

University of Groningen

Soil bacterial community assembly during succession

Jia, Xiu

DOI:
[10.33612/diss.156586683](https://doi.org/10.33612/diss.156586683)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Jia, X. (2021). *Soil bacterial community assembly during succession*. University of Groningen.
<https://doi.org/10.33612/diss.156586683>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 4

Community assembly processes of the microbial rare biosphere

Xiu Jia, Francisco Dini-Andreote, Joana Falcão Salles

Abstract

Our planet teems with microorganisms that often present a skewed abundance distribution in a local community, with relatively few dominant species coexisting alongside a high number of rare species. Recent studies have demonstrated that these rare taxa serve as limitless reservoirs of genetic diversity, and perform disproportionate types of functions despite their low abundances. However, relatively little is known about the mechanisms controlling rarity and the processes promoting the development of the rare biosphere. Here, we propose the use of multivariate cut-offs to estimate rare species and phylogenetic null models applied to predefined rare taxa to disentangle the relative influences of eco-evolutionary processes mediating the assembly of the rare biosphere. Importantly, the identification of the factors controlling rare species assemblages is critical for understanding the types of rarity, how the rare biosphere is established, and how rare microorganisms fluctuate over spatiotemporal scales, thus enabling prospective predictions of ecosystem responses.

Published in *Trends in Microbiology*, 2018, 26(9): 738-747

doi: 10.1016/j.tim.2018.02.011

The copious rare taxa in microbial communities

Advances in high-throughput sequencing technologies have provided extensive information of the vast majority of uncultured microbes across disparate systems (Locey and Lennon 2016). Despite technical limitations hampering a realistic molecular characterization of microbiomes, such as the distinction between truly living cells from relic DNA (Carini et al. 2016) and artificial methodological biases (e.g. chimeras generated by PCR and artificial taxa originated from *de novo* operational taxonomic unit (OTU) binning algorithms), it is common knowledge that microbial communities harbour a large number of low-abundant taxa, coexisting with a small fraction of dominant ones (Ashby et al. 2007, Nemergut et al. 2011). The fraction composed of low-abundant taxa is termed as the microbial ‘rare biosphere’ (Sogin et al. 2006).

Box 4.1 Highlights

- Increasing sequencing depth efforts and the number of replicates, and performing standardized benchmarking analyses, will improve the resolution of rare biosphere dynamics across spatiotemporal scales.
- The rare biosphere dominates microbial communities, and often accounts for important ecosystem services. Technical procedures limit our understanding of the rare biosphere.
- The detection of rare taxa should be based on multivariate cut-offs, by removing rare OTUs until removal influences community similarity.
- Different types of rarity are structured by interplay in community assembly processes. Determining assembly processes for rare species is essential to understand if they differ from processes governing abundant microbes.
- Understanding the ecological processes structuring each type of rarity is essential to know how the rare biosphere might respond to environmental changes.

While the ‘hidden’ diversity that constitutes the rare biosphere is often neglected due to low sequencing coverage and/or singletons removal step during data processing, it can influence community structure and play essential roles in community functioning and stability (Jousset et al. 2017). For example, Gobet et al. (2012) showed that, after removing up to 50% of the rare species from their dataset, the previously significant turnover in community composition was no longer observed, indicating that rare species significantly account for shifts in β -diversity. In addition, the rare biosphere serves as a reservoir of species — greatly contributing to the quantification of species richness within a given community — and functional genes. Both the number of species (phyloge-

netic reservoir) and the genetic diversity (functional potential) of these species provide an ‘ecological insurance’ for the community by contributing to important ecosystem processes and services. For example, using dilution-to-extinction approach in soil microcosms, van Elsas et al. (2012) found that removal of rare species leads to a negative relationship between microbial diversity and invasion. Their results provide evidence that rare species contribute to ecosystem stability, where more diverse systems hinder the establishment of an alien species through resource competition (Mallon et al. 2015). Moreover, given the fact that the rare biosphere is composed of functionally relevant taxa, changing local conditions can lead to fluctuations in abundance of hitherto rare taxa, which might influence ecosystem functioning. A recent study on lake ecosystems demonstrated that the presence of an organic pollutant led to a substantial increase in the abundance of rare microbial taxa capable of decomposing the pollutant (Wang et al. 2017). Finally, environmentally active rare taxa can have a disproportionately large functional effect in ecosystem. For instance, Pester et al. (2010) showed that *Desulfosporrosinus*, despite comprising less than 0.006% of the total microbial community, had a fundamental role in sulphate reduction in peatland soils.

