Title: **Macroecology to unite all life**

Authors

Ashley Shade[1], Robert R. Dunn[2,3,4], Shane A. Blowes[4], Petr Keil[4], Brendan J.M. Bohannan[5], Martina Herrmann[4,6], Kirsten Küsel[4,6], Jay T. Lennon[8], Nathan J. Sanders[3,9], David Storch[10,11], Jonathan Chase[4,12]

Affiliations

[1]Departments of Microbiology and Molecular Genetics and Plant, Soil and Microbial Sciences, Program in Ecology, Evolutionary Biology and Behavior, and The Plant Resilience Institute, East Lansing MI USA 48824

[2]Department of Applied Ecology, North Carolina State University

[3] Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen, Denmark.

[4] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany.

[5] Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA 97403.

[6] Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University, 07743 Jena, Germany

[8] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[9]Environmental Program, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington VT 05405

[10]Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic (storch@cts.cuni.cz)

[11]Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic

[12]Department of Computer Science, Martin Luther University, Halle-Wittenberg

Keywords

Microbial ecology, species-area relationship, rarefaction, diversity gradient, species-abundance distribution,distance-decay, metabolic theory of ecology, rarity, abundance-occupancy, biodiversity, metagenomics, singletons, belly buttons, microbiome

Abstract

Macroecology is the study of the mechanisms underlying general patterns of ecology across scales. Research in microbial ecology and macroecology have long been detached. Here, we argue that it is time to bridge the gap, as they share a common currency of species and individuals, and a common goal of understanding the causes and consequences of changes in biodiversity. Microbial ecology and macroecology will mutually benefit from a unified research agenda and shared datasets that span the entirety of life's biodiversity and the geographic expanse of Earth.

Length 3,642 als 16:00 27 Jul

Highlights

- Macroecology is the study of the mechanisms underlying general patterns of ecology across scales. A major focus of research within macroecology is understanding ***biodiversity*** patterns and their underlying processes. The field of macroecology has been biased towards charismatic "macroorganisms (a.k.a. *macrobes*)", and has largely ignored insights and breadth that can be gained by considering microorganisms.
- We argue that microbial ecology and macroecology are united by common currencies (individuals and species), as well as by comparable challenges of documenting their distributions and abundances.
- Future directions that would lead to a unified macroecology include: expansion of spatial and temporal scales to encompass the diversity of microbes; synthesis-driven, systematic comparisons of "macrobial" and microbial macroecological patterns and processes; and support of interdisciplinary approaches in training, publishing, and funding to equitably value macrobial and microbial insights into understanding life's rules and exceptions.

It's time to unite

Every individual, be it a mammoth, mule, marmot, or microbe, occupies a particular space and exists at a particular time. The number of marmots varies from place to place, as does the number of any particular microbial taxon. Identifying and counting individuals,

regardless of where they reside in the tree of life, is at the crux of understanding biodiversity and the natural world [1]. Decades of research have revealed that variation in the number of individuals of different species in space and time can give rise to a number of patterns, such as species abundance distributions and species-area relationships. These variables form the foundations of research in ***macroecology*****,** biogeography and community ecology. From the biodiversity patterns that emerge from counting individuals and species, many of ecology and evolution's most general rules emerge [2–4].

Until recently, the field of macroecology almost exclusively involved the study of large, multicellular organisms (a.k.a. macroorganisms or "macrobes"), especially plants, vertebrates and a few charismatic invertebrate groups like butterflies. However, in the early 2000's, the advent of new (and increasingly less expensive) molecular tools inspired some ecologists to ask the simple question: do microscopic forms of life play by the same rules as plants and animals? Initially, discussion centered around whether microbes exhibited macroecological patterns that were common in macrobes [5]. For example: Do microbes exhibit distance-decay relationships [6,7]? Are there elevational gradients in microbial diversity [8,9]? Do places with high macrobial diversity also have high microbial diversity [10,11]? An especially robust debated commenced around the ideas of dispersal limitation and whether microbial taxa where found "everywhere"[12] and then selected by the environment, which initiated new research on microbial biogeography (e.g., [13–15]). Despite these initial lines of inquiry, microbial ecology has evolved largely independently from macroecology and the two fields are not yet well integrated. Their continued separation seems to arise for historical and cultural reasons rather than inherent differences.

It is time to move on from asking whether microbes are different. Instead, there is a need to unify microbes and macrobes to ask overarching questions and to test general theories about the rules and mechanisms underpinning patterns in ecology across scales. The inclusion of microbial species into macroecological theory will extend and enrich our understanding of ecological patterns, not only to include a far greater range of spatial and temporal scales, evolutionary divergence, and organismal sizes, but also to provide insights into the fundamental processes that govern patterns of diversity and abundance across all types of organisms.

Microbes include the most phylogenetically and functionally diverse and abundant taxa on Earth [16–18]. Large advances in understanding microbial diversity have historically coincided with large advances in the technology used to observe microorganisms, from the invention of the microscope to the development of high-throughput DNA sequencing. At the beginning of the high-throughput sequencing revolution, about a decade ago, the technology was relatively expensive. Thus, large datasets to examine microbial diversity in space and time were not common. Calls for the study of "microbial biogeography" [14,19] would have to wait until there were more empirical data against which to test (and develop) theory. Although many microbial ecologists were using and applying concepts and methods from macroecology [13], there were few calls for microbial macroecology [20]. Meanwhile, macroecology has developed over recent decades with little reference to microbes, though, as discussed above, there are several key references that compare some patterns directly.

The rich data necessary to unify microbes into macroecology are now here. Microbial datasets that consider tens of thousands of microbial taxa observed over hundreds, thousands, or even tens of thousands of samples have become common, and these datasets are often

open access. Importantly, high-throughput, deeply sequenced datasets have made it possible to observe the important contribution of rare taxa to microbial community structure and diversity, leading to more precise analysis of biodiversity patterns. Furthermore, ecologists have begun to consider these microbial data in light of macroecological theory [15,21–24], or in direct comparisons to data on macrobes (e.g., [25–28]). As an exemplar case, the metabolic theory of ecology has especially benefited from the inclusion of microbial taxa to generally predict scaling of metabolic rates with of body size (**Box 1**). It is time for macroecology to forge ahead with unified currencies to count the number of individuals of the same or different species, distributed in space and time, for all of life's diversity. This accounting applies to moths, mammoths, and microbes - the bacteria, archaea, fungi, protists, and viruses that are all around us.

Unified currency: individuals and species

Considering all of life at once, be it macrobial or microbial, expands the breadth and reach of macroecology, if for no other reason than the reality that most individual organisms and species are microbes. The number of individuals of a single bacterial phylum Firmicutes in the guts of a single human, for instance, exceeds the total number of trees on Earth (3×10^{12}, [29]). There are close to 10^{29} or 10^{30} individual prokaryotic organisms (bacterial and archaea) on the globe [30–32]. These microorganisms derive from an astonishing diversity of taxa. Using scaling laws based on these abundances, Earth could be home to ~10^{12} microbial taxa, which far exceeds estimates of plant and animal diversity (~8×10^{6}, [33]). This suggests that we have only

131 inventoried one one-thousandth of one percent of all species on the planet [26], and that the
132 majority of these species have yet to contribute to our understanding of macroecology.
133 The idea that there are common macroecological currencies, individuals and species,
134 that apply to both macrobes and these numerous and diverse microbes has been controversial
135 for several reasons. Here, we argue against each of four challenges cited in support of
136 segregating microbes and macrobes in ecology: defining individuals, identifying individuals,
137 delimiting "species", and comparing methods.
138 Defining an individual. It is often assumed to be fairly straightforward to identify and
139 enumerate macrobial individuals, but, in practice, this is rarely the case (**Box 2**). As with some
140 macrobes, some microbes are modular (e.g., filamentous), which make identifying an individual
141 challenging. However, it is no harder to define the individual boundary of an ant supercolony,
142 for instance, than of a clonal or modular bacterium.
143 Identifying individuals. For a tiny fraction of microbial biodiversity, there is phenotypic
144 and genomic information that allows for robust identification of the species to which
145 individuals belong. Thus, ***genetic barcoding*** of ***marker genes*** [37] can be used to assign names
146 to microbial individuals that can be isolated through culture, or more recently through dilution
147 or physical capture. However, for the vast majority of yet-uncultivated microbial biodiversity,
148 identification of the species to which individuals belong is only possible *en masse* through
149 ***metabarcoding***. This might seem to be a situation very different from the case with macrobes,
150 but Identifying macrobial individuals to species is not always straightforward or precise (Box 2).
151 For example, many macrobial groups, such as insects, are often named as arbitrary and non-
152 monophyletic ***morphospecies***, especially in highly diverse ecosystems such as the tropics.

