

1
2
3
4
5
6
7
8
9
10
11
12
13

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, Scotland, AB24 3UU, UK

***Correspondence author:** James Prosser, School of Biological Sciences, University of
Aberdeen, Cruickshank Building, Aberdeen AB24 3UU, United Kingdom

Email: j.prosser@abdn.ac.uk; Telephone: +441224 273254;

Mob: +447802 959364; Fax: +441224 272703

14 Abstract

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

37 **Introduction**

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

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

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

72 **Description**

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

76 *Who is there?*

77 The most obvious examples of descriptive studies are surveys that catalogue the microbes
78 present in an environment. This approach is termed nature study, when performed by amateurs
79 with basic techniques, or natural history, if more advanced techniques and analysis are used by
80 professional scientists; it also applies to mining of large quantities of data. Until the 1990s, this
81 required laboratory cultivation, from an environmental sample, of as many microbes as feasible
82 and phenotypic classification of each (mainly through determination of numerous
83 physiological characteristics during laboratory growth). The current approach is high-
84 throughput sequencing of 16S rRNA genes, with identification and phylogenetic analysis
85 following comparison with database sequences. Sequence analysis of functional genes

86 provides similar description of organisms with the potential for a particular function, e.g.
87 nitrogen fixation, denitrification. These molecular techniques provide limited information on
88 phenotype but are relatively cheap and rapid; crucially, they do not involve laboratory
89 cultivation and may provide better phylogenetic information.

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

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

110 *What is meant by 'there'?*

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

119 *Who is doing what?*

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

132 Current attempts to describe microbial function involve metagenomics and other 'omics'
133 approaches or, for cultivated organisms, analysis of genomes. Omics approaches also have

134 major limitations [9] and, at best, provide information only on potential activity. Although
135 avoiding cultivation, descriptions of potential function again rely on the efficiency and
136 accuracy of techniques used. Many genes will be transcribed and translated only under specific
137 conditions, many will be in dormant or dying cells, the predicted function of many (often most)
138 will be unknown or inaccurate, quantitative functional information is lacking and many
139 important ecophysiological characteristics have no obvious genetic determinant.

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

146 *What is the effect of ...?*

147 Descriptive studies are, of course, not restricted to community surveys. For example, many
148 investigate the effect of a particular environmental factor on microbes or their activities.
149 Although these studies may have an objective and a question, e.g. ‘does temperature influence
150 soil microbial community composition or activity?’, observing and describing the effect does
151 not increase understanding or provide explanations that may be more broadly relevant. Again,
152 the objectives and questions are technical and not scientific. Unfortunately, these studies are
153 sometimes even presented as testing hypotheses (e.g. ‘we hypothesised that temperature affects
154 communities’), but these are not meaningful or scientific hypotheses and this represents misuse
155 of the term ‘hypothesis’ (see below), often in an attempt to make a study seem more ‘scientific’.
156 Descriptive studies therefore may address technical questions and challenges, but not scientific
157 challenges. The ease with which molecular, genomic or metagenomic can be obtained invites

158 its collection in the hope that something interesting may ‘fall out’ of the data¹. This may happen
159 but the probability of answering an important ecological question without first asking or having
160 a question is low. Similarly, the chances of finding an explanation for an environmental
161 phenomenon without first observing the phenomenon is low. As a consequence, these studies
162 lead to desperate attempts to find a question to which the data provide an answer, to provide
163 belated justification for the study and to ‘make a story out of the data’, with the unavoidable
164 bias and subjectivity that this entails.

165 *Are there purely descriptive studies?*

166 In practice, descriptive studies may not always be performed ‘in the dark’. Purely descriptive
167 studies are sometimes justified as providing baseline data, e.g. prior to monitoring following
168 environmental change, or when using new techniques or exploring previously unstudied
169 environments. However, it would be surprising, and worrying, if there was no reason for
170 expecting environmental change to influence a community or for having an interest in a new
171 site, and how it might differ from others. These reasons should provide the basis for scientific
172 questions or hypotheses. Similarly, every environmental characteristic cannot be measured and
173 every organism cannot be characterised. Decisions must therefore be made on which microbial
174 groups (bacteria, archaea, fungi, protozoa, functional groups) and characteristics are measured.
175 Again, it would be surprising and worrying if these decisions were not made on the basis of
176 underlying views of what is important and interesting. There is, therefore, an implicit, if not
177 stated reason for expecting something different or unusual that provides the basis for a

¹ This 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 realise when it had produced the works of Shakespeare and would continue typing randomly and aimlessly.

