INSTITUT PASTEUR

Research in Microbiology 159 (2008) 67–73

www.elsevier.com/locate/resmic

Tracking microbial biodiversity through molecular and genomic ecology

Abstract

Molecular ecology and metagenomics applied to the study of microbial biodiversity are changing our comprehension of the biosphere. An impressive diversity of archaea, bacteria and, more recently, protists has been uncovered by molecular tools. Efforts to couple function to the phylogenetic diversity observed in natural environments are leading to the discovery of novel metabolisms and to a re-evaluation of the global ecological impact of known ones. Interesting questions relating to mechanisms of speciation and evolutionary trends at the smallest and largest phylogenetic scales are emerging.

Keywords: Molecular ecology; Genomic ecology; Metagenomics; Microbial biodiversity

1. Introduction

One of the first and most successful applications of molecular phylogeny was the recognition of the Archaea and the building of a tripartite tree of life by C.R. Woese and collaborators from the late 1970s. Since then, microbiology lives under a permanent revolution, being one of the most fast-moving scientific disciplines. Vast fields of exploration are opening up for microbiology, and exciting questions whose answers were thus far largely inaccessible by only classical approaches can now be tackled by their combination with molecular and genomic tools.

How many microbial species are there? Over the last two decades, the widespread use of molecular methods based on the amplification, cloning and sequencing of small subunit ribosomal RNA (SSU rRNA) genes from the environment led us to open the black box of prokaryotic diversity in natural communities. Operational taxonomic units (OTUs) based on SSU rRNA sequence similarity have become indispensable proxies to the otherwise intangible concept of microbial species. What is the real extent of this diversity? Despite many thousands of environmental gene and (meta)genome sequences accumulating in databases, not even an approximate answer is in close view, and the application of the required ecological and statistical tools that could help to provide fair estimates is just starting. To the colossal diversity of archaea and bacteria detected by molecular tools can be added a wealth of microbial eukaryotic sequences, which are revealing an unsuspected variety of protist taxa that had escaped identification by classical protistology. Although they fall out of our scope here, viruses on their own represent a whole understudied world. DNA and RNA viruses outnumber by and large prokaryotes in oceans and soils, and their diversity is only beginning to be roughly drawn by metagenomic analyses.

What does all that microbial biodiversity do? How do ecosystems ultimately work? Despite massive advances in evaluating microbial diversity, its actual contribution to ecosystem functioning is still elusive. Substantial progress is being made though improved methodologies, including multidisciplinary approaches and metagenomic analyses, in the context of more discovery-prone conceptual frameworks, which are partly made possible by the finding of novel, sometimes unexpected metabolisms that are changing our understanding of global biogeochemical cycles.

How do species evolve? Why are there so many lineages? How did major taxa split and diversify? The study of microbial evolution is also largely benefiting from environmental sequencing by allowing access to reservoirs of genotypes from uncultured organisms, thus providing raw data to study micro- and macroevolutionary issues.

In the following, we very briefly summarize current lines and perspectives of research about these questions.

2. The extent of microbial diversity

The intensive application of molecular techniques to describe microbial diversity in natural environments is yielding

0923-2508/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2007.11.019

a massive amount of data that reveals interesting common points and differences between prokaryotic and eukaryotic diversity parameters.

2.1. Prokaryotes

In most environments, SSU rDNA surveys have unveiled a diversity of prokaryotes that is orders of magnitude larger than ever thought. Comparisons between classical culture-dependent and molecular methods have revealed huge gaps and biases in our appreciation of microbial diversity. Only a tiny fraction, about 1%, of the actual prokaryotic diversity appears to be amenable to culture [19], but these estimates evolve with each methodological improvement. Soils offer a good example. Several hundred bacterial species can be isolated from soils, a few thousand of different sequence types (phylotypes or OTUs, usually defined on the basis of 97% sequence identity) retrieved by SSU rDNA surveys and more than 25,000 phylotypes by direct massive sequencing [34]. Moreover, re-evaluation of classical DNA reassociation data with new analytical methods provides estimates of several million species [13]. This estimate's dependence on the methods (and underlying species concepts in some cases) makes any trial to quantify global microbial diversity too premature. In addition, accumulation curves of SSU rDNA surveys are usually far from saturation, suggesting that a far larger number of sequences would be required to fairly represent natural communities. However, interesting efforts to borrow parametric and mostly non-parametric estimators of species/OTU richness from classical ecology are arising both for microbial prokaryotes and eukaryotes, even though their present reliability seems limited [8].