153 Delimiting species. Identifying the species to which individual organisms belong,
154 assumes that species exist in the first place. It has been argued that the prevalence of
155 parasexuality among microbes precludes the use of a common species currency for macrobes
156 and microbes. Because of parasexuality, rates and extents of genetic recombination can vary
157 among microorganisms. The "rare but promiscuous" exchange of genes among unrelated taxa
158 has the potential to fundamentally alter the species currency for microbes because it can
159 decouple traits and lineages. Traits can spread across unrelated lineages if there is strong
160 selection, as can happen with the spread of antimicrobial resistance genes among pathogens.
161 However, recent studies have provided strong evidence that many ecologically important traits
162 are phylogenetically conserved within microbial lineages (e.g., [34]), suggesting that such
163 genetic exchange is not so widespread or frequent that it reduces the utility of microbial taxa.
164 As a result, while the definition of microbial taxa may depend on the question being asked, they
165 nonetheless represent stable and useful units of study, just as for macrobes.
166 Comparable methods. Some have suggested that contemporary microbial community
167 methods, which typically rely on sequencing from the environment, are fundamentally different
168 from those approaches used to observe individuals and species for macrobes. However, there
169 also are biases in approaches to observe macrobial communities (**Box** 2). Furthermore,
170 macrobial communities increasingly are observed with metabarcoding methods as sequencing
171 prices plummet. This approach is essentially identical to that used by microbial ecologists.
172 In short, although there are real challenges in counting both macrobes and microbes,
173 the challenges are more similar between these groups than they are different. As more

biologists studying macrobes use molecular (and, particularly, metagenomic) approaches, the differences between them will shrink further.

Unified accounting: understanding patterns in diversity over space and time

Regardless of real and perceived differences in tallying macrobes and microbes, there is a primary data structure that is universal to the analysis of biodiversity: a site-by-species matrix, (including presence-absence or abundances; **Figure 1A**). From this matrix, we can assess patterns of diversity and ask how these patterns scale over space or time [35]. Below, we consider six common patterns in macroecology that can be assessed using the site-by-species matrix. We selected examples from our collective works and the published literature to illustrate how these macroecological patterns of microbes and macrobes can be similar. These datasets (**Table S1**) are intended to serve as examples of the kinds of patterns that can be discovered, and are not representative of all macrobial and microbial communities. Later, we will discuss how these patterns are interconnected.

Species Abundance Distributions. One of the most fundamental patterns in community ecology and macroecology is the ***species abundance distribution*** (SAD). Typical SADs describe communities that have a few species that are highly abundant and many species that are rare; indeed, this has been suggested as one of the "true universal laws" in ecology (Lawton et al. 1998, McGill et al. 2007). Notably, every SAD represents a sampled subset of the "true" SAD for the whole community. There is some indication that spatial aggregation of species can inflate the representation of rare taxa in the sampled SADs [37]. Though we do not expect any aggregation bias to be different between microbes and macrobes, understanding differences in

aggregation among taxa (be they microbes and macrobes or just different kinds of microbes) will be key to truly generalizing SAD relationships. Here, we show examples of SADs for groundwater bacterial communities and moths, both of which show the characteristic pattern, albeit with some structural differences in the distributions of rarity which we discuss in more detail below (**Figure 1B**).

Abundance-occupancy. Another macroecological pattern is revealed when considering the relationship between species abundance and ***occupancy*** (**Figure 1C**). Here, we provide examples of ***abundance-occupancy relationships*** for microbiota sampled from human belly buttons and for birds observed in the Czech Republic. Both datasets generally show that species that tend to have high abundance within one site also tend to occupy many sites, while those that are locally rare tend to not be detected in many sites [4]. Abundance-occupancy patterns have been applied in microbial ecology to create null or neutral expectations about the drivers of community structure [38]. There are many factors that can influence abundance-occupancy relationships. Microbial laboratory models (**Box 3**) offer a useful approach to assessing the specific influences of biotic interactions and habitat heterogeneity in microbial abundance-occupancy patterns [39]. In the microbial ecology literature, some have argued that deviations from a null hypothesis are suggestive of deterministic drivers of community structure [21,38,40,41]. For example, taxa that are very abundant only in a few sites or very rare taxa that are consistently observed in many sites would be exceptions to the neutral expectation.

Species-area relationships. Species area relationships (SAR) assess the increase in species richness with increasing spatial area (**Figure 1D**). The shape and slope of the SAR can be derived from the knowledge of some properties of species distributions [42], such that the SAR

219 can be used to predict and compare changes in diversity over increasing spatial extent.
220 However, there are nuances to its application, especially for microbial communities, because of
221 practical challenges in sampling contiguous areas. In the nested SAR, larger areas should be
222 contiguous and encompass all the smaller areas therein. However, empirical SARs are often
223 constructed by a collection of samples from smaller areas (here, we call these "piecemeal" SARs
224 for clarity), which are assumed to be representative of the whole contiguous and mutually
225 adjacent area. SARs have been extensively examined in many microbial communities [43–46],
226 using the piecemeal approach because of the necessity of destructive sampling for DNA
227 extractions. Such piecemeal SARs are predicted to be more curvilinear in the log-log scale due
228 to the limited total number of individuals at small areas [42,47], and their slope is predicted to
229 be higher due to lower occupancies of individual species [42]. Thus, care is needed when
230 constructing and interpreting nested and piecemeal SARs. Our example shows increases in
231 fungal community richness at Barro Colorado Island (BCI) as compared to tree richness at the
232 same location (but note differences in x- and y- scales).
233 Distance-decay. (**Figure 1E**). Distance-decay relationships assess how community
234 similarity or beta-diversity [48] changes over space. Distance-decay is used to address
235 compositional turnover (using unweighted resemblance metrics, like Jaccard) or shifts in
236 relative abundance (using weighted resemblances, like Bray-Curtis) with distance from a
237 reference community. The slope of the distance-decay relationship is interpreted as a rate of
238 change over space, and there are macroecological studies as well as microbial-focused studies
239 that have compared these rates [6,7,49–51]. Our example shows the same BCI fungal and tree
240 communities from **Figure 1D**, but because the Jaccard metric can be calculated for both, their

241 rates of decay in similarity can be compared directly on the same y-axis scales, although some
242 caution is necessary when comparing trees with microbes, since the area (grain) of the samples
243 differs. [6,7,53–55].

244 Rarefaction. Rarefaction assesses how richness accumulates with the number of
245 individuals or samples observed (**Figure 1F**). Here, we use individual-based rarefaction curves
246 to compare how species richness accumulates with increasing numbers of individuals (after
247 eliminating spatial structure via randomizations, [56]). We show English Channel bacteria and
248 archaea to Celtic Sea fishes. In microbial ecology, rarefaction is commonly used to assess
249 completeness of sequencing effort for a dataset. The y-axis for a rarefaction of microbial
250 sequences reveals the number of taxa observed for each additional sequence collected within a
251 community (increasing sequencing depth – observations of individuals). This is distinct from a
252 sample-based rarefaction analysis that reveals the number of species observed for each
253 additional community observed (increasing sampling – observations of communities).

254 The first four features of diversity matrices we have described above are intrinsic to the
255 matrices. Each of these features can, as we have shown, be calculated just as readily for
256 microbes as for macrobes. Once these aspects of diversity are estimated, they can be compared
257 along geographic (e.g., latitude, elevation) and environmental (e.g., energy, disturbance)
258 gradients (**Figure 1G**). Moving forward from these comparative analyses, we can address
259 paramount questions in macroecology: If some patterns in biodiversity are the same for
260 microbes and macrobes, are the underlying processes also the same? Also, do similar
261 processes lead to different patterns?

The abovementioned macroecological patterns are related to each other, and each can be used to inform the others (e.g., [52]). When there is a predictable relationship between abundance and occupancy, there is also a link between the SAD and the probability distribution of the proportion of available area (or available set of sites). Species richness for a given area can be calculated as the sum of probabilities of occurrence across all species, and the SAR thus can be reconstructed using knowledge of species occupancy patterns in each spatial scale [42]. Therefore, if we know the SAD for some large area and the level of spatial aggregation of individuals of every species (which determines occupancy patterns across spatial scales), we can derive all the other macroecological patterns. Moreover, these links work in all possible directions. For example, it is possible to derive the SAD from scale-dependent patterns of species aggregation [53]. Although these links are complex, the general insight is that patterns of species rarity and occupancy are directly linked to scaling patterns in species richness. Indeed, the rarer the species are on average, the faster the number of species increases with area or number of samples, and the higher are the differences in community composition between neighboring areas or samples (i.e. higher beta diversity). A comprehensive understanding of patterns of diversity, distribution and abundance (which is one of the main goals of ecology) thus depends on understanding these links among major macroecological patterns.

Rarity: An exception or statistical inevitability?

