The importance of Marine genomics to Life

Popoola Raimot Titilade[1] and Elegbede, Isa Olalekan[2]

[1]Department of Marine Sciences, University of Lagos, Lagos, Nigeria
[2]Department of Environmental Planning, Brandenburg University of Technology, Cottbus-Senftenberg, Germany
Corresponding authors: titipopoola@gmail.com; isaelegbede@gmail.com

Abstract: Genomics is a field of study that is rapidly transforming many areas of biological and biomedical research which has enabled the transition from sequential studies of single genes to more ecological approach, involving the simultaneous study of many components and their interactions with the environment from pathways, through cell tissues to whole organisms and communities. Genomics application areas include clinical diagnostics, agro biotechnology, environmental biotechnology and pharmacogenomics. The focus of most genome research is on the nuclear genome, though mitochondrial genomes have been extremely useful for the identification of fish species and populations. Marine microbial assemblages are diverse and unique and the challenge is to discover what functions are played by these microorganisms. To provide adequate tools for marine biologists, therefore, one important aim will be to develop genomic approaches, such as whole genome sequencing and functional genomics, for key species across the evolutionary tree of marine organisms. Genomics is a highly dynamic research field. Hence, rapid developments in genomics can afford new opportunities for applications in marine environment, particularly in the areas of Fish genome resources conservation and genetic enhancement.
[Popoola Raimot Titilade and Elegbede, Isa Olalekan. **The importance of Marine genomics to Life.** *J Am Sci* 2018;14(5):40-55]. ISSN 1545-1003 (print); ISSN 2375-7264 (online). http://www.jofamericanscience.org. 8. doi:10.7537/marsjas140518.08.

Keyword: Genomes; genetic enhancement; genome sequencing; marine evolution; medicine

1.0 Introduction

Genomics is a field of study that is rapidly transforming many areas of biological and biomedical research which has enabled the transition from sequential studies of single genes to more ecological approach. It also involves the simultaneous study of many components and their interactions with the environment from pathways, through cell tissues to whole organisms and communities (Hollywood *et al.,* 2006; Joyce and Palsson, 2006). The importance of this field has been supported by the concurrent development of many new technologies and methods which are now used to address fundamental smaller-scale questions in areas such as ecology, biodiversity and evolution. With the exception of the use of genomics to address questions about the diversity and ecology of marine microbial communities, 'metagenomics fields' according to (Venter *et al.,* 2004; DeLong *et al.,* 2006; Sogin *et al.,* 2006) have not been broadly applied in marine ecology.

Genomics can be define as a discipline in genetics concerned with the study of genomes of organisms and the outburst of chromosomes, gene and DNA (fig 1) (Chandonia and Brenner, 2006). Genomes on the other hand are the entirety of an organism's hereditary information which is either encoded in the (DNA) Deoxyribonucleic acid for many type of virus or in (RNA) Ribonucleic acid and includes both the genes and the non-coding sequence of DNA or RNA (Joshua and Alexa, 2001). It can be seen as a branch of molecular biology that relates with the study of structure, function, expression, evolution and mapping of genomes of organisms. The field includes intensive efforts to determine the entire Deoxyribonucleic acid (DNA) sequence of organisms and fine-scale genetic mapping efforts. It always generates large data sets from cytogenetics, molecular genetics, quantitative genetics and population genetics, and has led to the development of bioinformatics, through which raw genome information links to meaningful biological information. Many sub-branches of genomics are emerging, including marine genomics and these demand new ways of data management.

Fig. 1: Structure of a Genome
Source: Shendure and Ji (2008).

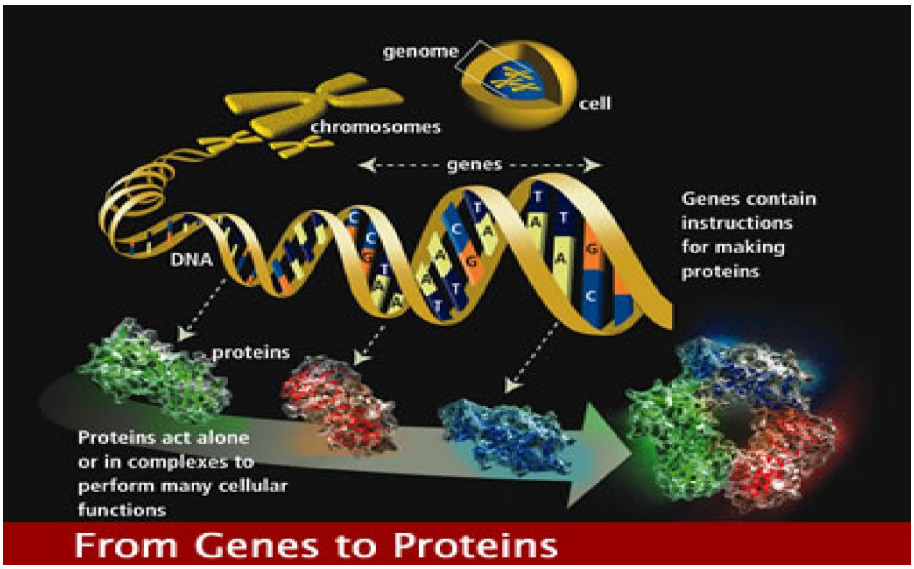

Fig. 2: Dimensions of a Genome
Source: Gracey and Cossins (2003).

Marine environment are the cradle of life containing 95% of the world's biomass and 38 (19 endemic) of the 39 known animal phyla. Apart from providing a third of the oxygen that we breathe and acting as moderators of global climatic change with a significant influence on the human population's terrestrial environment. These vast expanses are also an important source of high-protein food, contributing fundamentally to the planet's functioning, yielding some 60% of the total economic value of the biosphere and therefore comprise the largest untapped resource on earth. Marine and coastal environments include many diverse pelagic and benthic habitats such as open-ocean ecosystems, deep-sea communities including thermal vent ecosystems, kelp forests, mangroves, coral reefs and so on. Although these varied environments support a rich abundance of life, marine biodiversity has received much less attention than its terrestrial counterpart. This may be because the oceans have historically been thought of as regions of low biodiversity and because of difficulties with accessing marine environments. According to Ray (1988), by some measures, biodiversity in the oceans is greater than on land. Marine bio-systems have been evolving for an additional 2.7 billion years compared to terrestrial environments and almost all the currently described phyla are represented in the ocean while only about half have terrestrial members. The phylogenetic diversity of marine organisms is therefore, much broader than that of their terrestrial counterparts.

Presently, major advances are being made in oceanographic sciences. It is now technically possible to carry out remote and unmanned long-term monitoring of physical and biological processes across ocean basins with data updated in real time. Allied to this, high through-put genomics techniques have developed immaturity such that tools now exist to enable the monitoring of the links between (microbial) biodiversity and ecosystem function. It is possible to determine how life and ecosystems evolved and how the two function interactively. Therefore linking the latest developments in oceanographic hardware and genomics laboratory-based techniques provides opportunities for comprehensively understanding global marine biodiversity and ecosystem function at a time when both are severely threatened (Worm *et al.,* 2006). Since genome is all of a living thing's genetic material, it is the entire set of hereditary instructions for building, running and maintaining an organism and passing life on to the next generation, this is indeed an exciting time for marine scientists.

Given the professional excitement and veritable explosion of new marine genomics data, it is disconcerting to learn that still only 10% of publications are on marine-biodiversity-related studies. Less than 0.1% of the genomics-related papers (SCI 2002–2007) are devoted to the marine domain as compared with terrestrial or medical studies. It is believed that this discrepancy is fundamentally driven by insufficient public awareness and education. Marine ecosystems are far more important than that portrayed by popular Television images of coral reefs, sharks and extremophiles. It is also a mistaken popular notion that genomics is commercially restricted to human health, agriculture and nanotechnology.

