Marine microbial diversity: can it be determined?

Carlos Pedrós-Alió

Institut de Ciències del Mar, CMIMA-CSIC, P. Marítim de la Barceloneta 37-42, 08003 Barcelona, Spain

Estimates of the order of magnitude for the total number of microbial species on Earth range from $10^3$ to $10^8$. Despite global dispersal of microorganisms, this number is probably rather large. The total biodiversity of an ecosystem is composed of two elements: first, a set of abundant taxa that carry out most ecosystem functions, grow actively and suffer intense losses through predation and viral lysis. These taxa are retrievable with molecular techniques but are difficult to grow in culture. Second, there is a seed bank of many rare taxa that are not growing or grow extremely slowly, do not experience viral lysis and predation is reduced. Such taxa are seldom retrieved by molecular techniques but many can be grown in culture, which explains the dictum ‘everything is everywhere’.

How many species?

Ecologist Robert May remarked that “we have a catalog of all the celestial bodies our instruments can detect in the universe, but we ignore how many living beings share the Earth with us” (ISME meeting, Barcelona, 1992). This ignorance is particularly large for life in the oceans. Thus, the international initiative ‘Census of Marine Life’ [1] was launched to “assess and explain the changing diversity, the distribution and abundance of marine species from the past to the present, and to project future ocean life.” The purpose of the initiative is to make a census of the ‘known’ diversity in the oceans and to determine what portion of the ‘unknown’ can be determined to focus research in the most efficient way.

In the case of the diversity of microorganisms, even the right order of magnitude is unknown and the issue is highly controversial [2–4]. The total number of prokaryotic cells in the oceans has been estimated to be $\sim 10^{29}$ [5] but the question is: into how many species are these cells partitioned? This question is not of purely academic interest [6]. First, the unknown diversity hides novel metabolisms (such as the recently discovered photoheterotrophy in the sea) that force a re-evaluation of carbon and energy fluxes in the oceans [7–9]. These processes must be understood so that, for example, precise models of global change can be constructed. Second, the unknown microorganisms are the largest potential reservoir of useful genes for medicine and biotechnology. And, of course, knowledge of microbial diversity is essential to understand evolution [10].

During the past 15 years, molecular studies using the gene sequences that encode the small subunit rRNA (SSU rDNA) have revealed a wealth of new marine microorganisms that belong to the three realms of life: Bacteria [11], Archaea [12,13] and Eukarya [14–16]. The feeling among marine microbiologists is that of living through an age of discovery with no end in sight. However, most of this new diversity has not yet been described because pure cultures of the organisms behind the sequences are necessary to define a species (Box 1). How far has the goal of the Census of Marine Life been fulfilled? First, the known diversity has to be defined; second, the size of the unknown has to be estimated; and finally, I propose a framework within which both can be studied with different strategies.

The known and the unknown

Approximately 6000 species of prokaryotes (http://www.bacterio.cict.fr/number.html) and 100 000 species of protists [17] have been formally described. This is the known...
Box 1. What is a microbial species?

There are >20 definitions of species for eukaryotes [53], none of which is satisfactory for all living eukaryotes. It is no surprise that the definition for prokaryotes is also controversial [35]. The morphological species concept is clearly not applicable to prokaryotes and to many of the smallest eukaryotes [36,37,54]. Rosselló-Mora and Amann [35] defend the phylo-phenetic species definition for bacteria as pragmatic, useful, and applicable to all prokaryotes. This definition uses information from SSU rDNA. As a rule of thumb, it is accepted that two strains must show a similarity >97% to be considered part of the same species. However, SSU rDNA does not have enough resolution to delineate species [35,55]. A more powerful technique is DNA–DNA hybridization. Usually, two strains must show >70% hybridization to be considered as the same species. Strains that show >97% similarity in their SSU rDNA show a range of values in DNA–DNA hybridization between <20% and 70%. By contrast, no two strains with >70% DNA–DNA hybridization show a similarity of <97% in SSU rDNA. Obviously, taxa based on a 97% SSU rDNA similarity widely underestimate the number of species if the DNA–DNA hybridization criterion is used. In conclusion, to argue that the number of microbial species is high using the SSU rDNA criterion to count taxa (instead of vice versa) means always underestimating the true number of microbial species.