Our illustration of macroecological patterns among microbes and macrobes (**Figure 1B-G**), reveals similar shapes in general, as expected from major macroecological theories, but

notable differences that are all related to higher rarity in the microbial realm. The species abundance distribution has proportionally more singletons for microbes from groundwater compared to Fisher's moths (**Figure 1B**); the occupancy of bacteria in human belly buttons is lower than the occupancy of bird among sites in the Czech Republic (**Figure 1C**); the fungi continue with an appreciable slope as the trees have tapered in their species-area curves of the BCI data (**Figure 1D**), which is also reflected by the much lower similarity in species composition among even nearby fungal samples (**Figure 1E**); finally, the accumulation of new taxa with increasing numbers of marine microbes has not slowed as appreciably as the marine fishes (**Figure 1F**).

While the vignettes presented in **Figure 1** suggest possible differences in rarity between microbes and macrobes, they are anecdotal. Nevertheless, we illustrate a similar preponderance for rarity in microbes in a systematic comparison of >14000 macrobial and microbial SADs (**Figure 2**). As sequencing technologies have improved and coverage of microbial communities has increased, it has often been noted that many microbial communities have a high proportion of rare microbial taxa [54–56]. Subsequently, it was shown that some rare microbial taxa can provide specific and important functions within their communities [57].

To consider a particular aspect of rarity, microbial communities often include a large number of ***singletons***. It has been argued that singletons might not be real individuals (e.g., [58–60]) but an artifact of sequencing methods. As such, singletons are removed prior to analysis [22,61,62]. However, singletons are a general feature of ecological communities (e.g., [63,64]) and provide a potential quantitative point of comparison between microbes and macrobes. We argue that singletons from high-quality sequences should not be arbitrarily

306 removed. Study-to-study variability in whether to include microbial singletons presents a hurdle
307 to the common accounting required for cross-dataset comparisons in macroecology.
308 Communities become increasingly uneven with increasing numbers of individuals [65],
309 and rarity also increases with more individuals [26]. However, for a given community size,
310 microbial communities have more rarity than macrobial communities [26]. There are ecological
311 reasons to explain rarity, including transiency (vagabonds), recent speciation, local extinction,
312 and negative frequency dependence [63,64,66,67]. Future work should be directed to testing
313 ecological hypotheses concerning the mechanisms supporting rarity and singletons generally,
314 and specifically within microbial communities.
315
316 *A call for a unified macroecology of all life, great and small*
317 Moving forward from the understanding that species and individuals provide basic units
318 from which a unified macroecology can emerge, we must systematically observe and compare
319 macroecological patterns across macroorganisms and microorganisms. The next steps are to
320 understand the processes that underlie the patterns, determine their generality, and use them
321 to inform a grand, macroecological view of life's rules and exceptions (**Figure 3, Box 1**). It is
322 important to understand when microbes are distinct from macrobes in pattern, as these
323 distinctions can inform process. There are two particularly intriguing scenarios: one in which
324 divergent patterns result from the same process (**Figure 3-ii**), and one in which convergent
325 patterns mask distinct processes (**Figure 3-iii)**. Divergent and convergent scenarios
326 simultaneously offer a challenge and an opportunity towards a unified macroecology. The
327 challenge is that microbial ecologists often struggle with determining processes *in situ* because

328 observations are difficult and methods reliant on available technology and its limitations. The
329 opportunity is that laboratory microbial models offer the ability to manipulate and control
330 systems to explicitly test macroecological hypotheses of processes, an experimental luxury that
331 is relatively uncommon for communities of macrobes because of logistical constraints in scale,
332 expense, and, sometimes, ethics (**Box 3**). After standardizing language and a conceptual
333 framework, a priority should be to systematically determine which scenario in **Figure 2** applies
334 to which macroecological comparison. Microbial ecology especially will benefit from
335 advancement towards synthesis, and macroecology provides a foundation for this pursuit. A
336 unified synthesis of macroecology is needed and imminent.
337 There are also cultural and infrastructural silos to overcome before a truly unified
338 macroecology can be achieved. Patterns and processes typical of microbial communities
339 provide value and insights for macroecology, even when they are distinct from the patterns and
340 processes of macrobial communities. In publication and funding, microbial ecology should be
341 considered equitably and not as a subspecialty with limited scope or utility. Collaborations
342 between macrobial and microbial ecologists are key for advancing a unified macroecology, first
343 to understand jargon, culture, and methods and limitations, and next to together select
344 questions to tackle. Long-term working groups, focused workshops, and integrated sections in
345 professional societies can provide infrastructure for research efforts, and these should include
346 opportunities for trainees to contribute. Collaborative mentoring of students and post-docs,
347 who can bridge micro- and macro-leaning advisers and move forward working group research
348 initiatives, is another mechanism by which macroecology can aim to unify with the next
349 generation of inspired ecologists.

Let's move forward together, away from the artificial delineation in the ecological study of microorganisms and macroorganisms and towards an encompassing macroecology, inclusive of all biodiversity.

References

1 Wilson, E.O. (2017) Biodiversity research requires more boots on the ground: Comment. *Nat. Ecol. Evol.* 1, 1590–1591

2 Brown, J.H. (1995) *Macroecology*, University of Chicago Press.

3 Lawton, J.H. (1999) Are There General Laws in Ecology? *Oikos* 84, 177

4 Gaston, K.J. *et al.* (2000) Abundance-occupancy relationships. *J. Appl. Ecol.* 37, 39–59

5 Prosser, J.I. *et al.* (2007) The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 385–392

6 Soininen, J. *et al.* (2007) The distance decay of similarity in ecological communities. *Ecography (Cop.).* 30, 3–12

7 Astorga, A. *et al.* (2012) Distance decay of similarity in freshwater communities: Do macro- and microorganisms follow the same rules? *Glob. Ecol. Biogeogr.* 21, 365–375

8 Fierer, N. *et al.* (2011) Microbes do not follow the elevational diversity patterns of plants

and animals. *Ecology* 92, 797–804

9 Bryant, J.A. *et al.* (2008) Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proc. Natl. Acad. Sci.* 105, 11505–11511

10 Hillebrand, H. and Azovsky, A.I. (2001) Body size determines the strength of the latitudinal diversity gradient. *Ecography (Cop.).* 24, 251–256

11 Fuhrman, J.A. *et al.* (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7774–7778

12 De Wit, R. and Bouvier, T. (2006) “Everything is everywhere, but, the environment selects”; what did Baas Becking and Beijerinck really say? *Environ. Microbiol.* 8, 755–758

13 Green, J.L. *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature* 432, 747–750

14 Martiny, J.B.H. *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112

15 Hanson, C.A. *et al.* (2012) Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10, 497–506

16 Hug, L.A. *et al.* (2016) A new view of the tree of life. *Nat. Microbiol.* 1, 16048

17 Thompson, L.R.L.R. *et al.* (2017) A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* 551,

18 Parks, D.H. *et al.* (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542

19 O'Malley, M.A. (2007) The nineteenth century roots of "everything is everywhere." *Nat. Rev. Microbiol.* 5, 647–651

20 Noguez, A.M. *et al.* (2005) Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Glob. Ecol. Biogeogr.* 14, 241–248

21 Burns, A.R. *et al.* (2016) Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* 10, 655–664

22 Ramirez, K.S. *et al.* (2017) , Detecting macroecological patterns in bacterial communities across independent studies of global soils. , *Nature Microbiology*, 1–8

23 Stegen, J.C. *et al.* (2016) Aligning the measurement of microbial diversity with macroecological theory. *Front. Microbiol.* 7, 1–7

24 Shoemaker, W.R. *et al.* (2017) A macroecological theory of microbial biodiversity. *Nat. Ecol. Evol.* 1, 1–6

25 Shade, A. *et al.* (2013) A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J.* 7, 1493–1506

26 Locey, K.J. and Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* 113, 5970–5975

27 Baldridge, E. *et al.* (2016) An extensive comparison of species-abundance distribution models. *PeerJ* 4, e2823

28 Kieft, T.L. (2017) New Allometric Scaling Laws Revealed for Microorganisms. *Trends Ecol. Evol.* 32, 400–402

29 Crowther, T.W. *et al.* (2015) Mapping tree density at a global scale. *Nature* 525, 201–205

30 Kallmeyer, J. *et al.* (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci.* 109, 16213–16216

31 Whitman, W.B. *et al.* (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578–6583

32 Hughes, J.B. *et al.* (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67, 4399–4406

33 Mora, C. *et al.* (2011) How many species are there on earth and in the ocean? *PLoS Biol.* 9,

34 Martiny, J.B.H. *et al.* (2015) Microbiomes in light of traits: A phylogenetic perspective. *Science* 350, aac9323

35 Gotelli, N.J. and Graves, G.R. (1996) Null models in ecology. *Ecology* 14, 368

36 Lawton, J.H. *et al.* (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* 391, 72–76

37 Green, J.L. and Plotkin, J.B. (2007) A statistical theory for sampling species abundances. *Ecol. Lett.* 10, 1037–1045