Although the impact of the rare biosphere on ecosystem services is directly linked to the types of rarity — that is, how rare organisms fluctuate over time and space — we currently lack technical and theoretical frameworks that allow for a proper categorization of the rare biosphere, both in terms of the threshold of rarity and which types of rarity dominates in a given community. Moreover, we argue that understanding the processes mediating the community assembly of rare taxa is essential to advance fundamental knowledge of the rare biosphere. In this opinion article, we provide a synthesis on how statistical and ecological strategies can be applied to disentangle the assembly of the rare biosphere, under the assumptions that each type of rarity is driven by a specific interplay of known ecological and evolutionary processes.

The types of rarity

In general (macro-organismal) ecology, Rabinowitz et al. (1986) provided seven distinct categories of rarity based on the population sizes in a local community, habitat specificity and geographic distributions. In this context, only species with broad geographic distribution and habitat tolerance, as well as large local abundance, are not considered rare. Thus, the great majority of species that does not fit in this categorization will be classified into one of the types of rarity (Magurran and Henderson 2011). Moreover, the rare biosphere has both spatial and temporal dimensions. Therefore, the identification of rare species will depend on the location and frequency of sampling, as well as sampling effort and methodology (Magurran and Henderson 2011). This spatiotemporal concept allows for the delineation of two distinct types of rarity: (i) species that are rare through space within a given temporal window (spatially rare), and (ii) species that are rare through time within a given spatial domain (temporally rare).

Firstly, rarity can be characterized based on low population abundances within a local community, where rare species are represented by a long tail in rank–abundance

curves. Second, rarity can also be defined by habitat specificity, as some taxa with specific functions and metabolisms can thrive only in specific niches. Here, a strong role in abiotic selection is expected, as is the case for microbial extremophiles occurring in harsh environments under extreme temperature, radiation, pressure, salinity and pH (Rothschild and Mancinelli 2001). Third, rarity can be associated with taxa occurring at narrow geographic ranges, as a result of environmental selection and/or dispersal limitation (Sul et al. 2013). Temporal rarity can be categorized as *(i)* conditionally rare, that is, those taxa that stay at low abundances but become prevalent under specific circumstances (Shade et al. 2014), or *(ii)* permanently rare, that is, those taxa that occur at low abundances regardless of spatial or temporal scale (Lynch and Neufeld 2015). The persistence of these permanently rare organisms within a community is likely associated with particular life-history strategies and ecological trade-offs, whereas traits such as dormancy and sporulation might be relevant for conditionally rare taxa as they favor large-scale dispersal and longer temporal persistence under unfavorable environmental conditions. Although the current definition of rarity applies to both macro and microbial ecology, identifying the types of rarity in microbial communities remains a major challenge, in particular given the sheer number of species and methodological biases associated with it (see next section for details). For these reasons and other methodological constraints, we advocate that the microbial rare biosphere should be investigated in a community context rather than individual species as is usually applied in macro-organismal ecology. In doing so, the distinctiveness in the types of rarity and the processes underpinning them can be investigated using the fundamentals of community ecology.

Categorizing ‘the rare biosphere’ in microbial ecology

Despite advances in defining rarity in general ecology (Rabinowitz et al. 1986, Magurran and Henderson 2011), the operational definition of the rarity in microbial ecology is still challenging. The current sequencing coverages achieved in community profiling often fail to fully describe microbial communities, thus leading to insufficient detection and therefore an underestimation of rare microbial taxa (Zhou and Ning 2017). In addition, as reference databases do not encompass the diverse magnitude of microbial species, the use of reference-based OTU binning is known to exclude a large portion of rare taxa. Lastly, microbial ecologists also have to contend with the fact that taxa abundances obtained by community profiling are relative, rather than absolute values, which largely influences the identification of rare taxa (Weiss et al. 2015).

