Temporal patterns of rarity provide a more complete view of microbial diversity

Ashley Shade^{1*} and Jack A Gilbert^{2,3,4,5}

¹*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824*

²*Institute for Genomic and Systems Biology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439 USA.*

³*Department of Ecology and Evolution, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, U.S.A*

⁴*Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA*

⁵*College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China*

Corresponding author: Shade, A. (shadeash@msu.edu)

Keywords

rare biosphere, time series, next generation sequencing, microbial communities, ecology

Abstract

Recently, conditionally rare taxa (CRT), those taxa that are typically in very low abundance but occasionally achieve prevalence, were shown to contribute to patterns of microbial diversity because their collective dynamics explained a large proportion of temporal variability in microbial community structure. Here, the benefits and challenges of characterizing the presence and interpreting the role of CRT are further explored, along with questions about CRT ecology. We also introduce a conceptual model for thinking about microbial taxa as dynamic components along the dimensions of occurrence and abundance. Accounting for CRT in interpretations of microbial ecological dynamics is essential if we are to understand community stability and eco-evolutionary interactions.

Rare taxa and microbial community dynamics

Conditionally rare taxa (CRT) are those organisms that are usually very rare within a community, but can become prevalent under certain conditions. At any given time in a specific microbial community, CRT are usually at very low abundance (see Glossary) or below the limit of detection, but occasionally achieve higher abundance relative to their associated community members [1] (**Box 1**). The majority of CRT are poorly characterized, as their temporal rarity makes them inherently difficult to observe. Within an observation of a microbial community, a conditionally rare taxon may be prevalent, and so it may be considered to be important for an ecosystem function, yet an appropriate long-term characterization of the environment could reveal that its predicted importance was contextual. The increased availability and cost-efficiency in next generation sequencing tools have provided a framework for improved

observational investigations of ecosystem dynamics, [2] as well as yielded insights into unknown microbial taxa (e.g., microbial dark matter [3]) and the persistently rare biosphere [4]. Using statistical tools on an increasingly diverse series of temporal investigations into microbial dynamics, CRTs can begin to be characterized and investigative paths can be identified to determine the roles of these organisms in ecosystem resilience and stability.

Previously a simple algorithm was developed for detecting rare taxa that occasionally ‘bloom,’ and this algorithm was used to discover that CRT contribute substantially to changes in microbial communities over time (temporal beta diversity), and in a variety of distinct microbial ecosystems [1]. A key observation was that, though CRT comprised 1.5-28% of observed total community membership over time, they accounted for up to 97% of temporal variability in microbial community structure. Also, the majority of CRT were observed not only when they were prevalent, but consistently as a member of the community, providing evidence that they may contribute to the functionally active component of microbial communities. Given the ubiquity of CRT across vastly different habitats, we hypothesized that accounting for these blooming taxa that moved in and out of the ‘rare biosphere’ would provide an ecological mechanism for community resilience and stability, including responses to and recovery from disturbances. We expect that a better understanding of CRT will provide a more complete answer to ecological questions such as: Why are microbial communities so diverse? What are the benefits and trade-offs associated with the maintenance of such diversity? What are the relative contributions of specific ecological mechanisms in maintaining stability?

Abundance-occurrence continuums

Our exploration of the role of CRT in addressing these questions is grounded in the community ecology literature. Ecological definitions of rarity are usually from the perspective of spatial variance in abundance, rather than time [5]. There are dimensions of spatial rarity that have been very elegantly described by Rabinowitz in 1986 [6], and later reinforced by Magurran in 2004 [7]. These spatial dimensions of rarity include local population size, habitat specificity, and geographic distribution. Thus, though a species may be locally prevalent, it could also be very restricted in its habitat requirements and also geographically rare. The conceptual framework of Rabinowitz provided more precision to the use of the term 'rare' in spatial ecology.