For the purpose of the present article, however, a definition of a species is not essential. Fortunately, diversity can be calculated with other units besides species. Individuals must be grouped into non-overlapping classes according to a consistent classification criterion [56]. Two descriptors that fulfill these requirements (and are appropriate for microbes) are the number of DNA bands in denaturing gradient gel electrophoresis or the number of clones in a DNA clone library. These units are operational taxonomic units (OTU). The same arguments that are used in the main text for taxa (or OTUs) can be used for species if a satisfactory definition is eventually found [57].

diversity. Most information about these microorganisms comes from the study of pure cultures, at least in the case of prokaryotes, in which a strain deposited in a culture collection is necessary to describe a new species.

The unknown diversity is currently being explored with molecular techniques, particularly cloning and sequencing [6]. A library is constructed with the DNA from a particular sample. Each clone has the SSU rDNA of one member of the original community. As more clones are sequenced, new taxa arise. At the beginning of this process, almost every new clone will belong to a new taxon (Figure 1a) but, as more clones are examined, the curve of taxa versus sequenced clones should approach an asymptote. This would indicate that most diversity in the library has been retrieved. In some cases, the curve does not approach an asymptote but instead increases almost as a straight line. This indicates that many more clones must be sequenced to retrieve the diversity in the library. There are statistical tools to estimate the richness (see Glossary) of a library [18] and different authors have used results from clone libraries and statistical procedures to estimate a global number of microbial taxa in the ocean. Thus, Hagström et al. [19] suggested that only a few thousand bacterial taxa were present in the sea, judging from the sequences deposited in GenBank. More recently, these authors estimated that ~360 new bacterial taxa are submitted to GenBank every year. This will increase the total number of taxa but it does not seem to require a major revision of their order of magnitude (between 10^3 and 10^4 taxa) [20].

These estimates are based on the use of clone libraries but this approach is insufficient to retrieve all of the diversity present in nature (see later). Different authors have provided estimates that would compensate for the shortcomings of the cloning approach. Curtis et al. [21] used the taxon–abundance curve (Figure 1b). The assumption is that this relationship follows a lognormal
distribution. In this case, the total number of taxa can be estimated from the total number of individuals (\(N_t\)), and the abundance of the most common taxon (\(N_{\text{max}}\)). The authors' estimate for the whole ocean was \(10^8\) bacterial taxa. Dykhuisen [22] used a completely different approach and compared differences between microbial communities using total DNA hybridization. The author then assumed a certain number of different microbial communities and estimated that \(10^9\) different bacterial taxa are present on Earth. These estimates use different definitions of a microbial taxon but the essential argument is that, regardless of the particular definition, the total number is huge (Box 1).

A different school of thought holds that the global number of microbial taxa is low, perhaps in the order of \(10^3–10^4\) [2,23–25]. These studies refer to morphospecies of ciliates but then extrapolate to all microorganisms, basing their conclusion on a fact and its corollary. The fact is that because microorganisms are small and have many individuals, they are easily dispersed globally. This is an explanation for the classical dictum that ‘everything is everywhere, the environment selects’ [26]. Most, if not all, microbiologists will agree with this assertion [3,6]. The corollary states that because allopatric speciation will be unlikely, the global number of species will be low [23]. Allopatric speciation takes place when a geographical barrier separates two populations from the same species. The two populations will eventually evolve into separate species if there is no genetic exchange between them (Box 2).

