Chapter 12
Bioinformatics–Driven Big Data Analytics in Microbial Research

Ratna Prabha
Indian Council of Agricultural Research, India

Anil Rai
Indian Council of Agricultural Research, India

D. P. Singh
Indian Council of Agricultural Research, India

ABSTRACT
With the advent of sophisticated and high-end molecular biological technologies, microbial research has observed tremendous boom. It has now become one of the most prominent sources for the generation of “big data.” This is made possible due to huge data coming from the experimental platforms like whole genome sequencing projects, microarray technologies, mapping of Single Nucleotide Polymorphisms (SNP), proteomics, metabolomics, and phenomics programs. For analysis, interpretation, comparison, storage, archival, and utilization of this wealth of information, bioinformatics has emerged as a massive platform to solve the problems of data management in microbial research. In present chapter, the authors present an account of “big data” resources spread across the microbial domain of research, the efforts that are being made to generate “big data,” computational resources facilitating analysis and interpretation, and future needs for huge biological data storage, interpretation, and management.

INTRODUCTION
Microorganisms are the oldest and most abundant life forms on the earth. Entwined with the remarkable power of natural selection, they successfully developed evolutionary aptitude to adapt diverse and often hostile environments. Because of their occurrence in normal to extreme environments, microbes can be seen as a living population with remarkable genotypic and phenotypic properties intrinsic to them. When explored and identified for their natural behavior, these organisms can act as prototype for molecular, biochemical and physiological manipulations rendering their direct or indirect benefits to crops and soils. The microbial biology is of great importance as it can represent functional trends of adaptation and evolution of the old and primitive organisms.
that can mirror organisms in all kinds of natural ecosystems. Because of their primitive existence on the earth and the diverse behavior shown by them in diverse environmental conditions, these organisms are believed to possess unique genetic, molecular and biochemical mechanisms to cater their adaptation needs and address their evolutionary diversification (Keller & Zengler, 2004; Ouzounis, 2002). The journey of microorganisms parallel to the plants, animals and human being has led to their co-evolution in such a way that microbes are entwined with all kinds of biological systems. They have an associative biology which when deciphered, often leads to various levels of information about pathogenic, symbiotic and beneficial associations and the benefits associated with the function.

Microbial biology has many interesting lessons connecting the science to ecosystem function at basic to a very complex level. Molecular biology deals with the most fundamental aspect of biology and provides information about the internal functionalities of the organisms. The area basically finds its roots planted in the DNA which consists of all the genetic information pertaining to the creation, evolution and sustenance of life forms. Within the organisms, the information is usually encrypted in such a way that deciphering the same generates a large volume of data through many experimental platforms like whole genome sequencing projects, microarray technologies, mapping of single nucleotide polymorphisms (SNP), transcriptomics, proteomics, metabolomics, phenomics etc (Adnan, 2010; Fenstermacher, 2005; Waterman, 1995). For the management of the huge data and extraction of dynamic wealth of information from different experimental platforms, bioinformatics have come across the globe to solve the problems of biologists.

Microbial research in the present era is basically related to evolutionary diversification, phylogenetic lineages, identification and characterization, and adaptation of microbes in different extreme environmental conditions. Their interactions with biotic and abiotic entities including that with plants, microbes and other fauna, beneficial functions for plants, soil, environment and human being has led to explore the in-depth insights within the microbial genomes, proteomes and metabolomes (Goodman, 2002). During the past decades, the field of genomics has witnessed an explosion in huge data or ‘big data’ that of long strings of base pairs (A, T, G, C) which encode for all the genetic information required for life and sustainability of the organisms. Similar is true for –omics era in microbial research that has emerged as a major source of big data and need special and focused attention to bring out comprehensive, accurate and precise definitions of the life forms and their functions (Figure 1) (Casari et al., 1995).

In the present era of technological advancement, the microbial science is facing a phase of fast and dynamic data revolution. Since there is huge microbial diversity on the earth, the area is leading to the generation of large quantity of data on various aspects related to diverse microbial genomes and metagenomes, the impact on global climate change and genetic blueprints of microbes. Researchers are working to generate data, its refinement, analysis and visualization to make sense of the data to make it meaningful (McCulloch, 2013; Pearson & Lipman, 1988). Two decades back when the human whole genome sequencing project was taken up, it was thought to generate huge data (Watson, 1990). Now, in comparison to human genome, collection of the genomes of microbes within the human body and with other habitats like ocean, soil etc. has come-up rather more voluminous. The whole content of 3 billion base pairs with about 20,000 genes of human genome seems only a little part of the microbial genomes with a rough content of 100 billion base pairs and millions of genes that make up microbial world within the human body (Singer, 2013). Now, with the decipherization of many human genomes along with that of plants, animals and microbes, the public domain repertoire NCBI contains twenty petabytes of data (1 petabyte is
Related Content

Similarity Searching of Medical Image Data in Distributed Systems: Facilitating Telemedicine Applications
[www.igi-global.com/chapter/similarity-searching-medical-image-data/73071?camid=4v1a](www.igi-global.com/chapter/similarity-searching-medical-image-data/73071?camid=4v1a)

Facilitating and Augmenting Collaboration in the Biomedical Domain
Nikos Karacapilidis, Manolis Tzagarakis, Spyros Christodoulou and Georgia Tsiliki (2012). *International Journal of Systems Biology and Biomedical Technologies* (pp. 52-65).
[www.igi-global.com/article/facilitating-augmenting-collaboration-biomedical-domain/63046?camid=4v1a](www.igi-global.com/article/facilitating-augmenting-collaboration-biomedical-domain/63046?camid=4v1a)

Cathelicidins Revisited: Molecular Evolution, Structure and Functional Implications
[www.igi-global.com/article/cathelicidins-revisited/89398?camid=4v1a](www.igi-global.com/article/cathelicidins-revisited/89398?camid=4v1a)

Dis2PPI: A Workflow Designed to Integrate Proteomic and Genetic Disease Data
[www.igi-global.com/article/dis2ppi-workflow-designed-integrate-proteomic/77811?camid=4v1a](www.igi-global.com/article/dis2ppi-workflow-designed-integrate-proteomic/77811?camid=4v1a)