Later, Magurran and Henderson described 'occasional' species (1% or less of the total) that contributed to an interesting pattern of species abundance distribution in fish assemblages [8]. Species abundance distributions are statistical models to understand the evenness (equitability of abundances of taxa) of a community (**Figure 1A**). Species abundance distributions are an ongoing topic of debate because different models can be used to infer different ecological or statistical mechanisms (see review by [9]), but several models often fit equally well to the same distribution, confounding efforts to advance interpretation. However, it was from an ocean microbial species abundance distribution that the microbial rare biosphere was first described as the 'long tail' of rare membership [4,10], which remains one of the few generalities among the different analyses and observations of microbial communities over the past decade. In the end, it was concluded that the occasional taxa skewed the fish species abundance distribution, and that if these taxa were separated from the more persistent community, two distinct

distributions could be refined: one for the occasional taxa, which followed a log-series distribution, and one for the persistent taxa, which followed a log normal distribution.

An interesting observation of the Magurran and Henderson work was that the majority of occasional fish taxa were also rare, but not exclusively. There is an expectation that rarity and transience are correlated, however this is not always supported empirically. For example, Resh and colleagues investigated occurrence patterns of 'temporally rare' taxa in 8- and 20-year time series of stream benthic invertebrate communities using an analog of a species abundance distribution for occurrence patterns, called a species occurrence distribution (**Figure 1B**)[11]. They discovered that 17-33% of taxa were only observed in a single year of the full time series. Though the study reported that abundance (measured as density) increased with commonness overall, the densities of some temporally rare taxa were relatively high in the one year that they were observed (see Figures 4 and 5 in[11]).

There are several studies that have explored patterns of microbial rarity in space. For example, rare bacterial taxa have biogeography in the Arctic Ocean [12], which suggests that they are constrained by their environment and are not, as is sometimes assumed, cosmopolitan dispersers. Similar conclusions about biogeographic drivers of rare microbial taxa can be gleaned from an investigation of micrometer-scale patterns in anaerobic sediments of a sulfur-rich spring [13], from a survey of depth-discrete bacterioplankton populations in a meromictic lake [14], and from a comparison of hydrothermal vent microbial communities [15]. All of these recent studies and others provide evidence that rare taxa can exhibit biogeography along

environmental gradients. As observed by Anderson *et al.* [15], in many cases, rare taxa exhibit interesting biogeographic patterns while prevalent taxa appear cosmopolitan.

Here we focus on patterns of microbial rarity over time rather than in space. We present a conceptual model of temporal rarity within a local assemblage (**Figure 1C**). In this model, the first axis is the temporal dynamic along a transience-to-persistence occurrence continuum. The second axis is a continuum of local proportional abundance from rarity to prevalence. Our model can be nested into the existing Rabinowitz model of spatial rarity [6], within the local dimension. We expect that microbial taxa will move between classifications over time. In the particular cases of CRT that were previously discovered, most were persistently rare with occasional forays into the prevalence quadrat (**Figure 1C, ii.**), but ‘one-hit-wonder’ CRT were always transient (**Figure 1C, i.**).

A few microbial ecology investigations have explored the rare/prevalent or transient/persistent components of microbial communities. As part of this, patterns of microbial abundance and occurrence in different systems have been examined. In 2010, Shade and colleagues ([16], see Figures 3 and 4 therein) divided the membership of a lake bacterioplankton into ‘rare’, ‘common’, and ‘dominant’ abundance components defined using Ueland and Gray’s recommendation from the upper and lower quartiles of the species abundance distribution [17], and further into ‘transient’ and ‘persistent’ occurrence components defined using an arbitrary cut-off at the median number of occurrences over the time series. Each group had a distinct response to water column mixing, and the rare/transient and dominant/persistent

groups were likely driving the overall community response to the mixing. This work later inspired a conceptual framework for defining a 'core' microbiome using occurrence patterns within time series to find and eliminate those taxa that are not consistently detected [18]. This framework was then applied to a flower microbiota time series to discover that apple flower communities contain many members that are transient and in low abundance, and do not contribute to a successional pattern [19]. Rather, persistent members that fluctuated in their relative abundances drove the successional pattern; some of these taxa meet the definition of CRT within the context of primary succession.

