Tackling the diversity of the marine microbial rare biosphere: methodological challenges and ecological insights

Francisco Pascoal[1,2], Catarina Magalhães[2,3], Rodrigo Costa[1,4]
[1] Institute for Bioengineering and Biosciences (iBB), Department of Bioengineering, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
[2] Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, Porto, Portugal
[3] Faculty of Sciences, University of Porto, Porto, Portugal
[4] Centre of Marine Sciences (CCMAR), University of Algarve, Portugal

Abstract:

The microbial rare biosphere represents the bulk of microbial diversity in virtually all environments. This is historically recognized in general biology and confirmed in microbiology due to the emergence of high throughput sequencing of the small subunit of the ribosome gene. The number of studies on the microbial rare biosphere has been growing every year, allowing for the recognition that, despite their low abundance, they contribute to ecosystem functioning and are important to understanding microbial community assembly. Currently there is no coherent and unifying definition of microbial rarity, as the concepts in use vary greatly and commonly lack biological meaning. To approach this hurdle from a statistical standpoint, the Multivariate Cutoff Level Analysis (MultiCoLA) algorithm was recently proposed for determining abundance thresholds from where microbial rarity could be delineated across distinct microbiomes with their own unique community structures. This algorithm was tested in the present study, where it generated coherent results across independent marine datasets, but it is not able to give support to a non-subjective definition of microbial rarity. Nevertheless, using the rare prokaryotic communities identified with the later method, it was possible to explore how different metagenomic strategies and seawater sampling methodologies affect the structure of the so-defined marine microbial rare biosphere. Ecological insights from the Arctic Ocean data and from the *Spongia officinalis* (marine sponge) microbiome data corroborate existing knowledge on the marine prokaryotic rare community assembly processes. Furthermore, this study integrates both stochastic and deterministic mechanisms in the process of marine prokaryotic rare biosphere assembly, with water masses and host-associated relationships playing key roles. Finally, this work provides methodological guidelines for optimal sampling of the seawater rare biosphere.

Key-words: community assembly; microbial dark matter; microbial ecology; next-generation sequencing; rarity definition; seawater sampling.

Introduction

The microbial rare biosphere was first described by Sogin et al. (1) and, since then, there is an increasing number of studies in this field. Those studies agree that the microbial rare biosphere, that is, the pool of low abundance species present in a given microbial community, represents most of the existing microbial diversity across Earth's ecosystems (2–5). This high diversity is believed to function as a vast genetic reservoir contributing to the resistance and resilience of the ecosystem (4,5). First, the microbial rare biosphere was hypothesized to work as a backup system, in the form of a seed bank that could become abundant when necessary (2,6), with early evidence from Szabó et al. (7). Later, there was a debate on this hypothesis, regarding the finding that, in marine systems, most of the prokaryotic rare biosphere remained rare across different seasons (8). Despite that, by disentangling the complexity of the microbial rare biosphere into different types of rarity (9), the consensus is that some rare microbes are always rare, and thus do not

respond to changing variables (permanently rare taxa - PRT) while others do respond to changing conditions through shifts in abundance (conditionally rare taxa – CRT) (10). Besides CRT and PRT, some rare microbes are considered transiently rare, associated with the finding of alien species, that appear and disappear from the environment (11,12). The permanent or persistent prokaryotic rare biosphere usually displays biogeographical patterns while the transient rare communities are associated with a cosmopolitan distribution of species, thus combining stochastic and deterministic mechanisms (11–15).

By associating community assembly theory with the types of rarity described by Lynch and Neufeld (9), Jia et al. (15) proposed a model to determine the types of rarity, based on deterministic and stochastic ecological mechanisms. In that model, deterministic mechanisms include homogenous selection and variable selection, the first resulting in permanent rarity and the second resulting in conditional rarity. The reasoning is that the same selective pressure will result in the same pattern of abundance across time, leading to permanent rarity, whereas variable selection will result in variable abundances as well. Conversely, stochastic mechanisms would essentially include dispersal limitation and homogenizing dispersal. The first one resulting in transient rarity, where a microbe randomly appears and disappears from an environment and the second one results in permanent rarity, but with variation, meaning a microbe that remains in the environment, but does not grow abundant, receiving and losing members randomly.