38 Sloan, W.T. *et al.* (2007) Modeling taxa-abundance distributions in microbial communities using environmental sequence data. *Microb. Ecol.* 53, 443–455

39 Holt, A.R. *et al.* (2004) The importance of habitat heterogeneity, biotic interactions and dispersal in abundance-occupancy relationships. *J. Anim. Ecol.* 73, 841–851

40 Dini-Andreote, F. *et al.* (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci.* 112, E1326–E1332

41 Ayarza, J.M. and Erijman, L. (2011) Balance of Neutral and Deterministic Components in the Dynamics of Activated Sludge Floc Assembly. *Microb. Ecol.* 61, 486–495

42 Storch, D. (2016) The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *J. Veg. Sci.* 27, 880–891

43 Green, J. and Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* 21, 501–507

44 Horner-Devine, M.C. *et al.* (2004) A taxa-area relationship for bacteria. *Nature* 432, 750–753

45 Zhou, J. *et al.* (2008) Spatial scaling of functional gene diversity across various microbial taxa. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7768–7773

46 Liang, Y. *et al.* (2015) Over 150 years of long-term fertilization alters spatial scaling of microbial biodiversity. *MBio* 6,

47 Okie, J.G. *et al.* (2015) Niche and metabolic principles explain patterns of diversity and distribution: theory and a case study with soil bacterial communities. *Proc. R. Soc. B Biol. Sci.* 282, 20142630

48 Anderson, M.J. *et al.* (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* 14, 19–28

49 Nekola, J.C. *et al.* (1999) The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* 26, 867–878

50 Bell, T. (2010) Experimental tests of the bacterial distance-decay relationship. *ISME J.* 4, 1357–1365

51 Martiny, J.B.H. *et al.* (2011) Drivers of bacterial beta-diversity depend on spatial scale. *Proc. Natl. Acad. Sci.* 108, 7850–7854

52 McGill, B.J. (2011) Linking biodiversity patterns by autocorrelated random sampling. *Am. J. Bot.* 98, 481–502

53 Sizling, A.L. *et al.* (2009) Species abundance distribution results from a spatial analogy of central limit theorem. *Proc. Natl. Acad. Sci.* 106, 6691–6695

54 Sogin, M.L. *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120

55 Shade, A. *et al.* (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* 5, e01371-14

56 Lynch, Mi.D.J. and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229

57 Jousset, A. *et al.* Where less may be more: How the rare biosphere pulls ecosystems strings. , *ISME Journal*, 11. (2017) , 853–862

58 Reeder, J. and Knight, R. (2009) The "rare biosphere": a reality check. *Nat. Methods* 6, 636–637

59 Huse, S.M. *et al.* (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898

60 Quince, C. *et al.* (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12, 38

61 Gobet, A. *et al.* (2010) Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. *Nucleic Acids Res.* 38, e155–e155

62 Schloss, P.D. *et al.* (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* 6, e27310

63 Straatsma, G. and Egli, S. (2012) Rarity in large data sets: Singletons, modal values and the location of the species abundance distribution. *Basic Appl. Ecol.* 13, 380–389

64 Novotný, V. *et al.* (2000) Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* 89, 564–572

65 Locey, K.J. and White, E.P. (2013) How species richness and total abundance constrain the distribution of abundance. *Ecol. Lett.* 16, 1177–1185

66 Rabinowitz, D. *et al.* (1986) Seven forms of rarity and their frequency in the flora of the British Isles. In *Conservation Biology: The Science of Scarcity and Diversity* (Soulé, M. E., ed), Sinauer Associates, Inc

67 Preston, F.W.F. (1948) The commonness, and rarity, of species. *Ecology* 29, 254–283

68 McGlinn, D. *et al.* Rdataretriever: R interface to the Data Retriecer. . (2017)

69 Brown, J.H. *et al.* (2004) Toward a metabolic theory of ecology. Ecology. *Ecology* 85, 1771–1789

70 DeLong, J.P. *et al.* (2010) Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci.* 107, 12941–12945

71 Zhou, J. *et al.* (2016) Temperature mediates continental-scale diversity of microbes in forest soils. *Nat. Commun.* 7,

72 Makarieva, A.M. *et al.* (2008) Mean mass-specific metabolic rates are strikingly similar across life's major domains: Evidence for life's metabolic optimum. *Proc. Natl. Acad. Sci.* 105, 16994–16999

73 Andersen, K.H. *et al.* (2016) Characteristic Sizes of Life in the Oceans, from Bacteria to Whales. *Ann. Rev. Mar. Sci.* 8, 217–241

74 Kembel, S.W. *et al.* (2012) Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Comput. Biol.* 8, 16–18

75 Louca, S. *et al.* (2018) Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem.

76 Wen, C. *et al.* (2017) Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PLoS One* 12,

77 Zhou, J. *et al.* (2015) High-throughput metagenomic technologies for complex microbial community analysis: Open and closed formats. *MBio* 6,

78 Nayfach, S. and Pollard, K.S. (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* 16, 51

79 Miller, C.S. *et al.* (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 12, R44

80 Freitas, T.A.K. *et al.* (2015) Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic Acids Res.* 43,

81 Quince, C. *et al.* Shotgun metagenomics, from sampling to analysis. , *Nature Biotechnology*, 35. (2017) , 833–844

82 Fish, J.A. *et al.* (2013) FunGene: The functional gene pipeline and repository. *Front. Microbiol.* 4, 291

83 Roux, S. *et al.* (2011) Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiol. Ecol.* 78, 617–628

84 Fierer, N. Tilting at Windmills: A response to a recent critique of terminal restriction fragment length polymorphism data. , *Applied and Environmental Microbiology*, 73. (2007) , 8041

85 Rosselló-Móra, R. and Amann, R. Past and future species definitions for Bacteria and Archaea. , *Systematic and Applied Microbiology*, 38. (2015) , 209–216

86 Konstantinidis, K.T. *et al.* Uncultivated microbes in need of their own taxonomy. , *ISME Journal*, 11. (2017) , 2399–2406

87 Gaby, J.C. and Buckley, D.H. (2012) A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. *PLoS One* 7,

88 Rideout, J.R. *et al.* (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2, e545

89 Schloss, P.D. (2016) Application of a Database-Independent Approach To Assess the Quality of Operational Taxonomic Unit Picking Methods. *mSystems* 1, e00027-16

90 Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon

reads. *Nat. Methods* 10, 996–8

91 Storch, D. and Šizling, A.L. (2008) The concept of taxon invariance in ecology: Do diversity patterns vary with changes in taxonomic resolution? *Folia Geobot.* 43, 329–344

92 Graham, C.H. *et al.* (2018) Phylogenetic scale in ecology and evolution. *Glob. Ecol. Biogeogr.* 27, 175–187

93 Lenski, R.E. *et al.* (1991) Long-Term Experimental Evolution in Escherichia coli . I . Adaptation and Divergence During. *Am. Nat.* 138, 1315–1341

94 Belisle, M. *et al.* (2012) Flowers as Islands: Spatial Distribution of Nectar-Inhabiting Microfungi among Plants of Mimulus aurantiacus, a Hummingbird-Pollinated Shrub. *Microb. Ecol.* 63, 711–718

95 Bell, T. *et al.* (2005) Larger islands house more bacterial taxa. *Science (80-.).* 308, 1884

96 Svoboda, P. *et al.* (2018) Dispersal timing determines the importance of priority effects in bacterial communities. *ISME J.* 12, 644–646

97 Jessup, C.M. *et al.* (2004) Big questions, small worlds: microbial model systems in ecology. *Trends Ecol. Evol.* 19, 189–197

98 Blackburn, T.M. Method in macroecology. , *Basic and Applied Ecology*, 5. (2004) , 401–412

99 Peay, K.G. *et al.* (2012) Phylogenetic relatedness predicts priority effects in nectar yeast

communities. *Proc. R. Soc. B Biol. Sci.* 279, 749–758

100 Violle, C. *et al.* (2010) Experimental demonstration of the importance of competition under disturbance. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12925–9

101 Kassen, R. *et al.* (2000) Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* 406, 508–12

102 Widder, S. *et al.* (2016) Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568

103 Webb, T.J. Marine and terrestrial ecology: Unifying concepts, revealing differences. , *Trends in Ecology and Evolution*, 27. (2012) , 535–541

Figure legends

Figure 1. **Examples of macroecological patterns from the microbial (gray) and macrobial (black) realms.** (A) The site-by-species matrix, where samples/communities are provided in columns (sites) and species/taxonomic units (species) in rows. From this table, all subsequent patterns of diversity can be derived, such as (B) rank-abundance curves, (C) occupancy-abundance relationships, (D) species-area curves, (E) distance-decays of similarity, (F) rarefaction curves, and (G) elevational richness gradients. Thick lines in D and F are means of the simulated species-area and rarefaction curves, grey ribbons are 95% quantiles of the simulations. Thick lines in E and G are means modelled by GAM splines. Grey contours in E show density of the data, grey ribbons in G are 95% confidence intervals of the splines. Data sources for panels B-G are in Supporting Table 1. For licensing information on the inset icons see the acknowledgements.

