our knowledge.

There are many examples of induction in microbial ecology, e.g. inference that all ammonia oxidizers are bacteria, which was believed until 2005 because all cultivated ammonia oxidizers were bacteria; or that all plant rhizosphere microbiomes contain a particular phylotype because it has been observed in 50 plants; or that a physiological characteristic of a single isolate or a gene in a single genome or reconstructed metagenome

will exist in all closely related phylotypes. Similar extrapolation from properties of isolates to those of microbial relatives in natural environments has always been a concern of cultivation-based approaches. Model organisms are important, e.g. for in-depth physiological studies, but they will differ from the organism that was originally obtained from the environment, because of rapid physiological and genetic changes occurring during isolation, requiring care in predicting its ecology. (We would be very wary of predicting the ecology of all cats on the basis of the characteristics of a wild cat that had been domesticated for 100 generations. Nevertheless, we frequently see unqualified assumptions about the ecology of relatives of a microbial culture, after a similar number of generations in laboratory culture.) Single-cell genomics has also demonstrated considerable genomic diversity within individual rRNA-defined phylotypes (e.g. [11]), preventing prediction of all of the genetic characteristics of other phylotype members. Nevertheless, induction is used to infer, often with no qualification or reservation, the functions and metabolic pathways of relatives in the environment on the basis of a single gene, genome or chimeric metagenome-assembled genome (even when more than 50% of gene functions are unknown).

This is not to suggest that molecular, genomic or metagenomic approaches have no value, but they have very little value if based on induction alone. In fact, molecular techniques have themselves provided many examples of the dangers of induction and unquestioning 'knowledge'. This does not mean, for example, that we should ignore recurring observations but we should not predict outside the range of our experience in the absence of understanding.

(a) Induction and correlation

Correlation analysis is a powerful statistical technique that can be used to test theoretical predictions. However, it is most frequently used in descriptive or inductive studies, involving quantitative analysis of correlations or associations. Indeed, the discovery of high microbial diversity has fuelled many 'explorations' of correlations or associations between community composition (relative abundances of phylotypes or genes) and environmental characteristics.

I will illustrate this approach with an example from my own research area, where correlations are determined between soil ammonia oxidizer communities and soil characteristics. Communities are characterized by analysis of 16S rRNA or a functional gene (*amoA* genes for ammonia oxidizers) sequences and 5–10 soil characteristics, typically pH, moisture content, total C, total N, C:N ratio and sometimes soil P, ammonium and nitrate. After grouping sequences into arbitrary operational taxonomic units (OTUs), statistical methods are used to quantify correlations between phylotype relative abundance and soil characteristics, with attention focused on those associations considered to be statistically significant. Purely descriptive studies stop at this point and merely report the associations with no interpretation. Others invoke induction and infer that the associations observed apply to all soils and all ammonia oxidizer communities. Note that these inferences arise solely from the experimental data after they have been collected and analysed. Note, also, that these studies make no *a priori* predictions, e.g. there is no consideration of which ammonia oxidizer phylotypes might increase or decrease, or why. These studies only 'explore'. In providing only descriptions, they suffer the limitations described above: lack of

scientific aims and the absence of criteria for justification, experimental design, endpoints, etc.

A similar approach is adopted in biogeographic studies, in which these environmental characteristics are supplemented or replaced by environmental characteristics such as latitude, altitude, mean annual temperature (but not soil temperature), mean annual precipitation (but not soil moisture content), net primary production, vegetation. The characteristics are chosen to differentiate the different environments, regardless of their possible influence on microbial communities, which will be broad, frequently indirect, often overlapping and often negligible. For example, microbial activity will be affected by soil pH, but this may be through its influence on the availability of nutrients or on other microbes, plants and animals within the community and, thereby, microbial interactions. Further issues associated with the choice of environmental characteristics are discussed below.