There are three main mechanisms currently hypothesized for the microbial rare biosphere role in ecosystem functioning, the first one is through clonal amplification, as suggested from the seed bank perspective (6). Other rare microbes have been found to remain rare, while changing their activity rates (16), with the canonical example of Pester et al. (17), where a rare prokaryote (*Desulfosporosinus* genus) significantly contributed to the overall sulfate reduction in a peatland environment. Finally, recent evidence also suggests that permanent rare microbes might keep transmissible functional genes that are transferred to other abundant microbes, in response to specific environmental changes (e.g. Wang et al. (18)). Despite the growing number of studies on the microbial rare biosphere, the concept in itself is subjective and the definition of rarity remains problematic (15). Most studies use High Throughput Sequencing (HTS) methods based on the sequencing of the 16S or 18S rRNA gene amplicons, where high quality sequences are clustered into Operational Taxonomical Units (OTUs). In those studies, microbial rarity is defined using rarity thresholds, usually expressed as relative abundance per sample (e.g. 0.1% per sample (19)). That threshold can also be expressed as an absolute value, meaning the actual number of reads below which the OTU is considered rare. The problem with this approach relies on the lack of a biological principle underlying the definition of rarity, with thresholds remaining arbitrary. To circumvent this hurdle, Jia et al. (15) proposed the adaptation of the Multivariate Cutoff Level Analysis (MultiCoLA) algorithm to define rarity. MultiCoLA was originally constructed by Gobet et al. (20), with the objective of understanding what is the impact on

community composition after removal of rare OTUs. Briefly, the algorithm applied to define rarity uses different thresholds to produce rare communities, and then compares each truncated community with the original community. Assuming the rare community is different from the overall community, a sudden decrease in the correlation between the original and the rare community is expected. Thus, by analyzing the correlation values of each threshold, a rarity definition could be tailored for each specific sample, with an absolute abundance threshold. To the knowledge of the cited literature in this work, there is no proof of concept for this method.

Furthermore, the influence of different seawater sampling methodologies on the marine prokaryotic rare biosphere is overlooked. Although Liu et al. (21) reported that rare marine prokaryotes are highly influenced by different DNA extraction methodologies, there is little knowledge on the effect of seawater sampling methodologies on the analysis of the marine prokaryotic rare biosphere. With the EuroMarine Open Science (EMOSE) 2017 dataset, different seawater sampling methodologies were compared for the same marine environment. This work also used the EMOSE 2017 dataset to explore test the MultiCoLA algorithm to the definition of rarity. Besides this dataset, the *Spongia officinalis* 2014 dataset (22) and the Norwegian Young Sea Ice Expedition (NICE) 2015 dataset (23) were also used to define rarity using the MultiCoLA algorithm and to test questions regarding the ecological dynamics of the marine prokaryotic rare biosphere.

Methodology

This work used three different datasets: EMOSE 2017, *Spongia officinalis* 2014 and NICE 2015. The EMOSE 2017 team sampled seawater at a single site of the Mediterranean Sea, with different sampling methodologies (differing in the type of filter, size fractioning and filtered volume) and metagenomic approaches (16S and 18S rRNA gene amplicon sequencing and TC-DNA shotgun sequencing). The *Spongia officinalis* 2014 dataset has been previously described in Karimi et al. (22) it includes four sponge tissue samples, three seawater samples and three sediment samples, 1m away from the sponge specimens. The NICE 2015 dataset (24,25) sampled seawater using a vessel fixed on ice, north of Svalbard, in the Arctic ocean, with three samples from March, April and June, for surface (5m), middle depth (25m or 50m) and deeper layers (250m). Different samples also represent different water masses, namely: Polar Surface Water (PSW), warm Polar Surface Water (wPSW), Atlantic Water (AW) and Modified Atlantic Water (MAW). Raw reads from all datasets used in this work were processed by the MGnify platform, with the bioinformatic steps described in (26). The EMOSE 2017 accession key is MGYS00001935, the *Spongia officinalis* 2014 dataset accession key is MGYS00000563 and the NICE 2015 accession keys are MGYS00001922, for the 16S rRNA gene amplicon sequencing and MGYS00001869 for the TC-DNA shotgun sequencing data.

MultiCoLA scripts from Gobet et al. (20) were adapted in this work based on the model to define rarity by Jia et al. (15). Briefly, the

parameters selected differently from the original method (20) were the type of analysis (type parameter), which is “sample by sample” (type= SAM) instead of “all dataset” (type= ADS), and the analysis is for the rare component and not the abundant one, thus, the typem parameter is “rare” (typem= rare), instead of “abundant” (typem= abundant). Diversity analysis included alpha diversity metrics, where the number of OTUs, the number of reads and the Shannon index were calculated with custom R commands (27), using phyloseq (28). Ordination analysis was produced using the R Vegan package (29).