Genomics is a recent scientific discipline that strives to define and characterize the complete genetic makeup of an organism. Its primary approaches are to determine the entire sequence and structure of an

organism's Deoxyribonucleic acid (DNA) (its genome) and then to determine how that DNA is arranged into genes. This second goal is accomplished by determining the structure and relative abundance of all messenger Ribonucleic acids (mRNAs), the middlemen in genetics that encode individual proteins. It provides the tools to unravel key issues and processes related to climate change, loss of biodiversity and ecosystem function. The Tree of Life can be reconstructed, understand the evolution of genomes and the myriad of ways in which gene networks affect adaptive evolution. These are the more tractable challenges now awaiting today's marine genomic scientists.

Genomics emerged as a mature science with the sequencing of the human and other genomes along with the development of DNA microarrays and the computing power to analyze the multiple data points generated. These combined factors allow for fully comprehensive and rapid investigations of gene expression (Schena *et al.,* 1998).

Equally important is the understanding of the obtained insights into genome evolution and composition, along with the intrinsic mechanisms. It paved the way for other subsets of functional scientific study such as pharmacogenomics, proteomics, metabolomics and nutrigenomics, which are sought to improve the human conditions. Among the objectives of the study of the human genes is to help the biologists to work out on the several different molecular interactions leading to the normal development of the organisms.

The genetic sequence can also serve as a reference base to investigate other members of the same species and other matters of interests such as identifying expressions of proteins and the inherited conditions. Furthermore, even the molecular medicine benefits from the Human Genome Project as they can conduct more precise and specific diagnostic tests that will help determine the stage of the disease, along with its accurate drug therapy. There will be enhanced therapeutic regimens tailored after a specific genetic makeup, immunotherapy techniques and the opportunity to choose which lifestyle is best for an individual with hope to avoid the medical conditions to which he is pre-disposed.

Generally, From its first conception, Genomics has grown and influenced the development of associated technology which has made an impact across the globe by improving the success rates quickens the research pace of the life sciences. In turn, it has created many applications in the socio-economic aspects including the ability to determine a person's risks in developing some diseases, initiating drug discovery, identifying the accurate medical treatments, developing drought-resistant crops for optimal yields of nutritious foods and healthier poultry and livestock animal foods, and finding ways to preserve the biodiversity.

Therefore, Marine genomics is the application of genomic sciences which attempt to understand the structure and function of marine organisms in other to provide biological information of the organisms that underlie the ecology of oceanic ecosystems. The term ‘genomics’ appeared in the 1980s as the name of a new journal (McKusick and Ruddle, 1987), but the genomics revolution really began in 1990 with the Human Genome Project and since then, rapid developments in molecular biology technologies, genomics based discovery has grown exponentially. The new sequencing system developed by Margulies *et al.,* (2005) will be capable of sequencing 25 million bases in a 4 h-period—about 100 times faster than current state-of-the-art systems—with the same reliability and accuracy. The genomes of more than 300 organisms have been sequenced and analyzed since the publication of the first complete genome in 1995, and today a new organism is sequenced nearly every week (Rogers and Venter, 2005; Van Straalen and Roelofs, 2006).

The current challenge is no longer to collect sequence information but rather to analyze the data. Genomic approaches combine molecular biology with computing sciences, statistics and management. The intellectual infrastructure in genomics must be extended into bioinformatics (data storage and data query), computational biology (more complex, often hypothesis-driven analyses that may require the development of new algorithms and tools), and information technologies to share software and data.

With the Human Genome project, the scientists determined the complete genetic information found in the human DNA, which composed of billions of genes having four chemical bases Adenine, Guanine, Cytosine, and Thymine A, G, C and T which are repeated billion times throughout a genome (completely sequenced genes). They better understood genetic variations, varying genetic expressions, the roles of genes, including the defective in the human body and their relationship with one another.

2.0 Functional And Structural Genomics

The term genome refers to the complete genetic material of an organism which includes the nuclear and mitochondrial genomes for plant and animals, also chloroplast genomes for plants. Mitochondrial and chloroplast genomes are small and contain only a limited number of genes. The focus of most genome research is on the nuclear genome, though mitochondrial genomes have been extremely useful for the identification of fish species and populations. Genomics is the science that studies the genome. The genetic information stored in DNA cannot be used

without being transcribed into RNA which then, with very few exceptions, must be translated into proteins in order to have biological functions. The term genomics often is used to cover not only this narrow sense genomics, but also transcriptomics, and in many cases proteomics as well. As (Fig. 3) shows, the entire DNA content of an organism (the genome) is transcribed into RNA (the entire RNA content of the organism is called the transcriptome), and the RNA is translated into proteins (the proteome). Genomics, transcriptomics, and proteomics are sciences that study the genome, transcriptome, and proteome respectively.

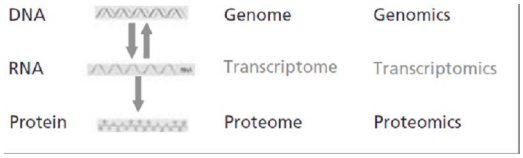

Fig. 3: The entire DNA Content of an Organism Source: NAS, 2007.

Thus, Genomics can be divided into structural genomics, which studies the structures, organization and evolution of genomes, and functional genomics, which studies expression and functions of the genomes.

2.1 Functional Genomics

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data such as genome sequencing projects produced by genomic projects to describe gene (and protein) functions and interactions. Unlike genomics, functional genomics focuses on the dynamic aspects such as gene transcription, translation and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts and protein products. Functional genomics includes all the steps of transferring information in DNA to functional gene product and assessing the effects that are occurring referred to as complete gene expression. Gene expression can be controlled at transcription, RNA processing (RNA transcript to mRNA), RNA transport and localization (nucleus versus cytosol), mRNA stability, translation and protein stability or structure (Alberts *et al.,* 2008).

The scope of functional genomics ranges from expression profiling, the relationship between genome expression and functions, discovery of gene functions and their interrelationships, understanding networking among genes in relation to carrying out their functions, to proteomics and protein-protein interactions.

The goal of functional genomics is to understand the relationship between an organism's genome and its phenotype. The term functional genomics is often used broadly to refer to the many possible approaches to understanding the properties and function of the entirety of an organism's genes and gene products. This definition is somewhat variable; Gibson and Muse, (2007) define it as "approaches under development to ascertain the biochemical, cellular, and/or physiological properties of each and every gene product" while Pevsner, (2009) includes the study of non-genetic elements in his definition: "the genome-wide study of the function of DNA (including genes and non-genetic elements), as well as the nucleic acid and protein products encoded by DNA". Functional genomics involves studies of natural variation in genes, RNA and proteins over time (such as an organism's development) or space (such as its body regions), as well as studies of natural or experimental functional disruptions affecting genes, chromosomes, RNAs or proteins.

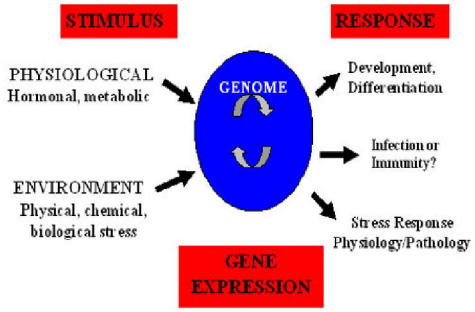

Fig. 4: Functional Genomics Source: Gracey and Cossins (2003).

The promise of functional genomics is to expand and synthesize genomic and proteomic knowledge into an understanding of the dynamic properties of an organism at cellular and/or organismal levels. This would provide a more complete picture of how biological function arises from the information encoded in an organism's genome. The possibility of understanding how a particular mutation leads to a given phenotype has important implications for human genetic diseases, as answering these questions could point scientists in the direction of a treatment or cure.

Potential application areas include clinical diagnostics, agro biotechnology, environmental biotechnology and pharmacogenomics. Although functional genomics remains young enough that people argue over their definition, few squabble over the value of this field. Advances in areas from gene expression to proteomics promise to push ahead basic research, biotechnology, and medicine. In fact, some experts predict an annual compound growth rate of 28% for the next six years in commercial sectors of functional genomics. As functional genomics moves

forward, it will provide many options for applications in aquaculture and capture fisheries.