More commonly, results will be compared to those from other studies, to gain support for the inference that the associations and patterns observed apply generally. This requires objectivity but, unfortunately, it is easier to search for data that support an inference than data that do not. It is also easy to discount lack of support through, e.g. different soils, methodologies, experimental design, etc., while similar criticisms are not made of supportive data. In addition, while the strength of associations is quantified, comparisons with previous data are usually made on the basis of qualitative or at best semi-quantitative comparisons, such that interpretations become matters of degree and opinion and lack precision.

There are also fundamental issues associated with correlation studies, e.g. what is the minimum number of characteristics required to discriminate the number of phylotypes being considered, what is the 'best' spatial scale, how can this be assessed without aims and criteria? The only (apparent, but usually unstated) aim of these studies is to explore and look for associations. The approach is based on the premise that the inferences are correct and aims to accumulate evidence in support of the inferences. It, therefore, becomes like a football match, or basketball game (depending on the number of studies), with scores for and against depending on the numbers of supportive and conflicting studies. As for descriptive studies, these studies are unbounded in that there is no scientific aim or objective with which to determine the number of studies required for confirmation of the inferences.

Inductive studies, therefore, range from naive induction, which is no more advanced than superstition or 'belief', through attempts to correlate communities with environmental factors to large scale correlation-based analyses of biogeographic patterns. Although many induction studies highlight the need to increase understanding, they are not designed or able to do so. They provide knowledge but not understanding. We might 'know', and we might even be correct in 'knowing' that the relative abundance of a particular ammonia oxidizer is always favoured by a particular combination of the 5–10 soil characteristics measured, but this approach gives no information on why, i.e. there is no mechanistic information. A dramatic illustration of this distinction is the example of a turkey being reared for Thanksgiving Day [12]. Turkey contentment increases during rearing, as the turkey is well-fed, warm, dry, disease-free and much more content than wild turkeys. There will be a strong correlation between turkey contentment and contact with the farmer, who provides food and shelter. This high correlation, however, does not allow

the turkey to infer or predict the future correctly, and contentment falls abruptly when it is slaughtered by the farmer. The turkey had only knowledge, while the farmer had both knowledge and understanding.

4. Inference to best explanation

Inference to best explanation describes interpretation of inferences, arising from data, in the context of existing or new hypotheses or mechanisms in attempts to find the hypothesis that best explains the data, concluding that this hypothesis is true. For example, selection of a particular phylotype under certain environmental conditions may be explained through existing knowledge of physiological or genetic characteristics of its relatives. In fact, many community studies are implicitly, if unconsciously, testing the concept of niche specialization and differentiation. (See [13] for steps involved and some consideration of its application to microbes.) Briefly, this posits that environmental characteristics will lead to evolution and selection of strains whose physiological characteristics are best adapted to those environmental characteristics. The relative abundance of these strains will increase and they will be dispersed, colonizing other similar environments, again through selection based on their physiology. Both 16S rRNA- and functional gene-analyses enable tracking of phylogenetic groups and correlation is predicted between phylotypes and ecological conditions. The validity of this concept depends on the validity of the assumptions on which the concept is based. In particular, it assumes strong links between phylogeny and function, which is considerably less in prokaryotes than in animals and plants for which the concept was developed (see [13]).

(a) Niche specialization and microbiomes

In fact, this concept is incorporated, though not intentionally, in the term 'microbiome'. 'Biome' has been used for many years to describe plant or animal communities that have common characteristics for the particular environment in which they are found, e.g. temperate forests. This implies links between physiological and environmental characteristics, for which there is evidence in plant and animal communities. The 'micro' prefix represents microbial community and, for example, the term soil microbiome not just refers to soil microbial community composition but also suggests that the community is special in some way, with physiological characteristics selected by the physical, chemical and biological characteristics of soil. In addition, the 'ome' suffix implies a holistic description, and resonates with terms such as genomics, transcriptomics and proteomics, although microbiome studies rarely characterize total microbial communities, usually being restricted to bacteria, omitting archaea and viruses, and even more rarely including microbial eukaryotes. Microbiome studies, therefore, assume implicitly that there is a relationship between the phenotypic characteristics of the community members and the characteristics of the environment in which they are found. For eukaryotes, sexual reproduction increases the strength of links between phylogeny and function and reasonably consistent (but by no means perfect) definitions of species are provided as units of diversity. For bacteria and archaea, species cannot be defined and links are much weaker and poorly understood. Despite these major limitations in applying niche theory to microbial communities, the concept provides the (usually unstated) basis for

correlation studies and for exploring links between communities and their environments.

