Institut Pasteur
Research in Microbiology 166 (2015) 831−841
www.elsevier.com/locate/resmic

Rarity in aquatic microbes: placing protists on the map

Ramiro Logares*, Jean-François Mangot, Ramon Massana

Institute of Marine Sciences, CSIC, Passeig Maritim de la Barceloneta 37-49, ES08003, Barcelona, Catalonia, Spain

Received 6 April 2015; accepted 28 September 2015
Available online 24 October 2015

Abstract

Most microbial richness at any given time tends to be represented by low-abundance (rare) taxa, which are collectively referred to as the "rare biosphere". Here we review works on the rare biosphere using high-throughput sequencing (HTS), with a particular focus on unicellular eukaryotes or protists. Evidence thus far indicates that the rare biosphere encompasses dormant as well as metabolically active microbes that could potentially play key roles in ecosystem functioning. Rare microbes appear to have biogeography, and sometimes the observed patterns can be similar to what is observed among abundant taxa, suggesting similar community-structuring mechanisms. There is limited evidence indicating that the rare biosphere contains taxa that are phylogenetically distantly related to abundant counterparts; therefore, the rare biosphere may act as a reservoir of deep-branching phylogenetic diversity. The potential role of the rare biosphere as a bank of redundant functions that can help to maintain continuous ecosystem function following oscillations in taxonomic abundances is hypothesized as its main ecological role. Future studies focusing on rare microbes are crucial for advancing our knowledge of microbial ecology and evolution and unveiling their links with ecosystem function.

Keywords: Rare-biosphere; Microeukaryotes; Abundance; High-throughput sequencing

1. Unveiling the rare biosphere in planktonic microbes

Already in 1859, Charles Darwin wrote that "rarity is the attribute of a vast number of species in all classes, in all countries". Since then, until today, a variety of classical ecological studies have shown that most communities of animals and plants are composed of a few abundant species and many rare ones [1]; a pattern explored thoroughly in multiple studies [2]. Determining whether or not microbial communities followed similar species abundance distributions was a question that researchers were able to start investigating only a few decades ago, with the introduction of molecular tools in microbial ecology during the late 1980s and early 1990s. In particular, tools like PCR, DGGE, Sanger sequencing and FISH enabled recovering culturable and unculturable microbes from natural communities [3], providing an approximation of their in situ abundances [4]. Yet, these techniques led to recovery of only the most abundant members of the community, and the exploration of rare microbial species had to wait until the arrival of high-throughput DNA sequencing (HTS) technologies during the late 2000s [5]. HTS caused a revolution in microbial diversity studies, since it led to recovering a number of taxa from communities far beyond what was possible with earlier molecular tools at a fraction of the cost (€ per sequenced nucleotide). Interestingly, the initial studies, sequencing with Roche-454 [5], the ribosomal RNA gene from marine bacterial communities, indicated not only that bacterial communities followed species-abundance distribution patterns similar to those observed in multicellular organisms, but also that the number of rare species was unexpectedly large. This large amount of rare taxa was called the "rare biosphere" [6]. Despite the fact that initial estimations of operational taxonomic units (OTUs; a proxy of species) based on HTS were inflated due to errors during PCR and sequencing [7−9], later

---

* Corresponding author.
E-mail addresses: ramiro.logares@gmail.com (R. Logares), mangot@icm.csic.es (J.-F. Mangot), ramonm@icm.csic.es (R. Massana).

http://dx.doi.org/10.1016/j.resmic.2015.09.009
0923-2508/

HTS studies indicated that most OTUs in marine bacterial communities were rare [10] and that at least a fraction of them were metabolically active [11,12]. The exploration of the rare biosphere in protists followed the advance observed in prokaryotes [13,14]. Eventually, HTS studies started to indicate that protists also have a rare biosphere [15–20], which includes taxa that can be metabolically active [21].