178 meaningful study that, if defined and discussed prior to measurements, would provide the basis
179 for a scientific question and rational experimental design. Indeed, results arising from
180 descriptions are often described as surprising or unexpected and ‘effect of’ studies are often
181 described as interesting or even ‘successful’. All of these terms indicate that there was prior
182 information that could and should have been formulated as a hypothesis. Another perceived
183 benefit of descriptive studies is in extending databases that others might find useful in
184 answering ecological questions. However, critical testing of ecological questions generally
185 requires well considered experimental design, rather than analysis of data that has been
186 collected randomly.

187 **Induction**

188 In its simplest form (enumerative) inductive reasoning involves creation of a general rule from
189 a number of observations for a particular class and inference that all members of that class will
190 follow the same rule. To employ a common example of induction, we examine 50 swans,
191 observe that all are white and then infer that all swans are white. This approach was criticised
192 by the 18th century philosopher David Hume, who questioned whether “instances of which we
193 have had no experience resemble those of which we have had experience” [10] i.e. he
194 questioned whether the past could tell us about the future. He concluded that this process was
195 not based on reason and was interested in why it was adopted (the problem of induction),
196 suggesting that it was based on imagination, custom and habit. It is not difficult to find support
197 for this view. There are many cases in which the past has not predicted the future and previous
198 success in predicting the future does not logically mean that future predictions will be

199 successful². Induction can be defined more broadly, e.g. relaxing the need for generalisations
200 and considering the number of ‘supportive’ observations and the lack of contradictory
201 observations (naïve induction). However, the inductive approach is based solely on
202 observations and does not provide information on causes or mechanisms. We might be correct
203 in inferring that all swans are white (although black swans do exist), but we have no
204 information on why this might be the case, i.e. we may have knowledge but we have no
205 understanding and no way of confirming our knowledge.

206 There are many examples of induction in microbial ecology, e.g. inference that all ammonia
207 oxidisers are bacteria, which was believed until 2005 because all cultivated ammonia oxidisers
208 were bacteria; or that all plant rhizosphere microbiomes contain a particular phylotype because
209 it has been observed in 50 plants; or that a physiological characteristic of a single isolate or a
210 gene in a single genome or reconstructed metagenome will exist in all closely related
211 phlotypes. Similar extrapolation from properties of isolates to those of microbial relatives in
212 natural environments has always been a concern of cultivation-based approaches. Model
213 organisms are important, e.g. for in-depth physiological studies, but they will differ from the
214 organism that was originally obtained from the environment, because of rapid physiological
215 and genetic changes occurring during isolation, requiring care in predicting its ecology. (We
216 would be very wary of predicting the ecology of all cats on the basis of the characteristics of a
217 wild cat that had been domesticated for 100 generations. Nevertheless, we frequently see

² The 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 forty 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.’

218 unqualified assumptions about the ecology of relatives of a microbial culture, after a similar
219 number of generations in laboratory culture.) Single-cell genomics has also demonstrated
220 considerable genomic diversity within individual rRNA-defined phylotypes (e.g. [11]),
221 preventing prediction of all of the genetic characteristics of other phylotype members.
222 Nevertheless, induction is used to infer, often with no qualification or reservation, the functions
223 and metabolic pathways of relatives in the environment on the basis of a single gene, genome
224 or chimeric metagenome-assembled genome (even when >50% of gene functions are
225 unknown).

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

231 *Induction and correlation*

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

237 I will illustrate this approach with an example from my own research area, where correlations
238 are determined between soil ammonia oxidiser communities and soil characteristics.
239 Communities are characterised by analysis of 16S rRNA or a functional gene (*amoA* genes for
240 ammonia oxidisers) sequences and 5 – 10 soil characteristics, typically pH, moisture content,
241 total C, total N, C:N ratio and sometimes soil P, ammonium and nitrate. After grouping

242 sequences into arbitrary OTUs, statistical methods are used to quantify correlations between
243 phylotype relative abundance and soil characteristics, with attention focused on those
244 associations considered to be statistically significant. Purely descriptive studies stop at this
245 point and merely report the associations with no interpretation. Others invoke induction and
246 infer that the associations observed apply to all soils and all ammonia oxidiser communities.
247 Note that these inferences arise solely from the experimental data after it has been collected
248 and analysed. Note, also, that these studies make no *a priori* predictions, e.g. there is no
249 consideration of which ammonia oxidiser phylotypes might increase or decrease, or why. These
250 studies only 'explore'. In providing only descriptions, they suffer the limitations described
251 above: lack of scientific aims and the absence of criteria for justification, experimental design,
252 end-points, etc.

