International Journal of Environmental & Earth Sciences Research Article

Current Status of Marine Genomics

Leena Grace Beslin

Department of Aquatic Biology and Fisheries, University of Kerala, Kariavattom, Thiruvananthapuram-695581, INDIA

***Corresponding author**
Leena Grace Beslin, Department of Aquatic Biology and Fisheries, University of Kerala, Kariavattom, Thiruvananthapuram - 695581, India. E-mail:drblgrace@rediffmail.com.

Received: 06 June 2021; **Published**: 29 June 2021

Abstract

Marine environment is the cradle of life containing 95% of the world's biomass and 38 (19 endemic) of the 39 known animal phyla. The fundamental understanding of the biotechnological potential of marine organisms is the assessment of their genetic capabilities, i.e. sequencing of their genome and annotation of the genes which is the focus of genomics. Currently, about 1000 prokaryotic genomes have been sequenced and annotated. More than half of these genomes are of medical or industrial relevance and no phylogenetic systematic genome sequencing has been carried out until recently. Though mitochondrial genomes are useful for the identification of fish species and populations the focus of most genome research displayed on the nuclear genome. Diverse and unique Marine microbial assemblages challenged to discover special functions occupied by these microorganisms. In the evolutionary tree of marine organisms, key roles were played by whole genome sequencing and functional genomics. This elaborated the determination of the total Deoxyribonucleic acid sequence of organisms and fine-scale up of genetic mapping efforts. It led to the development of bioinformatics by producing large data sets of cytogenetics, molecular genetics, quantitative genetics and population genetics. This links the raw genome information to meaningful biological information. Marine fishes also exhibit high levels of gene flow mainly due to the pelagic larval phase and consequent dispersal of many pelagic and demersal resources. In view of the fact that marine microbes are important for the Earth system, marine genomics resources development has been primarily decided on marine microbes which include both prokaryotic and eukaryotic plankton, because they involved in significant mineral cycles of the oceans. The marine ecosystem plays a great role in the sustenance of the global environment, for the reason that half of the annual primary production of the earth happens in the ocean. The present day technologies and the development of high-throughput technology for sequencing DNA from the natural marine environment and its resources paved way to generate enormous sequences. However, genomics resources of marine taxa are currently limited to the full genome sequence of the 'model species'. It will be important to actively channel this process in the future to ensure the coverage of groups in particular to most important eukaryotes.

Keywords: Marine, Environment, Model Species, Genomics, Sequencing, Annotation

Introduction

An interdisciplinary science highlight the structure, function, evolution, mapping and editing of genomes is **Genomics**. A genome is the total set of DNA, including all the genes in an organism. Collective characterization and quantification of genes were aimed by genomics which direct the production of proteins with the assistance of enzymes and messenger molecules [1]. Proteins make up the body structures such as organs, tissues, control chemical reactions and carry signals between cells. Genomics also involves the sequencing and analysis of genomes by using high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes [2].

Discovery based research and systems biology facilitated the understanding of the most complex biological systems such as the brain which triggered the advances of genomics. The intragenomic (within the genome) trends such as epistasis (effect of one gene on another), pleiotropy (few traits expressed by a gene), heterosis (hybrid vigour), and other interactions linking loci and alleles within the genome were also represent the genome advances [3]. Central to the understanding of the biotechnological potential of marine organisms is the assessment of their genetic capabilities, i.e. sequencing of their genome and annotation of the genes. This understanding is the focus of genomics [5]. Currently, about 1000 prokaryotic genomes have been sequenced and annotated. More than half of these genomes are of medical or industrial relevance and no phylogenetic systematic genome sequencing has been carried out until recently [6]. phylogenetically diverse microbial genomes Sequencing results in the discovery of many novel proteins per genome demonstrating the existence of a huge reservoir of undiscovered proteins [7].

About 7500 bacterial species have been validly described; it follows that still thousands of new proteins will be discovered by sequencing in a systematic manner from all cultured bacterial species [8]. Another level of diversity has to be expected from the uncultured prokaryotes which make up about 70% from 100 bacterial phyla. This uncultured diversity became apparent when the first whole [4] genome analysis of marine microbial communities revealed as many new clusters of ortholog groups (COGs) as were already known at the time [9]. On the other end of the phylogenetic diversity, i.e. comparing different strains of a bacterial species, it is becoming clear that each new strain can add hundreds of new genes [10].