(b) Application to correlation-based studies

To apply inference to best explanation to the above examples, of ammonia oxidizers and biogeographic patterns, the implicit assumption is that phylotype relative abundance will be related to soil characteristics. Correlations lead to inferences and different hypotheses can be explored that might explain these correlations, but these examples illustrate problems with this approach.

Firstly, and fundamentally, niche specialization suggests a mechanism or cause for differences in community composition. Soil characteristics will determine the abundance of ammonia oxidizers as a functional group and differences in relative abundance of different phylotypes. Crucially, however, correlation analyses do not distinguish cause and effect and we, therefore, cannot suggest that an environmental characteristic causes, explains or predicts, or is a driver of the presence or relative abundance of a phylotype. (This false interpretation of data is exacerbated by ambiguous terminology. To a statistician, pH may be described as a driver or predictor or explanatory factor of relative abundance, or as explaining relative abundance with which it is correlated. For an ecologist, this would wrongly imply a cause and effect relationship, rather than a mere statistical relationship.) It is not valid to consider that correlation demonstrates direct links between phenotypic and environmental characteristics. Internet searches of bizarre or spurious correlations provide many examples of correlations with no imaginable rationale. They also provide countless examples of football supporters, players and, even, managers following rituals on the basis of past correlations. The significance of these correlations usually dwarfs those that we can hope for from community ecology studies, but are based solely on superstition. Nevertheless, we routinely see examples of inference or 'prediction' of future events based on correlations between a few environmental characteristics and relative abundance of phylotypes with no evidence of causality. This can reflect a desperate, last resort, arising from ignorance of physiological characteristics, but usually it reflects a lack of desire to identify scientific questions and consider potential mechanisms prior to data collection.

Secondly, measured soil characteristics are usually chosen on the basis of custom (measuring what other people measure), habit (cf. Hume [10]), cheapness, availability of equipment and expertise, ease of use, fashion, etc. The choice is not based on *a priori* consideration of characteristics that might be expected to influence community composition, e.g. through knowledge of physiological differences between phylotypes. Ammonia oxidizers are autotrophs and use CO₂, making measurement of organic C irrelevant. It might be possible to think of a rationale for measuring C:N ratio (it may influence production of ammonia by mineralization), but these arguments are never made. Characteristics are often irrelevant in other ways. Soil moisture content, if determined by rainfall, will vary temporally at scales from minutes to months, and changes in community composition will only occur for those organisms that react at the same time scales. Moisture content, measured at the bulk scale, does not have a direct effect on microbes but has many indirect effects: decreased diffusion of oxygen; increased mobility of soluble nutrients and cells, including predators; changes in root growth; leaching of nutrients, etc. Some

of these factors will interact, e.g. release of a soluble nutrient will increase activity of aerobic microbes, which will decrease oxygen concentration. Increased mobility of predators will increase predation and nutrient turnover. The lack of relevance of measured characteristics is illustrated by the fact that the same, very limited number of soil characteristics are measured regardless of the organisms being studied, their environments or the scale of study. Niche specialization assumes a link between physiological and environmental characteristics, but the characteristics that are routinely measured do not relate the physiological characteristics that might be expected to lead to differences in community composition.