Results

Defining microbial rarity with MultiCoLA results

The EMOSE 2017 dataset was divided into groups, according with the metagenomic strategy used, namely: TC-DNA shotgun sequencing (MetaG 16S), amplicon sequencing for 18S V9 region of rRNA gene (MetaB 18S), amplicon sequencing of 16S V4-V5 region of rRNA gene, without sizing (MetaB 16S nS), amplicon sequencing of 16S V4-V5 region of rRNA gene, with sizing for 400bp (MetaB 16S small) and amplicon sequencing of 16S V4-V5 region of rRNA gene, with sizing for 600bp (MetaB 16S large). The resulting thresholds were listed in Table 1. Despite the discrepancies in the absolute abundance thresholds, when they were converted into relative abundance, per sample, on average, they were relatively close to 0.1% (Table 1). Thus, the thresholds obtained for the different EMOSE 2017 dataset were consistent with each other, despite the differences in number of reads, and number of samples. For other independent datasets, *Spongia officinalis* 2014 resulted in a threshold of 13 reads per sample, equivalent to a relative abundance threshold of 0.44%. For the NICE 2015 dataset, the analysis was divided in TC-DNA shotgun sequencing and 16S rRNA gene amplicon sequencing groups. The rarity threshold for the first was of 42 reads per sample, equivalent to 1.25% relative abundance, per sample, on average (results on the thesis main text, section 3.1.2). For the second group of the NICE 2015 data, the rarity threshold in absolute abundance was 1054 reads, per sample, equivalent to an average relative abundance per sample of 0.6% (results on the thesis main text, section 3.1.3).

Table 1. MultiCoLA based rarity thresholds for the EMOSE 2017 dataset.

Sub dataset	Absolute abundance threshold	Relative abundance, per sample, threshold (average)	Number of reads clustered into OTUs (average)	Number of samples
MetaG	6	0.097%	60 014	50
MetaB 18S	197	0.047%	1 492 134	47
MetaB 16S nS	154	0.055%	1 920 794	68
MetaB 16S small	972	0.094%	1 367 111	53
MetaB 16S large	7899	0.514%	1 810 973	53

Seawater sampling methodologies effect on the prokaryotic rare biosphere

The MetaB 16S nS group of samples, from the EMOSE 2017 dataset, were selected to compare how different seawater sampling methodologies influence the marine prokaryotic rare biosphere. The variables compared included a range of volumes from 2.5L to 1000L, Sterivex and Membrane filtering units and size fractioning (small, medium and large) and whole water filtering. Alpha metrics and significance values were listed in the main thesis text, where the most significant differences were on the size fractions used during filtering. This pattern was confirmed by the PCA plot (Figure 1), where different variables overlap with each other, except for the filtration size fractions, that were separated in different areas, with the small size filtration fraction overlapping with the whole water filtering.

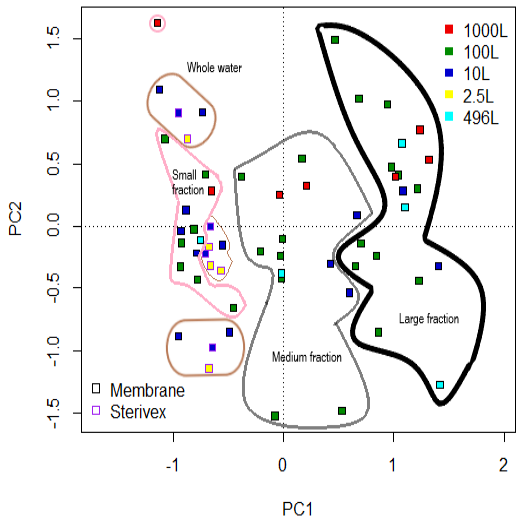

Figure 1. PCA of MetaB 16S nS data, from the EMOSE 2017 dataset, for prokaryotic OTUS.

***Spongia Officinalis* 2014 dataset**

The prokaryotic rare biosphere was more diverse than the abundant biosphere for all samples, as most OTUs were rare, despite the low number of reads. Overall, the sediment samples had the most diverse prokaryotic rare biosphere, followed by the seawater and sponge tissue samples. By analyzing the number of shared and specific prokaryotic OTUs (both rare, abundant and total), it was possible to understand some patterns, namely that some OTUs were specific to one environment (sponge tissue, seawater or sediment) and others were shared (Figure 2). Seawater had more shared OTUs than specific OTUs for the rare and total biosphere, but not for the abundant biosphere, where most OTUs were specific. For the rare and total communities, within the seawater shared OTUs, most were shared with sediment or with sediment and sponge simultaneously. The shared OTUs between seawater and sediment were a minority. For the rare and total microbial community, sediment OTUs were mostly specific, and most of the sediment shared OTUs were shared with seawater. Sponge tissue rare OTUs were mostly shared with sediment. Total community OTUs from the sponge tissue were mostly shared with sediment and seawater simultaneously (Figure 2).

A

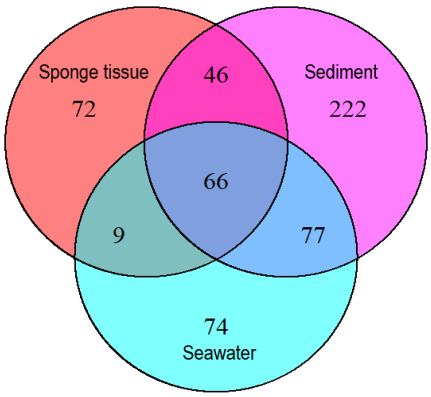

Figure 2 continues on next page.