In addition to bacteria, aquatic ecosystems contain viruses which are the most common biological entities in the marine environment. This means that the pan-genome of a microbial

species, comprising all genes of all strains of that species is several times larger than the core genome [11]. The abundance of viruses exceeds that of prokaryotes at least by factor of ten and they have an enormous impact on the other microbiota, lysing about 20% of its biomass each day. Recent metagenomic surveys of marine viruses demonstrated their unique gene pool and molecular architecture [12]. Their host range covers all major groups of marine organisms from archaea to mammals. Algal genome sizes can even vary about 20 fold within a genus, as illustrated with *Thalassiosira species* [13].

The overall size range for microalgal genomes is 10 Mb to 20 Gb, with an average size of around 450 Mb, except for *Chlorophyta,* that are on the average four times larger. Many marine microalgae are highly complex single celled organisms containing chromosomal DNA as well as mitochondrial and chloroplast DNA [14]. They have a complex nucleus that has been subjected to extensive exchange of genes between the organelles and the nucleus (endosymbiotic gene transfer) as well as horizontal gene transfer during their hundred million years of evolution. In addition, the first genome of a macroalgae (*Ectocarpus*) has been sequenced and several others are being completed. The challenge is to investigate this novel 'terraincognita' through post-genomics, biochemical approaches and genetic developments [15]. The reward for taking on this challenge is an improved understanding of the biochemical functioning of key players in aquatic ecosystems with new insights into the regulatory genetic network of eukaryotes and their early evolution with great potential for the production of a huge variety of bioproducts [16].

Marine Species and Genomics

The vast majority of marine microbes cannot be cultured in the laboratory and so were not amenable to study by the methods that had proved so successful with medically important microorganisms throughout the 20th century [15]. It was only with the development of high throughput technology to sequence DNA from the natural marine environment and this information demonstrated the exceptional diversity of microbes in the marine environment. In reality, most marine microbes are exclusively novel in their characteristics. Marine microbial assemblages are diverse and unique and the challenge is to discover what functions are displayed by these microorganisms [17].

At present gene resources for other marine taxa are limited to the full genome sequence at the level of 'model species', the purple sea urchin *Strongylocentrotus purpuratus*. For other model and non-model species such as surf clam *Spisula solidissima*, the sea squirts *Ciona intestinalis* and Ciona savignyi, the tunicate Oikopleura dioica, the little skate *Leucoraja erinacea* and the mollusk parasite *Perkinsus marinus*, the sequencing experiments are in progress. Gordon and Betty Moore Foundation done the Marine Microbial Genome Sequencing Project in 2004, sequenced nearly 180 marine microorganisms, of which 80% were already published. Microorganisms are known to be the "gatekeepers" of these processes. So their catalytic activities and interaction with the environment will enhance the ability to monitor, model and predict changes in the marine ecosystem [18].

Case Studies in Marine Genomics

Because of the vast phyletic diversity of marine organisms, existing genomic model organisms are often with limited relevance, because there is an enormous evolutionary distance separating these models from an organism of interest. Genome sequencing has been completed in the unicellular green algae *Chlamydomonas reinhardtii* [19].Genome projects are in progress in the marine key species such as *Emiliania huxleyi* (a pelagic coccolithophore), Hydra *magnipapillata, Litopenaeus vannamei* (the pacific white shrimp) and Amphioxus (the closest living invertebrate relative of the vertebrates).In the prokaryotes several marine organisms such as multiple strains of the pelagic photosynthetic bacteria S*ynechococcus and Prochlorococcus*, rapid progress in sequencing was achieved with many sequenced genomes [20].

Whole Genome Sequence of Diatoms

Based upon the studies of Sogin et al. the sequencing of the genomes of environmentally important organisms such as the diatom *Thalassiosira pseudonana* provided the first complete genome from the heterodont lineage [13]. In addition to this environmental and the phylogenetic importance silicate metabolism also gained attention. Like most diatoms the biotechnological potential of silicate metabolism constructs a silicate exoskeleton called the frustule (the production of this structure has great applications in nanotechnology).

Whole Genomes of Aquatic Animals

Full genome sequences of some fishes such as zebrafish, fugu, tetraodon, medaka and three spined stickle back are most valuable. Some non-model organism must be used for answering many questions, because there are close to 30,000 species of fish with maximally more than 300,000,000 years of independent evolution between groups. The species occur in different habitats from arctic streams to marine tropical areas including underground caves and hypoxic tropical lakes. Consequently, what may be suitable for a small tropical freshwater cyprinid "zebrafish" may not be valid for marine tuna which has weighing several kilograms. Whenever a new method becomes available for genomic studies, its utility for non-model organisms should be evaluated which has been done for microarray methodology [21].