Many phylotypes appear to define novel high-ranked lineages, namely, new phyla intermixed in phylogenetic trees with those having cultured members. However, SSU rDNA-based phylogenies are often characterized by the lack of resolution of the deepest nodes, which can partly explain the discrepancies in the number of prokaryotic phyla estimated by different authors. A phylotype can be misplaced in trees because of insufficient phylogenetic signal and other tree reconstruction artifacts (deficient taxonomic sampling, mutational saturation, long branch attraction), so that it can be erroneously considered a ''novel phylum''. The case of the archaeon *Nanoarchaeum equitans* illustrates this point. SSU rDNA phylogenies suggest that it defines a new archaeal kingdom, but multi-marker phylogenies have shown that it is actually a fast-evolving member of the Thermococcales, misplaced in SSU rDNA trees by long-branch attraction [4]. This is probably the case for many candidate ''novel phyla''. Among current classifications, the number of prokaryotic phyla varies from 50 to 88, reflecting not only phylogenetic artifacts but also opposite taxonomic practices. Like zoologists and botanists several decades ago, microbial taxonomists currently range between two extremes: ''splitters'', who define new phyla for any divergent phylotype, and ''lumpers'', who make large units to keep as stable a taxonomy as possible. A compromise between the two is likely more realistic: several prokaryotic phylotypes cannot be ascribed to any known group and therefore define novel phyla, but many others can, and should not be used to artificially inflate the real number of phyla. Recent initiatives, such as Greengenes (http://greengenes.lbl.gov), allow a very useful cross-comparison of different taxonomic schemes. Fig. 1 shows only the 53 bacterial phyla that are acknowledged by at least 3 of the 5 taxonomic frameworks used in Greengenes, and 17 major archaeal lineages.

2.2. Microbial eukaryotes come into play

Compared with prokaryotes, molecular inspection of the diversity of microbial eukaryotes (protists) is still in its infancy, but it already shows contrasting tendencies to those seen in prokaryotes. Environmental protist phylotypes usually differ from known species sequences, but only very few of them define potential novel divergent lineages [27]. This suggests that the traditional protist description has been much more exhaustive than the prokaryotic one, something easy to explain taking into account the relatively large size and complex morphology of many protists. However, molecular approaches are decisive in demonstrating that not only are the protists of ''traditional'' size (>5 μm) very diverse, but also small nano- and pico-eukaryotes (<5 μm). Though understudied by traditional methods, small protists turn out to be the most speciose in many SSU rDNA molecular surveys [27].

Up to now, the majority of these surveys have been focused on marine planktonic communities, although several other ecosystems, such as freshwater, sediments and extreme environments, have also been studied [8]. In marine plankton, the largest proportion of protist phylotypes belongs either to the Heterokonta or the Alveolata (Fig. 1). Heterotrophic heterokonts dominate in surface waters together with photosynthetic picoalgae, whereas two groups of unidentified alveolates, Marine Alveolate Groups I and II, dominate in deep, aphotic waters [27]. Group II alveolates were soon recognized as relatives of the genus *Amoebophrya* and, very recently, species of the genus *Duboscquella* have been shown to branch within Group I [17]. Both *Amoebophrya* and *Duboscquella* belong to the Syndiniales, a poorly known group of parasitic dinoflagellates. Parasitism may therefore be extremely common in marine protist planktonic communities, most likely playing an unexpected major role in population control.

3. Microbial biodiversity and ecosystem functioning

To understand how microbial and, ultimately, all ecosystems work, identifying the different components of the community is but a very preliminary step that needs to be complemented with data about their functions (metabolism, lifestyle), interactions, spatial and temporal dynamics and environmental parameters. Developing ecosystem models that accommodate all this information is fundamental, but this is a long-term objective, as primary data connecting most microbes to their functions are still missing.