In addition to these issues, we still lack a unified method to distinguish rare from abundant taxa. Currently, artificial thresholds are often used to delineate rare organisms — for example, species with relative abundances below 0.01% or even 0.001% are placed within the rare biosphere (Galand et al. 2009). Similarly, the number of sequences in a dataset is used to determinate rarity, with either OTUs that appear only once in a given sample (local community) or in the entire dataset (metacommunity) determining the group to which they are classed (Gobet et al. 2012). However, an important drawback of these two approaches is the strong correlation between abundance or frequency of

occurrence and sequencing coverage, that is, the greater the number of sequences (sequencing effort), the higher the chance of detecting low-abundant OTUs. In some cases, researchers characterize rarity by shifts in taxa abundances, as rare species may drop below the detection limit under unfavourable environmental conditions, but bloom when the environment becomes favourable, although this approach cannot detect permanently rare species (Aanderud et al. 2015). An alternative is the use of multivariate cut-offs which sets rarity cut-offs after the systematic removal of rare OTUs from the dataset, at the point where removal has an influence on community similarity (Gobet et al. 2010). This can provide a community-specific definition of rarity, and allow the discrimination, per sample, of the rare OTUs, which can be further tested regarding assembly processes (see, below, the section ‘Using community assembly theory to infer the types of rarity’), by focusing only on the rare subset of the dataset. It is important to note that the threshold of rare species is dependent on the β -diversity and therefore influenced by the number and identity of samples. Thus, the removal of samples from the analyses or the addition of new ones requires running new simulations [systematic application of multiple cut-off values, generating truncated communities, which are used to generate sample-by-sample dissimilarity matrices that are further compared to the original matrix through nonparametric Spearman correlations and Procrustes analyses (Gobet et al. 2010)]. Instead of relying on artificial thresholds, by exploring the effect of rarity on community structure, this method is driven by ecological interpretation, which can provide conceptual inferences across a continuum of abundance. We thus recommend this approach, as it provides a comprehensive understanding of the ecological processes regulating the abundant and rare fractions of microbial communities.

Disentangling the assembly processes of the microbial rare biosphere

The discipline of microbial community ecology is progressing from the description of patterns towards a mechanistic understanding of how communities are assembled, by seeking to investigate how ecological processes structure and maintain the composition of species from local to global spatiotemporal scales (Ovaskainen et al. 2017). Historically, ecologists have debated along two distinct theoretical lines to examine and interpret community assembly: the niche-based theory and the neutral theory (Box 4.2). Nowadays, this polarized dichotomy has been suppressed, as it has become more broadly accepted that both stochastic (‘neutral’) and deterministic processes influence community assembly simultaneously (Stegen et al. 2012). The challenge lies in quantifying their relative influences across community types and unveiling the mechanisms underpinning their variations across space and time. In line with that, Vellend’s theory of ecological communities proposes that four high-level ecological and evolutionary processes, namely selection, dispersal, drift and diversification, drive all principles in community ecology (Vellend 2010, Vellend 2016). These processes interplay across community types, shaping the microbial community structure and their biogeographic patterns, and influencing both abundant and rare species across any given settings (Hanson et al. 2012). Although distinguishing some of these processes might

still be a challenge in microbial system, such as in the measurement of in situ diversification rates, Stegen et al. have proposed a framework that enables the relative quantification of these processes in microbial communities (Stegen et al. 2013, Stegen et al. 2015). We suggest that a variation of this framework can be applied to disentangle how these processes operate in distinct fractions of microbial communities, thus enabling the quantification of the interplay between ecological processes involved in the assembly of the rare biosphere (Box 4.3 and 4.4).

