

COMMENTARIES

Metagenomics and the global ocean survey: what's in it for us, and why should we care?

Kenneth H Nealson and J Craig Venter

The ISME Journal (2007) 1, 185–187; doi:10.1038/ismej.2007.43; published online 14 June 2007

Recently, a special Oceanic Metagenomics Collection of articles from the J Craig Venter Institute was published in PLoS Biology, available at: <http://collections.plos.org/plosbiology/gos-2007>. At first glance, the publication represents a very large (and very welcome) addition of data to the nascent field of marine microbial metagenomics. These data, consisting of more than 7.7 million sequencing reads (>6 billion base pairs of sequence), reveal more new genes, more new proteins, more diversity and a more complex ocean than might have been thought: yet they do not begin to touch the real complexity of the ocean ecosystem(s). The data are gathered from 41 sites, primarily marine, covering a transect that includes a sample about every 330 km for more than 8000 km, from the North Atlantic, southwards along the eastern edge of North America, through the Panama Canal, and onward towards the South Pacific. In addition, there is some extensive coverage near and around the Galapagos Islands. Included in the dataset are previously studied samples from the Sargasso (Venter *et al.*, 2004).

A deeper look, however, reveals that these impressive numbers are the tip of an intellectual iceberg of fascinating inconsistencies with regard to marine microbial diversity. Indeed, it may well be that what is not in the dataset may offer opportunities for future studies that transcend the opportunities lying in the dataset itself. To understand what is not there, one needs to keep in mind where and how the samples were collected: these are all near-surface (within a few meters) samples that were filtered multiple times to yield a size fraction in the 0.2–0.8 μm range. Thus, the sample can be aptly characterized as the near-surface marine planktonic niche, consisting mostly of unattached, single cells. Other organisms should have been removed on the larger 0.8 μm filters, which remain as a resource for further study.

As for what is contained in the dataset, there is something for almost everyone. Rusch *et al.* (2007) lead off with a synopsis of the gene data – new genes

galore, new phylotypes galore and the conclusion that in this niche there is still to be found an impressive array of diversity at both the taxonomic and biochemical levels. This being said, however, the dominant species are remarkably few in number. If one simply removes all ‘abundant’ species that occur at only one site, as well as those that are found only in the non-marine (hypersaline, mangrove and freshwater) sites, the number of dominant groups that characterize this marine planktonic niche decreases to about 10–20 (depending on whether you are a splitter or a grouper). This is quite remarkable, perhaps the paradox of the plankton is not a paradox at all, but is hidden in the way that microbiologists define diversity, and our understanding of what is being competed for in the so-called uniform ocean. Of these, only three (*Synechococcus*, *Prochlorococcus* and *Pelagibacter ubique*, a SAR-11 type) have been cultivated and have genomic sequences available.

However, among these abundant species can be found an impressive array of diversity – so impressive that in no case was it possible to assemble a genome from any of them. Thus, while taxonomic/phylogenetic diversity was quite limited, the diversity at the gene level was remarkably high, an observation fitting with several previous studies of localized sites, but apparently a general feature of the marine planktonic environment. Given these challenges, some new approaches were adopted to try and understand this immense diversity. For example, 584 sequenced genomes in finished or draft form were used for ‘fragment recruitment’ of the entire database. Remarkably, only 30% of the database revealed recruitment to any of the 584 genomes: 15% recruited to three genomes of the ‘marine planktonic niche’ (*Pelagibacter*, *Prochlorococcus* and *Synechococcus*), while 15% recruited to two genomes that appeared at only one site in the global ocean survey (GOS) (*Shewanella* and *Burkholderia*). In terms of understanding the nature of diversity in the marine planktonic niche, such information tells us that the sequencing of the other dominant species should be a high-priority item – one that will allow retrospective fragment

recruitment studies that will begin to unravel this conundrum.