2.2 Structural Genomics

Structural genomics seeks to describe the 3-dimensional structure of every protein encoded by a given genome. This genome-based approach allows for a high-throughput method of structure determination by a combination of experimental and modeling approaches. Structural genomics attempts to determine the structure of every protein encoded by the genome, rather than focusing on one particular protein. With full-genome sequences available, structure prediction can be done more quickly through a combination of experimental and modeling approaches, especially because the availability of large number of sequenced genomes and previously-solved protein structures allows scientists to model protein structure on the structures of previously solved homologs.

Because protein structure is closely linked with protein function, the structural genomics has the potential to inform knowledge of protein function. In addition to elucidating protein functions, structural genomics can be used to identify novel protein folds and potential targets for drug discovery. Structural genomics involves taking a large number of approaches to structure determination, including experimental methods using genomic sequences or modeling-based approaches based on sequence or structural homology to a protein of known structure or based on chemical and physical principles for a protein with no homology to any known structure. The determination of a protein structure through a structural genomics effort often (but not always) comes before anything is known regarding the protein function. This raises new challenges in structural bioinformatics, which is determining protein function from its 3D structure.

Structural genomics emphasizes high throughput determination of protein structures. This is performed in dedicated centers of structural genomics. While most structural biologists pursue structures of individual proteins or protein groups, specialists in structural genomics pursue structures of proteins on a genome wide scale. This implies large scale cloning, expression and purification. One main advantage of this approach is economy of scale. On the other hand, the scientific value of some resultant structures is at times questioned. A Science article from January 2006 analyzes the structural genomics field, (Chandonia and Brenner, 2006).

One advantage of structural genomics, such as the Protein Structure Initiative, is that the scientific community gets immediate access to new structures, as well as to reagents such as clones and protein.

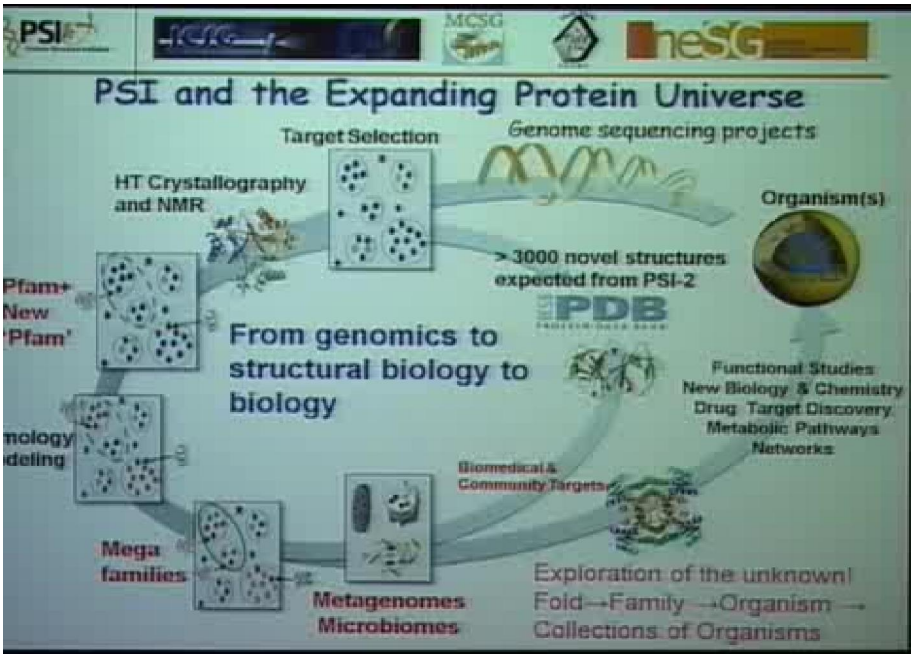

Fig. 5: Structural Genomics
Source: NAS (2007).

A disadvantage is that many of these structures are of proteins of unknown function and do not have corresponding publications. This requires new ways of communicating this structural information to the broader research community. The Bioinformatics core of the Joint Center For Structural Genomics (JCSG)

has recently developed a wiki-based approach namely, The Open Protein Structure Annotation Network (TOPSAN) for annotating protein structures emerging from high-throughput structural genomics centers.

3.0 Application Of Genomics

Genomics application areas include marine biology, preclinical drug safety evaluation, genetics, controlling infectious disease, health care, clinical diagnostics, agro biotechnology, environmental biotechnology and pharmacogenomics.

3.1 Application of Genomics in Marine Biology

Genomic approaches are expected to provide essential information for studying, monitoring and exploiting biodiversity in the oceans. According to Davis, 2004, the field of genomics was initially developed by biologists working on the biology of terrestrial species and the key factor in the emergence of this discipline was the existence of well-defined and intensely studied model organisms such as baker's yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila melanogaster,* a nematode worm *Caenorhabditis elegans,* mouse ear cress *Arabidopsis thaliana* and more recently, the mouse *Mus musculus*. These model organisms were developed to study animal and terrestrial plant biology and, although much of the information obtained from them is of fundamental interest, their study has often been justified and driven by their use as tools to address concrete, applied problems. These problems are essentially of two types:

- ❖ The one directly relevant to the human population: disease (including both infectious diseases and diseases of a developmental nature such as cancer) for the animal models.
- ❖ The other for food production (in a wide sense, including the effects of both development factors and disease on plant production) for plant models.

In order to facilitate the transfer of knowledge to humans or crop plants, work on these model organisms has concentrated primarily on conserved traits that are, in many cases, understood in considerable depth. This approach allowed the establishment of large research communities and the development of extensive resources around these model systems and this was a key factor in the transition to genome-scale biology.

The context for marine biology is significantly different, the accent being more on understanding how organisms function in the context of their particular ecosystem than on asking general questions about their biology. This does not mean that genomic approaches are not relevant to marine biology but rather that they need to be applied in a different way. For instance, for marine biologists, the concept of a model organism is used in a much more flexible manner. In some contexts it could be useful to have a very complete model organism for which both genome sequence data and functional genomics tools are available, whereas for other questions models may not need to allow such in-depth analysis.

In some situations, even the organism-level approach itself is not relevant, hence the development of metagenomic approaches in which marine biosystems are directly sampled and sequenced (Beja *et al,*. 2000) also, Venter *et al*., 2004 who carried out high-throughput sequencing on DNA from microplankton obtained by filtering water from the Sargasso Sea through a 3 μm filter. This type of approach not only represents a very interesting method of obtaining a snapshot of the genetic complexity of a particular biosystem, but also obviates the need for culture methodologies for the constituent organisms. Metagenomics was pioneered in marine biology and provides a good example of how genomic approaches can be adapted to address the questions posed by marine biologists. However, whilst metagenomics provides a broad overview of the genetic composition of ecosystems, more detailed analyses will require organism-level approaches.

3.2 Application of Genomics in Preclinical Drug Safety Evaluation

The age of genomics has opened up many opportunities in drug research and development. Expectations have been high for the usage of genomics and the benefits that genomics can provide to reinvigorate many stages of drug discovery and development. In drug safety assessment, the application of genomics and especially gene expression profiling has been heralded as a means to minimize late stage drug attrition by improving pharmaceutical companies' abilities to predict toxicity and to assess safety risks (Ulrich & Friend 2002; Suter *et al.,* 2004). Beyond the early opinions and promises, there has been a pragmatic and collegial effort amongst molecular toxicologists to fully evaluate the practical uses of gene expression data in drug safety evaluation.

Genomics emerged as a mature science with the sequencing of the human and other genomes along with the development of DNA microarrays and the computing power to analyze the multiple data points generated. These combined factors allow for fully comprehensive and rapid investigations of gene expression (Schena *et al.,* 1998). The sequencing of genomes of the mouse and rat, in particular, enabled the expansion of the use of DNA microarrays in toxicological studies. A concept recognized for a number of years has been that changes in gene expression can be predictive of toxicity by revealing early biological responses to xenobiotics (Mac-Gregor

et al., 1995; Farr and Dunn, 1999) since many biological responses to xenobiotics are manifest at the transcriptional level. Hence, differential gene expression studies are highly applicable to the academic study of toxicology and to practical applications in assessing toxicity.

