

Chapter 25

Applications of Supercomputers in Sequence Analysis and Genome Annotation

Gerard G. Dumancas

Oklahoma Medical Research Foundation, USA

ABSTRACT

In the modern era of science, bioinformatics play a critical role in unraveling the potential genetic causes of various diseases. Two of the most important areas of bioinformatics today, sequence analysis and genome annotation, are essential for the success of identifying the genes responsible for different diseases. These two emerging areas utilize highly intensive mathematical calculations in order to carry out the processes. Supercomputers facilitate such calculations in an efficient and time-saving manner generating high-throughput images. Thus, this chapter thoroughly discusses the applications of supercomputers in the areas of sequence analysis and genome annotation. This chapter also showcases sophisticated software and algorithms utilized by the two mentioned areas of bioinformatics.

INTRODUCTION

Bioinformatics is often regarded as a discipline in its infancy. However, this interdisciplinary field had its historical start in 1960s when computers emerged as a vital tool in molecular biology. With the notable efforts of Margaret O. Dayhoff, Walter M. Fitch, Russell F. Doolittle among others, this area emerged as an approach to managing and interpreting massive data generated by genomic research. Bioinformatics today represent a convergence of various fields, which involve modeling of biological phenomena, genomics, biotechnology and information technology, analysis and interpretation of data, and the development of novel algorithms for analyzing biological datasets. With the advent of the emergence of these large amount of biological datasets, scientists are often confronted with the issues of analyzing and interpreting these massive information and datasets in a less amount of time, requiring high accuracy, and cost-saving. In the last few decades, this has been made possible with the emergence of supercomputers. The wide array of available supercomputers has made it possible to analyze and interpret biological datasets and systems

DOI: 10.4018/978-1-5225-8903-7.ch025

in a more convenient manner. Nowadays, because of supercomputers, groundbreaking bioinformatics research is made possible. A good example is the discovery of novel genes associated with different diseases. With the discovery of these genes, scientists have come up to a deeper understanding of the etiology of various unexplained diseases caused genetically. Consequently, various drugs and treatments were discovered to counteract such diseases. Within the area of bioinformatics, sequence analysis and genome annotation are among the two of the emerging and most important branches. In the recent years, supercomputers play very important roles in the successes of such branches.

The objective of this chapter is to provide the readers a clear understanding of the specific applications of supercomputers in the two most emerging areas of bioinformatics, sequence analysis and genome annotation. Though supercomputers play critical roles in such areas, the audience is not often aware of the potential applications that may arise from them. A universal understanding that constitute both fundamental and experimental methodologies will enhance the development and progress of such areas. Thus, the major motivation of this chapter is to provide the abovementioned understanding by discussing and analyzing the fundamentals of several examples centered on the various applications of supercomputers in sequence analysis and genome annotation. While the content of this chapter may be technical to some readers, we encourage them to review some basic concepts of genetics and biochemistry as well as to look at the definition of terms to better understand this chapter.

BACKGROUND

Genotype analysis involves studying the association between genotype and phenotype, and the genotype frequencies. Genetic association studies are aimed primarily in identifying genetic variants that explain differences in phenotypes among individuals in a study population. Once association is found between the gene(s) and the phenotype, scientists would be able to understand the mechanism of action and disease etiology in individuals and consequently characterize the relevance and importance of such in the general population. The long-term goal of these studies is to identify better treatment and prevention strategies. Association or any genetic analyses usually require highly intensive mathematical calculations. Supercomputers play a critical role in the success of such calculations. Prior to genetic association analyses, any genotype information needs to undergo two critical steps—sequence analysis and genome annotation. Applications of supercomputers in genotype analyses involve a wide array of applications and will be discussed in the mentioned areas below.

SUPERCOMPUTERS IN SEQUENCE ANALYSIS

Sequence analysis is the most commonly performed task in bioinformatics. It was one of the first bioinformatics techniques founded in ~1970 (Webb-Roberts, 2004). DNA sequencing is simply any process used to map out the sequence of the nucleotides that comprise a strand of DNA. After the discovery of the double helix shape of DNA in 1953, and seeing how it is comprised of a series of ladder like units known as DNA nucleotides, the primary goal has been to find out just how the sequence of those little nucleotides leads to the physical characteristics of an organism, that is, whether what your hair color, your skin color, and every other detail from your bone marrow to the tip of your hair. Thus, DNA sequenc-

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/applications-of-supercomputers-in-sequence-analysis-and-genome-annotation/228642

Related Content

Paratexts and Documentary Practices: Text Mining Authorship and Acknowledgment from a Bioinformatics Corpus

Nicholas M. Weber and Andrea K. Thomer (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 597-624).

www.irma-international.org/chapter/paratexts-and-documentary-practices/228641

Potential of Bio-Inspiration in 3- and 4-D Printing

(2021). *Inspiration and Design for Bio-Inspired Surfaces in Tribology: Emerging Research and Opportunities* (pp. 294-347).

www.irma-international.org/chapter/potential-of-bio-inspiration-in-3--and-4-d-printing/257604

Competitive Advantage, Open Innovation, and Dynamic Capabilities: Is Sanofi Employing an Open Innovation Strategy?

Geoffroy Labrousse and Med Kechidi (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1556-1580).

www.irma-international.org/chapter/competitive-advantage-open-innovation-and-dynamic-capabilities/228684

Recent Advancements in Microalgae-Biofuel Generation Employing Nano-Additives

Mahesh Pattabhiramaiah, Bhargavi Rajarathinam and Shanthala Mallikarjunaiah (2023). *Biomass and Bioenergy Solutions for Climate Change Mitigation and Sustainability* (pp. 188-205).

www.irma-international.org/chapter/recent-advancements-in-microalgae-biofuel-generation-employing-nano-additives/314364

Scaffolds and Tissue Engineering Applications by 3D Bio-Printing Process: A New Approach

Ranjit Barua, Sudipto Datta, Pallab Datta and Amit Roy Chowdhury (2019). *Design, Development, and Optimization of Bio-Mechatronic Engineering Products* (pp. 78-99).

www.irma-international.org/chapter/scaffolds-and-tissue-engineering-applications-by-3d-bio-printing-process/223408