Genomics information of many aquatic invertebrates' species is even less satisfactory than on fishes shown that more than one third of the genes of the recently sequenced genome of *Daphnia pulex* have no complements in former sequenced genomes [22]. Especially these Daphnia specific genes respond rapidly to environmental disturbances. Altogether, the genome data on fish and Daphnia suggest both rapid evolution and rapid development of genetic responses to environmental changes [23]. Recently, scientists from Norway have examined and presented the genome sequence of Atlantic cod *Gadus morhua*. The entire genome assembly was 454 sequencing of shotgun and paired-end libraries and automated annotation identified 22,154 genes. Atlantic cod has missing the genes for MHC II, CD4 and invariant chain (Ii) that was the conserved trait of jawed vertebrates in the adaptive immune system [24].

Genomes of Crustaceans

Amazing group of organisms satisfying all types of habitats in the ocean with a wide array of adaptations are crustaceans (lobster, shrimp, crab, etc.) which hold the supreme species diversity among marine animals. They are not only plentiful in number, but also the most commercially exploited food species for human utilization [25]. However, they are not as well studied like their terrestrial arthropod relatives (insects). Especially in the Indo-Pacific region, the tiger shrimp (*Penaeus monodon)* has been one of the most important captured and cultured marine crustaceans [26]. Disastrously, the tiger shrimp industry has been weighed down by viral diseases lead to economic losses [27]. Developments in shrimp genomics have been limited although a reasonably good EST database is available [10]. Tiger shrimp genomic analysis will make a key contribution to decipher the

evolutionary history representing the crustacean lineages. The genomic sequences information will benefit the shrimp industry by contributing genomic tools to discover the viral diseases and to build up the breeding program [28].

Puffer Fish Genome Features in Draft Sequences

Fugu is the delicious fish and the liver is poisonous. The liver has to be removed in a peculiar manner before preparing it into a cuisine. The poison of a single liver of a fugu is able to kill 30 persons. The fugu genome was the first vertebrate genome sequenced after human, used the whole genome shotgun method. These draft sequences unravel many interesting diversities in specific protein families sandwiched between human and fugu [29].

The Tetraodon genome sequence was subsequently produced with the whole-genome shotgun method albeit with a higher redundancy in sequence reads (8.3 vs. 5.6). Puffer fish possess about 70 different families of transposable elements against only 20 for human or mouse [23]. Interestingly in Tetraodon, SINE and LINE families are distributed in opposite regions of the genome compared to human or mouse genome. In mammals SINEs are rich in G + C sequences and in Tetraodon more A + T regions and vice versa for LINE elements. More surprisingly, these initial studies of Tetraodon and fugu showed a number of differences in their genomes. G + C rich region in both Tetraodon and mammal genomes is absent in fugu [30].

Challenges in computations

Bioinformatics tools used in the latest computing technology deal with challenges of data analysis in less time. The steep fall in sequencing cost and the concomitant increase in sequencing speed outpaces the improvement in computational power, this will likely continue for several years. Until recently, the main repository for DNA sequences, GenBank, grew at about the same rate as computing power, following Moore's law and doubling every 18 months. GenBank contained about 300 Gb of sequence unto 2009. Today 2 Gb of sequences were generated due to the availability of sophisticated next generation sequencers. Current algorithms for sequence data analysis have their roots in the early times of sequence acquirement. Algorithms that are used for aligning genome sequences tend to have an exponential computing time requirement due to increase of analysis that leads raise of analysed sequences. New concepts and approaches will be necessary to reduce this into a linear requirement [31-37].

Conclusion

The function of genes has so far largely been studied in a very limited number of species and in the context of individual organisms. In the next decade an immense modification towards the addition of the ecological and evolutionary context in gene function analysis will have to be inserted with the genomics. As such, genetics will move on from a largely biomedical perspective to an ecological perspective with special relevance for global change questions [34]. A full understanding of the ecosystem, its services and its stability will not be possible without understanding the genetics of adaptations and community interactions.