The application of genomics, including transcriptional profiling, in toxicology has been termed toxico-genomics (Nuwaysir *et al.,* 1999; Aardema and Mac-Gregor 2002; Mac-Gregor, 2003). Experimental studies using DNA microarray technology have led to an increasing knowledge of gene expression responses. This body of information along with the technology brings about the ability to screen chemicals and drugs for toxic potential earlier in their development than previously possible (Rodi *et al.,* 1999; Johnson and Wolfgang, 2000).

As an industry, pharmaceutical companies are evaluating and using gene expression profiling approaches for several purposes: in drug candidate selection and drug development decision making, from predictive toxicity screening or profiling studies; and investigative studies aimed at risk assessment (Rodi *et al.,* 1999; Ulrich and Friend, 2002; Lord, 2004; Peterson *et al.,* 2004; Suter *et al.,* 2004). Strong support for the application of genomics in drug development has also come from the regulatory agencies, particularly the Food and Drug Administration in the United States as evidenced in their document‘‘Innovation or Stagnation: Challenge and opportunity on the critical path to new medical products’’ (March 2004). In the future it is expected that submissions of drugs for approval by regulatory bodies will in one form or another incorporate genomics data to support claims for efficacy, safety or for selection of responsive patients (Petricoin *et al.,* 2002; Kasper *et al.,* 2005).

3.3 Practical Applications of Genomics in Genetics

Genome sequence data now provide tools for the development of practical uses for genetic information. Deoxyribonucleic acid is an invaluable tool in forensics because - aside from identical twins - every individual has a uniquely different DNA sequence. Repeated DNA sequences in the human genome are sufficiently variable among individuals that they can be used in human identity testing. The FBI uses a set of thirteen Short Tandem Repeat (STR) DNA sequences for the Combined DNA Index System (CODIS) database, which contains the DNA fingerprint or profile of convicted criminals. Investigators of a crime scene can use this information in an attempt to match the DNA profile of an unknown sample to a convicted criminal. DNA fingerprinting can also identify victims of crime or catastrophes, as well as many family relationships, such as paternity. While we think of forensics in terms of identifying people, it can also be used to match donors and recipients for organ transplants, identify species, establish pedigree or even detect organisms in water or food.

The basis of many diseases is the alteration of one or more genes. Testing for such diseases requires the examination of DNA from an individual for some change that is known to be associated with the disease. Sometimes the change is easy to detect, such as a large addition or deletion of DNA or even a whole chromosome. Many changes are very small, such as those caused by SNPs. Other changes can affect the regulation of a gene and result in too much or too little of the gene product. In most cases if a person inherits only one mutant copy of a gene from a parent, then the normal copy is dominant and the person does not have the disease; however, that person is a carrier and can pass the disease on to offspring. If two carriers produce a child and each passes the mutant allele to the child (a one-in-four probability), that individual will have the disease.

Several different mutations in a gene often lead to a particular disease. Many diseases result from complex interactions of multiple gene mutations, with the added effect of environmental factors. Heart disease, type-2 diabetes and asthma are examples of such diseases. Many diseases do not show simple patterns of inheritance. Although not everyone with the mutation develops the disease, the risk is much higher than for individuals without the mutation.

Newborns commonly receive genetic testing. The tests detect genetic defects that can be treated to prevent death or disease in the future. Apparently normal adults may also be tested to determine whether they are carriers of alleles for cystic fibrosis, Tay-Sachs disease (a fatal disease resulting from the improper metabolism of fat), or sickle cell anemia. This can help them determine their risk of transmitting the disease to children. These tests as well as others (such as for Down's syndrome) are also available for prenatal diagnosis of diseases. As new genes are discovered that are associated with disease, they can be used for the early detection or diagnosis of diseases such as familial adenomatous polyposis (associated with colon cancer) or p53 tumor-suppressor gene (associated with aggressive cancers). The ultimate value of gene testing will come with the ability to predict more diseases, especially if such knowledge can lead to the disease's prevention.

Gene therapy is a more ambitious endeavor: its goal is to treat or cure a disease by providing a normal copy of the individual's mutated gene. The first step in gene therapy is the introduction of the new gene into the cells of the individual. This must be done using a vector (a gene carrier molecule), which can be

engineered in a test tube to contain the gene of interest. Viruses are the most common vectors because they are naturally able to invade the human host cells. These viral vectors are modified so that they can no longer cause a viral disease.

Gene therapy using viral vectors does have a few drawbacks. Patients often experience negative side effects and expression of the desired gene introduced by viral vectors is not always sufficiently effective. To counter these limitations, researchers are developing new methods for the introduction of genes. One novel idea is the development of a new artificial human chromosome that could carry large amounts of new genetic information. This artificial chromosome would eliminate the need for recombination of the introduced genes into an existing chromosome. Gene therapy is the long-term goal for the treatment of genetic diseases for which there is currently no treatment or cure.

3.4 Application of Genomics in Controlling Infectious Disease

Genomics has brought us to the threshold of a new era in controlling infectious diseases. These studies will likely lead to the development of new disease prevention and treatment strategies for plants, animals and humans alike. For instance, understanding pathogen genes, their expression and their interaction will lead to new antibiotics, antiviral agents and designer immunizations. These new DNA-based immunizations are by-products of genomic research and will undoubtedly eventually replace the traditional vaccines made from whole, inactivated microorganisms. This is highly relevant to domesticated animals, where viruses still kill billions of dollars worth of livestock every year.

Understanding the genomes of plants and animals has additional benefits. Gene mapping should allow for the understanding the basis for disease resistance, disease susceptibility, weight gain and determinants of nutritional value. The use of genomic information provides the opportunity to select optimal environments for the healthy growth of plants and animals, to develop disease-resistant strains and to achieve improved nutritional value. Success in these species may well provide important insights needed to improve the health of humans.

3.5 Global Applications of Genomics in Healthcare

This is a selection of international medical research successfully demonstrating the potential for genetic science and technology in healthcare innovation. These illustrations are drawn from countries around the world and present achievements in diverse areas including the creation of new drugs, better health policies and new research methodology.

Through these illustrations the Human Genetic Programme (HGP) aims to give visibility to the current benefits of genetics, facilitate the exchange of ideas and encourage innovation in public healthcare and services. Each case in this expanding list has been chosen according to a specific set of criteria. In order to highlight the wide scope and relevance of genomics, special effort has been made to provide a balanced geographic representation of health research that addresses a range of issues in innovative ways.

Also, scientists who study diseases were given better and more in-depth understanding of the relationship between genes and diseases including Cancer, Diabetes, Alzheimer's disease and Heart disease. Furthermore, genomics allowed medical experts to utilize the genetic information to identify which genes are susceptible to diseases which we are all pre-disposed and identify which genes are defective and how they affect the human body.

4.0 Genomics And Marine Species

The genomic revolution is rooted in medicine and biotechnology, but marine genomics currently delivers a great quantity of data in its own right. At the time of publication, the marine metagenome sequencing of the Global Ocean Sampling (GOS) campaign doubled the content in the public sequence repositories (Yooseph, *et al.,* 2007) and confirmed the astonishing diversity of microbes. The development of genomics resources for marine organisms has been primarily focused on marine microbes which include both prokaryotic and eukaryotic plankton (Thomas and Klaper, 2004), since Marine microbes are important for the Earth System because they control the cycling of elements in the oceans. Autotrophic processes fix carbon and release oxygen; heterotrophic processes result in the recycling of nitrogen, sulfur and phosphorus and other elements. Bacterial metabolism is involved in the chemical transformation of most elements. About half of the annual primary production of the planet occurs in the ocean so the marine ecosystem plays a very important role in maintaining the wellbeing of our global environment.