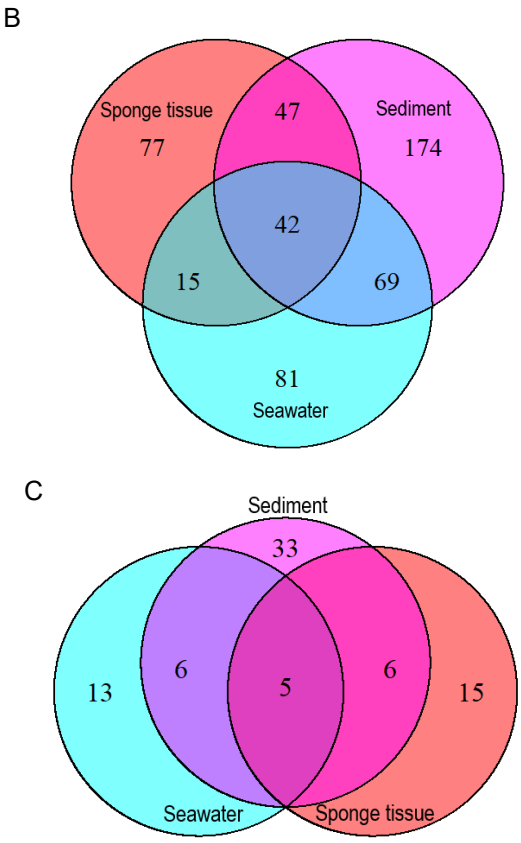

Figure 2. Venn diagrams for shared and specific prokaryotic OTUs across different samples, in the *Spongia officinalis* 2014 dataset. A – Total community; B – Rare community; C – Abundant community.

NICE 2015 dataset

The same diversity analysis was performed for the TC-DNA shotgun sequencing and 16S rRNA gene amplicon sequencing data from NICE 2015. Alpha diversity for the total, rare and abundant communities are available in the main thesis text, as well as significance values for the comparison of alpha diversity metrics across variables (sections 3.4.1 and 3.4.2). From those metrics, the patterns were generally the same for both metagenomic strategies, with an increase of rare and abundant prokaryotic reads from March to June, although not significant. The most significant differences were across different water masses, for both the number of rare OTUs, reads and their Shannon index. As corroborated by the PCA plots from both metagenomic strategies used in the NICE 2015 (Figure 3 and Figure 4). The grouping into different water masses was more evident in the TC-DNA shotgun sequencing data (Figure 3). When comparing both metagenomic approaches, the sequencing power of the marker gene did not reflect on the actual rare prokaryotic biodiversity, since the equilibrium of species, as determined by the Shannon index, was superior in the TC-DNA shotgun sequencing approach. The number of different rare OTUs collected was similar for both strategies (results in the main thesis text, section 3.4.3, figure 20). The different types of rarity were also calculated for both groups, in the NICE 2015 dataset, across spatiotemporal and depth variables. From that analysis, the vast majority of the rare OTUs were transiently rare, followed by PRT (both with and without variation) and only a small fraction of CRT (results in the main thesis text, sections 3.4.1 and 3.4.2, figures 15 and 17).

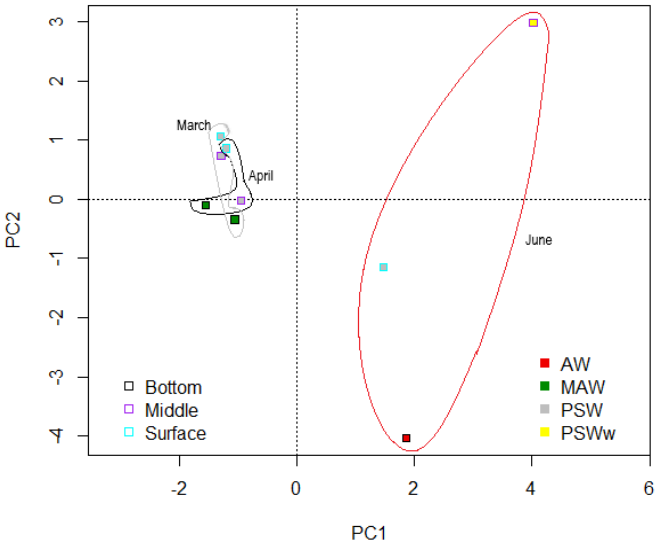

Figure 3. PCA of TC-DNA shotgun sequencing data from the NICE 2015 dataset, for rare prokaryotes.