Box 4.2 Ecological processes mediating community assembly using phylogenetic inferences

To infer on the influence of ecological processes structuring community composition, Webb and colleagues laid out a heuristic framework to examine deviations between phylogenetic community structure and the null expectation by randomization procedures, referred to as ‘phylogenetic null model’ (Webb 2000, Webb et al. 2002). This framework is based on the assumption that closely related species have similar ecological niches, and that species coexistence patterns depend on competition or physiological tolerance (Webb et al. 2002). An extended method, which combines phylogenetic community structure and standardized Bray-Curtis matrix with null models, builds on Webb’s framework by allowing the identification of the ecological processes mediating community assembly (Box 4.4) (Stegen et al. 2013, Stegen et al. 2015). Within this model, ecological processes are divided into five variants: variable selection, homogeneous selection, dispersal limitation, homogenizing dispersal and undominated processes. Variable selection results in high compositional turnover, where environmental factors (biotic and abiotic) shift temporally or spatially. In contrast, homogeneous selection leads to low compositional turnover, where the selection is consistent through time and space (Dini-Andreote et al. 2015). Dispersal also shows two variants: at low dispersal rate, dispersal limitation combined with drift leads to a large community turnover; whereas, at high dispersal rate, it causes a homogenization of community composition (i.e. homogenizing dispersal) (Leibold et al. 2004). Finally, when selection, dispersal and drift contribute equally to community composition and no one dominant, the variant is called an undominated process (Stegen et al. 2015).

Evidence suggests that rare species are susceptible to local elimination by drift or competition by other abundant species (Melbourne and Hastings 2008); however, it remains largely unclear how other assembly processes operate in the rare biosphere – knowledge that is critical to foster our understanding of the types of rarity. Evidence from macro-organisms shows that the occurrence of rare and abundant species is often regulated by different ecological processes (Bell 2000, Magurran and Henderson 2003). In the case of microbial communities, relatively little is known about how these ecological processes regulate the assembly of the rare biosphere. One of the few studies addressing this issue

showed that, in microbial communities from the Arctic Ocean, abundant species follow a log-normal distribution, while rare species fit a log-series distribution (Galand et al. 2009). These results indicate a more deterministic assembly of abundant taxa, while the assembly of rare species is mediated by stochastic processes. However, rank abundance distribution is sensitive to the size of the ‘tail’ in species distribution, that is, the number of rare species (Magurran and Henderson 2011), and therefore, partitioning rare individuals from abundant ones might undermine the approach and revoke the results. This suggests that rank abundance curves, although broadly used to investigate assembly processes of the entire community (Woodcock et al. 2007), cannot be properly extended to study subsets of communities. Alternatively, the use of phylogenetic null models (Box 4.3) would represent a scientifically sound alternative to contrast assembly processes driving rare versus abundant fractions of microbial communities, as the basic assumption relies on phylogenetic signal (ecological traits correlating with phylogeny) rather than solely on the number and distribution of individuals.

Box 4.3 Niche and neutral theories in community ecology

The niche theory emphasizes the importance of determinism, that is, it raises ecological selection as a unique ecological process that influencing the identities and abundance of species. This theory assumes both the role of selection imposed by abiotic factors (also termed as ‘environmental filters’) and species interactions (antagonistic and synergistic). In microbial ecology, such theory fits well within two postulates: (i) the Bass Backing hypothesis that ‘Everything is everywhere, but the environment selects’ (O’Malley 2007), and (ii) the Gause (1934) postulate that ‘the niche that is occupied by one organismal type is very unlikely to be also occupied by another one’. A large amount of research has supported the importance of selection via abiotic factors such as pH, salinity and the concentration and composition of organic carbon, in determining microbial community structure (Fierer and Jackson 2006, Lozupone and Knight 2007, Fierer et al. 2009). Beyond such a paradigm, Hubbell’s neutral theory stresses that neutral and often stochastic processes can largely mediate the assembly and dynamic changes in community structures (Hubbell 2001). This theory assumes that random demographic aspects, probabilistic dispersal, speciation and unpredictable disturbance are the mechanisms generating divergence across community types (Tilman 2004, Gravel et al. 2006, Leibold and McPeck 2006, Adler et al. 2007).