Despite the obvious importance of marine microbes very little is known of their diversity and its ecological function. Until recently, there were no appropriate techniques available to answer these important questions. The vast majority of these organisms 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. It was only with the development of high-throughput technology to sequence DNA from the natural marine environment that information began to accumulate that demonstrated the exceptional diversity of microbes in the marine environment in fact, most marine microbes

are entirely novel and have not previously been described. Even less is known about their function in the ecosystem or metabolic activity since no function can be assigned to the major part of their genes. Marine microbial assemblages are diverse and unique and the challenge is to discover what functions are played by these microorganisms.

However, genomics resources for other marine taxa are at present limited to the full genome sequence of the 'model species', the purple sea urchin *Strongylocentrotus purpuratus*. Genomics sequencing efforts for other model and non-model species, including the diatom *Thalassiosira pseudonana*, the 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*, are in progress.

Currently, the Gordon and Betty Moore Foundation Marine Microbial Genome Sequencing Project, founded in 2004, have sequenced nearly 180 marine microorganisms, of which 80% are already published. The project is motivated by the fact that marine ecosystems cover more than 70% of the Earth's surface, host the majority of biomass and significantly contribute to global organic matter and energy cycling. Microorganisms are known to be the "gatekeepers" of these processes. Therefore, insights into the genomic basis of their catalytic activities and interaction with the environment will enhance the ability to monitor, model and predict changes in the marine ecosystem.

The impressive number and size of marine genome and metagenome projects are driven by astonishing advancements in sequencing technologies. Current and predicted trends in the development of new sequencing technologies show that the sheer pace of sequence data growth is unlikely to slow (Hall, 2007; Bosch and Grody, 2008; Gupta, 2008; Shendure and Ji, 2008; Metzker, 2010). Thus, genomics including ecological genomics is being transformed into a data‐intensive science with an exponential increase of data (Szalay and Gray, 2006). The rapid development of platforms for high‐throughput experiments at lower costs can be observed in the fields of transcriptomics, proteomics and metabolomics as well, providing scientists with a more holistic view of microbes in their natural, environmental context through multiomic studies. These multiomics studies are extended to metatranscriptomics and metaproteomics, involving analysis of entire microbial communities. Indeed, multiomic studies not only significantly increase the size and complexity of genomic data; they demand the integration of diverse data to maximize scientific insights. The paradigm shift toward high‐throughput experiments also shifts the workload toward bioinformatics and computational resources, which have become a critical factor for success. Indeed, the rate of sequence data generation is far outpacing the rate of increase in CPUs, and the cost of analyzing large datasets produced by, for example, Solexa, already exceeds the cost of generating them (Meyer, 2006; Wilkening, *et al*., 2009; Metzker, 2010). This situation is often characterized by fear‐inducing metaphors such as data tsunami, data avalanche and data deluge. Rather than being a threat to humankind, however, the technological improvements open challenging, but excellent opportunities for marine biology and biotechnology.

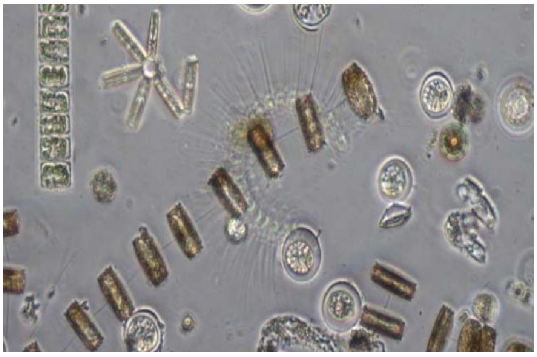

Plate 1: Marine Algae
Source: NAS (2007).

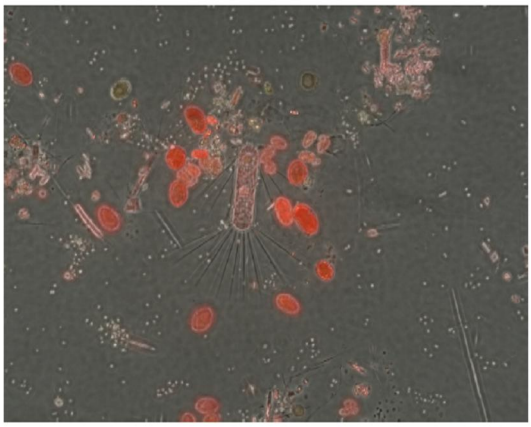

Plate 2: Microbial Biodiversity in Southern Ocean
Source: De-Long *et al.* (2006).

The current exponential data production is a universal fact in biology and will enable a new kind of research limited only by our computing power and bioinformatic capacity (Szalay and Gray, 2006; Committee on Metagenomics: Challenges and Functional Applications, 2007). The metagenomic approach may open a wide door to the rich metabolic and enzymatic repertoire of bacterial and archaeal communities for research in molecular ecology,

ecological genomics and marine biotechnology. However, the ability to make scientific use of the raw sequencing data heavily depends on the bioinformatic resources available to the marine genomics community.

5.0 Functional Marine Genomics Approaches

5.1 Metagenomics

It is generally expected that genomics and metagenomics will provide the diversity and ecological function of marine microbes. Metagenomics refers to all of the genetic information of a natural assemblage that is equivalent to the genomes of all of the organisms in the sample (NAS, 2007). There have been rapid advances in the technology of DNA sequencing which has resulted in an explosion of information on marine microbes. Basically, the first part of the Global Ocean Survey (GOS), which sampled the North Atlantic, Caribbean and a small part of the Pacific Ocean, added DNA sequence information that was equivalent to 50% of all protein-encoding sequences that had previously been deposited in GenBank. Global Ocean Survey confirmed that marine microbes are diverse; indeed it revealed how little is known about the genetic information of natural assemblages. This study highlighted the difficulties of making sense of metagenomic sequence data. A significant proportion of the open reading frames (ORFs, which are presumed to equate to genes) could not be characterized because there were no similar sequences in the databases. This difficulty of interpreting the Global Ocean Survey sequence data exists despite the large number of marine microbes whose whole genome sequences are already in the databases. Largely as a consequence of funding from the Gordon and Betty Moore Foundation (www.moore.org), 155 marine bacterial genomes have so far been sequenced.

Although marine bacteria are well represented in the genomic databases, this information was still not sufficient to decipher the metagenomic information coming from the Global Ocean Survey. The situation is even more complex for eukaryotic microorganisms, which have larger and more complex genomes than bacteria. Also, fewer genome sequences are available for phytoplankton than bacterial species, which increases the difficulties of ascribing function to genetic sequence for eukaryotic microbes.

5.2 Metatranscriptomics

Nevertheless, there is optimism in the oceanographic community that metagenomics will provide new insights into the microorganisms that are present in the oceans. In order to obtain information on microbial function, and especially the ways in which microbial assemblages might respond to changing environmental conditions, researchers are applying the same high-throughput sequencing techniques that have worked so well with metagenomics, to the study of metatranscriptomics. This involves sequencing mRNA (Messenger Ribonucleic acid) isolated from complex communities and synthesizing cDNA (Complemenntary Deoxyribonucleic acid) that can then be sequenced. Metatranscriptomics has the potential to describe how metabolic activity of an assemblage will change under different conditions in the ocean by revealing differences in both known and previously unknown transcripts in natural communities. Methods have now been published (Frias-Lopez *et al.,* 2008; Gilbert *et al.,* 2008) that allow synthesis of high quality cDNA from mRNA extracted from natural assemblages. The cDNA can then be sequenced to indicate how the transcription profile (the metatranscriptome) of communities differs and allows the immediate response to be determined of an assemblage to environmental change. Furthermore, the use of metatranscriptomics to explore gene function in eukaryotes is preferable to metagenomics because the method focuses on expressed gene repertoires rather than whole genomes, which typically contain large amounts of non-coding and therefore difficult to interpret sequences. Coupled with time-series experiments, preferential at long term ecological research sites, metatranscriptomics can help to unravel the functionality of microbial communities and to monitor seasonal changes.