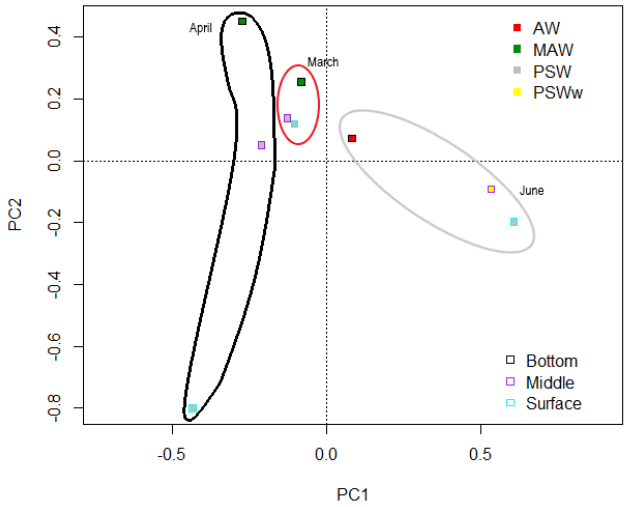

Figure 4. PCA of 16S rRNA gene amplicon sequencing data, from the NICE 2015 dataset, for rare prokaryotes.

Discussion

There is an increasing number of microbial rare biosphere studies, also in the marine environment (e.g. (8,30,31)). Despite that, the actual definition of the microbial rare biosphere remains ambiguous (15). For instances, definitions range from as low as 0.001% relative abundance per sample (e.g. (32)), to up to 1% relative abundance per sample (e.g. (33)). Although most thresholds used are of 0.1% relative abundance per sample (e.g. (19)). And many studies do not provide a specific rarity definition (e.g. (1)). Besides the lack of coherence in the different existing definitions, there is no biological basis to decide in favor of one or another threshold, to solve that, the MultiCoLA algorithm was proposed (15,34). This work tested MultiCoLA for independent datasets, using different metabarcoding and metagenomic approaches. In the EMOSE 2017 dataset, MultiCoLA was able to provide absolute abundance thresholds that were consistent with the sequencing power of the respective metagenomic methods. Those absolute values were very different from each other, but became similar when converted to their relative abundance, per sample, equivalents. The relative abundance values were not far from 0.1%, the most used threshold in the literature. Thus, it can provide a rarity threshold. Notwithstanding, some problems were found, namely, the behavior of the correlation values in the MultiCoLA output, where there is no evident sudden decrease in correlation. Therefore, the selected threshold will differ across different research groups, using the same data. For the same metagenomic strategies, for independent datasets (comparing EMOSE 2017, *Spongia officinalis* 2014 and NICE 2015 datasets), the MultiCoLA values were similar, but in the TC-DNA shotgun sequencing data from NICE 2015, the threshold was higher than expected (1.2% relative abundance, per sample), probably because of the lower number of sequences used for taxonomic assignment.

Regarding the marine prokaryotic rare biosphere assessment, there is no knowledge of the effect of the different sampling steps on the view of rarity. In marine prokaryotic rare biosphere studies using 16S rRNA gene amplicon sequencing, whole water and pre filtration are methodologies used in the literature(1,11,32,35–37). Pre filtration prior to bacterial cells filtration is usually used to lower eukaryotic contamination (30,31,38–45) by the usage of a mesh with pore size of 200µm (31,42,44,45), or membranes for pore sizes of 3µm (8,39–41,43) or 0.8µm (30). The bacterial cells, in the marine prokaryotic rare biosphere studies, independently of pre filtration or not, are mostly filtered with Sterivex filter units (8,11,30,31,39–44), but membrane filter units are also used (1,32,36,37,45), both with pore sizes of

0.22μm. Regarding volume, marine prokaryotic rare biosphere studies use: less than 1L (32,45), 1L (1,35,37), 2L (42), 5-7L (8,40,41), 20L (39) and 170L (36). From the EMOSE 2017 results, it was found that volume is not very relevant and that the range of volumes most commonly used in the marine prokaryotic rare biosphere studies, from 1L to 20L, are enough. The type of filter did not induce significant differences, probably due to the pore size of the filters (0.22μm). The most important variable in the seawater sampling methodology is the filtration size fractioning, due to the selection of different communities according with cellular size. Despite that, it is contradictory to find an excess of diversity in the larger size fractions, that can be explained by the existence of host associated prokaryotes, that get stuck in the larger fractions, but this work does not have results to support that hypothesis.

Seawater works as a reservoir of sponge symbionts, that remain viable and rare outside the host, until being filtrated by the marine sponge (46,47). This is further supported by the finding that abundant sponge microbial symbionts are essentially generalists and specialists (48) and by the finding that some rare microbial symbionts are species specific (49). An understudied component in the tradeoff of prokaryotic OTUs and sessile hosts, is the sediment (22). It was suggested that sponge cellular shedding could work as a source of sponge associated microbial OTUs to the sediment environment (22). If specialized (host microbiome) OTUs sink in the sediment, then it is expected that they become rare in the sediment. It was also suggested that particle intake by sponge (50) can justify the existence of shared OTUs between sediment and sponge associated communities (22). From this work results, influx of seawater and sponge tissue cells shedding to sediment randomly transports prokaryotes across different types of environment (sponge tissue, sediment and seawater). This stochastic component explains the high numbers of transient rarity in the sponge tissue. The deterministic component is within each environment, where a group of conditions are maintained through time, resulting in a constant selective pressure, that allows some of the randomly distributed cells to persist. For example, permanently rare symbionts in the sponge tissue, that are viable and with possible functional redundancy (51). Regarding CRT, they result from deterministic mechanisms, in this context they can remain viable in the surrounding, non-optimal environment, and wait to (randomly) get in the optimal environment, where they are able to grow. Thus, CRT, in the host associated landscape, can be considered opportunistic. Whereas dominant symbionts are generalists and specialists (48).