Using community assembly theory to infer the types of rarity

Revealing the relative contribution of the ecological processes structuring the rare biosphere will not only add substantial knowledge towards the identification of biotic and abiotic drivers of rarity, but will also assist in the classification of the types of rarity proposed by Lynch and Neufeld (2015) (Figure 4.1). Within this framework, if deterministic processes have a higher influence on the assembly of the rare biosphere, the

existence of the rare biosphere is largely caused by narrow niches and a set of abiotic and biotic interactions that affect specific population dynamics within a community (Jousset et al. 2017). This phenomenon leads to rare populations that are driven by trade-offs and will persist at low-abundances, that is, permanently rare (Figure 4.1A). This type of rarity occurs due to the strong homogeneous selection, as these specifically rare taxa occupy narrow niches and can often interact with other species. For instance, oligotrophic organisms can successfully persist under low nutrient conditions, as they cannot reach high abundances within a community due to intrinsic metabolic constraints (Fierer et al. 2007). Moreover, low-abundant species can be suppressed/controlled by competitors, that is, through antibiosis or outcompeted in the exploitation of nutritional resources, leading to the phenomenon of negative frequency dependency (Yenni et al. 2012). In addition, the ‘kill the winner’ theory indicates that bacteria and archaea face strong predation pressures, mostly by protozoan grazing and viral attack (Rodriguez-Valera et al. 2009), which mainly occurs in a species-specific manner, affecting the abundance of taxa preferentially exploited by predators.

On the other hand, rare taxa can periodically switch between abundant and rare (Figure 4.1D, green line). The dynamics of their abundances depends on variable selection through periodic changes in environmental conditions, such as seasonal fluctuations. Similar to the periodically dynamic group, species that persist at relatively low abundances can become more prevalent (Figure 4.1D, yellow line) by responding to occasional events such as precipitation and nutritional pulses. Under stressful conditions, these taxa remain at low growth rates and consequently low densities. In extreme cases this can lead to microbial dormancy, in which the organism enters a reversible state of decreased metabolic activity that can persist for long periods of time under unfavourable or harsh local environmental conditions (Lennon and Jones 2011). However, it is disputable whether these organisms should be considered part of the rare biosphere (see next section). Once favourable conditions arise, both active but slow-growth and dominant rare species gradually recover their population densities. Shade et al. (2014) described that conditionally rare taxa are usually at low abundance but can occasionally increase their abundances under suitable conditions. They surveyed nine different ecosystems and reported that 1.5–28% of all detected taxa are conditionally rare taxa. Importantly, the conditions described above indicate the role of deterministic processes in driving the assemblage of both permanently and conditionally rare organisms, through homogeneous and variable selection, respectively (Shade et al. 2014).

Alternatively, stochastic processes can also play a role in the assembly of the rare biosphere. Passive dispersal is the common mechanism of microbial dispersal, for example, through wind, water and/or by hitch-hiking on other organisms (Ai et al. 2013). It has been assumed that microbes have ubiquitous dispersal capability due to their size and are not restricted by geographical barriers, as asserted by Baas Becking dictum ‘*everything is everywhere*’ (O’Malley 2007). However, it has recently been shown, through a modelling approach, that microbes are in fact dispersal-limited (Wilkinson et al. 2012). In any case, it is likely that dispersal has a larger influence on the microbial rare biosphere than on rare macro-organisms, as their smaller size facil-

itates displacement by passive dispersal. Thus, taxa can persist in the rare biosphere by (unrestricted) dispersal processes, with populations fluctuating by drift process while remaining below a rarity threshold (Figure 4.1B). Alternatively, transiently rare taxa could arise when recently immigrated species enter a community by a dispersal process (Figure 4.1C). The biogeographic patterns of the rare biosphere in the Arctic Ocean support that dispersal limitation, together with selection and drift, potentially promote cross-systems rare biosphere divergences (Galand et al. 2009). It is worth mentioning that, along with dispersal, the diversification process can also introduce new species into a community (Nemergut et al. 2013). Due to the relatively small genome sizes and often high growth rates, microorganisms evolve faster than macro-organisms through mutations or via horizontal gene transfer (Rainey and Travisano 1998, Ochman et al. 2000). Thus, the generation of new species/genotypes contributes to the rare biosphere, although assessing the relative contribution of speciation to rarity remains challenging in view of the current molecular approaches used for species identification (see next section).