5.3 Proteomics

Another key technique to investigate functional genome analyses of marine microorganisms is proteomics (Schweder *et al.,* 2008). In contrast to metagenomics and metatranscriptomics, proteomics has so far been most useful with bacteria that can be cultured under defined environmental conditions. It gives valuable information on the physiology of individual species and has been widely used to investigate how bacteria respond to stress and starvation conditions. The approach has improved understanding of regulatory networks and physiological strategies which ensure the survival under life-threatening environmental conditions. Proteomics has been successfully applied to physiological analysis of an uncultured bacterial endosymbiont from a deep sea tube worm (Markert *et al.,* 2007). New proteomic techniques allow direct determination of the physiology of key marine bacteria and are thus valuable tools for future functional genome analyses. A proteomic view of cell physiology reaches beyond the mere prediction of putative metabolic functions as coded in the genome sequence. The extreme conditions of the Polar Regions also provide examples of environments in which proteins have evolved to operate efficiently at very low temperatures, ensuring that microorganisms survive in

extreme habitats. However "polar" genomics and proteomics studies are still in their infancy, and there is a very small database of DNA sequences from Polar Regions, but there is much information on protein structure and function.

Acquiring data on the genome, gene expression, protein structure and function in polar species is the basis for understanding the evolutionary forces operating at sub-zero temperature. Any prediction of the physiological costs and evolutionary consequences of global warming is strictly dependent on the knowledge gained on the structure and functioning of polar ecosystems. More important, life sciences are not the only area gaining key insights from studying biological communities inhabiting the poles, because of the strong linkage between organisms and the oceans and atmosphere. The field of proteomics is complementary to genomics in that it provides additional information on gene expression and regulation and also enables the analysis of other biological processes that lead to the production of proteins.

6.0 Case Studies Of Marine Organism Genomics

A genome is the entire set of hereditary instructions for building, running and maintaining an organism and passing life on to the next generation. Because of the vast phyletic diversity of marine organisms, existing genomic model organisms are often of limited relevance, because there is an enormous evolutionary distance separating these models from an organism of interest. To provide adequate tools for marine biologists, therefore, one important aim will be to develop genomic approaches, such as whole genome sequencing and functional genomics, for key species across the evolutionary tree of marine organisms. As genome projects have become cheaper, it has been possible to finance more diverse projects.

Genome projects for additional key species are in progress, including quite a number of marine species, such as *Emiliania huxleyi* (a pelagic coccolithophore), *Hydra magnipapillata, Strongylocentrotus purpuratus* (purple sea urchin), *Litopenaeus vannamei* (the pacific white shrimp), and Amphioxis (the closest living invertebrate relative of the vertebrates), together with key species from other environments such as *Phytophthora infectans* (an oomycete) and the unicellular green algae *Chlamydomonas reinhardtii* (a green algae, for which the genome sequence has been completed). The prokaryotes, progress is even more rapid and many sequenced genomes are available including genomes of several marine organisms such as multiple strains of the pelagic photosynthetic bacteria *Synechococcus* and *Prochlorococcus*. Hence progress is being made towards coverage of all the major eukaryotic and prokaryotic groups.

However, it will be important to actively channel this process in the future, to ensure that coverage extends to all the most important groups and especially to key groups for marine biologists, in particular the eukaryotes, many of which have large genomes.

❖ **Whole Genome Sequence Of Diatom**

The sequencing of the genomes of environmentally important organisms, such as the diatom *Thalassiosira pseudonana*, provides the first complete genome from the heterodont lineage. Of course environmental importance was not the only argument put forward for *Thalassiosira pseudonana*, and the phylogenetic argument itself was also important in addition to other factors such as the biotechnological potential of silicate metabolism in this species *(Thalassiosira pseudonana,* like most diatoms, constructs a silicate exoskeleton, the frustule, and the processes involved in the production of this structure are of great interest for applications in nanotechnology). In this respect, *Thalassiosira pseudonana* is an interesting example for marine biologists of how phylogenetic arguments can be combined with other arguments.

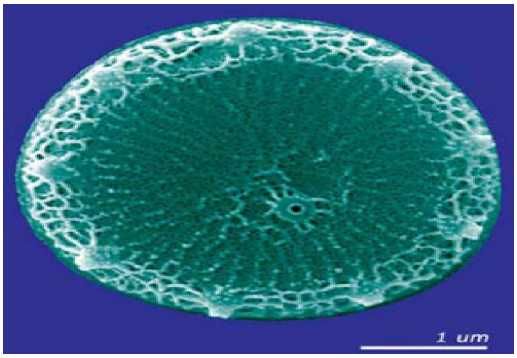

PLATE 3: Diatom *Thalassiosira pseudonana*
Source: Hall (2007)

❖ **Whole Genomes Of Aquatic Animals**

While full genome sequences of some fish: zebrafish, fugu, tetraodon, medaka and three-spined stickleback 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 very different habitats: spanning virtually all aquatic environments from arctic streams to marine tropical areas including underground caves and hypoxic tropical lakes. Consequently, what may be true for a small, tropical freshwater cyprinid, zebrafish, may not be valid for marine tuna weighing

several kilograms. Whenever a new method becomes available for genomic studies, its utility for non-model organisms should be evaluated, as has been done for microarray methodology (Gracey and Cossins, 2003).

The situation with regard to the genomic information on the many species of aquatic invertebrates is even less satisfactory than on fish, especially now that one knows that more than one third of the genes of the recently sequenced genome of Daphnia pulex (Davidson, 2010) do not have counterparts in other so far sequenced genomes. Especially these Daphnia-specific genes respond rapidly to environmental disturbances (Davidson, 2010). Taken together, the genome data on fish and Daphnia suggest both rapid evolution and rapid development of genetic responses to environmental changes.

- **Whole Genome Sequence Of Atlantic Cod**

Recently, scientists from Norway have investigated and present the genome sequence of Atlantic cod *Gadus morhua*. The genome assembly was obtained exclusively by 454 sequencing of shotgun and paired-end libraries, and automated annotation identified 22,154 genes. Genome sequence provided evidence for complex thermal adaptations in its haemoglobin gene cluster and an unusual immune architecture compared to other sequenced vertebrates. Atlantic cod has lost the genes for MHC II, CD4 and invariant chain (Ii) that are conserved feature of the adaptive immune system of jawed vertebrates and, are essential for the function of this pathway. These observations affect fundamental assumptions about the evolution of the adaptive immune system and its components in vertebrates.

Plate 4: **Atlantic Cod**
Source: Gracey and Cossins, (2003).

- **Puffer Fish Genome Features In Draft Sequences**

The fugu genome, the first vertebrate genome to be sequenced after human, was obtained using the whole-genome shotgun method (Lighner and Redman, 1998). This sequence draft enabled a number of interesting observations, such as differences in specific protein families between human and fugu. The Tetraodon genome sequence was subsequently produced (Davidson, 2010), also with the whole-genome shotgun method albeit with a higher redundancy in sequence reads (8.3 vs. 5.6). Both puffer fish possess about 70 different families of transposable elements against only 20 for human or mouse, but in puffer fish they comprise two to three orders of magnitude fewer copies. Interestingly in Tetraodon, SINE and LINE families are distributed in opposite regions of the genome compared to human or mouse: SINEs are more abundant in G + C-rich sequences in mammals, and in A + T-rich regions in Tetraodon, and vice versa for LINE elements.

More surprisingly, these initial studies of Tetraodon and fugu showed a number of differences in their genomes. For instance a G + C-rich region present in both Tetraodon and mammal genomes is absent in fugu. Also some gene families such as type I cytokines and their receptors, present in all vertebrates studied so far, were notably difficult to find in fugu, while over 30 members of the family could be identified in Tetraodon. These discrepancies are most likely attributable to biases in clone libraries or differences in methodologies, and hopefully should be resolved as the genomes reach completion. When comparing fish and mammal gene catalogs, surprisingly few major differences could be documented when using the Gene Ontology (Harris *et al.,* 2004) classification system. More striking differences could be seen using protein domain comparisons: Proteins involved in sodium transport are more abundant in fish, which also contain an allantoin pathway for purine degradation that is absent, in humans. Neutral nucleotidic sequence evolution per year was found to be twice as fast in pufferfish as between human and mouse, and protein evolution also appears to proceed at a faster rate in fish, although the reasons for this are still unclear. It should be noted that these results depend on the dating of the divergence between Tetraodon and fugu (18–30 Mya) (Gracey and Cossins, 2003).