A third issue is that microbes themselves will influence many of the characteristics measured. If metabolic activity reduces pH, as occurs with ammonia oxidation, a negative correlation between relative abundance of a phylotype and pH may be due to its preference for low pH, or due to a reduction in pH resulting from its growth at a higher pH. Some characteristics, but surprisingly few, involve measurement of substrates, but does a positive correlation between abundance of a particular ammonia oxidizer phylotype and ammonia concentration indicate its tolerance of, and preference for high ammonia concentration? If so, why has it not already oxidized the ammonia, reducing ammonia concentration, leading to a negative correlation? Other soil characteristics change temporally and at different spatial scales and correlations are often due to two-way interactions, and not simple cause and effect. This applies also, of course, to microbial (and other) communities, which may also be evolving. We should also consider characteristics that, objectively, are important. Most studies consider community changes only in terms of microbial growth, while any growth must be balanced by death, unless total biomass changes significantly. Differences in survival and death rates within community members will, therefore, be equally important in determining community composition and diversity, but are rarely considered. Similarly, versatility, flexibility and speed of response to environmental change may be more important than growth rate or substrate affinity, but only these parameters are considered because they are more easily measured, even if irrelevant. Even if communities have been selected because their physiological characteristics are perfectly aligned with the environmental characteristics, both will be changing and organisms will presumably be continually evolving in response to the new conditions. This applies particularly to microbes, for whom ecological and evolutionary time scales can converge.

This does not mean that we should despair, but it does mean that we need to think carefully before we begin studies, rather than blindly measuring what others measure, define specific and better thought-out scientific questions and test hypotheses even more critically. Employing correlation-based studies to increase understanding is equivalent to thinking with a mental straightjacket in which any potential explanations or mechanisms are constrained by the organisms, genes, genomes or metagenomes and environmental characteristics that have been measured. The straps of the straightjackets are tightened if the environmental characteristics are chosen merely because they are 'those that everyone else measures', as this will constrain entire fields of study, and not just that of the individual researcher. Any explanation or hypothesis arising from the data, or from studies generating similar types of data, will be restricted by the characteristics measured. The above example would not detect any influence on ammonia

oxidizers through increased predation or the abundance of worms, because these organisms are not measured.

Correlation studies, and other look-see or 'effect of' studies, are sometimes the last resort. They may provide a starting point when investigating the function or ecology of organisms that have never been cultivated and about which nothing is known because closest relatives have not been characterized. This is effectively admitting defeat, in terms of intellectual effort and imagination, but a survey of effects of environmental characteristics on abundance of this organism might provide hints. If so, then the focus shifts from the organism to the environmental characteristics which, in the absence of prior knowledge, should be chosen randomly, and as many as possible should be measured. In these cases, rather than collecting yet more sequence data, resources should be expended on measuring a greater range of environmental characteristics.

5. Deduction and hypothesis testing

Induction approaches are based on data that, in some cases, are then used to assess which hypothesis provides the best explanation. The deductive approach, by contrast, is closest to the accepted view of scientific method. This begins with a scientific question or observation of a phenomenon that cannot be explained and proposals of a hypothesis, or hypotheses, based on assumptions regarding the cause or mechanism that can answer the question or explain the phenomenon. If these assumptions are true, then predictions of the hypothesis will also be true. Experiments are then designed to generate data that can be compared with predicted observations, to test the hypothesis and the assumptions on which it is based. This process is also termed hypothetical-deductivism.

(a) Hypothesis construction

The processes involved in hypothesis construction are difficult to characterize. They involve analysis, synthesis and integration of current knowledge, but also creativity, imagination and innovation. Crucially they require thought and intellectual effort and, even more crucially, they involve thinking before experimental work is even considered. These hypotheses are driven by attempts to explain phenomena, and are not derived after data have been collected. As a consequence, for example, in trying to explain why a particular phylotype is associated with a root, the researcher is not immediately focused on characteristics that are easy to measure (total soil C, pH, plant species, etc.) but maybe considers how conditions around a root might differ from the bulk soil, and from any other environment, which physiological characteristics might be important, whether oxygen will be limiting, whether predators might be more abundant, etc.

(b) Assumptions

Assumptions fall into two categories. The first are those associated with the particular mechanism being proposed. For example, a phylotype in the rhizosphere of one plant may increase in relative abundance because the root produces a substrate that is specific for this phylotype, or through resistance to an antibiotic to which others are sensitive. In the example above, an ammonia oxidizer phylotype may decrease in relative abundance if it is more sensitive to high ammonia

concentration or increased plant growth results in production of inhibitors.