Regarding the Arctic ocean, using the results from the NICE 2015 dataset, water masses are the main factors influencing the diversity of the prokaryotic rare biosphere, as previously described (8,30). Water masses induce dispersal limitation, a stochastic mechanism that results in transient rarity, that is also the main type of rarity found in this study. Other studies have reported permanent rarity in the Arctic Ocean, associated with deterministic patterns (8). But the concept of transient rarity can be considered a sub group of the permanent rarity, with the difference that transient OTUs disappear eventually. Thus, our results suggested that the marine prokaryotes in the Arctic region

studied are stochastically distributed by dispersal limitation, provoked by the water masses. But within the water masses, there are a set of specific conditions, leading to some rare OTUs survival at low abundances (PRT) and to the death of others (transiently rare). This framework is consistent with Jia et al. (15) by considering both stochastic and deterministic components in the description of the marine microbial rare biosphere.

Acknowledgments

This work was possible due to Prof. Rodrigo da Silva Costa, from the Biological Sciences Research Group, Department of Bioengineering, Instituto Superior Técnico (Lisbon, Portugal) and Prof. Catarina Maria Pinto Mora de Magalhães, from the Interdisciplinary Center of Marine and Environmental Research, University of Porto (Porto, Portugal). The work was financially supported by the Portuguese Science and Technology Foundation (FCT) through a grant to Francisco Pascoal on behalf of NITROLIMIT project (PTDC/CTS-AMB/30997/2017). This research was also partially supported by the "Programa Operacional Regional de Lisboa" (Project N. 007317) and the Strategic Funding UID/Multi/04423/2013 and UID/BIO/04565/2013 through national funds provided by FCT and European Regional Development Fund (ERDF), in the framework of the "PT2020" program.

References

1. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci [Internet]. 2006 Aug 8;103(32):12115–20. Available from: http://doi.wiley.com/10.1002/9781118010549.ch24
2. Caron D, Countway P. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquat Microb Ecol [Internet]. 2009 Nov 24;57(3):227–38. Available from: http://www.int-res.com/abstracts/ame/v57/n3/p227-238/
3. Coveley S, Elshahed MS, Youssef NH. Response of the rare biosphere to environmental stressors in a highly diverse ecosystem (Zodletone spring, OK, USA). PeerJ [Internet]. 2015 Aug 20;3:e1182. Available from: https://peerj.com/articles/1182
4. Pedrós-Alió C. The Rare Bacterial Biosphere. Ann Rev Mar Sci [Internet]. 2012 Jan 15;4(1):449–66. Available from: http://www.annualreviews.org/doi/10.1146/annurev-marine-120710-100948
5. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J [Internet]. 2017 Apr 10;11(4):853–62. Available from: http://www.nature.com/articles/ismej2016174
6. Pedrós-Alió C. Dipping into the Rare Biosphere. Science (80-) [Internet]. 2007 Jan;315(5809):192–3. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1135933
7. Szabó KÉ, Itor POB, Bertilsson S, Tranvik L, Eiler A. Importance of rare and abundant populations for the structure and functional potential of freshwater bacterial communities. Aquat Microb Ecol. 2007;47:1–10.
8. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. Ecology of the rare microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci [Internet]. 2009 Dec 29;106(52):22427–32. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0908284106
9. Lynch MDJ, Neufeld JD. Ecology and exploration of the rare biosphere. Nat Rev Microbiol [Internet]. 2015 Apr 2;13(4):217–29. Available from: http://www.nature.com/articles/nrmicro3400
10. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, et al. Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. MBio [Internet]. 2014 Aug 29;5(4):1–9. Available from: http://mbio.asm.org/cgi/doi/10.1128/mBio.01371-14
11. Vergin K, Done B, Carlson C, Giovannoni S. Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. Aquat Microb Ecol [Internet]. 2013 Nov 15;71(1):1–13. Available from: http://www.int-res.com/abstracts/ame/v71/n1/p1-13/
12. Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA. The Western English Channel