Many other groups also have been tackling the question of the roles of abundance and occurrence for microbial community dynamics. Caporaso *et al.* investigated the persistent and transient components of time series of human-associated microbiota and classified them by their median abundances [20]. The authors found that persistent and transient groups overlapped substantially in their class-level membership, and noted that there was no evidence supporting a persistent 'temporal core' microbiome. Vergin *et al.* showed that the occurrence patterns of many rare taxa in the Sargasso Sea correlated with seasonal drivers [21], which suggests that some rare taxa may have dynamics that are predictable with expected changes in their environment. Also, if synthesized across a variety of studies, there is some evidence that CRT dynamics may vary substantially by ecosystem type. For example, Gobet *et al.* dissected discrete temporal patterns of members in coastal sand communities sampled six times over 13 months, and discovered that the rare taxa did not often achieve abundance [22]. Meanwhile, Kim *et al.* found that rare taxa were less stable (fluctuated more over time) than the common

taxa in an activated sludge community [23]. These studies, and others, collectively suggest that a persistent but rare component of many microbial communities exists, but that more work is needed to understand the ecology of their dynamics.

Benefits of and challenges to considering CRT

A moving pictures perspective of microbial consortia, those that include time series and longitudinal host–microbiota series, is more complete than a snapshot of a single moment of an assemblage. Thus, there has been a generation of more and longer time series datasets (e.g., [20,24–28]) and beginnings of synthesis of microbial time series [29], which will improve our ability to discover CRT, describe their dynamics, and determine the mechanisms underlying their patterns. Accounting for CRT among existing time series will help to better define core microbiota by allowing us to remove transient and rare taxa that do not contribute to deterministic community dynamics. It will also improve our ability to detect keystone taxa, those that are not highly abundant, but provide an essential function to the community.

Placing a scientific value on accounting for CRT will change the way we analyze and interpret next-generation sequencing data in microbial ecology. First, there are many technical hurdles and ambiguities to observing and quantifying the rare members of communities assessed using high-throughput tag-sequencing (e.g., [30]), and overcoming these hurdles will improve sensitivity and accuracy of CRT discovery. For example, some studies report that PCR biases communities against observing rare members [31], while others report that some high-

throughput sequencing technologies result in an overestimate of richness as a result of noise and errors (e.g., [32–34]). Furthermore, during short read sequence processing, it is common practice to remove some proportion of low-abundance taxa at an arbitrary threshold of abundance. This is done to reduce noise/enhance signal and to simplify the community to improve analysis efficiency. Arguably, this practice reduces our ability to observe CRT and synthesize across studies because it is likely that some rare taxa that are omitted may become prevalent at future time points. Also, alternatives to operational taxonomic unit definitions (e.g., oligotyping or subpopulations [35–37]) will improve our ability to maintain high quality sequencing data and avoid artifacts while defining ecologically relevant taxonomic units. Additionally, there are several tools available and in development for discovering novel and rare taxa in high-throughput sequencing datasets (e.g., [38,39]) and these will improve CRT discovery.

Another challenge to understanding CRT is the limitation in how available microbial time series studies were designed. Time series and host-associated longitudinal studies should be designed to have study durations and sampling intensities over the dynamic range of important drivers of microbial community structure. For example, temperate soil microbial communities should be observed to encompass annual seasonal cycles, at sampling intensities relevant to average growth or community turnover in the soil. Collecting well-designed time series will improve documentation of persistent CRT because including the environmental and biological fluctuations that may drive CRT dynamics increases the likelihood that CRT are observed. This will allow for better distinction between persistent and transient CRT and describe the dynamic

species abundance and species occurrence distributions, which will help inform ecological roles and future research into metabolic mechanisms of CRT (e.g., [40]). For example, well-designed time series will be important for distinguishing successful immigration events from blooms from persistent seed banks. In general, time series that are designed to encompass periods of primary succession, secondary succession, and eco-evolutional scales are needed to provide ecological context to CRT dynamics (**Box 2, Figure 1.**).