Once multiple lines of evidence supported the existence of a rare biosphere in both microeukaryotic (i.e. unicellular eukaryotes) and prokaryotic communities, then several questions emerged related to its relevance for: 1) community structure; 2) ecosystem functioning; and 3) evolution [13,22]. For example, from a community structure perspective, researchers may ask: Why do microbial communities harbor so many species and why are a handful of them abundant, while most are rare? Do spatio-temporally different microbial communities harbor similar proportions of rare and abundant taxa? If so, what processes contribute to stable proportions across communities? How often do taxa oscillate between abundant and rare categories? Could kill-the-winner dynamics explain transitions from abundance to rarity and promote highly diverse communities with a high turnover of abundant taxa? [23]. In terms of ecosystem functioning, the discovery of the rare biosphere also had important implications and generated key questions. Firstly, as most species are rare, then only a handful of taxa would be central at any given time to general community metabolisms (e.g. carbon fixation, nutrient cycling, respiration) [10]. Secondly, could rare taxa have specific key functions in the community? Thirdly, could rare taxa replace abundant ones after a disturbance, thus contributing to uninterrupted ecosystem functioning? [13] And lastly, which proportion of rare species is metabolically active? (that is, not dormant [24]). The rare biosphere also poses evolutionary questions: Is it possible that some species evolve towards permanent rarity after positive selection for slow-growth traits? Or would all rare taxa eventually start growing under optimal environmental conditions?

In the following sections, we will discuss the above-mentioned questions, focusing in protistan communities (of which we know the least) in a context of what is known about prokaryotes.

2. Exploring the rare biosphere

2.1. What is abundant or rare?

The concept of abundance seems to have different meanings for ecologists and taxonomists [25]. Ecologists normally define abundant or rare taxa in terms of local relative abundance (that is, numbers of individuals per location). Alternatively, taxonomists may consider one species as abundant or rare if it is present in more or less than *x* % of the investigated locations (that is, frequency of occurrence). This could lead to confusion, as one species that, for example, is present in >80% of the sampled locations, with an abundance <0.01% of the community in each location, could be considered as rare by an ecologist and abundant by a taxonomist. In this review, we will concentrate on the ecological perspective of abundance, considering species abundant or rare depending on local estimates of individual relative abundance.

Another important consideration when exploring rarity is related to spatial scales. Let's define an oceanic microbe as rare if, in every single location, it is found in abundances of about 1 cell every 100 L (this microbe would seldom be captured by most sampling designs, which collect 0.5–20 L of marine water per location). Even though we agree that *locally* (that is, in one specific location) this species is rare, in a global oceanic perspective it could reach an important absolute population size. In 1 km^3 of oceanic water, our hypothetical species would have 1×10^{10} cells (that is, ten thousand million cells per cubic kilometer of oceanic water). A volume of 1 km^3 is insignificant in the ocean. For example, the upper 200 m (epipelagic zone) of the water column in the North Pacific Subtropical Gyre encompass a volume of 4 million km^3 [26]. In this volume of water, our theoretical species would reach a population size of 4×10^{16} cells. Then, it is important to contemplate the fact that local rarity does not necessarily imply global rarity, and that even though a global population of 4×10^{16} cells is low compared to the total number of estimated bacterial cells in the open ocean (1.2×10^{29}) [27], such a population could still have importance for ecosystem function [22]. These considerations are important for the interpretation of results as well as for sampling and experimental design. Yet, in this work, we will consider oceanic taxa as rare as in most studies thus far, which means taxa with local abundances that do not surpass an arbitrary threshold of rarity in a volume of water that is normally 0.5–20 L, and from which only a small aliquot is subsampled.

Before moving into deeper explorations of the rare biosphere, one important aspect that needs to be discussed is how it is defined [28]. Typically, arbitrary abundance cut-offs are used for determining abundant or rare species. A number of studies have used local abundances >1% and <0.01% to define abundant or rare OTUs, respectively [21,29], leaving a range of abundances (0.01% < OTU*x* <1%) to accommodate OTUs with intermediate abundances. Other studies included more relaxed cut-offs for rarity (e.g. <0.1%) [22,30]. Furthermore, abundant/rare cut-offs should be considered in terms of space or time. When using data from a single sampling event, one gets a snapshot of the community composition at a given time, and what is rare or abundant at one time may not be rare or abundant at another time. For these cases, we define the term *instantly rare or abundant*, referring to OTUs that are rare or abundant at a given time, and for which there is no evidence they were or they will be in the same category in the past or future, respectively (Table 1). In any case, when investigating a set of communities within a given geographical region, it is possible to explore rarity and abundance in the entire regional community [21] (Table 1). In that case, the regional abundance of OTUs would be represented by the average of local abundances including zero values [31]. Here, one could detect cases of OTUs that are instantly rare in most locations but instantly abundant in one or vice versa, leading to regionally rare or abundant OTUs, respectively. Since