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

263 More commonly, results will be compared to those from other studies, to gain support for the
264 inference that the associations and patterns observed apply generally. This requires objectivity
265 but, unfortunately, it is easier to search for data that support an inference than data that do not.
266 It is also easy to discount lack of support through, e.g. different soils, methodologies,

267 experimental design, etc., while similarly criticisms are not made of supportive data. In
268 addition, while the strength of associations is quantified, comparisons with previous data are
269 usually made on the basis of qualitative or at best semi-quantitative comparisons, such that
270 interpretations become matters of degree and opinion and lack precision.

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

281 Inductive studies therefore range from naïve induction, which is no more advanced than
282 superstition or 'belief', through attempts to correlate communities with environmental factors
283 to large scale correlation-based analyses of biogeographical patterns. Although many induction
284 studies highlight the need to increase understanding, they are not designed or able to do so.
285 They provide knowledge but not understanding. We might 'know', and we might even be
286 correct in 'knowing' that the relative abundance of a particular ammonia oxidiser is always
287 favoured by a particular combination of the 5 – 10 soil characteristics measured, but this
288 approach gives no information on why, i.e. there is no mechanistic information. A dramatic
289 illustration of this distinction is the example of a turkey being reared for Thanksgiving Day
290 [12]. Turkey contentment increases during rearing, as the turkey is well-fed, warm, dry,
291 disease-free and much more content than wild turkeys. There will be a strong correlation

292 between turkey contentment and contact with the farmer, who provides food and shelter. This
293 high correlation, however, does not allow the turkey to infer or predict the future correctly, and
294 contentment falls abruptly when it is slaughtered by the farmer. The turkey had only
295 knowledge, while the farmer had both knowledge and understanding.

296 **Inference to best explanation**

297 Inference to best explanation describes interpretation of inferences, arising from data, in the
298 context of existing or new hypotheses or mechanisms in attempts to find the hypothesis that
299 best explains the data, concluding that this hypothesis is true. For example, selection of a
300 particular phylotype under certain environmental conditions may be explained through existing
301 knowledge of physiological or genetic characteristics of its relatives. In fact, many community
302 studies are implicitly, if unconsciously, testing the concept of niche specialisation and
303 differentiation. (See [13] for steps involved and some consideration of its application to
304 microbes.) Briefly, this posits that environmental characteristics will lead to evolution and
305 selection of strains whose physiological characteristics are best adapted to those environmental
306 characteristics. The relative abundance of these strains will increase and they will be dispersed,
307 colonising other similar environments, again through selection based on their physiology. Both
308 16S rRNA- and functional gene-analyses enable tracking of phylogenetic groups and
309 correlation is predicted between phylotypes and ecological conditions. The validity of this
310 concept depends on the validity of the assumptions on which the concept is based. In particular,
311 it assumes strong links between phylogeny and function, which is considerably less in
312 prokaryotes than in animals and plants for which the concept was developed (see [13]).

313 *Niche specialisation and microbiomes*

314 In fact, this concept is incorporated, though not intentionally, in the term ‘microbiome’.
315 ‘Biome’ has been used for many years to describe plant or animal communities that have