Finally, as suggested by Resh and colleagues in their benthic invertebrate study [11], ecologists have been indirectly observing temporal rarity by asking questions about its complement: persistence. There are comparably many more studies about persistence of microbial community members (e.g., [41]), and so it is possible that we could glean more about conditionally rare taxa by synthesizing what has been documented as far as persistence. Given the ever-expanding scientific interest in core microbiome studies and their extension to sometimes include a longitudinal series, host-associated ecosystems may be a prime target for meta-analyses of persistence.

Interesting questions

Moving forward, there remain many unanswered, interesting questions about CRT, their places in the microbial community abundance-occurrence continuum (**Figure 1C**), and their roles in maintaining microbial community diversity and stability (**Box 3**). First and perhaps most obviously, empirical evidence is needed from well-designed time series to understand how often-rare taxa remain rare, common taxa remain common, and prevalent taxa remain

prevalent. We must query both dimensions of the abundance-occurrence continuum and recognize that many taxa will be dynamic in both dimensions. Then, it will be useful to assign the observed CRT dynamics to either stochastic or deterministic processes, which can be investigated in part by fitting the appropriate models. Though both are ecologically interesting, those CRT dynamics that are deterministic are arguably more useful for predicting community changes. However, several studies, including the classic work of Fernández *et al.* [42], show that temporal changes in microbial community structure are not always deterministic, and could instead be stochastic or chaotic. Thus, if possible, CRT should be interpreted within the context of the community trajectory. For example, CRT patterns within a stochastic or chaotic trajectory could lead to hypotheses about functional redundancy among CRT [43]. Ultimately, querying both dimensions of abundance and occurrence will also provide direct insight into the roles of CRT for microbial community resilience and stability.

After CRT are better documented, it will be possible to ask whether there are common ecological traits of taxa that exhibit the rare-to-prevalent dynamic, and if these traits are different than those of common taxa. Kunin and Gaston critiqued the notion that rare ‘macrobial’ taxa have distinct biological traits, and proposed that rare taxa must remain rare over evolutionary time scales for adaptation and evolution to be substantiated [44]. Over ten years later, Resh and colleagues presented empirical evidence for different traits among temporally rare and common benthic invertebrates [11]. However, many of the benthic invertebrate traits found to be important are not directly transferable to microorganisms (e.g., aquatic adults and aerial respiration). Though attempts were made to explore evidence of

shared ecological traits among CRT by asking whether closely related lineages were represented across ecosystems (thereby suggesting shared traits among related taxa, which is not universally true), the data were too limited to draw conclusions [1]. We hypothesize that dormancy strategies may be a common trait of microbial CRT. Other CRT traits to investigate include copiotrophy, stress tolerance, multiple life strategies (e.g., planktonic and attached/biofilm), and opportunistic pathogenesis.

Another relatively unexplored question is whether CRT achieve social behavior. Density-dependent behaviors, such as quorum sensing, are important for emergent properties such as biofilm production. Are the transitions of persistent taxa from the rare biosphere to prevalence driven by group behavior (e.g., collective awakening from dormancy [45]) or by the tremendous success of one individual? Are quorum sensing or other signaling pathways of rare taxa more sensitive or specific than those of relatively common taxa? These questions can best be addressed if rare taxa and their extracellular products can be queried in the laboratory.

From an eco-evolutionary perspective, it is also interesting to ask how genetic variation among CRT compares to that of persistent and prevalent taxa. This has implications for thinking about the how microorganisms evolve to interact with each other or with a host, or to adapt to changes in the environment. A hypothesis that gained popularity in the 1990s was that rare plants would have less genetic variation than common plants, which presented a conservation challenge for threatened species. However, some studies found equivalent levels of genetic variation among geographically restricted (rare) and widespread plant species, while others

reported that widespread species have more genetic variation (e.g., [46,47]). Investigating temporal dimensions of rarity in the context of these hypotheses, and then using microbial communities to test them could be advantageous, as relatively quick microbial generation times allow for these questions to be addressed across evolutionary time scales. To add yet another point of intrigue, community eco-evolutionary dynamics are often driven by interactions with phages [48]. Understanding phage host-specificity for CRT, their potential roles in creating CRT genetic variation, and their contribution to CRT population crashes will likely provide much insight into the mechanisms and maintenance of CRT within microbial communities.