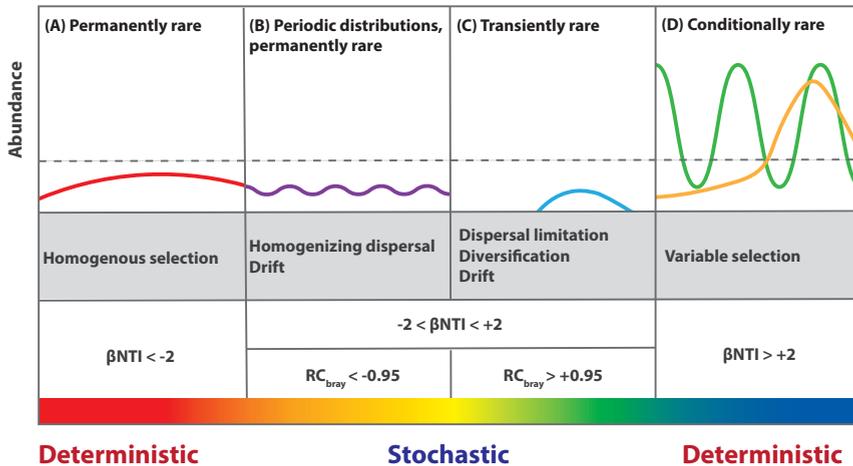


Figure 4.1 The ecological processes mediating the structure of low-abundant species according to the type of rarity. The upper four panels show the (temporal or spatial) abundance profiles of different rare types [adapted from (Lynch and Neufeld, 2015)]. The dashed line is the abundance threshold of the rare biosphere. The outcome from Stegen’s quantitative framework (Box 4.3 and 4.4) can be used to deduce the main ecological processes (grey panels) responsible for each type of rarity, as well as the between-community nearest taxon index (βNTI) values associated with each of the assembly processes (deterministic or stochastic; lower panel). (A) Permanently rare taxa are species whose abundance remains steadily low, due to strong homogeneous selection, and assembled in a deterministic manner. In (B) the abundance of rare species fluctuates, but always below the rarity threshold. In this case, selection is weak, and communities are assembled in a stochastic manner by processes associated with homogenizing dispersal and ecological drift. (C) Transiently rare species are those that have immigrated to the community or emerged due to diversification processes. These species can occasionally go extinct as a result of drift and/or dispersal limitation. (D) Conditionally rare taxa are those that periodically or occasionally reach high abundances, in response to environmental fluctuations. These taxa are strongly driven by variable selection where selective pressure changes through time and/or space.

Based on the lines of evidence presented, we argue that different ecological processes are associated with the different types of rarity. In order to quantify the importance of each process in driving the assembly of the rare biosphere, and the biotic and abiotic mechanisms underpinning their relative influences, we advocate using the approach proposed by (Stegen et al. 2013), which combines the use of phylogenetic inferences and null modelling approaches (Box 4.4) to quantitatively probe major ecological processes (e.g. variable selection, homogeneous selection, dispersal limitation, homogenizing dispersal, and undominated processes; Box 4.3) involved in structuring the rare biosphere (Stegen et al. 2015).

Box 4.4 Modelling community assembly and ecological processes

The use of phylogenetic information enables the elucidation of the phylogenetic relatedness of species within and across communities. The basic requirement for using such a method is the conservation of a phylogenetic signal, which indicates that ecological traits are not distributed randomly across a phylogenetic tree (Losos 2008). Between-community mean nearest taxon distance (β MNTD), which is based on the mean distance to the nearest neighbour, is a measure of the clustering of closely related species (Webb et al. 2002). Null model distribution of β MNTD can be built by shuffling species among the tips of the phylogenetic tree or between different communities using randomizations and permutation analyses. The β -nearest taxon index (β NTI), a standardized measure of the MNTD, is then generated by comparing the observed and the null distribution of β MNTD, using the following formula:

$$\beta NTI = \frac{\beta MNTD_{obs} - \overline{\beta MNTD_{null}}}{sd(\beta MNTD_{null})} \quad (4.1)$$