The second category comprises a number of simplifying or qualifying assumptions, which are crucial for two main reasons. Firstly, they ensure that the hypothesis is well thought-through and is stated with clarity and precision. Secondly, they determine the experimental approach, techniques and design required to test predictions of the hypothesis. If the proposed mechanism is likely to be affected by different environmental factors, then this should be stated and should determine the experimental design. For example, the relative abundance of a rhizosphere phylotype may increase through provision of a specific nutrient by the plant but this effect will be difficult to test if temperature or oxygen concentration are varying significantly and influencing the phylotype for other reasons. The experimental system should, therefore, be designed to eliminate these additional, potentially complicating or confounding factors to enable focus on the specific mechanism being tested. If, however, the mechanism is hypothesized as the only influence on relative abundance (which is unlikely), then some of these simplifying assumptions may not be necessary and experimental design can be relaxed.

(c) 'Good' hypotheses

A 'good' hypothesis has a number of desirable properties. It should be bold, risky and meaningful, addressing an important issue and not stating the obvious. A good hypothesis should have explanatory power, unifying previously unrelated problems and observations, and great predictive power. It must be testable and should have generality, with relevance outside the system on which it is based; e.g. although derived from published data on one system or for one phylotype, it should be relevant to other systems or phylotypes.

Unfortunately, and frequently, the desire to give the impression of performing hypothesis-driven research leads to use of the term hypothesis to suggest something that is either obvious, untestable or meaningless. This is common for 'effect of' studies. For example, the hypothesis that ammonia fertilization will affect soil ammonia oxidizer communities is not meaningful and is not bold or risky, as we would be very surprised if addition of a substrate did not influence communities of organisms using that substrate. The hypothesis is also imprecise; will changes occur immediately, or only after a period of incubation; will ammonia influence plant growth and that of other microbes, leading to indirect effects on ammonia oxidizers; will ammonia effects themselves be influenced by other factors, e.g. pH, and will these factors be controlled? The hypothesis is also difficult to falsify. If the communities do not change, the researcher could claim that the hypothesis is correct but that deeper sequencing is required or a longer incubation period. There is also no information on assumptions on which the hypothesis is based, e.g. mechanisms by which ammonia concentration or supply might differentially affect ammonia oxidizer phylotypes through inhibition or other mechanisms. There is no consideration of mechanisms, there are no mechanistic assumptions and, consequently, there is nothing to suggest that the findings could be generalized or are specific to this soil and this community. In other words, the lack of mechanistic assumptions prevents qualitative and quantitative predictions, with no information on the magnitude or speed of change, preventing critical testing of the hypothesis. Such hypotheses are, therefore, not meaningful,

they are not scientific hypotheses and experimental testing will not provide any advance in understanding.

(d) Experimental testing

Experimental testing can take two forms. The first involves accumulation of supporting evidence to verify a hypothesis, but this suggests that hypotheses or theories can be proved if sufficient evidence can be collected. It is also easily subject to bias, as it is usually relatively easy to design experiments that will provide supportive evidence and to think of reasons why data do not fit predictions, as discussed in the previous section. This problem is greatest for hypotheses that are vague, poorly defined and non-quantitative.