Another general unknown is the form and consistency of species occurrence distributions across communities. We suggest that the shape of species occurrence distributions and their best-fitted models (**Figure 1B**) will depend on whether a community is observed over primary succession or temporally stable state. Thoughtfully observed microbial community time series could contribute empirical data to, and advance theory within, this active area of ecological research.

Concluding remarks

Accounting for the dynamic patterns of rare taxa will only improve our understanding of the ecology of microbial communities. Recognizing that many, if not most, members of microbial communities exhibit abundance changes over time will help to move microbial community

282 ecology from static laundry lists of taxa to dynamic models that will allow us to better predict,
283 manage, and remodel microbial consortia.

284

285 **Acknowledgements**

286 This work was supported by Michigan State University. JAG was generously supported by the
287 Alfred P Sloan foundation. This work was supported in part by the U.S. Dept. of Energy under
288 Contract DE-AC02-06CH11357.

289 **References**

- 290 1 Shade, A. *et al.* (2014) Conditionally rare taxa disproportionately contribute to temporal
291 changes in microbial diversity. *MBio* 5, e01371–14
- 292 2 Knight, R. *et al.* (2012) Unlocking the potential of metagenomics through replicated
293 experimental design. *Nat. Biotechnol.* 30, 513–520
- 294 3 Rinke, C. *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark
295 matter. *Nature* 499, 431–437
- 296 4 Sogin, M.L. *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare
297 biosphere.” *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120
- 298 5 Preston, F.W.F. (1948) The commonness, and rarity, of species. *Ecology* 29, 254–283
- 299 6 Rabinowitz, D. *et al.* (1986) Seven forms of rarity and their frequency in the flora of the
300 British Isles. In *Conservation Biology: The Science of Scarcity and Diversity* (Soulé, M. E.,
301 ed), Sinauer Associates, Inc
- 302 7 Magurran, A.E. (2004) *Measuring biological diversity*, Blackwell Science Ltd.
- 303 8 Magurran, A.E. and Henderson, P.A. (2003) Explaining the excess of rare species in
304 natural species abundance distributions. *Nature* 422, 714–716
- 305 9 McGill, B.J. *et al.* (2007) Species abundance distributions: Moving beyond single
306 prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–
307 1015
- 308 10 Pedrós-Alió, C. (2012) The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* 4, 449–466
- 309 11 Resh, V. *et al.* (2005) How common are rare taxa in long-term benthic macroinvertebrate
310 surveys? *J. North Am. Benthol. Soc.* 24, 976–989
- 311 12 Galand, P.E. *et al.* (2009) Ecology of the rare microbial biosphere of the Arctic Ocean.
312 *Proc. Natl. Acad. Sci. U. S. A.* 106, 22427–22432
- 313 13 Youssef, N.H. *et al.* (2010) Fine-scale bacterial beta diversity within a complex ecosystem
314 (Zodletone Spring, OK, USA): The role of the rare biosphere. *PLoS One* 5,
- 315 14 Gies, E.A. *et al.* (2014) Illuminating Microbial Dark Matter in Meromictic Sakinaw Lake.
316 *Appl. Environ. Microbiol.* 80, 6807–6818

317 15 Anderson, R. *et al.* (2014) Biogeography and ecology of the rare and abundant microbial
318 lineages in deep-sea hydrothermal vents. *FEMS Microbiol. Ecol.*
319 <http://dx.doi.org/10.1093/femsec/fiu016>

320 16 Shade, A. *et al.* (2010) Seasonal and episodic lake mixing stimulate differential planktonic
321 bacterial dynamics. *Microb. Ecol.* 59, 546–554