contains a persistent microbial seed bank. ISME J [Internet]. 2012 Jun 10;6(6):1089–93. Available from: http://dx.doi.org/10.1038/ismej.2011.162
13. Ai D, Chu C, Ellwood MDF, Hou R, Wang G. Migration and niche partitioning simultaneously increase species richness and rarity. Ecol Modell [Internet]. 2013 Jun;258:33–9. Available from: http://dx.doi.org/10.1016/j.ecolmodel.2013.03.001
14. Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev [Internet]. 2017 Dec 11;81(4):e00002-17. Available from: http://mmbr.asm.org/lookup/doi/10.1128/MMBR.00002-17
15. Jia X, Dini-Andreote F, Salles JF. Community Assembly Processes of the Microbial Rare Biosphere. Trends Microbiol [Internet]. 2018 Sep;26(9):738–47. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0966842X18300477
16. Jones SE, Lennon JT. Dormancy contributes to the maintenance of microbial diversity. Proc Natl Acad Sci [Internet]. 2010 Mar 30;107(13):5881–6. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0912765107
17. Pester M, Bittner N, Deevong P, Wagner M, Loy A. A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland. ISME J [Internet]. 2010 Dec 10;4(12):1591–602. Available from: http://dx.doi.org/10.1038/ismej.2010.75
18. Wang Y, Hatt JK, Tsementzi D, Rodriguez-R LM, Ruiz-Pérez CA, Weigand MR, et al. Quantifying the Importance of the Rare Biosphere for Microbial Community Response to Organic Pollutants in a Freshwater Ecosystem. Stams AJM, editor. Appl Environ Microbiol [Internet]. 2017 Apr 15;83(8):e03321-16. Available from: http://aem.asm.org/lookup/doi/10.1128/AEM.03321-16
19. Hausmann B, Pelikan C, Rattei T, Loy A, Pester M. Long-Term Transcriptional Activity at Zero Growth of a Cosmopolitan Rare Biosphere Member. Bailey MJ, editor. MBio [Internet]. 2019 Feb 12;10(1):1–16. Available from: http://mbio.asm.org/lookup/doi/10.1128/mBio.02189-18
20. Gobet A, Quince C, Ramette A. Multivariate Cutoff Level Analysis (MultiCoLA) of large community data sets. Nucleic Acids Res [Internet]. 2010 Aug;38(15):e155–e155. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq545
21. Liu M, Xue Y, Yang J. Rare Plankton Subcommunities Are Far More Affected by DNA Extraction Kits Than Abundant Plankton. Front Microbiol [Internet]. 2019;10(March):1–12. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5632291/
22. Karimi E, Ramos M, Gonçalves JMS, Xavier JR, Reis MP, Costa R. Comparative Metagenomics Reveals the Distinctive Adaptive Features of the Spongia officinalis Endosymbiotic Consortium. Front Microbiol [Internet]. 2017 Dec 14;8(2499). Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2017.02499/full
23. de Sousa AGG, Tomasino MP, Duarte P, Fernández-Méndez M, Assmy P, Ribeiro H, et al. Diversity and Composition of Pelagic Prokaryotic and Protist Communities in a Thin Arctic Sea-Ice Regime. Microb Ecol. 2019;78(2):388–408.
24. Granskog M, Assmy P, Gerland S, Spreen G, Steen H, Smedsrud L. Arctic Research on Thin Ice: Consequences of Arctic Sea Ice Loss. Eos (Washington DC) [Internet]. 2016 Jan 26;97. Available from: https://eos.org/project-updates/arctic-research-on-thin-ice-consequences-of-arctic-sea-ice-loss
25. Sousa AG. Arctic microbiome and N-functions during the winter-spring transition. Porto University; 2017.
26. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, et al. EBI Metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res [Internet]. 2018;46(D1):D726–35. Available from: https://academic.oup.com/nar/article-abstract/46/D1/D726/4561650
27. R Core Team. R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria [Internet]. 2018; Available from: https://www.r-project.org/
28. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. PLoS One [Internet]. 2013 Apr 22;8(4):e61217. Available from: https://dx.plos.org/10.1371/journal.pone.0061217
29. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Community Ecology Package. R Package Version 2.5-3 [Internet]. 2018. Available from: http://cran.r-project.org/package=vegan
30. Kirchman DL, Cottrell MT, Lovejoy C. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environ Microbiol [Internet]. 2010 Feb 3;12(5):1132–43. Available from: http://doi.wiley.com/10.1111/j.1462-2920.2010.02154.x
31. Ruiz-González C, Logares R, Sebastián M,