- **Genomes Of Crustacean**

Crustaceans (lobster, shrimp, crab and so on), a remarkable group of organisms filling up all types of habitats in the ocean with a wide array of adaptations, possess the greatest species diversity among marine animals. They are not only abundant in number, but also are among the most commercially exploited food species for human consumption (FAO, 2006). Given their primarily aquatic habitats, however, they are not

as well studied as insects, their terrestrial arthropod relatives. The tiger shrimp (*Penaeus monodon*) has been one of the most important captured and cultured marine crustaceans in the world, especially in the Indo-Pacific region (Lucien-Brun, 1997; FAO, 2006). However, the tiger shrimp industry has been plagued by viral diseases (Lighner and Redman, 1998; Supungul *et al.,* 2002; Leu *et al.,* 2007) resulting in substantial economic losses. Developments in shrimp genomics have been limited although a reasonably good EST database is available (Leu *et al.,* 2010). A genomic analysis for the tiger shrimp will make a key contribution to deciphering the evolutionary history representing the crustacean lineages, especially those living in the ocean. The information contained in the genomic sequences will also benefit the shrimp industry by offering genomic tools to fend off the viral diseases and to improve the breeding program.

❖ **Genome Sequence Of Marine Virus**

Marine viruses, the majority of which are phages, have enormous influences on global biogeochemical cycle's microbial diversity and genetic exchange. Despite their importance, virtually nothing is known about marine viral biodiversity or the evolutionary relationships of marine and non-marine viruses. Addressing these issues is difficult because viruses must be cultured on hosts, the majority of which cannot be cultivated by using standard techniques. In addition, viruses do not have ubiquitously conserved genetic elements such as rDNA that can be used as diversity and evolutionary distance markers. To circumvent these limitations, we developed a method to shotgun clone and sequence uncultured aquatic viral communities.

7.0 Current Challenges In Marine Genomics

1. Lack of Data from Relevant Model Organisms

The wealth of sequence data from both marine microbes and diverse marine provinces is presenting considerable challenges. There are huge numbers of putative genes, the function of which is often unknown and at best only deduced from sequence comparisons. Because more is often known about the genetics and physiology of terrestrial organisms, the number of unknown or putative genes is overwhelming for marine samples because there is so little experimental data on marine model organisms. For instances, all phyla of marine algae synthesise sulphated polysaccharides that have no equivalent in land plants and most of these enzymes constitute completely new protein families: i- and l-carrageenases (Michel *et al.,* 2003), a-agarases (Flament *et al.,* 2007) or fucanases (Colin, 2006). It is not possible to gain any useful information on these proteins by genomics approaches because the sequence data do not exist. It was only through the application of standard biochemical approaches that these enzymes have been identified; otherwise, they would have been annotated as "conserved hypothetical proteins" or given incorrect substrate specificities in genome annotation.

2. The Need for More Relevant Marine Model Organisms

There is an urgent need for more cultures of marine bacteria, archaea, viruses, protozoa and phytoplankton. Most culture collections are based on readily cultivated microbes. When these organisms were isolated, there were no techniques to establish if the isolate was abundant in the natural environment or even if it had any relevant function. Molecular biology has changed that and the isolation of new cultivable microbes can now be based on their abundance and relevance in defined marine habitats. There are a number of novel and innovative approaches to the isolation of new potential-model microorganisms. Rappé *et al.,* (2002) used a dilution approach to isolate SAR11 – the bacterium whose 16S sequence has world-wide distribution (the isolates are now referred to as "*Candidatus Pelagibacter ubique*"). Zengler *et al.,* (2002) described a method of encapsulation of individual bacterial cells, which meant that slow-growing cells could be cultured without being overgrown by rapidly dividing species. So methods exist for isolating useful model microbes from the natural environment; but these are not high-throughput systems and are labour-intensive. Nevertheless, they are probably the only way in which relevant bacteria can be brought into culture since classical microbiology methods have not proved to be useful for difficult-to-cultivate microbes. There is also a need to develop forward and reverse genetics techniques and other molecular resources for relevant marine model organisms. There are still too few examples of phytoplankton that can be manipulated in this way, and without such methods it will be difficult to explore the function of the thousands of genes found only in these organisms.

3. Genomics of Novel Model Microbes

Rhodopirellula baltica provides an example of an environmentally relevant marine bacterium whose genome has recently been sequenced (Glöckner *et al.,* 2003). *Rhodopirellula baltica* is a marine planctomycete isolated from the water column in the Baltic Sea. Genomic analysis has revealed many fascinating and rare features, such as a high number of sulphatases, genes for a C1 metabolism and a global mechanism of gene regulation. But, as with all genomes so far sequenced, function is unknown for a large proportion (~50%) of the genes. Being in culture, it is possible to change growth conditions in a much defined way and investigate how the transcriptome changes in response. Hence it is possible to unravel gene functions and add valuable

information about how the microorganism adapts to changing environmental influences.

***4.* Bioinformatics is Currently a Bottleneck**

Although the giga-base amounts of microbial DNA sequences and other high-throughput approaches have made fundamental improvements to our understanding of uncultivated marine microbes, bioinformatics is often the limiting factor in metagenomic and metatranscriptomic studies. The major hurdles are still;

- The computational aspects of data archiving, analysis and visualization of thousands of millions of DNA sequences which are released to databases and
- Integrating sequences from environmental samples with experimental studies so that unknown genes can be assigned a function.

Novel techniques are required that would allow a numerical description of the specific biological functions unique to specific niches and acting against particular elements.

8.0 Conclusion

Genomics is a highly dynamic research field, currently dominated by human genomics but rapid developments in genomics can afford new opportunities for applications in marine environment, particularly in the areas of Fish genome resources conservation and genetic enhancement.

References

1. Aardema, M. and MacGregor, J. (2002). Toxicology and genetic toxicology in the new era of ‘‘toxicogenomics’’: impact of ‘‘-omics’’ technologies. *Mut. Res*. 499: 13 - 25.
2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008). Molecular biology of the cell garland. *Science*. 64: 69 - 114.
3. Beja, O., Aravind, l., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, l. P. and Jovanovich, S. B. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science.* 289: 1902 - 1906.
4. Chandonia, J. M. and Brenner, S. E. (2006). "The impact of structural genomics: expectations and outcomes". *Science*. 311 (5759): 347 – 351.
5. Colin, S., Deniaud, E., Jam, M., Descamps, V., Chevolot, Y., Kervarec, N., Yvin, J. C., Barbeyron, T., Michel, G. and Kloareg, B. (2006). Cloning and biochemical characterization of the fucanase FcnA: definition of a novel glycoside hydrolase family specific for sulfated fucans. *Glycobiology.* 16: 1021 - 1032.
6. Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. *Nature.* 468: 911 - 920.
7. Davis, R. H. (2004). The age of model organisms. *Nat. Rev. Genet.* 5: 69 - 76.
8. DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N-U., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R., Chisholm, S. W. and Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean’s interior. *Science*. 311: 496 - 503.
9. F. A. O. (2006). State of World Aquaculture: FAO Fisheries Department. Rome. 24pp.
10. Farr, S. and Dunn, R. T. (1999). Gene expression applied to toxicology. *Toxicol. Sci.* 50: 1 - 9.
11. Flament, D., Barbeyron, T., Jam, M., Potin, P., Czjzek, M., Kloareg, B. and Michel, G. (2007). Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families. *Appl. Environ. Microbiol.* 73: 4691 - 4694.
12. Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L. and Schuster, S. C. (2008). Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci.* 105: 3805 - 3810.
13. Gibson, G. and Muse, S. V. (2007). A primer of genome science (3rd Ed.). Sunderland, MA: Sinauer Associates, 122pp.
14. Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P. and Joint, I. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE.* 3(8): 3042.
15. Glöckner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T. and Ludwig, W. (2003). Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc. Natl. Acad. Sci.* 100: 8298 - 8303.
16. Gracey, A. Y. and Cossins, A. R. (2003). Application of microarray technology in environmental and comparative physiology. *Annu. Rev. Physiol*. 65: 231 -258.
17. Gupta, P. K. (2008). Single‐molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol.* 26: 602 ‐ 611.
18. Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *J. Exp. Biol.* 210: 1518 ‐ 1525.
19. Hollywood, K., Brison, D. R. and Goodacre, R. (2006). Metabolomics: current technologies and future trends. *Proteomics* 6: 4716 - 4723.
20. Johnson, D. E. and Wolfgang, H. I. (2000). Predicting human safety: screening and computational approaches. *Drug Discov. Today.* 5: 445 - 454.
21. Joshua, L. and Alexa, T. Mc. (2001). "'Ome Sweet 'Omics -- A genealogical treasury of words". *The Scientist*. 15: p7.
22. Joyce, A. R. and Palsson, B. Ø. (2006). The model organism as a system integrating ‘omics’