Popper [5,6] argued against this approach and proposed that for science to be truly unbiased, objective and dispassionate, the researcher should design experiments to falsify or reject a hypothesis. In fact, he used falsifiability as a means of demarcating science from pseudoscience, i.e. for a hypothesis or statement to be considered scientific, it must be possible to think of observation or argument that would refute it.³ He argued, and it is generally accepted, that it is never possible to prove a scientific theory but it is possible to disprove a theory. Experiments should, therefore, be designed with the aim of falsifying a hypothesis and failure to falsify increases confidence in that hypothesis. The fundamental problem of this approach, when adopted strictly, is that it lacks an endpoint, as there could always be a further experiment that has not yet been considered that might falsify a hypothesis. Indeed, Popper suggested that you should have no more confidence in a hypothesis that you have failed to reject 100 times than one that you have rejected only once. A partial solution is to introduce the concept of corroboration, in which increasing failures to reject a hypothesis are taken as corroboration, similar to increased confidence. In addition, rejection does not necessarily mean that a hypothesis is worthless. It may be that parts are useful and others not and the data may highlight ways, or the researcher may consider ways in which the discrepancy between predictions and experimental data can be corrected by modification of assumptions. This modified hypothesis would then need to be tested with new experiments. Nevertheless, this approach can be seen as truly objective and dispassionate and it highlights the inability to prove a hypothesis or theory.

The key feature of the deduction approach is that it begins with a scientific question, i.e. with a phenomenon that cannot be explained by existing hypotheses and requires construction of a new hypothesis that is then tested experimentally. It, therefore, avoids a major criticism of most ecological studies, which is a lack of scientific aims or questions and reliance on data and techniques. The approach overtly aims at increasing understanding and, in defining a question, provides clear and assessable criteria for beginning a study and directs and structures experimental work. It determines which experiments and which techniques are required, rather than just choosing those that are available or fashionable, presenting clear criteria for judging the success of a study. Importantly, it determines which techniques are not required, avoids unnecessary wastage of resources, avoids attempts at the impossible, or discovery of impossibility when it is too late, and avoids the need to make stories out of data. This approach is also more intellectually challenging, which in itself should be attractive and interesting, and leads to explanations that are not based

solely on available data and available techniques; it removes the mental straightjacket and, indeed, requires that we think freely and broadly about how microbial communities interact with their environment.

6. Analysis

The above discussion considers four approaches adopted by microbial ecologists, but each contains a range of approaches and boundaries that are sometimes not clear. Nevertheless, this classification can be used to analyse published work to provide an indication of current practice. To achieve this, papers published in a single issue of five leading microbial ecology journals (*Applied and Environmental Microbiology*, *Environmental Microbiology*, *FEMS Microbiology Ecology*, *ISME Journal* and *Microbial Ecology*) were examined. The analysis was restricted to articles on microbial ecology, rather than biotechnology or other applied aspects. A total of 100 papers were analysed, with similar numbers from each journal, although I am not claiming that this can be considered a rigorous study.

Of these papers, 67 were descriptive, of which 36 were purely descriptive, with no indication that the study aimed to do more than observe and measure, usually community composition. The remaining 31 were effectively descriptions, in that they did not pose a scientific question and made little attempt to explain findings. Many of these were 'effect of' studies, exploring the effect of a particular factor on microbes. While some compared findings with those already published, they did not use these comparisons to seek explanations or mechanisms. The introductions to many of these papers highlighted the need to increase understanding, but none adopted an approach that could achieve this.

Of the remaining 33 papers, 23 could be classified as inference to the best explanation, 10 of which led to new hypotheses or variations of existing hypotheses to explain observations. Only 10 papers aimed to test a hypothesis. Across all papers, only 22 overtly based their study on a question, of which only nine could be considered significant scientific questions.

This analysis illustrates the degree to which the criticisms and limitations of non-scientific approaches, discussed above, are limiting scientific advances in microbial ecology. These journals are likely to attract the majority of high-quality microbial ecology articles and their scope requires papers that provide a scientific advance. It is likely that analysis of microbial ecology papers in all journals would demonstrate a much larger proportion of descriptive studies and even fewer based on hypotheses. It is, therefore, reasonable to ask why such a small minority of studies are driven by questions and hypotheses, when these are the basis of scientific method and are designed to increase understanding, while the majority of studies are descriptive and 'question-free'.