However, this corollary does not hold. First, allopatry is not necessary for speciation [27]. Several mechanisms of sympatric speciation (with the two populations coexisting in the same area; Box 2) have been described for sexually reproducing organisms such as birds [28] or arthropods [29]. Sympatric speciation should be easier for asexual organisms such as bacteria and many eukaryotic microorganisms. Second, studies of bacteria in continuous populations such as bacteria and many eukaryotic microorganisms. Speciation can also occur when the two populations are separated by a geographical barrier (for example, each population colonizes a different island in an archipelago). This is called allopatric speciation. Speciation can also occur when the two populations coexist. This is called sympatric speciation. It is easier to envision how two populations can differ in allopatry. They might find different resources in the two islands or there could be genetic drift in different directions, for example. However, several mechanisms that enable speciation in sympatry have been described [27]. In fact, studies of the mechanisms of speciation in sympatry were considered part of the ‘breakthrough of the year 2005’ by Science [58].

An example among marine microorganisms is that of *Micromonas pusilla*, an abundant and widespread prasinophyte. Isolates from different parts of the ocean are morphologically undistinguishable and, thus, have been characterized as one cosmopolitan morphospecies. Slapeta et al. [59] studied a suite of nuclear, mitochondrial and chloroplast genes in *M. pusilla* isolates from different oceans. They found three or four lineages that were sufficiently different to merit species status. Some of these lineages were distributed globally and several of them co-existed in the same places. The authors concluded that these lineages were diversifying “in the absence of apparent barriers (sympatric speciation).”

A case among bacteria was experimentally studied in the laboratory [30]. A culture of *Escherichia coli* was grown on minimal medium in a continuous culture where the only source of carbon and energy was glucose. In this situation, *E. coli* takes up the glucose and degrades it to acetate and some glycerol, which are both excreted into the medium and can be used later. Yet after ~800 generations, three different ecosspecies were found to coexist. One was specialized in the use of glucose, another in acetate and the third in glycerol (the ecosspecies consisted of 80%, 18% and 2% of the cells, respectively). Thus, three niches had appeared where there was only one and three ecosspecies had evolved in one of the most constant and featureless environments possible. It is hard to believe that this phenomenon has not been taking place in the much more complex real world to create uncountable niches for different microbial species to develop.

### Box 2. Allopatric and sympatric speciation

Speciation requires that two populations of the same species become reproductively isolated from each other. Then each population can diversify from the other until they become two species. Speciation can occur when the two populations are separated by a geographical barrier (for example, each population colonizes a different island in an archipelago). This is called allopatric speciation. Speciation can also occur when the two populations coexist. This is called sympatric speciation. It is easier to envision how two populations can differ in allopatry. They might find different resources in the two islands or there could be genetic drift in different directions, for example. However, several mechanisms that enable speciation in sympatry have been described [27]. In fact, studies of the mechanisms of speciation in sympatry were considered part of the ‘breakthrough of the year 2005’ by Science [58].
Compostela, Spain, 2005) used their extensive collection of P. delicatissima strains isolated from the Bay of Naples to compare morphology, gene sequences and the ability to reproduce sexually. A strength of this system is that whether two strains interbreed or not determines if they belong to the same biological species. It was found that some strains with the same large subunit rDNA did not mate, whereas others did. However, when the more variable ITS region (interspace region between the different rRNA genes) was compared in these strains, strains that mated always had the same ITS sequence and strains with different ITS sequences never mated. Thus, rDNA sequences are a conservative estimate of species, not the other way around. Similar studies have been carried out with freshwater green algae [37].

If the number of microbial species is close to the lower estimates proposed by Finlay and Fenchel, a description of a majority of microbial species has already been carried out. If the true numbers are close to the upper estimates, however, the challenge ahead is formidable. From the previous discussion I think it is likely that the number of microbial taxa approaches the upper estimates. The difference between the known diversity (the $10^9$ described bacterial species) and the estimated total diversity ($10^6–10^9$ bacterial taxa) is the unknown. Obviously, the isolation and characterization of $10^6$ or $10^9$ bacteria are beyond the capabilities of current microbiologists. Thus, one might be inclined to agree with the pessimistic thought that “determining the number of species today is like reaching for the stars; there is no way with the data available today that we are even going to get reasonably close” [38].