A value of $|\beta NTI| > 2$ indicates that observed turnover between a pair of communities is governed primarily by selection, in which $\beta NTI > +2$ is consistent with variable selection, while $\beta NTI < -2$ indicates homogeneous selection (Dini-Andreote et al. 2015). As a corollary, $|\beta NTI| < 2$ means that the turnover of a group of communities is governed by dispersal limitation, homogenizing dispersal or undominated processes (Stegen et al. 2015). To disentangle these processes, the Raup-Crick matrix (RC_{bray}), built based on the standardized Bray-Curtis matrix of communities, provides information on whether the observed degrees of turnover deviates significantly from the expectation (Chase et al. 2011). This value equals the deviation between observed Bray-Curtis and the null distribution, ranging from -1 to +1. Within this, $|RC_{bray}| < 0.95$ can be interpreted as an influence of undominated processes. In turn, dispersal limitation coupled with drift leads to greater than expected turnover ($RC_{bray} > +0.95$), whereas $RC_{bray} < -0.95$ indicates that turnover in community composition is governed primarily by homogenizing dispersal (Table 4.1).

(Continued from previous page)

In order to determine the ecological processes controlling the assembly of rare communities, we propose using a variant of the method described above which would use only the rare species present in each sample (determined by multivariate cut-off analyses) to generate the phylogenetic tree. This tree will serve as the basis for calculating the observed and null β MNTD, generating β NTI values for each rare community, which can be < -2 or $> +2$, indicating that rare communities are driven by deterministic processes and can be considered to be permanently rare (homogeneous selection) or conditionally rare (variable selection), respectively. Alternatively, the rare communities can be driven by stochastic processes ($-2 < \beta$ NTI < 2), in which case RC_{bray} should be used to determine whether communities are transiently rare ($RC_{\text{bray}} > +0.95$), permanently rare with periodic distribution ($RC_{\text{bray}} < -0.95$), or driven by undominated processes ($-0.95 < RC_{\text{bray}} < +0.95$).

Table 4.1 Criteria to quantify the importance of ecological processes responsible for turnover between communities (Stegen et al. 2015).

| β NTI | RC_{bray} | Processes |
|------------------------|------------------------------------|--|
| < -2 | | Homogeneous selection |
| | < -0.95 | Homogenizing dispersal |
| $-2 < \beta$ NTI < 2 | $-0.95 < RC_{\text{bray}} < +0.95$ | Undominated processes |
| | $> +0.95$ | Dispersal limitation acting with drift |
| $> +2$ | | Variable selection |

Challenges in the investigation of rare assemblages

To unravel the assembly processes of the rare biosphere, it is important to obtain information on the entire community at multiple time points (time series), for instance by increasing sequencing depth efforts, replicates and benchmarking analyses. The development of techniques to remove artificial sequences from a dataset will help us to accurately investigate microbial communities. In addition, a finer taxonomic resolution will decrease the omission of rare taxa. For instance, it has been argued that exact sequence variants defined by methods, such as UNOISE3, DADA2, and Deblur, provide greater resolution in marker gene analysis and improve reusability, reproducibility and comprehensiveness for microbial community analyses (Callahan et al. 2017).

A potential bias lies in the use of relative abundances (i.e. standardizing the number of taxa within microbial communities) to facilitate the comparison across communities. While this is common practice, Props et al. (2017) showed that the use of absolute abundances leads to different community patterns than those observed by relative abundances. Thus, whether abundances are displayed by relative or absolute

values may influence the conclusions we draw on the rare assemblages (Magurran and Henderson 2011). Other alternatives could rely on measuring microbial communities based on the number of sequences per mass of substrate (Stegen et al. 2016) or by quantifying absolute taxon abundances using flow cytometry, quantitative PCR, or fluorescence in situ hybridization (Props et al. 2017).