One explanation for the many molecular and omic surveys and descriptive studies of microbial communities is that the availability of a new technique often leads to descriptions, but that these are then followed by scientific studies, as the techniques are used to address questions and test theory. In most cases, however, theory already exists that allows the descriptive studies to be by-passed. More worryingly, new techniques are continually appearing, leading to the view that attempts to increase understanding should be delayed until the next new technique becomes available. This suicidal approach focuses solely on what can be measured and not on

what needs to be measured, when the value of a technique is determined solely by its ability to assist in testing hypotheses and answering scientific questions. It is also suggested that new techniques identify new phenomena and questions, but there is no shortage of questions to be answered; there is already much that we cannot explain. More importantly, questions and phenomena do not rely on descriptive studies. Hypothesis-driven studies using molecular techniques are just as likely to lead to new discoveries as random molecular or metagenomic surveys. The nature of hypothesis-driven studies is also more likely to identify truly interesting and unusual observations and to employ experimental design needed to assess their significance.

A further justification for induction, particularly of correlation-based, pattern-searching studies, is that it can generate hypotheses. This may happen, particularly when studies involve controlled manipulations or treatments, rather than for unstructured studies. Similarly, induction studies involving inference to best explanation can compare hypotheses. In all of these cases, however, hypotheses are considered after data have been obtained and are solely dependent on what has been measured. These data cannot, therefore, be used to test the hypotheses; this requires further experimental work. Many of these studies, however, could have been approached as hypothesis-driven studies with valid hypothesis testing, often with no additional experimental work and usually with less. All involve analysis of published work, certainly when discussing and trying to explain data, and often when providing background information in introductory sections. This analysis, if performed prior to experimental work, would have provided hypotheses that could then have been tested in rationally designed experiments to enable critical testing of predictions of the hypothesis. This would avoid collection of irrelevant data and would follow scientific method, with potentially critical testing of the same hypotheses, but with much fewer resources.

Unfortunately, the dominance, popularity and undemanding nature of descriptive studies can lead to alternative approaches, such as hypothesis testing, being seen as idealistic, particularly given the complexity and difficulties in studying microbial interactions with natural environments. However, this complexity in itself demands that a more scientific approach is adopted. The more difficult it is to explain a phenomenon, the greater the need to clearly define hypotheses and test them critically, rather than hoping that something interesting will arise from essentially random observations or, more worryingly, attempting to turn data into answers and then searching for relevant questions.

The approach adopted in scientific studies is, of course, influenced by many factors that are outside the scope of this article. My discussion of the philosophy of science has focused on those philosophers whose thinking aims to improve the scientific process and to assess the validity of different approaches and interpretation of experimental data. Others are more concerned with how scientific research proceeds, rather than how it maybe should proceed. In these respects, we can be influenced by many factors. It is often suggested that microbial ecology is driven by techniques. Certainly, microbial ecologists, rather than microbial ecology, are (like many other scientists) driven and seduced by new techniques. Techniques wrongly become the focus of studies and the availability of new techniques can change the direction of research, even when they have no scientific value. This is partly through

pressure, when applying for funding and publishing, for research to be seen as 'cutting-edge' and there are many examples of research proposals or papers being rejected through lack of use of modern techniques, despite established techniques being adequate.

It is also often suggested that microbial ecology is limited by techniques. Techniques are obviously important for making observations that lead to questions and identification of unexplained phenomena, and for testing theoretical predictions. We have a vast array of techniques. It is relatively easy to think of more than 100 techniques available for the analysis of microbial growth, activity and interactions in natural environments, even before the introduction of currently available molecular techniques. We, therefore, need very good justification for investment of valuable time and money in learning yet another new technique. That justification is rarely presented in terms of the ability of the techniques to test an ecological theory or increase understanding. We have generated a plethora of observations and phenomena but with a dearth of explanations. It could, therefore, be argued that techniques do, in fact, limit scientific progress in microbial ecology by diverting time and money to development of new techniques that would be better spent on generating and testing new ideas and theories.