data sets. *Nature Rev. Mol. Cell. Biol*. 7: 98 - 210.
23. Kasper, P., Oliver, G., Silva-Lima, B., Singer, T. and Tweats, D. (2005). Joint EFPIA/CHMP SWP workshop: the emerging use of omic technologies for regulatory non-clinical safety testing. *Pharmacogenomics*. 6: 181 - 184.
24. Leu, J. H., Chang, C. C., Wu, J. L., Hsu, C. W., Hirono, I., Aoki, T., Juan, H. F., Lo, C. F., Kou, G. H. and Huang, H. C. (2007). Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *penaeus monodon*. *BMC Genomics*. 8: 120 - 133.
25. Leu, J. H., Chen, S. H., Wang, Y. B., Chen, Y. C., Su, S. Y., Lin, C. Y., Ho, J. M. and Lo, C. F. (2010). A review of the major penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four penaeid shrimp. 65pp.
26. Lighner, D. V. and Redman, R. M. (1998): Shrimp disease and current diagnostic methods. *Aquaculture*. 164: 201 - 220.
27. Lord, P. G. (2004). Progress in applying genomics in drug development. *Toxicol. Lett*. 149: 371 - 375.
28. Lucien-Brun, H. (1997). Evolution of world shrimp production. Fisheries and Aquaculture. *World Aquaculture*. 28: 21 - 33.
29. MacGregor, J. T. (2003). The future of regulatory toxicology: impact of the biotechnology revolution. *Mutat. Res*. 75: 236 - 248.
30. MacGregor, J. T., Farr, S., Tucker, J. D., Heddle, J. A., Tice, R. R. and Turteltaub, K. W. (1995). New molecular endpoints and methods for routine toxicity testing. *Fundam. Appl. Toxicol*. 26: 156 - 173.
31. Margulies, M., Egholm, M., Altman, W. E. and Attiya, S. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 437: 376 – 380.
32. Markert, S., Arndt, C. and Felbeck, H. (2007). Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science*. 315: 247 - 250.
33. McKusick VA, Ruddle FH (1987) A new discipline, a new name, a new journal. *Genomics*. 1: 1 – 2.
34. Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet*. 11: 31 - 46.
35. Meyer, F. (2006). Genome Sequencing vs. Moore's Law: Cyber challenges for the next decade. *CTWatch Quarterly*. 2: 14 -17.
36. Michel, G., Helbert, W., Kahn, R., Dideberg, O. and Kloareg, B. (2003). The structural bases of the processive degradation of iota-carrageenan, a main cell wall polysaccharide of red algae. *J. Mol. Biol*. 334: 421 - 433.
37. N. A. S. (2007). The new science of metagenomics: revealing the secrets of our microbial planet. ISBN: 0-309-10677-X 170pp.
38. Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C. and Afshari, C. A. (1999). Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinogenesis*. 24: 153 - 159.
39. Peterson, R. L., Casciotti, L., Block, L., Goad, M. E., Tong, Z., Meehan, J. T., Jordan, R. A., Vinlove, M. P., Markiewicz, V. R., Weed, C. A. and Dorner, A. J. (2004). Mechanistic toxicogenomic analysis of WAY-144122 administration in Sprague-Dawley rats. *Toxicol. Appl. Pharmacol*. 196: 80 - 94.
40. Petricoin, E. F., Hackett, J. L., Lesko, L. J., Puri, R. K., Gutman, S. I., Chumakov, K., Woodcock, J., Feigal, D. W., Zoon, K. C. and Sistare, F. D. (2002). Medical applications of microarray technologies: a regulatory science perspective. *Nat. Genet*., 32: 474 - 479.
41. Pevsner, J. (2009). Bioinformatics and functional genomics (2nd Ed.). *Hoboken*, *NJ,* 7: Wiley-Blackwell, 15pp.
42. Rappé, M. S., Connon, S. A., Vergin, K. L. and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*. 418: 630 - 633.
43. Ray, G. C. (1988). Ecological diversity in coastal zones and oceans. In E. O. Wilson, ed. Biodiversity. Washington, D.C.: National Academy Press, 36 - 50.
44. Rodi, C. P., Bunch, R. T., Curtiss, S. W., Kier, L. D., Cabonce, M. A., Davila, J. C., Mitchell, M. D., Alden, C. L. and Morris, D. L. (1990). Revolution through genomics in investigative and discovery toxicology. *Toxicol. Pathol*. 27: 107 - 110.
45. Rogers, Y. H. and Venter, J. C. (2005). Genomics: massively parallel sequencing. *Nature*. 437: 326 – 327.
46. Schena, M., Heller, R. A. Theriault, T. P., Konrad, K., Lachenmeier, E., and Davis, R. W. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol*. 16: 301 - 306.
47. Schweder, T., Markert, S. and Hecker, M. (2008). Proteomics of marine bacteria. *Electrophoresis*. 29: 2603 - 2616.
48. Shendure, J., and Ji, H. (2008). Next - generation DNA sequencing. *Nat. Biotechnol*. 26: 1135 - 1145.
49. Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. and Herndl, G. J. (2006). Microbial

diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl. Acad. Sci*. USA 103: 12115 - 12120.

50. Supungul, P., Klinbunga, S., Pichyangkura, R., Jitrapakdee, S., Hirono, I., Aoki, T. and Tassanakajon, A. (2002): Identification of immune-related genes in hemocytes of black tiger shrimp (Penaeus monodon). *Mar Biotechnol*. 4: 487 - 494.
51. Szalay, A., and Gray, J. (2006). 2020 Computing: Science in an exponential world. *Nature*. 440: 413 - 414.
52. Ten Bosch, J. R. and Grody, W. W. (2008). Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J. Mol. Diagn*. 10: 484 - 492.
53. Thomas, M.A. and Klaper, R. (2004). Genomics for the ecological toolbox. *Trends Ecol. Evol*. 19: 439 - 445.
54. Ulrich, R. and Friend, S. H. (2002). Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat. Rev. Drug Discov*. 1: 84 -88.
55. Van Straalen, N. M. and Roelofs, D. (2006). An introduction to ecological genomics. Oxford University Press, Oxford. Pp85.
56. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. W. U. (2004). Environmental genome shotgun sequencing of the Sargasso Sea". *Science*. 304: 66 - 74.
57. Wilkening, J., Wilke, A. N. D. and Folker, M. (2009). Using Clouds for Metagenomics: A Case Study. In: Proceedings IEEE Clouds. 12 - 19.
58. Worm, W., Barbier, E. B. and Beaumont, N. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science* 314: 787 - 790.
59. Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J. and Remington, K. (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS. Biol*. 5: 16pp.
60. Zengler, K., Toledo, G., Rappé, M. S., Mathur, E. J., Short, J. M. and Keller, M. (2002). Cultivating the uncultured. *Proc. Natl. Acad. Sci*. U. S. A. 99: 15681 - 15686.

5/24/2018