Figure 2. Rarity is a distinctive ecological feature of microbial communities. Microbial data (gray) are from [26]; macrobial data (black) in panels were downloaded using the R data retriever [68] (n = 14,980 for both microbes and macrobes). In general, microbial communities have proportionally more singletons (A) than macrobial communities. Doubletons (B) are more comparable, with a wider observed range and more bias observed in microbial doubletons. ***Fisher's alpha*** (C) is notably much higher in microbial communities as compared to macrobial communities.

Figure 3. **Conceptual framework for comparing the relationships between patterns and processes across macroorganisms and microorganisms.** Different relationships are represented by different letters. (i) *Universal* is when macroecological patterns agree between microbes and macrobes, and result from the same processes despite nuances or variability in exact mathematical properties, like the exponents of SARs and metabolic scaling (**Box 1**). (ii) *Divergent* is when equivalent underlying processes result in different patterns for microbes and macrobes. (iii) *Convergent* is when microbes and macrobes exhibit the same patterns, but the patterns are attributable to fundamentally different mechanisms. Divergent and convergent relationships are difficult to characterize without measurement of potential processes. Convergent patterns especially can be overlooked because similar patterns are often assumed to be underpinned by similar processes when the processes are yet-uncharacterized. (iv). *Independent* is when microbes and macrobes exhibit distinct patterns that are also underpinned by distinct processes. In independent relationships, both sets of patterns and processes are equally valuable in informing a unified macroecology.

603 Boxes

604 ~~~~~~

605 **Box 1. Metabolic scaling across macrobes and microbes**

606 One macroecological pattern that was considered universal across both microorganisms and
607 macroorganisms is the scaling of metabolic rate (and many other biological rates) with body
608 size. It was generally believed that the relationship is linear when both the body mass and
609 metabolic rate axes are logarithmic, and that this line spans all organisms from microbes to
610 whales with a universal slope ¾ (and thus can be represented as a power law with the exponent
611 of 0.75 [69]. However, [70] have shown that a more detailed data analysis provides a different
612 picture. While multicellular organisms indeed reveal ¾ scaling, metabolic rate in protists scale
613 proportionally to body size (i.e. the scaling coefficient is close to 1) and bacteria and archaea
614 reveal scaling coefficient close to 2, i.e. a quadratic increase of metabolic rate with body size.
615 The authors attributed these differences to different constraints on metabolic rate across
616 microorganisms and macroorganisms. While in bacteria the metabolic rate is assumed to be
617 limited by number of genes and proteins involved in metabolism (so that bigger bacteria have
618 disproportionately higher number of molecules participating in metabolic reactions), in protists
619 it is supposedly limited by the number of mitochondria within the cell, leading to approximate
620 proportionality between cell size and metabolic rate. Multicellular organisms, in contrast, are
621 limited by their ability to provide resources to all metabolically active cells, so that their
622 metabolic rate is constrained by the structure of their transportation system, which leads to
623 sublinear scaling, with coefficient close to ¾ [69]. There has been recent work to determine the

624 utility of metabolic scaling in explaining soil microbial community responses to global warming

625 [71], and microbes have been integrated into macroecology energetics (e.g., [72,73]).

Box 2. Primary currencies of individuals and species.

Counting the individual. Even though counting individuals can at first seem straight-forward for macrobial biologists, counting of animals or plants relies on simplifying assumptions made within taxonomic subfields (Table **i**). However, these challenges have not prevented progress in understanding the global patterns in the distribution and diversity of species or the general rules that drive them.

Assessment of individuals is similarly challenging for microbiologists. Counting individual cells was traditionally performed with microscopy, which does not accurately reveal taxonomic identity. Individual microbes and their taxonomic identity are often estimated using molecular approaches like marker gene studies, such as those amplifying and sequencing of bacterial and archaeal *16S rRNA genes*. Quantitative PCR of 16S rRNA genes is used as an estimate of community size, though this value is imprecise because different taxa can have different numbers of 16S rRNA operons. A recent meta-analysis similarly estimated a mean community 16S rRNA gene copy number of 2.2 among free-living bacteria and archaea [17], which supports a trend towards low 16S rRNA gene operon copies per the "average" cell. Although not widely applied, there are bioinformatics methods to correct for the number of operons per genome (e.g., [74]), though some argue that there is still too limited information to apply such corrections accurately [75]. Alternatively, quantification of a single-copy *housekeeping gene* can be used enumerate community size.

Despite the limitations of using 16S rRNA genes or similar to count individuals [76,77], macroecological patterns emerge from these types of data. However, with new tools for counting individuals from shotgun metagenomes [78–80], improvements in coverage and

quality of high-throughput sequencing and analysis [81] and the use of single-copy *marker genes* for diversity [82,83], microbial ecologists are poised to increase precision. It is time to no longer be distracted by the limitations of today's methods [84], adopt standard best practices in sequence analysis, and move forward in using the best quantifications currently available to boldly count individual microbes within their communities.

Counting the species. "Species" has historically been chosen as the primary unit in studies of plant and animal communities because it is believed to be the smallest consistent unit of variety representing important ecological differences (in life history, optimal growth conditions, resource use, etc.), although these assumptions have been challenged for plants and animals. For macroorganisms, species are often based on morphological characteristics and mating capacity, but still, there are many "cryptic" species.

Defining a microbial species is also challenging [85,86]. Therefore, microbial ecologists that use molecular approaches, such as sequencing of the 16S rRNA gene, apply an *operational taxonomic unit* (OTU) definition in lieu of "species". OTUs are just that: operational, and so they can be defined using whatever method is biologically or statistically defensible. There are examples in which OTU definitions matter for microbial macroecology (e.g., [44]), and others in which they do not (e.g., [26]). In addition, although the 16S rRNA gene is the most common target, microbial functional genes [82], such as the nitrogen fixation gene, *nifH* [87], are also used in microbial ecology to count taxa in terms of their functional traits. OTUs can be created from any gene that has nucleotide variation.

There are different methods employed to “cluster” similar sequences together into an OTU. Most require that a sequence identity cut-off be chosen for the out (97% is standard, but 98%, 99% and 100% cut-offs – exact sequence variants - have also been applied). There are a variety of clustering methods available, from those that rely on a well-curated reference database to those that define OTUs *de novo* for every study [88], and it is beyond our scope to discuss them all here, except to say that it has important consequence for OTU definitions [88–90]. Regardless of which OTU definition is applied, a consistent OTU definition is necessary in comparative or meta- analyses among datasets.

Notably, if a 97% sequence identity definition was applied to a similar gene in mammals, it would result in grouping all of the primates (from lemurs to humans) into one taxon. But we disagree that this suggests that the species currency is fundamentally different for microbes. Macroecological processes function at multiple taxonomic scales and macroecological patterns have been documented for macrobes at various taxonomic [91] and phylogenetic levels [92], including genera and families. As mentioned above, changing the sequence similarity cut-off (essentially sliding from subspecies through species to genera and families), can provide important macroecological information. Macroecologists should view this example set by microbial ecologists as an encouragement towards taxonomic agnosticism. Such agnosticism would support integration around patterns (instead of unmatched “species” definitions), inform as to which resolution of taxonomic units are most ecologically meaningful, and provide a full understanding of biodiversity patterns across phylogenetic scales.

Box 3. Microbial systems in macroecology: Advantages, contributions, and frontiers

Microbial systems, which include *in situ* communities and controlled laboratory models, boast an often-understated legacy of providing foundational insights into ecology and evolution. Microbial systems have contributed to our understanding of, among other topics, long-term evolutionary processes [93], island biogeography [94,95], and dispersal limitation and metacommunities [96]. The utility of microbial systems for ecology has been detailed previously [97]. They offer several advantages, including: efficient observations at temporal and spatial scales that are compressed relative to their "macrobial" equivalents; molecular tools for characterizing population dynamics; and controlled manipulations of experimental treatments and community biodiversity. Microbial laboratory models include synthetic or simplified microbial communities and **mesocosms**, and have been suggested as an important tool for advancing macroecology [98]. There is an especially rich legacy of using microbial mesocosms in community and population ecology (e.g., [99–101]). The capability to complement *in situ* observations and reductionist models can provide a rich understanding of macroecological patterns and their underlying processes [102]. In addition, because related lineages or similar functional guilds of microorganisms are found across otherwise disparate habitats, microbial systems also offer a common denominator that can be leveraged for cross-ecosystem comparisons and in support of a unified macroecology (e.g., [103]). In summary, microbial systems continue to offer exciting methods that yield insights for macroecology.