The real limitation to our understanding of microbial ecology lies, not in a lack of techniques, but in a lack of motivation, enthusiasm, desire and courage to identify and ask significant scientific questions in advance of experimental work, and a lack of testable hypotheses and theory, i.e. lack of adoption of the basic scientific method. In this respect, it is worth considering, as a microbial ecologist, if you were to be given the answer to a single scientific question, or given a theory that explained a single phenomenon, what would be your question or phenomenon; in other words, what drives your science? This questioning is essential if ecological research is to go beyond mere descriptions and natural history. Identification of

important scientific questions provides criteria by which to assess potential scientific value, a framework for research, assessment of tractability and feasibility, identification of experimental systems and techniques required to test hypotheses and, ultimately, criteria for the assessment of success in advancing microbial ecology.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. This work was supported by the Natural Environment Research Council (grant no. NE/L006286/1).

Acknowledgements. I am indebted to Dr Cécile Gubry-Rangin and Professor Graeme Nicol for invaluable comments on the manuscript.

Endnotes

¹This process is analogous to the idea that a sufficiently large number of monkeys, typing randomly, will eventually produce the works of Shakespeare. The probability of this occurring is obviously vanishingly small, but of equal significance is the fact that a monkey would not realize when it had produced the works of Shakespeare and would continue typing randomly and aimlessly.

²The dangers of inductive reasoning are illustrated dramatically by the following quote from Captain Edward J. Smith 1907, 5 years prior to his captaining the Titanic on its final voyage: 'When anyone asks me how I can best describe my experiences of nearly 40 years at sea, I merely say uneventful. ... I have never been in an accident of any sort worth speaking about ... I never saw a wreck and have never been wrecked, nor was I ever in any predicament that threatened to end in disaster of any sort.' (See <https://www.encyclopedia-titanica.org/titanic-captain-smith-a-captains-career.html>.)

³It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations. Confirmations should count only if they are the result of risky predictions. A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice. Every genuine test of a theory is an attempt to falsify it, or refute it' [6, p. 36].

References

1. Curd M, Psillos S (eds). 2013 *The Routledge companion to philosophy of science*, 2nd edn. London, UK: Routledge.
2. Godfrey-Smith P. 2003 *Theory and reality: an introduction to the philosophy of science*. Chicago, IL: University of Chicago Press.
3. Fieser J, Dowden B (eds). *Internet Encyclopedia of Philosophy*. See <https://www.iep.utm.edu/> (accessed 12 February 2020).
4. Zalta EN (ed.). *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Stanford University. See <https://plato.stanford.edu/>. (accessed 12 February 2020)
5. Popper K. 1959 *The logic of scientific discovery*. New York, NY: Basic Books.
6. Popper K. 1963 *Conjectures and refutations*, 1st edn. London, UK: Routledge & Kegan Paul.
7. Morris A, Meyer K, Bohannan B. 2020 Linking microbial communities to ecosystem functions: what we can learn from genotype–phenotype mapping in organisms. *Phil. Trans. R. Soc. B* **375**, 20190244. (doi:10.1098/rstb.2019.0244)
8. Isobe K, Bouskill NJ, Brodie EL, Martiny JBH, Sudderth EA. 2020 Phylogenetic conservation of soil bacterial responses to simulated global changes. *Phil. Trans. R. Soc. B* **375**, 20190242. (doi:10.1098/rstb.2019.0242)
9. Prosser JI. 2015 Dispersing misconceptions and identifying opportunities for the use of 'omics' for soil microbial ecology. *Nat. Rev. Microbiol.* **13**, 439–446. (doi:10.1038/nrmicro3468)
10. Hume D. 1978 *A treatise of human nature* (eds LA Selby-Bigge, PH Nidditch), vol. 1. Oxford, UK: Oxford University Press.
11. Kashtan N *et al.* 2014 Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**, 416–420. (doi:10.1126/science.1248575)
12. Taleb NN. 1960 *The black swan: the impact of the highly improbable*. New York, NY: Random House.
13. Prosser JI. 2012 Ecosystem processes and interactions in a morass of diversity. *FEMS Microbiol. Ecol.* **81**, 507–519. (doi:10.1111/j.1574-6941.2012.01435.x)