316 common characteristics for the particular environment in which they are found, e.g. temperate
317 forests. This implies links between physiological and environmental characteristics, for which
318 there is evidence in plant and animal communities. The ‘micro’ prefix represents microbial
319 community and, for example, the term soil microbiome refers not just to soil microbial
320 community composition but also suggests that the community is special in some way, with
321 physiological characteristics selected by the physical, chemical and biological characteristics
322 of soil. In addition, the ‘ome’ suffix implies a holistic description, and resonates with terms
323 such as genomics, transcriptomics and proteomics, although microbiome studies rarely
324 characterise total microbial communities, usually being restricted to bacteria, omitting archaea
325 and viruses, and even more rarely including microbial eukaryotes. Microbiome studies
326 therefore assume implicitly that there is a relationship between the phenotypic characteristics
327 of the community members and the characteristics of the environment in which they are found.
328 For eukaryotes, sexual reproduction increases the strength of links between phylogeny and
329 function and reasonably consistent (but by no means perfect) definitions of species are
330 provided as units of diversity. For bacteria and archaea, species cannot be defined and links are
331 much weaker and poorly understood. Despite these major limitations in applying niche theory
332 to microbial communities, the concept provides the (usually unstated) basis for correlation
333 studies and for exploring links between communities and their environments.

334 *Application to correlation-based studies*

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

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

361 Secondly, measured soil characteristics are usually chosen on the basis of custom (‘measuring
362 what other people measure’), habit (cf. Hume), cheapness, availability of equipment and
363 expertise, ease of use, fashion, etc. Choice is not based on *a priori* consideration of
364 characteristics that might be expected to influence community composition, e.g. through

365 knowledge of physiological differences between phylotypes. Ammonia oxidisers are
366 autotrophs and use CO₂, making measurement of organic C irrelevant. It might be possible to
367 think of a rationale for measuring C:N ratio (it may influence production of ammonia by
368 mineralisation), but these arguments are never made. Characteristics are often irrelevant in
369 other ways. Soil moisture content, if determined by rainfall, will vary temporally at scales from
370 minutes to months and changes in community composition will only occur for those organisms
371 that react at the same time scales. Moisture content, measured at the bulk scale, does not have
372 a direct effect on microbes but has many indirect effects: decreased diffusion of oxygen;
373 increased mobility of soluble nutrients and cells, including predators; changes in root growth;
374 leaching of nutrients, etc. Some of these factors will interact, e.g. release of soluble nutrient
375 will increase activity of aerobic microbes, which will decrease oxygen concentration. Increased
376 mobility of predators will increase predation and nutrient turnover. The lack of relevance of
377 measured characteristics is illustrated by the fact that the same, very limited number of soil
378 characteristics are measured regardless of the organisms being studied, their environments or
379 the scale of study. Niche specialisation assumes a link between physiological and
380 environmental characteristics, but the characteristics that are routinely measured do not relate
381 the physiological characteristics that might be expected to lead to differences in
382 community composition.

383 A third issue is that microbes themselves will influence many of the characteristics measured.
384 If metabolic activity reduces pH, as occurs with ammonia oxidation, a negative correlation
385 between relative abundance of a phylotype and pH may be due to its preference for low pH, or
386 due to a reduction in pH resulting from growth at a higher pH. Some characteristics, but
387 surprisingly few, involve measurement of substrates, but does a positive correlation between
388 abundance of a particular ammonia oxidiser phylotype and ammonia concentration indicate its
389 tolerance of, and preference for high ammonia concentration? If so, why has it not already

390 oxidised the ammonia, reducing ammonia concentration, leading to a negative correlation?
391 Other soil characteristics change temporally and at different spatial scales and correlations are
392 often due to two-way interactions, and not simple cause and effect. This applies also, of course,
393 to microbial (and other communities), which may also be evolving. We should also consider
394 characteristics that, objectively, are important. Most studies consider community changes only
395 in terms of microbial growth, while any growth must be balanced by death, unless total biomass
396 changes significantly. Differences in survival and death rates within community members will
397 therefore be equally important in determining community composition and diversity, but are
398 rarely considered. Similarly, versatility, flexibility and speed of response to environmental
399 change may be more important than growth rate or substrate affinity, but only these parameters
400 are considered because they are more easily measured, even if irrelevant. Even if communities
401 have been selected because their physiological characteristics are perfectly aligned with the
402 environmental characteristics, both will be changing and organisms will presumably be
403 continually evolving in response to the new conditions. This applies particularly to microbes,
404 for whom ecological and evolutionary time scales can converge.

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

415 restricted by the characteristics measured. The above example would not detect any influence
416 on ammonia oxidisers through increased predation or the abundance of worms, because these
417 organisms are not measured.