322 17 Uglund, K.I. and Gray, J.S. (2011) Lognormal distributions and the concept of community
323 equilibrium. *Oikos* 39, 171–178

324 18 Shade, A. and Handelsman, J. (2012) Beyond the Venn diagram: the hunt for a core
325 microbiome. *Environ. Microbiol.* 14, 4–12

326 19 Shade, A. *et al.* (2013) Unexpected diversity during community succession in the apple
327 flower microbiome. *MBio* 4, e00602–12

328 20 Caporaso, J.G. *et al.* (2011) Moving pictures of the human microbiome. *Genome Biol.* 12,
329 R50

330 21 Vergin, K.L.K. *et al.* (2013) Spatiotemporal distributions of rare bacterioplankton
331 populations indicate adaptive strategies in the oligotrophic ocean. *Aquat. Microb. Ecol.*
332 71, 1–13

333 22 Gobet, A. *et al.* (2012) Diversity and dynamics of rare and of resident bacterial
334 populations in coastal sands. *ISME J.* 6, 542–553

335 23 Kim, T.S. *et al.* (2013) General and rare bacterial taxa demonstrating different temporal
336 dynamic patterns in an activated sludge bioreactor. *Appl. Microbiol. Biotechnol.* 97,
337 1755–1765

338 24 Flores, G. *et al.* (2014) Temporal variability is a personalized feature of the human
339 microbiome. *Genome Biol.* 15, 531

340 25 Gilbert, J.A. *et al.* (2012) Defining seasonal marine microbial community dynamics. *ISME*
341 *J.* 6, 298–308

342 26 Grice, E.A. *et al.* (2009) Topographical and temporal diversity of the human skin
343 microbiome. *Science* 324, 1190–1192

344 27 Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. , *Science*
345 *Translational Medicine*, 4. (2012) , 132ra52–132ra52

346 28 Faith, J.J. *et al.* (2013) The long-term stability of the human gut microbiota. *Science* 341,
347 1237439

- 348 29 Shade, A. *et al.* (2013) A meta-analysis of changes in bacterial and archaeal communities
349 with time. *ISME J.* 7, 1493–1506
- 350 30 Brown, S.P. *et al.* (2015) Scraping the bottom of the barrel: are rare high throughput
351 sequences artifacts? *Fungal Ecol.* 13, 221–225
- 352 31 Gonzalez, J.M. *et al.* (2012) Amplification by PCR artificially reduces the proportion of the
353 rare biosphere in microbial communities. *PLoS One* 7, e29973
- 354 32 Quince, C. *et al.* (2011) Removing noise from pyrosequenced amplicons. *BMC*
355 *Bioinformatics* 12, 38
- 356 33 Amend, A.S. *et al.* (2010) Quantifying microbial communities with 454 pyrosequencing:
357 Does read abundance count? *Mol. Ecol.* 19, 5555–5565
- 358 34 Kunin, V. *et al.* (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to
359 artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123
- 360 35 Eren, A.M. *et al.* (2014) Oligotyping analysis of the human oral microbiome. *Proc. Natl.*
361 *Acad. Sci. U. S. A.* 111, E2875–E2884
- 362 36 Tikhonov, M. *et al.* (2014) Interpreting 16S metagenomic data without clustering to
363 achieve sub-OTU resolution. *ISME J.* DOI: 10.1038/ismej.2014.117
- 364 37 Preheim, S.P. *et al.* (2013) Distribution-based clustering: using ecology to refine the
365 operational taxonomic unit. *Appl. Environ. Microbiol.* 79, 6593–6603
- 366 38 Lynch, M.D.J. *et al.* (2012) Targeted recovery of novel phylogenetic diversity from next-
367 generation sequence data. *ISME J.* 6, 2067–2077
- 368 39 Gagic, D. *et al.* (2014) Improving the genetic representation of rare taxa within complex
369 microbial communities using DNA normalisation methods. *Mol. Ecol. Resour.* DOI:
370 10.1111/1755-0998.12321
- 371 40 Samo, T.J. *et al.* (2014) Broad distribution and high proportion of protein synthesis active
372 marine bacteria revealed by click chemistry at the single cell level. *Front. Mar. Sci.* 1, 1–
373 18
- 374 41 Caporaso, J.G. *et al.* (2012) The Western English Channel contains a persistent microbial
375 seed bank. *ISME J.* 6, 1089–1093
- 376 42 Fernández, A. *et al.* (1999) How stable is stable? Function versus community
377 composition. *Appl. Environ. Microbiol.* 65, 3697–3704