Mestre M, Rodríguez-Martínez R, Galí M, et al. Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. Mol Ecol. 2019;28(8):1930–45.
32. Gonnella G, Böhnke S, Indenbirken D, Garbe-Schönberg D, Seifert R, Mertens C, et al. Endemic hydrothermal vent species identified in the open ocean seed bank. Nat Microbiol [Internet]. 2016 Aug 13;1(8):1–7. Available from: http://dx.doi.org/10.1038/nmicrobiol.2016.86
33. Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. Activity of abundant and rare bacteria in a coastal ocean. Proc Natl Acad Sci [Internet]. 2011 Aug 2;108(31):12776–81. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1101405108
34. Gobet A, Böer SI, Huse SM, van Beusekom JEE, Quince C, Sogin ML, et al. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. ISME J [Internet]. 2012 Mar 6;6(3):542–53. Available from: http://www.nature.com/doifinder/10.1038/ismej.2011.132
35. Bowen JL, Morrison HG, Hobbie JE, Sogin ML. Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. ISME J [Internet]. 2012 Nov 28;6(11):2014–23. Available from: http://dx.doi.org/10.1038/ismej.2012.47
36. Anderson RE, Sogin ML, Baross JA. Biogeography and ecology of the rare and abundant microbial lineages in deep-sea hydrothermal vents. FEMS Microbiol Ecol [Internet]. 2015 Jan 1;91(1):1–11. Available from: https://academic.oup.com/femsec/article-lookup/doi/10.1093/femsec/fiu016
37. Baltar F, Palovaara J, Vila-Costa M, Salazar G, Calvo E, Pelejero C, et al. Response of rare, common and abundant bacterioplankton to anthropogenic perturbations in a Mediterranean coastal site. FEMS Microbiol Ecol [Internet]. 2015 Jun 1;91(6):1–12. Available from: https://academic.oup.com/femsec/article-lookup/doi/10.1093/femsec/fiv058
38. Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. ISME J [Internet]. 2009 Jul 26;3(7):860–9. Available from: http://dx.doi.org/10.1038/ismej.2009.23
39. Sjöstedt J, Koch-Schmidt P, Pontarp M, Canbäck B, Tunlid A, Lundberg P, et al. Recruitment of Members from the Rare Biosphere of Marine Bacterioplankton Communities after an Environmental Disturbance. Appl Environ Microbiol [Internet]. 2012 Mar 1;78(5):1361–9. Available from: http://aem.asm.org/lookup/doi/10.1128/AEM.05542-11
40. Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci [Internet]. 2013 Apr 9;110(15):6004–9. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1216863110
41. Alonso-Sáez L, Zeder M, Harding T, Pernthaler J, Lovejoy C, Bertilsson S, et al. Winter bloom of a rare betaproteobacterium in the Arctic Ocean. Front Microbiol. 2014;5(AUG):1–9.
42. Quero GM, Luna GM. Diversity of rare and abundant bacteria in surface waters of the Southern Adriatic Sea. Mar Genomics [Internet]. 2014 Oct;17:9–15. Available from: http://dx.doi.org/10.1016/j.margen.2014.04.002
43. Crespo BG, Wallhead PJ, Logares R, Pedrós-Alió C. Probing the Rare Biosphere of the North-West Mediterranean Sea: An Experiment with High Sequencing Effort. Martinez-Abarca F, editor. PLoS One [Internet]. 2016 Jul 21;11(7):e0159195. Available from: http://dx.doi.org/10.1371/journal.pone.0159195
44. Hamasaki K, Taniguchi A, Tada Y, Kaneko R, Miki T. Active populations of rare microbes in oceanic environments as revealed by bromodeoxyuridine incorporation and 454 tag sequencing. Gene [Internet]. 2016 Feb;576(2):650–6. Available from: http://dx.doi.org/10.1016/j.gene.2015.10.016
45. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S. Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. ISME J [Internet]. 2018;12(9):2198–210. Available from: http://dx.doi.org/10.1038/s41396-018-0153-6
46. Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, et al. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. Environ Microbiol [Internet]. 2009 Oct;12(8):2070–82. Available from: http://doi.wiley.com/10.1111/j.1462-2920.2009.02065.x
47. Taylor MW, Tsai P, Simister RL, Deines P, Botte E, Ericson G, et al. 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. ISME J [Internet]. 2013 Feb 4;7(2):438–43. Available from: http://www.nature.com/articles/ismej2012111
48. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun [Internet]. 2016 Dec 16;7(1):11870. Available from: http://www.nature.com/articles/ncomms11870

49. Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, Sogin ML, et al. Host-specificity among abundant and rare taxa in the sponge microbiome. ISME J [Internet]. 2014 Jun 9;8(6):1198–209. Available from: http://www.nature.com/articles/ismej2013227
50. Schönberg CHL. Happy relationships between marine sponges and sediments – a review and some observations from Australia. J Mar Biol Assoc United Kingdom [Internet]. 2016 Mar 4;96(2):493–514. Available from: https://www.cambridge.org/core/product/identifier/S0025315415001411/type/journal_article
51. Hardoim CCP, Costa R. Temporal dynamics of prokaryotic communities in the marine sponge Sarcotragus spinosulus. Mol Ecol [Internet]. 2014 Jun;23(12):3097–112. Available from: http://doi.wiley.com/10.1111/mec.12789