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

428 **Deduction and hypothesis testing**

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

438 *Hypothesis construction*

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

449 *Assumptions*

450 Assumptions fall into two categories. The first are those associated with the particular
451 mechanism being proposed. For example, a phylotype in the rhizosphere of one plant may
452 increase in relative abundance because the root produces a substrate that is specific for this
453 phylotype, or through resistance to an antibiotic to which others are sensitive. In the example
454 above, an ammonia oxidiser phylotype may decrease in relative abundance if it is more
455 sensitive to high ammonia concentration or increased plant growth results in production of
456 inhibitors.

457 The second category comprises a number of simplifying or qualifying assumptions, which are
458 crucial for two main reasons. Firstly, they ensure that the hypothesis is well thought-through
459 and is stated with clarity and precision. Secondly, they determine the experimental approach,
460 techniques and design required to test predictions of the hypothesis. If the proposed
461 mechanism is likely to be affected by different environmental factors, then this should be stated

462 and should determine experimental design. For example, relative abundance of a rhizosphere
463 phylotype may increase through provision of a specific nutrient by the plant but this effect will
464 be difficult to test if temperature or oxygen concentration were varying significantly and
465 influencing the phylotype for other reasons. The experimental system should therefore be
466 designed to eliminate these additional, potentially complicating or confounding factors to
467 enable focus on the specific mechanism being tested. If, however, the mechanism is
468 hypothesised as the only influence on relative abundance (which is unlikely), then some of
469 these simplifying assumptions may not be necessary and experimental design can be relaxed.

470 *'Good' hypotheses*

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

477 Unfortunately, and frequently, the desire to give the impression of performing hypothesis-
478 driven research leads to use of the term hypothesis to suggest something that is either obvious,
479 untestable or meaningless. This is common for 'effect of' studies. For example, the hypothesis
480 that ammonia fertilisation will affect soil ammonia oxidiser communities is not meaningful and
481 is not bold or risky, as we would be very surprised if addition of a substrate did not influence
482 communities of organisms utilising that substrate. The hypothesis is also imprecise; will
483 changes occur immediately, or only after a period of incubation; will ammonia influence plant
484 growth and that of other microbes, leading to indirect effects on ammonia oxidisers; will
485 ammonia effects themselves be influenced by other factors, e.g. pH, and will these factors be
486 controlled? The hypothesis is also difficult to falsify. If the communities do not change, the

487 researcher could claim that the hypothesis is correct but that deeper sequencing is required or
488 a longer incubation period. There is also no information on assumptions on which the
489 hypothesis is based, e.g. mechanisms by which ammonia concentration or supply might
490 differentially affect ammonia oxidiser phylotypes through inhibition or other mechanisms.
491 There is no consideration of mechanisms, no mechanistic assumptions and, consequently,
492 nothing to suggest that the findings could be generalised or are specific to this soil and this
493 community. In other words, the lack of mechanistic assumptions prevents qualitative and
494 quantitative predictions, with no information on the magnitude or speed of change, preventing
495 critically testing of the hypothesis. Such hypotheses are therefore not meaningful, are not
496 scientific hypotheses and experimental testing will not provide any advance in understanding.

497 *Experimental testing*

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

504 Popper [5, 6] argued against this approach and proposed that for science to be truly unbiased,
505 objective and dispassionate, the researcher should design experiments to falsify or reject a
506 hypothesis. In fact, he used falsifiability as a means of demarcating science from
507 pseudoscience, i.e. for a hypothesis or statement to be considered scientific, it must be possible

508 to think of an observation or argument that would refute it³. He argued, and it is generally
509 accepted, that it is never possible to prove a scientific theory but it is possible to disprove a
510 theory. Experiments should therefore be designed with the aim of falsifying a hypothesis and
511 failure to falsify increases confidence in that hypothesis. The fundamental problem of this
512 approach, when adopted strictly, is that it lacks an end-point, as there could always be a further
513 experiment that has not yet been considered that might falsify a hypothesis. Indeed, Popper
514 suggested that you should have no more confidence in a hypothesis that you have failed to
515 reject 100 times than one that you have rejected only once. A partial solution is to introduce
516 the concept of corroboration, in which increasing failures to reject a hypothesis are taken as
517 corroboration, similar to increased confidence. In addition, rejection does not necessarily mean
518 that a hypothesis is worthless. It may be that parts are useful and others not and the data may
519 highlight ways, or the researcher may consider ways in which the discrepancy between
520 predictions and experimental data can be corrected by modification of assumptions. This
521 modified hypothesis would then need to be tested with new experiments. Nevertheless, this
522 approach can be seen as truly objective and dispassionate and it highlights the inability to prove
523 a hypothesis or theory.