Glossary

- *16S rRNA gene* – In microbial ecology, the structural gene that encodes the 16S small subunit of the ribosome. It includes both highly conserved and hypervariable regions, which are used for primer design to capture broad phylogenetic diversity and for assessing phylogenetic divergence, respectively.
- *Abundance-occupancy relationships*- The (generally positive) relationship between the mean abundance a species attains at individual sites, and the number or proportion of all sampled sites at which it is found.
- *Biodiversity*- the variety of life's species. Biodiversity can be measured using the currencies of individuals and species. These currencies can be used to estimate biodiversity for local communities, planet Earth, and every scale of spatial observation in between.
- *Diversity gradients* – the assessment of how the number of species changes as function of an environmental gradient.
- *Exact sequence variants* – the practice of defining highly resolved microbial taxonomic units by identical nucleotide sequences of marker genes. Also called "amplicon sequence variants", "sequence variants", "oligotypes", and "zero-radius OTUs".
- *Fisher's alpha* – an alpha diversity metric that considers the relationship between the number of species and the number of individuals within species.

- *Functional redundancy* - the concept that, within a microbial community, there are several microbial taxa that are capable of performing the same function in the same conditions, and, presumably, at the same rate.
- *Genetic barcoding* - the sequencing of taxonomically informative marker genes amplified from individuals.
- *Housekeeping gene* – in microbial ecology, a gene that is present in only one copy within a microbial genome and encodes a function necessary for life (typically involved in central metabolism).
- *ITS*—Intergenic spacer, a.k.a. intergenic transcribed spacer. A marker sequence flanked by ribosomal operons that is used to phylogenetically distinguish eukaryotic microorganisms, especially fungi.
- *Macroecology* is the study of the rules and mechanisms (processes) underpinning general patterns of ecology across scales [2].
- *Marker genes* – in microbial ecology, genes and their sequences that have been used as a signature of microbial diversity. An example are the 16S rRNA gene for bacteria and archaea and the ITS region for fungi.
- *Mesocosm* – a small container containing organisms and substrate that can be replicated and manipulated in the laboratory. Microbial mesocosms can have natural or artificial substrate, like soil or microbiological medium, respectively, and can be seeded with "wild" communities from a particular habitat or inoculated with specified cultivable members. It is expected that the influences of captivity away from nature (sometimes called "container effects") can be minimized in microbial mesocosms. This is because

752 microbial individuals, and their expected effective ranges for interactions with each
753 other and with their environment, are small relative to the container's volume,
754 • *Metabarcoding* - the sequencing of taxonomically informative marker genes amplified
755 from an environmental sample that contains mixed populations or communities.
756 "General" primers that target a conserved nucleotide sequence are used to amplify the
757 signal of marker genes from a mixed microbial community. These sequences are
758 typically multiplexed for sequencing, and then they can be used with databases of
759 known sequences to build phylogeny, assign taxonomy, assess alpha diversity, and
760 create an species-by-sample table (*OTU table*, as in **Fig. 1A**) for community analysis.
761 • *Metagenomics* – the sequencing of all nucleic acid extracted from an environmental
762 sample, without targeted amplification. Also known as "shotgun" metagenome
763 sequencing, this method is commonly applied to microbial communities to assess
764 functional potential by annotating sequences against a database of known functional
765 genes.
766 • *Microorganisms* – Broadly defined as those organisms too small to be visible with the
767 naked eye, including viruses, bacteria, archaea, protists, a subset of fungi or even the
768 smallest arthropods (such as face mites). When evolutionarily defined, microorganisms
769 include the domains of bacteria and archaea (previously, the prokaryotes), which were
770 the first evolved lineages that through endosymbiosis gave rise to eukarya.
771 • *Morphospecies* – a species concept that is based on morphology, and is commonly used
772 in the fields of entomology and botany. Unidentifiable individuals with shared physical

characteristics are grouped artificially into an operational taxonomic unit without reference to other distinguishing traits.

- *Occupancy* – the number or proportion of sites in which a species is detected.
- *Operational taxonomic unit (OTU)* – approximations of species that are commonly used in the field of microbial ecology, arbitrarily defined as informed by the technology used to observe the microorganisms. For example, 16S rRNA gene amplicon sequencing datasets often define OTUs at 97% gene sequence identity. Thus, all sequences that are 97% similar would be counted towards a single OTU.
- *Parasexual*- nonsexual mechanisms for transferring genetic material, common among single-celled organisms like bacteria, archaea, protists, and fungi.
- *Singletons* – within a dataset, taxa that are observed only one time and in an abundance of one individual. In microbial ecology, this often refers to a singly observed unique sequence of a marker gene.
- *Species-abundance distribution* – depicts the number of individuals (N) of each species in a sample, and is often expressed as a relationship between the logarithm of N plotted against species rank (from the most to the least abundant species)
- *Species area relationship* – relates the number of species (S) to the area of the plot (gray squares) in which species richness is sampled (A). In the nested SAR, larger areas should be therefore contiguous and should encompass all the smaller areas. However, empirical SARs are often constructed based on much smaller samples, which are assumed to be representative of the whole contiguous and mutually adjacent areas.

Acknowledgements

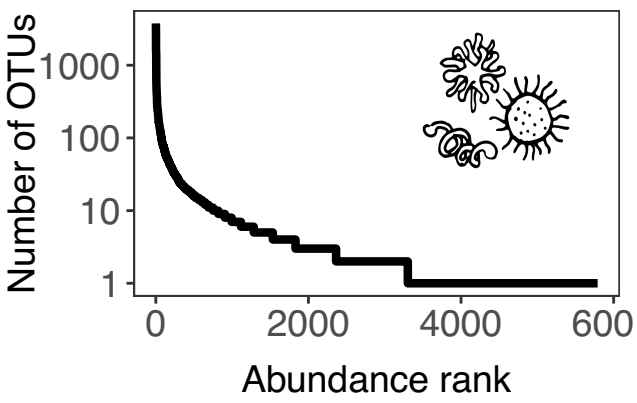
This piece is the product of a “Micro-Macro” working group initiated by JC and KK that met at and was supported by the German Centre for Integrative Biodiversity Research (iDiv) in September 2017. JC acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) FZT 118. AS acknowledges support in part from the National Science Foundation under Grants DEB #1655425 and DEB #1749544, from the USDA National Institute of Food and Michigan State AgBioResearch, and from the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. NJS acknowledges support from a Semper Ardens grant from the Carlsberg Foundation. DS acknowledges support from the Czech Science Foundation (grant no. 16-26369S).

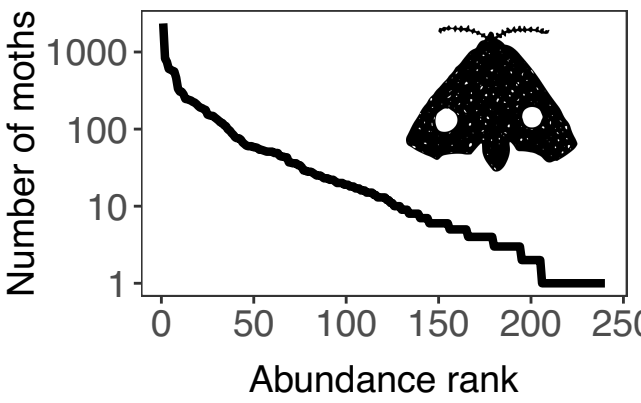
With the exception of the fungi in panels D and E, all inset icons in Fig. 1 are from the Noun Project under CC license: Microbes by Dima Lagunov, moth by Carpe Diem, bird by Ian Graham, fish by Andy Mc, tree by Rayhan Maulana Rikzan, and ant by Cédric Stéphane Touati.