Yooseph *et al.* (2007) then present a paper dealing with the study of protein families gleaned from analysis of the dataset. In this study, intensive analysis of protein sequences led to the conclusion that at this level there is immense diversity and variation; 1700 new protein families were found with no apparent homology to existing protein groups. The study not only identifies new proteins, but also adds a much-needed input of data with regard to diversity of known protein families. What can be done with a dataset like this is then illustrated in the paper by Kannan *et al.* (2007), in which diversity of protein kinases was studied, resulting in a tripling of information with regard to ELK (eukaryotic protein kinase-like) proteins. The intriguing observation that prokaryotic ELKs are now more numerous than the prokaryotic histidine kinases, which have been considered to be the major regulatory elements for prokaryotic metabolism, begs the question of whether there are completely new regulatory pathways waiting to be discovered in this very interesting realm.

Finally, an overview article by Eisen (2007) discusses the ups and downs of the various approaches to studying microbial communities – a nice article to read before diving into the three articles discussed above. In addition, Seshadri *et al.* (2007) present an introduction and description of CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis), a community database system for the deposition and analysis of data related to marine microbial ecology.

So, this would seem to be the end – a fantastic journey into the world of bioinformatics, an immense amount of data being made available to the community for detailed work on their systems of interest, and a workable interactive datasystem with which to do it. However, as mentioned above, the Rusch *et al.* (2007) paper suggests that there is much more, and that this may be lurking in some of the things that are not seen here. What do we mean by this?

The GOS survey, as noted above, focuses on the planktonic niche, and as such, misses certain parts of the marine microbial ecosystem, notably the larger single cells (small eukaryotes and large prokaryotes), multicells, attached cells and symbionts, to name a few. Yet, plating of seawater often yields low but consistent numbers of such microbes that for many years were known as the dominant oceanic species – genera like *Vibrio*, *Shewanella* (a.k.a. *Alteromonas*) and *Pseudomonas*, to name a few. In many cases, these bacteria have well-defined niches – disease causation, gut symbionts of marine fish, light organ symbionts of fish and squids, food spoilage, and so on, and there is no doubt that they play a role in marine ecosystems, almost certainly as attached forms (Visick and Ruby, 2006). In fact, on the basis of many studies of such genera (all of

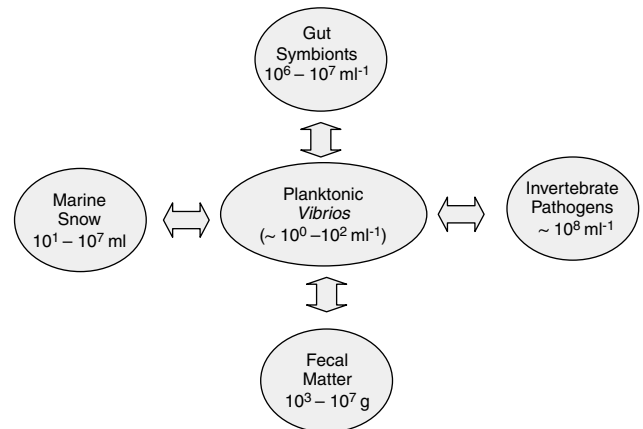


Figure 1 A simplistic diagram of the distribution of some marine *Vibrios*. This shows the simple case for the coast of California, where the major niches for these organisms involve a heterotrophic travel through the stomachs of marine fish, acting as catalysts for chitin degradation (Visick and Ruby, 2006). Moving to other locations could add more high-concentration environments such as the light organs of fishes and/or squids, where cell populations reach 10^{10} ml^{-1} or higher and constantly inoculate the waters.

which were used for recruitment studies, and all of which proved negative with regard to recruitment) a model similar to that shown in Figure 1 can be proposed, in which the planktonic populations are simply a reflection of the various high-density niches for the attached forms. Such a picture stresses the importance of examining a variety of size classes, looking, perhaps, for those organisms that might account for those occasionally abundant microbes that are clearly not part of the planktonic niche.