524 The key feature of the deduction approach is that it begins with a scientific question, i.e. with
525 a phenomenon that cannot be explained by existing hypotheses and requires construction of a
526 new hypothesis that is then tested experimentally. It therefore avoids a major criticism of most
527 ecological studies, which is a lack of scientific aims or questions and reliance on data and

³ 'It 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]

528 techniques. The approach overtly aims at increasing understanding and, in defining a question,
529 provides clear and assessable criteria for beginning a study and directs and structures
530 experimental work. It determines which experiments and which techniques are required, rather
531 than just choosing those that are available or fashionable, presenting clear criteria for judging
532 the success of a study. Importantly, it determines which techniques are not required, avoids
533 unnecessary wastage of resources, avoids attempts at the impossible, or discovery of
534 impossibility when it is too late, and avoids the need to make stories out of data. This approach
535 is also more intellectually challenging, which in itself should be attractive and interesting, and
536 leads to explanations that are not based solely on available data and available techniques; it
537 removes the mental straight-jacket and, indeed, requires that we think freely and broadly about
538 how microbial communities interact with their environment.

539 **Analysis**

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

549 Of these papers, 67 were descriptive, of which 36 were purely descriptive, with no indication
550 that the study aimed to do more than observe and measure, usually community composition.
551 The remaining 31 were effectively descriptions, in that they did not pose a scientific question
552 and made little attempt to explain findings. Many of these were 'effect of' studies, exploring

553 the effect of a particular factor on microbes. While some compared findings with those already
554 published, they did not use these comparisons to seek explanations or mechanisms. The
555 introductions to many of these papers highlighted the need to increase understanding, but none
556 adopted an approach that could achieve this.

557 Of the remaining 33 papers, 23 could be classified as inference to the best explanation, 10 of
558 which led to new hypotheses or variations of existing hypotheses to explain observations. Only
559 10 papers aimed to test a hypothesis. Across all papers, only 22 overtly based their study on a
560 question, of which only 9 could be considered significant scientific questions.

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

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

578 assist in testing hypotheses and answering scientific questions. It is also suggested that new
579 techniques identify new phenomena and questions, but there is no shortage of questions to be
580 answered; there is already much that we cannot explain. More importantly, questions and
581 phenomena do not rely on descriptive studies. Hypothesis-driven studies using molecular
582 techniques are just as likely to lead to new discoveries as random molecular or metagenomic
583 surveys. The nature of hypothesis-driven studies is also more likely to identify truly interesting
584 and unusual observations and to employ experimental design needed to assess their
585 significance.

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

601 Unfortunately, the dominance, popularity and undemanding nature of descriptive studies can
602 lead to alternative approaches, such as hypothesis-testing, being seen as idealistic, particularly

603 given the complexity and difficulties in studying microbial interactions with natural
604 environments. However, this complexity in itself demands that a more scientific approach is
605 adopted. The more difficult it is to explain a phenomenon, the greater the need to clearly define
606 hypotheses and test them critically, rather than hoping that something interesting will arise
607 from essentially random observations or, more worryingly, attempting to turn data into answers
608 and then searching for relevant questions.

609 The approach adopted in scientific studies is, of course, influenced by many factors that are
610 outside the scope of this article. My discussion of the philosophy of science has focused on
611 those philosophers whose thinking aims to improve the scientific process and to assess the
612 validity of different approaches and interpretation of experimental data. Others are more
613 concerned with how scientific research proceeds, rather than how it maybe should proceed. In
614 these respects, we can be influenced by many factors. It is often suggested that microbial
615 ecology is driven by techniques. Certainly, microbial ecologists, rather than microbial ecology,
616 are (like many other scientists) driven and seduced by new techniques. Techniques wrongly
617 become the focus of studies and the availability of new techniques can change the direction of
618 research, even when they have no scientific value. This is partly through pressure, when
619 applying for funding and publishing, for research to be seen as ‘cutting-edge’ and there are
620 many examples of research proposals or papers being rejected through lack of use of modern
621 techniques, despite established techniques being adequate.

