Computing Platforms for Big Biological Data Analytics: Perspectives and Challenges

Zekun Yin, Haidong Lan, Guangming Tan, Mian Lu, Athanasios V. Vasilakos, Weiguo Liu

PII: S2001-0370(17)30035-1
DOI: doi:10.1016/j.csbj.2017.07.004
Reference: CSBJ 192

To appear in: *Computational and Structural Biotechnology Journal*

Received date: 11 March 2017
Revised date: 30 June 2017
Accepted date: 28 July 2017

Please cite this article as: Yin Zekun, Lan Haidong, Tan Guangming, Lu Mian, Vasilakos Athanasios V., Liu Weiguo, Computing Platforms for Big Biological Data Analytics: Perspectives and Challenges, *Computational and Structural Biotechnology Journal* (2017), doi:10.1016/j.csbj.2017.07.004

Computing Platforms for Big Biological Data Analytics: Perspectives and Challenges

Zekun Yin[a], Haidong Lan[a], Guangming Tan[b], Mian Lu[c], Athanasios V. Vasilakos[d], Weiguo Liu[a,*]

[a]*School of Software, Shandong University, Jinan City, Shandong Province, China*
[b]*Institute of Computing Technology, Chinese Academy of Sciences.*
[c]*Huawei Singapore Research Centre, Singapore.*
[d]*Department of Computer Science, Electrical and Space Engineering Luleå University of Technology, SE-931 87 Skellefteå, Sweden.*

Abstract

The last decade has witnessed an explosion in the amount of available biological sequence data, due to the rapid progress of high-throughput sequencing projects. However, the biological data amount is becoming so great that traditional data analysis platforms and methods can no longer meet the need to rapidly perform data analysis tasks in life sciences. As a result, both biologists and computer scientists are facing the challenge of gaining a profound insight into the deepest biological functions from big biological data. This in turn requires massive computational resources. Therefore, high performance computing (HPC) platforms are highly needed as well as efficient and scalable algorithms that can take advantage of these platforms. In this paper, we survey the state-of-the-art HPC platforms for big biological data analytics. We first list the characteristics of big biological data and popular computing platforms. Then we provide a taxonomy of different biological data analysis applications and a survey of the way they have been mapped onto various computing platforms. After that, we present a case study to compare the efficiency of different computing platforms for handling the classical biological sequence alignment problem. At last we discuss the open issues in big biological data analytics.

---

[*] Corresponding author
Email address: `weiguo.liu@sdu.edu.cn` (Weiguo Liu)

Preprint submitted to Computational and Structural Biotechnology Journal August 12, 2017