Diversity and biodiversity

A way out can be found in the distinction between diversity and biodiversity proposed by Margalef [39,40]. Biodiversity would be the total genetic information on Earth or in any part of it. Diversity, in turn, would be the components that are active and abundant at one particular time and place. A literary analogy would equate biodiversity with a dictionary of all the words in a language, whereas diversity would be the particular set of words (and their frequency of use) chosen for a particular book. This distinction makes a lot of biological sense. Magurran and Henderson [41], for example, analyzed an assemblage of fish in an estuary over a period of 21 years. Some species were present every year and their abundance was high. They called these ‘core’ species. The remaining species had low abundance and were not found every year. These were ‘occasional’ species. The authors proposed that core species are responsible for carbon and energy flow in the ecosystem whereas occasional species wander sporadically to areas at their limits of tolerance and eventually disappear.

These ideas can be applied to microorganisms (Figure 2). The most abundant taxa are the core taxa. They are maintained through active growth because they are well adapted to that particular ecosystem. At the same time, they are subject to predation and viral lysis, thus fuelling carbon and energy flow. These taxa constitute the diversity of the ecosystem. The long tail of rare taxa (or ‘seed bank’) would complete the biodiversity. The taxa in the seed bank would be the equivalent of occasional species. However, in the case of microorganisms, being a member of a rare taxon has fascinating consequences.

Delights of a diluted life

The rare taxa are recruited through immigration, which is dependent on dispersal from other ecosystems. It has already been argued that because of their small size and large numbers, microbes should be easily dispersed everywhere [2,3,6,23,26], thus, immigration rates are expected to be large. By contrast, death and subsequent loss will eliminate taxa from the rare part of the curve, however; these taxa can be expected to have extremely low loss rates. The two main loss factors for microorganisms are viral lysis and predation by protists [42], so by being rare, a microbial taxon is protected from both.

Viruses depend on encounter probabilities to find a prey. If the prey taxon is found below a minimal threshold concentration it will become extremely improbable that the virus finds any of its members. All of the taxa with an abundance lower than this threshold will be protected from this loss factor. Viruses function by a ‘kill the winner’ strategy [43]. When a microbial taxon grows actively and increases its abundance above the threshold, its specific virus will infect it and reduce its numbers back to the threshold level.

In the case of predation, if the predators are not selective, both abundant and rare taxa will be culled down proportionally and being rare will not be advantageous. My assumption is that rare taxa are rare because they are not growing (or growing slowly) compared with the abundant members of the assemblage. It is well known that bacterivores selectively prey on the largest and most active bacteria [42,44,45]. In fact, a predation resistance strategy is to become small, and actively growing bacteria are known to be bigger than starving bacteria. Rare taxa, therefore, will be represented by smaller cells and will be more resistant to grazing than the abundant, actively growing cells. In effect, the slow-growing taxa will suffer far less predation losses than their more abundant competitors.

As a consequence of the small losses, a rather long tail of rare taxa in microbial assemblages can be expected (Figure 2). An important property of this curve is that rare taxa can be promoted from the seed bank to the core zone of the curve. This might happen in a few days or weeks if conditions become adequate for the growth of the particular taxon. Likewise, a member of the core can be grazed down or reduced by viral lysis beyond the threshold level to become a member of the rare taxa without disappearing from the ecosystem. If it were true that all microbes are easily dispersed globally, the long tail of any ecosystem would include all of the microorganisms on Earth. In effect, the biodiversity of any ecosystem would be identical to the biodiversity on the planet. Ecosystem functions, however, would be carried out by those taxa that make up the diversity of that particular ecosystem at any given time point.
What portion of the unknown can be determined?