622 It is also often suggested that microbial ecology is limited by techniques. Techniques are
623 obviously important for making observations that lead to questions and identification of
624 unexplained phenomena, and for testing theoretical predictions. We have a vast array of
625 techniques. It is relatively easy to think of >100 techniques available for the analysis of
626 microbial growth, activity and interactions in natural environments, even before the
627 introduction of currently available molecular techniques. We therefore need very good

628 justification for investment of valuable time and money in learning yet another new technique.
629 That justification is rarely presented in terms of the ability of the techniques to test ecological
630 theory or increase understanding. We have generated a plethora of observations and
631 phenomena but with a dearth of explanations. It could therefore be argued that techniques do,
632 in fact, limit scientific progress in microbial ecology by diverting time and money to
633 development of new techniques that would be better spent on generating and testing new ideas
634 and theories.

635 The real limitation to our understanding of microbial ecology lies, not in a lack of techniques,
636 but in a lack of motivation, enthusiasm, desire and courage to identify and ask significant
637 scientific questions in advance of experimental work and a lack of testable hypotheses and
638 theory, i.e. lack of adoption of basic scientific method. In this respect, it is worth considering,
639 as a microbial ecologist, if you were to be given the answer to a single scientific question, or
640 given a theory that explained a single phenomenon, what would be your question or
641 phenomenon; in other words, what drives your science? This questioning is essential if
642 ecological research is to go beyond mere descriptions and natural history. Identification of
643 important scientific questions provides criteria by which to assess potential scientific value, a
644 framework for research, assessment of tractability and feasibility, identification of
645 experimental systems and techniques required to test hypotheses and, ultimately, criteria for
646 assessment of success in advancing microbial ecology.

647 **Acknowledgements**

648 I am indebted to Dr Cécile Gubry-Rangin and Professor Graeme Nicol for invaluable
649 comments on the manuscript and support is acknowledged from Natural Environmental
650 Research Council grant NE/L006286/1.

651 **References**

- 652 1. Curd M, Psillos S. (Eds.) 2013 The Routledge Companion to Philosophy of Science
653 (Second Edition), London: Routledge. (doi: 10.4324/9780203744857).
- 654 2. Godfrey-Smith P. 2003 Theory and Reality: An Introduction to the Philosophy of Science.
655 Chicago: University of Chicago Press.
- 656 3. Internet Encyclopedia of Philosophy. <https://www.iep.utm.edu/>
- 657 4. Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/>
- 658 5. Popper K. 1959 The logic of scientific discovery. New York: Basic Books.
- 659 6. Popper K. 1963 Conjectures and refutations. 1st ed. London: Routledge & Kegan Paul.
- 660 7. Morris A, Meyer K, Bohannan B. 2020 Linking microbial communities to ecosystem
661 functions: what we can learn from genotype-phenotype mapping in organisms. *Phil. Trans.*
662 *R. Soc. B* (in review).
- 663 8. Kazuo I, Bouskill NJ, Brodie EL, Martiny JBH (2020) Phylogenetic conservation of soil
664 bacterial responses to simulated global changes. *Phil. Trans. R. Soc. B* (in review).
- 665 9. Prosser JI. 2015 Dispersing misconceptions and identifying opportunities for the use of
666 ‘omics for soil microbial ecology. *Nature Rev. Microbiol.* **13**, 439-446.
667 (doi:10.1038/nrmicro3468)
- 668 10. Hume D. (1739–40) 1975 A treatise of human nature (eds LA Selby-Bigge, PH Nidditch),
669 Oxford: Clarendon Press.
- 670 11. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. 2014.
671 Single-cell genomics reveals hundreds of coexisting subpopulations in wild
672 *Prochlorococcus*. *Science* **344**, 416-20. (doi: 10.1126/science.1248575)
- 673 12. Taleb NN. 1960 The black swan: the impact of the highly improbable. New York: Random
674 House.

- 675 13. Prosser JI. 2012 Ecosystem processes and interactions in a morass of diversity. *FEMS*
676 *Microbiol Ecol* **81**, 507-519. doi: 10.1111/j.1574-6941.2012.01435.x)
- 677