A further challenge refers to microbial-specific issues, which cause microbial ecology to deviate from general (macro-organismal based) ecological principles (Prosser et al. 2007), such as the distinction between inactive or dormant cells and active ones. The combined use of DNA- and RNA-based sequencing could provide the first evidence to distinguish active from inactive taxa in the rare biosphere. However, to overcome the variation in RNA copy number and microbial activity, new high-throughput methods should be developed, which will ultimately provide answers to questions such as the following. (i) How active is the rare biosphere, that is, does it encompass only inactive cells? (ii) An even more conceptual question: can both active and inactive fractions be considered to be equal contributors to the rare biosphere? Or should a conceptual partition between functionally rare and dormant (nonfunctional) rare be proposed?

Another potential discrepancy between macro- and micro-organismal ecology lies in the identification of speciation rates and extinction events. Speciation is an important evolutionary process that cannot currently be measured at the resolution used in microbial community studies, given the high phylogenetic signal and the low evolutionary rate associated with the rRNA gene, which is often used for phylogenetic purposes. Extinction, which has been defined by Vellend as the outcome of selection and drift rather than being considered a process in itself, cannot be taken into consideration until the extent of microbial diversity has been fully scrutinized. While extinction might happen, the same species can recolonize the same patch if dispersal from a nearby patch takes place, thus rendering its detection currently impossible (Nemergut et al. 2013).

Concluding remarks and future perspectives

Although rare species are key constituents of microbial communities' diversity and functionality, the processes driving rarity remain unclear, both from a conceptual and technical perspective. Regarding the latter, most of the literature is based on the dominant species in microbial communities because low-abundance species are routinely removed from data sets in order to avoid noise caused by sequencing artifacts (Jousset et al. 2017). In order to gather information on the entire microbial community, low-abundant species must be included and not discarded arbitrarily. Moreover, as sequencing technologies advance and become more affordable, it is expected that sequencing depths will continue to increase in the near future, alongside proper sample acquisition and benchmarked computational analyses. Collectively, these practices will reveal the temporal and spatial dynamics of the rare biosphere at a higher resolution. From a practical standpoint, we argue that, by combining strategies of increased sequencing depth and multivariate cut-off values with theoretical models of commu-

nity assembly, it will be possible to determine the ecological processes controlling the rare biosphere, and by doing so, identify the types of rarity. A next step will consist of linking the rare fraction of communities to their ecosystem functioning. A recent modelling attempt indicated that communities that are stochastically assembled due to high dispersal rates (homogenizing dispersal), could harbour microbes that are maladapted to the environment, potentially leading to a reduction in function when compared to a community driven largely by selection (Graham and Stegen 2017). It remains to be tested whether rare communities driven by stochastic processes (transiently rare or permanent rare with periodic distributions) negatively influence the ecosystem processes they are associated with. Or whether deterministically assembled rare communities (permanently or conditionally rare) contribute to the ecosystem processes as much as do the abundant taxa. Moreover, future research should focus on understanding whether the rare biosphere is simply a subset of the abundant or total community or whether it follows a specific path either in terms of community assembly processes or functioning.

To a large extent, a more comprehensive knowledge of rare biosphere assembly rules will enhance our understanding of seminal questions in microbial ecology about why microbial communities are so diverse and what the main processes underpinning the observed diversity are (see Outstanding questions in Box 4.5). It will also pave the way towards guiding future experiments and environmental sampling in searching for generalized patterns and processes mediating the assembly and ecological role of the microbial rare biosphere.

Box 4.5 Outstanding questions

- What are the relative influences of the community assembly processes regulating the copious rare species? Do these processes differ from those mediating the abundant organisms or total biosphere?
- What is the fraction of each rarity type within the rare biosphere? How does that vary across community types?
- Is there a difference between the ecological processes regulating the active and inactive fractions of the rare biosphere?
- Do the influences of community assembly processes structuring the rare biosphere vary between ecosystem types? If so, what are the biotic/abiotic mechanisms mediating these variations across distinct community types?
- What are the functional implications of shifting mechanisms mediating the community assembly of distinct types of rarity?
- Are the relative influences of community assembly processes consistently associated with changes in relative or absolute taxa abundances?

Acknowledgments

We thank Stefanie Vink, James C. Stegen, and an anonymous reviewer for their constructive comments on the manuscript. X.J. was supported by a scholarship from China Scholarship Council and the University of Groningen scholarship program.