A

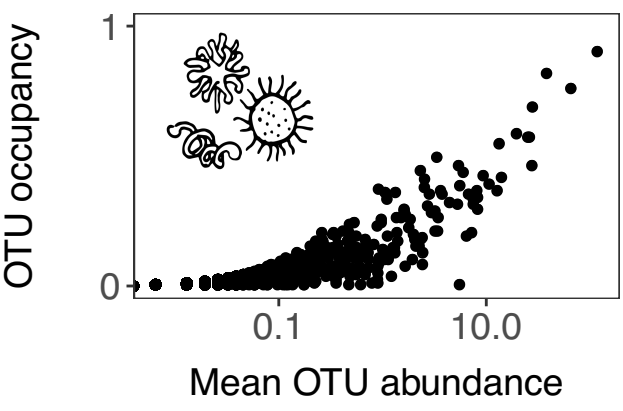
	Sample					
	A	B	C	D	E	F
species 6	0	0	0	0	1	2
species 5	0	1	0	0	0	0
species 4	4	3	3	0	1	2
species 3	25	11	23	8	25	10
species 2	10	19	9	20	10	12
species 1	0	0	0	0	5	6

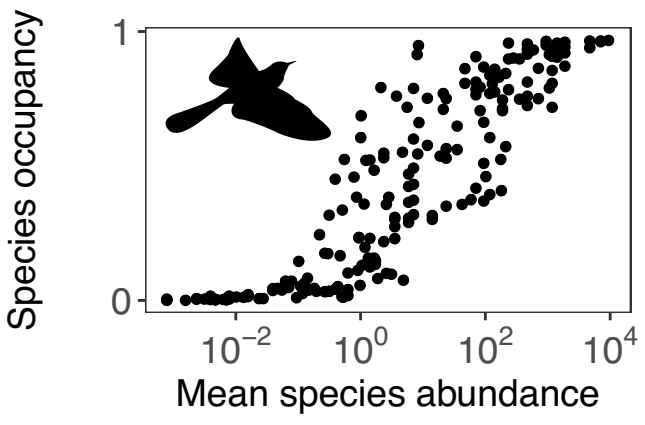
B



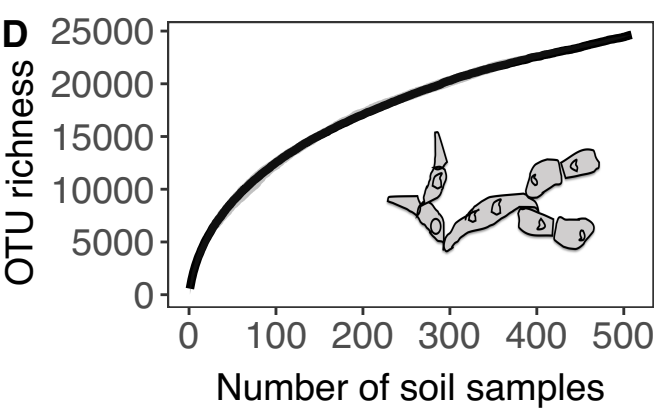


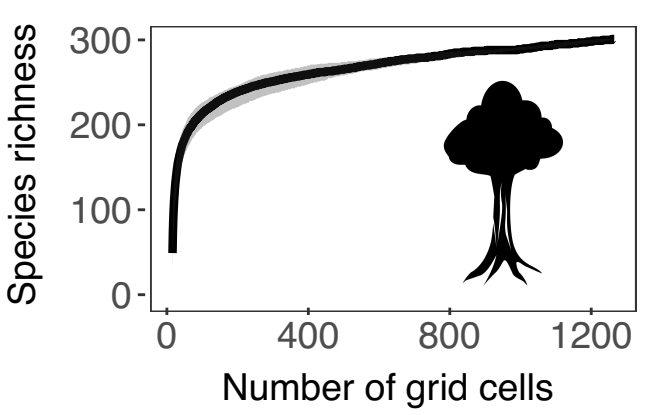
C



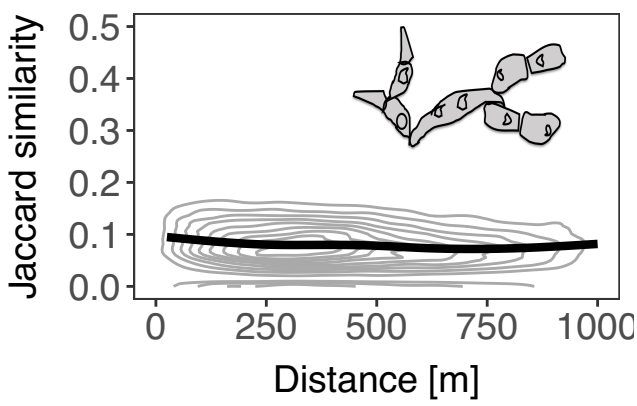


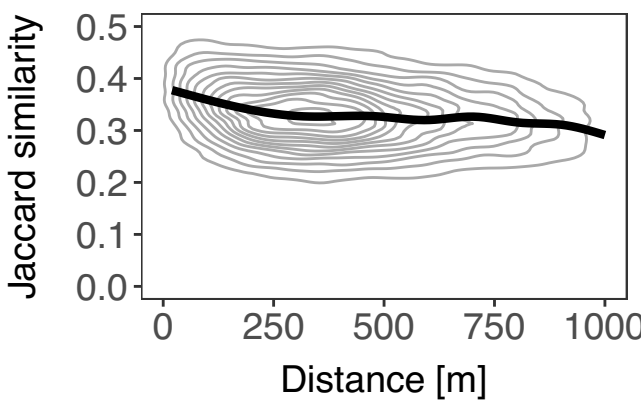
D



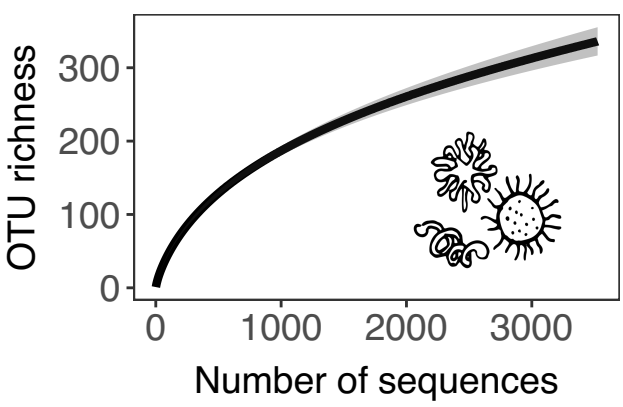


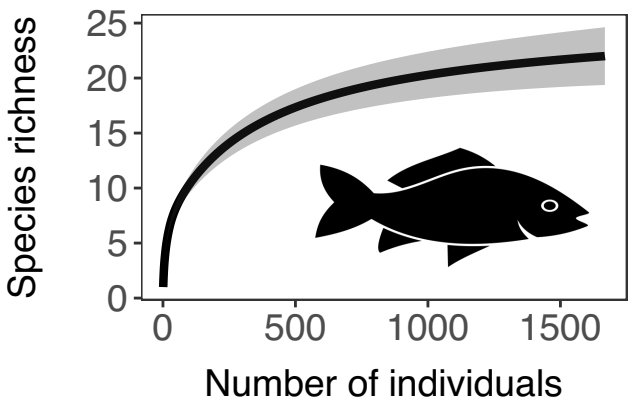
E



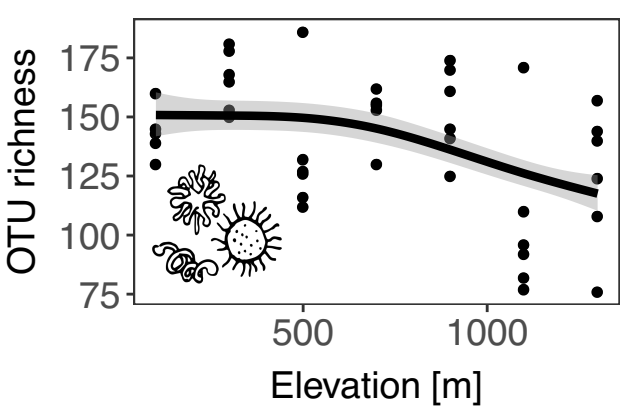


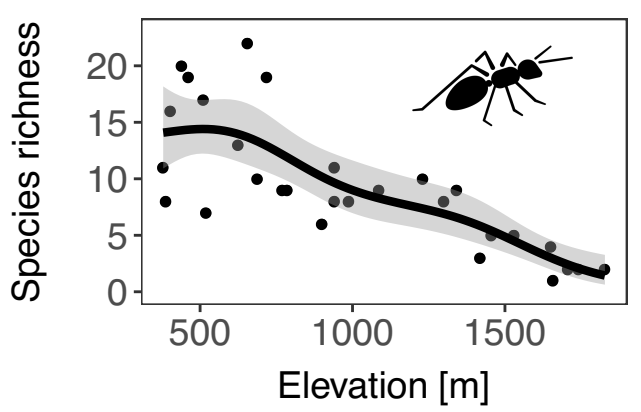
F

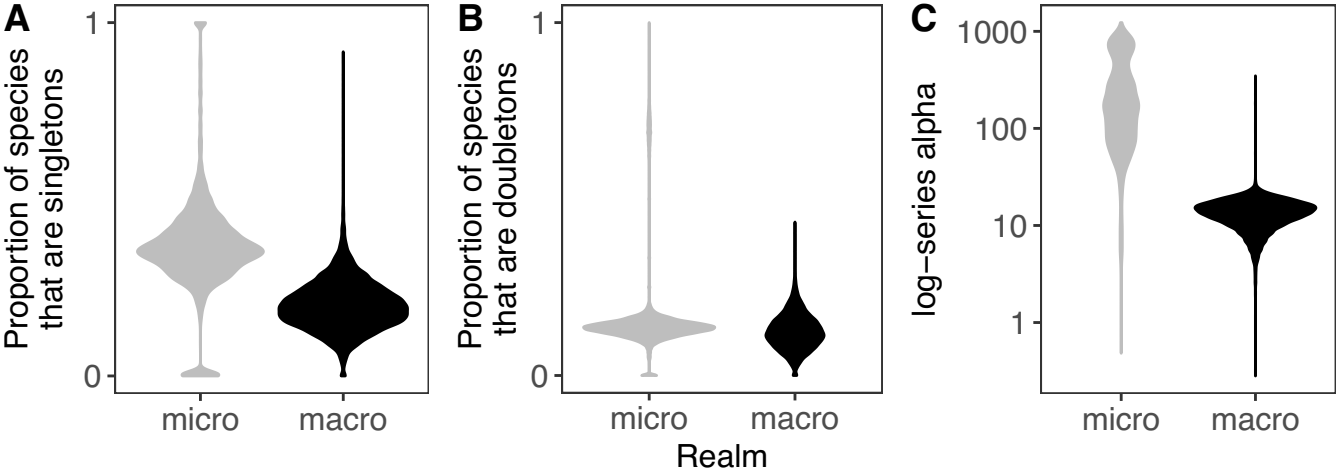




G






A
Proportion of species that are singletons
1
0
micro
macro
B
Proportion of species that are doubletons
1
0
micro
macro
Realm
C
log−series alpha
1000
100
10
1
micro
macro

Macroorganisms		Microorganisms			Comparative Macroecology Relationship	
Pattern	Process	Pattern	Process			
	a		a		Equivalent processes underpin same patterns *Universal*	i.
	a		a		Equivalent processes underpin different patterns *Divergent*	ii.
	a		b		Different processes underpin same patterns *Convergent*	iii.
	a		b		Different processes underpin different patterns *Independent*	iv.

Box 2. Table i. Examples of biases in counting macrobial individuals.

Macrobial community	Challenge in counting the individual
Trees	Seed banks and seedlings less than an arbitrary diameter excluded from surveys; clonal or modular individuals are difficult to distinguish (e.g., *Populus*)
Birds	Arbitrary decisions are made about when and where to count migratory birds
Social insects (e.g., ants, bees)	Trade-off in deciding to practically count individuals versus more precisely count colonies, which are the biological unit on which natural selection acts
Benthic invertebrates	Arbitrary decisions made about mesh size for sieving prior to counting individuals (e.g., all individuals under a certain size are excluded)

Supplementary Material

Macroecology to unite all life, large and small

Ashley Shade[1], Robert R. Dunn[2,3,4], Shane A. Blowes[4], Petr Keil[4], Brendan J.M. Bohannan[5], Martina Herrmann[4,6], Kirsten Küsel[4,6], Jay T. Lennon[8], Nathan J. Sanders[9], David Storch[10,11], Jonathan Chase[4,12]

[1]Departments of Microbiology and Molecular Genetics and Plant, Soil and Microbial Sciences, Program in Ecology, Evolutionary Biology and Behavior, and The Plant Resilience Institute, East Lansing MI USA 48824
[2]Department of Applied Ecology, North Carolina State University
[3] Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen, Denmark
[4] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, Leipzig, Germany.
[5] Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA 97403.
[6] Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University, 07743 Jena, Germany
[8] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA
[9]Environmental Program, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington VT 05405
[10]Center for Theoretical Study, Charles University and the Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Praha 1, Czech Republic (storch@cts.cuni.cz)
[11]Department of Ecology, Faculty of Science, Charles University, Viničná 7, 128 44 Praha 2, Czech Republic
[12]Department of Computer Science, Martin Luther University, Halle-Wittenberg

Corresponding author: Shade, A (shade.ashley@gmail.com)

Supporting Table 1. Studies included in this work in Fig. 1.

Study (as in Fig. 1)	Micro or macro	Taxon	Method used to get the *in situ* data	Details on calculation of patterns in Fig. 1	Number of individuals	Number of taxa	Reference
Groundwater microbes (panel **B** left)	micro	Bacteria	Counts of OTUs based on V3-V4 region of 16S rRNA. Samples were taken from groundwater wells.	Counts of OTUs were simply ordered and plotted.	276,809	16,383	unpublished data provided by M.H. & K.K.
Fisher's moths (panel **B** right)	macro	Moths	Counts of moth individuals in a light trap.	Counts over 4 years were simply ordered and plotted.	15,609	240	[1]