References

1. Colin S, Deniaud E, Jam M, Descamps V, Chevolot Y, et al. (2006) Cloning and biochemical characterization of the fucanase FcnA: definition of a novel glycoside hydrolase family specific for sulfated fucans. Glycobiology 16: 1021-1032.
2. Chandonia JM, Brenner SE (2006) The impact of structural genomics: expectations and outcomes. Science 311 (5759): 347-351.
3. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW (1998) Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol 16: 301-306.
4. Metzker M L (2010) Sequencing technologies-the next generation. Nat Rev Genet 11: 31-46.
5. Gupta PK (2008) Single-molecule DNA sequencing technologies for future genomics research. Trends Biotechnol 26: 602-611.
6. Rodi CP, Bunch RT, Curtis SW, Kier LD, Cabonce M A, et al.(1990) Revolution through genomics in investigative and discovery toxicology. Toxicol Pathol 27: 107-110.
7. Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol 210: 1518-1525.
8. Pevsner J (2009) Bioinformatics and functional genomics (2nd Ed.). Hoboken, NJ, 7: Wiley-Blackwell, London.
9. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26: 1135-1145.
10. Leu JH, Chang CC, Wu J L, Hsu C W, Hiron I, et al. (2007) Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected Penaeus monodon. BMC Genomics 8: 120-133.
11. DeLong EF, PrestonCM, Mincer T, Rich V, Hallam SJ, et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311: 496-503.
12. Frias Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC (2008) Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci 105: 3805-3810.
13. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al.(2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. Proc Natl Acad Sci 103: 12115-12120.
14. Gilbert JA, Field D, Huang Y, Edwards R, Li W, et al. (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS ONE 3: 3042.
15. Ten Bosch JR, Grody WW (2008) Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J Mol Diagn 10: 484-492.
16. Hollywood K, Brison DR, Goodacre R (2006) Metabolomics: Current technologies and future trends. Proteomics 6: 4716-4723.
17. Szalay A, Gray J (2006) 2020 Computing: Science in an exponential world. Nature 440: 413-414
18. Thomas MA, Klaper R (2004) Genomics for the ecological tool box. Trends Ecol Evol 19: 439-445
19. Valenzuela-Quinonez F (2016) How fisheries management can benefit from genomics? Brief Funct Genomics 15:352-357
20. Reyes-Prieto A, Yoon HS, Bhattacharya D (2019) Marine Algal Genomics and Evolution. Encyclopedia of Ocean Sciences 1: 561-568.
21. Ferrie DEK (2016) The origin of the Hox/ParaHox genes, the Ghost Locus hypothesis and the complexity of the first animal. Brief Funct Genomics 15: 333-341.
22. Meyer F (2006) Genome Sequencing vs. Moore's Law: Cyber challenges for the next decade. CTWatch Quarterly 2: 14-17.
23. Davidson EH (2010) Emerging properties of animal gene regulatory networks. Nature 468: 911-920.
24. Venter JC, Remington JF, Heidelberg AL, Halpern D, Rusch JA, Eisen DWU (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66-74
25. Wilkening J, Wilke AND, Folker M (2009) Using Clouds for Metagenomics: A Case Study. In: Proceedings IEEE

Clouds. 12-19.
26. Supungul P, Klinbunga S, Pichyangkura R, Jitrapakdee S, Hirono I, Aoki T, Tassanakajon A (2002) Identification of immune-related genes in hemocytes of black tiger shrimp (Penaeus monodon). Mar Biotechnol 4: 487-494.
27. Van Straalen NM, Roelofs D (2006) An introduction to ecological genomics. Oxford University Press, Oxford.
28. Margulies M, Egholm M, Altman WE, Attiya S (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380.
29. Worm W, Barbier EB, Beaumont N (2006) Impacts of biodiversity loss on ocean ecosystem services. Science 314: 787-790.
30. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5: 16.
31. Zengler K, Toledo G, Rappé MS., Mathur EJ, Short JM, Keller M (2002) Cultivating the uncultured. Proc Natl Acad Sci 99: 15681-15686.
32. Domazet-Loso M, Haubold B (2009) Efficient estimation of pairwise distances between genomes. Bioinformatics https://doi.org/10.1093/bioinformatics/btp590.
33. Davis RH (2004) The age of model organisms. Nat Rev Genet 5: 69-76.
34. Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, et al. (2003) Complete genome sequence of the marine planctomycete Pirellula sp. strain 1. Proc Natl Acad Sci 100: 8298-8303.
35. Gracey AY, Cossins AR (2003) Application of microarray technology in environmental and comparative physiology. Annu Rev Physiol 65: 231-258.
36. Petricoin EF, Hackett JL, Lesko LJ, Puri RK, Gutman SI, et al. (2002) Medical applications of microarray technologies: a regulatory science perspective. Nat Genet 32: 474-479.
37. Rogers YH, Venter JC (2005) Genomics: massively parallel sequencing. Nature 437: 326-327.

Citation: Leena Grace Beslin. Current Status of Marine Genomics. Int J Envi & Eart Scie 2021, 2: 1-4.