A careful look at the genes needed for each niche might be very revealing in terms of distinguishing what defines planktonic versus attached lifestyles. With regard to the example chosen here, the luminous *Vibrios* were surely present in the samples analyzed, but were cryptic to the methods used, and in sufficiently low abundance that not a single *luxABCD* or *E* gene sequence was seen in the database. It is an interesting exercise for each of us to take our own idea of where our organism fits into the marine ecosystem and ask where one might look for evidence in the data or samples of the GOS expedition.

In closing this brief overview of the GOS volume, one does not want to detract from what has been (and will be) learned from this magnificent dataset. Each of us should sit down with the data and add our private interests and expertise to its analyses, thus using this as a landmark system for marine microbiological systems studies. This being said, one must keep in mind that this is one niche of perhaps hundreds in the ocean, and similar studies will be needed for each of them. These niches can be defined by size fractionation, by physicochemical properties of the environment, by depth of samples and perhaps by many other parameters. The

important thing now is to seize the moment and move forward gathering more data, depositing it in the central CAMERA databank (<http://camera.calit2.net/>), and working to describe the marine ecosystem as the complex system it is.

*KH Neelson is at University of Southern California,
Department of Earth Sciences,
Los Angeles, CA, USA and
JC Venter is at J Craig Venter Institute,
Rockville, MD, USA.
E-mail: knealson@usc.edu*

References

- Eisen JA. (2007). Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *PLoS Biol* **5**: e82.
- Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biol* **5**: e17.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.
- Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. (2007). CAMERA: a community resource for meta-genomics. *PLoS Biol* **5**: e75.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso sea. *Science* **304**: 66–74.
- Visick KL, Ruby EG. (2006). *Vibrio fischeri* and its host: it takes two to tango. *Curr Opin Microbiol* **9**: 632–638.
- Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K *et al.* (2007). The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol* **5**: e16.

Soaking it up: the complex lives of marine sponges and their microbial associates

Michael W Taylor, Russell T Hill, Jörn Piel, Robert W Thacker and Ute Hentschel

The ISME Journal (2007) **1**, 187–190; doi:10.1038/ismej.2007.32; published online 24 May 2007

Marine sponges (phylum Porifera) are among the oldest multicellular animals (metazoans), the sea's most prolific producers of bioactive metabolites, and of considerable ecological importance due to their abundance and ability to filter enormous volumes of seawater. In addition to these important attributes, sponge microbiology is now a rapidly expanding field.

In the 1970s and 1980s, Clive Wilkinson and Jean Vacelet set the benchmark for sponge microbiology, describing dense and diverse microbial communities comprising up to 40% of total 'sponge' volume (Vacelet and Donadey, 1977; Wilkinson and Fay, 1979; Wilkinson, 1983). Advances in molecular techniques have, over the past decade, taken the field to a new level, revealing new and exciting research opportunities. The catalyst for this article was a recent roundtable session, 'Marine sponges as microbial fermenters', held under the auspices of the *11th International Symposium on Microbial Ecology*, Vienna, Austria. Scientists from many different backgrounds (biotechnology, ecology, evolution, microbiology, organic chemistry) were brought together, stimulating wide-ranging

discussions across normal discipline boundaries. Based on these discussions and key recent findings, we chart a course for future research in the field, identifying as key focal points the evolution of symbiont diversity, microbial metabolism, host–microbe interactions and the potential biotechnological implications (Table 1).

The microbial consortia present in many sponges (Figure 1) span all three domains of life, with at least 18 bacterial and archaeal phyla now known from these hosts (Taylor *et al.*, 2007) (Figure 2). Interactions between sponges and microbes are also diverse, ranging from mutualistic to commensalistic and exploitative (parasitism/pathogenesis). A major challenge in sponge symbiont biology, and a key theme to emerge from the roundtable, is how best to handle this diversity. Although the study of simple systems offers valuable insights into host–microbe interactions (the squid–*Vibrio* symbiosis (Nyholm and McFall-Ngai, 2004) is a prime example), we suggest that this scenario is not the norm in nature, at least for the marine environment. Rather, most prokaryote–eukaryote associations in the oceans are probably a complex mix of stable and transient associates, as exemplified by the sponge–microbe model. Embracing the complexities of the sponge system will thus lead to its establishment as an