The New Science of Metagenomics: Fourth Domain of Life

Rup Lal

Published online: 28 April 2011
© Association of Microbiologists of India 2011

Metagenomics, signifying ‘genomics on a huge scale’ is a powerful approach that enables a massive assessment of the plethora of microorganisms present in the environment. Metagenomics is based on the genomic analysis of microbial DNA that is extracted directly from communities in environmental samples. It is a new approach that has revolutionized our understanding of microbial life present on Earth. The application of metagenomic analysis has accelerated the rapid rate of advancement in the study of uncultured microbes that began with the advent of rRNA analysis. It has now enabled the phylogenetic characterization of many entire communities that were otherwise difficult to analyze. In the coming years, metagenomic approaches are expected to become a powerful tool that will help developing alternate strategies for harnessing the power of microbial communities for more sustainable and environmentally friendly, biologically based energy sources, environmental remediation, agricultural biodefense and forensic sciences. Metagenomics could also unlock the massive uncultured microbial diversity present in the environment to provide new molecules for therapeutic and biotechnological applications. While there seem to be several benefits in the offerings from this approach, the analysis of metadata has raised a new debate about the possibility of a fourth domain of life. The current concept based on the rRNA gene sequences only predicts three domains of life, which are eukarya, eubacteria and archaea.

The metagenomic data generated by Craig Venter from the Sargasso Sea on the round the world yatching trip in 2005 [1] has been recently re-analyzed by Prof. Jonathan Eisen, an evolutionary biologist at the University of

California [2]. The J. Craig Venter Institute’s Global Ocean Sampling Expedition revealed an enormous amount of previously unexplored diversity in the world’s oceans. Prof. Eisen primarily used two genes *recA* and *rpoB*, both of which are old and conserved throughout the living kingdom. These two genes predict a different lineage also called the fourth domain of life. His team designed and implemented new methods for analyzing metagenomic data and used them to search the Global Ocean Sampling Expedition data set for novel lineages in three gene families commonly used in phylogenetic studies of known and unknown organisms: small subunit rRNA and the *recA* and *rpoB* superfamilies. Though the methods available could not accurately identify very deeply branched ss-rRNAs, the analysis revealed the existence of multiple novel branches in the *recA* and *rpoB* gene families (these code for proteins that are universal and very widely studied). Many hypothesis have been proposed to explain these unique findings in the metadata. It may be recombinations or duplications leading to representations of distant paralogs within these families. However, the best explanation that has been conceived is the origin of these sequences from uncharacterized viruses. This is widely supported by recent unexpected discoveries coming from viral genomics [3]. However, the four domain concept needs to be authenticated by the scientific community.

References

1. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Tillson HB, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66

R. Lal (✉)
University of Delhi, Delhi, India
e-mail: ruplal@gmail.com

2. Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, Frazier M, Venter JC, Eisen J (2011) Stalking the Fourth Domain in Metagenomic Data: Searching for, Discovering, and Interpreting Novel, Deep Branches in Marker Gene Phylogenetic Trees. *PLoS ONE* 6(3). doi:[10.1371/journal.pone.0018011](https://doi.org/10.1371/journal.pone.0018011)
3. Boyer MMM-A, Gimenez G, La Scola B, Raoult D (2010) Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. *PLoS ONE* 5:e15530