Rare taxa will seldom be retrieved by cloning and sequencing with universal primers. PCR primers will predominantly hybridize with the common taxa and the rare ones will remain unknown for the same reasons that viruses cannot find rare bacterial taxa in a reasonable amount of time. The situation can be improved by using primers specific for some groups of bacteria (such as ammonia oxidizers [46]) by techniques such as nested PCR [47] and by shotgun cloning, in which PCR is not used [48,49]. However, a large portion of the seed bank will remain beyond the reach of current molecular tools. Essentially, only the abundant taxa will be accessible (Figure 2b). Some studies indicate that taxa which make up $\geq 1\%$ of the total cell number can be detected with PCR-dependent techniques, and taxa that comprise $<0.1\%$ are difficult to retrieve [50,51].

I propose that the taxa retrieved by PCR are considered to form the diversity of an ecosystem. Then, by definition, diversity can be determined with current molecular techniques. Abundant taxa such as SAR11 (a group of bacteria that comprise 20–30% of the total count in many oceans) will certainly be found in this group. Taxa that are usually less abundant but important in nutrient cycles, such as ammonia oxidizers, will fall within this ‘abundant’ group in most instances because their concentrations are close to 1% of the total [52]. Thus, most bacteria that are relevant and active in carbon, energy and nutrient flows will be within the diversity retrievable by molecular techniques.

Concluding remarks and future perspectives

Studies of bacterial diversity within this framework make sense from an ecological point of view. Thus, for example,
latitudinal patterns in the distribution of the bacterial taxa that form the diversity have been found by cloning and sequencing large clone libraries from different oceans (T. Pommier et al., unpublished). If the seed bank from such samples could have been analyzed simultaneously, it is likely that ‘everything would be found everywhere’.

Some rare taxa will eventually be discovered with conventional molecular techniques. However, most of the rare taxa in the seed bank will only be accessible if they can be grown in enrichment or, even better, pure cultures. As has been learnt the hard way, this is only possible with a few of the microorganisms that form biodiversity [6].

Within this framework, microbial ecologists who are interested in functional aspects of carbon flow only need to be concerned with diversity (i.e. with the taxa retrievable with a clone library). From the point of view of general ecologists, the comparison of the peculiar situation of the rare microbial taxa with that of rare species of animals and plants is thought-provoking. And for microbiologists who are interested in metabolisms or industrially interesting genes, the realization that a large tail of biodiversity is not currently accessible represents a tremendous challenge to develop ingenious molecular and culturing methods. It is suggestive that in both the cases of stars in an expanding universe and rare bacteria in a diluted microscopic world, the depth of our knowledge depends on the acuity of our tools, whether the tools are a microscope or a telescope.

Acknowledgements

My work is currently supported by grant CTM2004–02586/MAR from the Spanish Ministry of Education and Science and the EU network of excellence ‘Marine Genomics Europe’ EU-FP6–505403. Ramon Massana, Josep M. Gasol, Jarone Pinhasi, Thomas Pommier and three anonymous reviewers provided useful comments on earlier versions of the manuscript.

References


---

AGORA initiative provides free agriculture journals to developing countries

The Health Internetwork Access to Research Initiative (HINARI) of the WHO has launched a new community scheme with the UN Food and Agriculture Organization.

As part of this enterprise, Elsevier has given 185 journals to Access to Global Online Research in Agriculture (AGORA). More than 100 institutions are now registered for the scheme, which aims to provide developing countries with free access to vital research that will ultimately help increase crop yields and encourage agricultural self-sufficiency.

According to the Africa University in Zimbabwe, AGORA has been welcomed by both students and staff. ‘It has brought a wealth of information to our fingertips’ says Vimbai Hungwe. ‘The information made available goes a long way in helping the learning, teaching and research activities within the University. Given the economic hardships we are going through, it couldn’t have come at a better time.’

For more information visit:
http://www.healthinternetwork.net

---

Elsevier.com – Dynamic New Site Links Scientists to New Research & Thinking

Elsevier.com has had a makeover, inside and out.

As a world-leading publisher of scientific, technical and health information, Elsevier is dedicated to linking researchers and professionals to the best thinking in their fields. We offer the widest and deepest coverage in a range of media types to enhance cross-pollination of information, breakthroughs in research and discovery, and the sharing and preservation of knowledge. Visit us at Elsevier.com.