378 43 Allison, S.D. and Martiny, J.B.H. (2008) Resistance, resilience, and redundancy in
379 microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11512–11519

380 44 Kunin, W. and Gaston, K. (1993) The biology of rarity: patterns, causes and
381 consequences. *Trends Ecol. Evol.* 8, 298–301

382 45 Epstein, S.S. (2009) Microbial awakenings. *Nature* 457, 1083

383 46 Huenneke, L.F. (1991) Ecological implications of genetic variation in plant populations. In
384 *Genetics and Conservation of Rare Plants* (Falk, D. A. and Holsinger, K. E., eds), pp. 31–44,
385 Oxford University Press, Inc

386 47 Cole, C.T. (2003) Genetic variation in rare and common plants. *Annu. Rev. Ecol. Evol. Syst.*
387 34, 213–237

388 48 Wilmes, P. *et al.* (2009) The dynamic genetic repertoire of microbial communities. *FEMS*
389 *Microbiol. Rev.* 33, 109–32

390 49 Olivas, A.D. *et al.* (2012) Intestinal tissues induce an SNP mutation in *Pseudomonas*
391 *aeruginosa* that enhances its virulence: possible role in anastomotic leak. *PLoS One* 7,
392 e44326

393

394

395 **Boxes**

396 **Box 1. Defining temporal rarity**

397 The scientific community has yet to agree upon a consistent definition of rarity, and this is
398 because rarity can have multiple spatial and temporal contexts. Thus, definitions of rarity are
399 often arbitrary. However, within a body of work, the applied definition of rarity should be
400 ecologically defensible. We provide some considerations for defining temporal rarity for
401 microbial communities. We encourage researchers to explore a range of rarity cut-offs to their
402 community time series with these considerations in mind, and to justify their ultimate decision.

- 403 1. How taxonomically diverse is the community? More diverse communities will have
404 more taxa, which will proportionally diminish the contribution that any one taxon can
405 have to the total richness.
- 406 2. How comprehensively was the community observed? Using sequencing technologies,
407 many complex microbial communities (e.g., soils, guts) fall short of reaching saturation
408 with standard sequencing efforts. The closer the community is to saturation, the more
409 confidence can be placed in the ability to estimate overall richness and the contribution
410 of rare members to that richness.
- 411 3. If applicable, what sequencing technology was used, and how have the known errors
412 and biases in the method been addressed in both the data handling and in the
413 interpretation of patterns? In particular, this consideration is important when
414 interpreting singleton and doubleton reads as rare community members.
- 415 4. Has the community been observed over ecologically meaningful temporal scales? The
416 frequency and extent of sampling should be relevant to the community turnover and

generation times and informed by the expected maximum environmental fluctuations in the ecosystem.

5. Could species abundance distribution (SAD) models inform an improved definition of rarity for the community? Fitting SADs to microbial datasets is not yet mainstream practice, but considering these models and their assumptions can inform appropriate cut-offs for and interpretations of the rare community members.

Box 2. Ecological interpretation of CRT requires context

What may drive microbial transitions into and out of the rare biosphere? Previously several reasons have been mentioned about how a rare-to-prevalent dynamic may be manifested, including awakening from dormancy, escape from predation, and release from competition [1]. However, these mechanisms fall within the ecological and evolutionary drivers of community change, and interpreting CRT dynamics depends on the community context (**Figure 1**). For example, CRT discovery during primary succession should not be interpreted the same as CRT discovery during secondary succession. The first scenario falls within the context of community assembly processes, while the second falls within the context of ‘mature’ community variability. In the first scenario, we would expect to observe ‘one-hit-wonder’ CRT as pioneer lineages that are outcompeted and potentially replaced by subsequent colonizers. In the second scenario, we may hypothesize that persistent CRT are exhibiting either stochastic or deterministic changes, and that deterministic changes should be linked to environmental or biological drivers. However, it is also likely that a combination of both factors plays a role in CRT dynamics under both primary and secondary successional systems. If we can well-observe these drivers, we should improve our ability to document CRT and understand their ecological roles.