Bacteria and archaea in human belly buttons (panel **C** left)	micro	Bacteria and archaea	Pyrosequencing (Roche 454) of the V4 region of the 16S rRNA gene. 1 sample from each human individual was taken using a cotton tip.	Mean number of occupied humans was plotted against mean abundance across all humans (each dot in the plot is an OTU).	109,910	3,574	Unpublished data provided by R.D. and [2]
Birds of Czech Republic (panel **B** right)	macro	Birds	Counts of observed birds in approx. 10 km x 10 km quadrats in Czech Republic	Mean number of occupied quadrats was plotted against mean abundance across the quadrats (each dot in the plot is a species).	42,771,392	197	[3]
Fungi in 50 ha BCI plot, Panama (panels **D** and E left)	micro	Soil fungi	Counts of OTUs based on ITS1 region of rRNA operon. The samples were 6.25 cm cores, 20 cm deep, each at a center of a 20 x 20 m cell.	The cores were compared using Jaccard similarity index (for the distance decay), and they were aggregated along a spatial proximity gradient to create a spatially-explicit sample-area relationship.	11,147,285	24,666	[4]
Trees in 50 ha BCI plot, Panama (panels **D** and **C** right)	macro	Trees	Counts of trees larger than 1 cm DBH, in 20 m x 20 m cells in a 50 ha forest plot.	Species composition among cells was compared using Jaccard similarity index (for the distance decay). A nested spatial increments were used to calculate the species-area relationship.	235,343	306	[5–7]

Marine microbes in English Channel (panel **F** left)	micro	Bacteria and archaea	Counts of OTUs based on V6 region of 16S rRNA	R package iNEXT was used to calculate the rarefaction curve for a sample taken on 8. Dec 2008.	3526	336	[8]
Marine fishes in Celtic Sea (panel **F** right)	macro	Fish	Counts of fish along 50 x 5 m underwater visual transects (9 sites, 16 transects).	R package iNEXT was used to calculate the rarefaction curve for a sample that is spatially closest to the microbial sample site (row above).	1669	22	[9]
Bacteria in Antarctica (panel **G** left)	micro	Bacteria	Counts OTUs based on 16S rRNA. Samples were taken from 10 cm x 10 cm quadrats, 5 cm deep in soil.	Numbers of OTUs in the samples were plotted against sample elevation.	Not determined	Mean = 139 (SD= 28.8)	[10]
Ants in Great Smoky Mountains, USA (panel **G** right)	macro	Ants	Counts of ants in $1m^2$ quadrats in leaf litter.	Numbers of species were plotted against sample elevation.	5310	45	[11]

Supplemental References

1 Fisher, R.A. *et al.* (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12, 42–58

2 Hulcr, J. *et al.* (2012) A Jungle in There: Bacteria in Belly Buttons are Highly Diverse, but Predictable. *PLoS One* 7,

3 Storch, D. and Šizling, A.L. (2002) Patterns of commonness and rarity in central European birds: Reliability of the core-satellite hypothesis within a large scale. *Ecography (Cop.).* 25, 405–416

4 Barberán, A. *et al.* (2015) Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. *Ecol. Lett.* 18, 1397–1405

5 Hubbell, S.P. (1999) Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science (80-.).* 283, 554–557

6 Hubbell, S.P. *et al.* (2005) , Barro Colorado Forest Census Plot Data. . [Online]. Available: http://ctfs.si.edu/webatlas/datasets/bci

7 Condit, R. (1998) *Tropical Forest Census Plots*, Springer-Verlag and R.G. Landes Company.

8 Caporaso, J.G. *et al.* (2012) The Western English Channel contains a persistent microbial seed bank. *ISME J.* 6, 1089–1093

9 Edgar, G.J. and Stuart-Smith, R.D. (2014) Systematic global assessment of reef fish communities by the Reef Life Survey program. *Sci. Data* DOI: 10.1038/sdata.2014.7

10 Okie, J.G. *et al.* (2015) Niche and metabolic principles explain patterns of diversity and distribution: theory and a case study with soil bacterial communities. *Proc. R. Soc. B Biol. Sci.* 282, 20142630

11 Sanders, N.J. *et al.* (2007) Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. *Glob. Ecol. Biogeogr.* 16, 640–649