Another distinct scenario is when a community member evolves into two divergent lineages, which may manifest as an initially undetected CRT that becomes persistent. These are by their definition difficult to identify and define. A recent study that described the putative induction of a specific SNP mutation in a strain of *Pseudomonas aeruginosa*, provides potential evidence of this event [49]. Prior to gastrointestinal surgery only one genotype of *P. aeruginosa* could be

observed, following surgery the vast majority of the community comprised a new genotype with only a single SNP in the genome. This new genotype had enhanced virulence and converged on the damaged tissue [49]. Though there is less known about eco-evolutionary dynamics within other community contexts, characterizing the dynamics of CRT may help us to document and better investigate the implications of member speciation and other more specific adaptations for community structure, dynamics, and function.

Box 3. Outstanding questions

- What are the contributions of CRT to maintaining community stability, especially after disturbance?
- What is the variability in both dimensions of microbial abundance and occurrence?
- Do CRT have ecological traits that are different than those of prevalent taxa? How consistent are CRT traits across different ecosystems and within different phylogenetic lineages?
- Do CRT exhibit specialized social behavior (e.g., quorum sensing)?
- Is the genetic variation of CRT less than that of prevalent taxa?
- What is the relationship between CRT and phages for driving rare-to-prevalent dynamics?
- What forms do species occurrence distributions take, and how do the dynamics of rare taxa shape these distributions?

Glossary

Abundance: the numerical count or relative contribution of a taxon to a community observation. Includes a continuum (from most to least abundant) of dominant, prevalent, and rare taxa.

Dominant: the most numerically abundant taxon (or highest relative contributor) to a single community observation.

Occurrence: the presence of a taxon (based on presence/absence information) over a series of community observations. Includes a continuum (from high to low occurrence) of persistent to transient taxa.

Persistent: a taxon that is observed frequently over a time series of community observations.

Prevalent: a taxon that has high abundance in a single community observation.

Rare: a taxon that has low abundance in a single community observation. Spatial dimensions of rarity examine the biogeographic distributions of rare taxa, and include local population size, habitat specificity, and geographic distribution [6].

Transient: a taxon that is observed infrequently over a time series of community observations.

Figure Legends

Figure 1. Dynamic species abundance and species occurrence continuums provide insight into microbial community structure. (A) Hypothetical species abundance distribution for a microbial community, where the majority of membership comprises rare taxa. (B) Hypothetical species occurrence distribution for a microbial community, where the majority of occurrences are transient taxa. Percent occurrence is calculated as the number of time points in which a taxon is observed over the total time points in the study. (C) An overlay of species abundance and species occurrence for microbial communities. Persistent but rare taxa are shaded in green, and transient and rare taxa are shaded in orange. Blue arrows are important taxa transitions that contribute to conditionally rare patterns: (i.) is the transition of transient rare taxa to persistence, and (ii.) is the transition of persistent rare taxa to prevalence, and back again.

Figure I. Three generalized scenarios for exploring and interpreting CRT patterns. Each line represents a different taxon. (A) Primary succession. Different taxa bloom consecutively during community assembly of a 'blank slate' habitat. (B) Secondary succession and routine variability. Taxa transitions along species abundance and species occurrence continuums are manifested as collective community variability around a mean community structure. (C) Eco-evolutionary dynamics. In this example, a prevalent taxon (i.) speciates into two related taxa (ii. and iii.), which may exhibit conditionally rare patterns as they transition from previously undetected to prevalent.

Figure 1

Figure I