



## BIG Data in Biology: A Review

Chahat Kubba, Tanya Sharma, Aroshi Panwar, Kamal Rawal \*

Department of Bioethnology, Jaypee Institute of Information Technology, A-10, Sec 62, Noida 201307, Uttar Pradesh, India

Address for Correspondence: Kamal Rawal, [kamal.rawal@gmail.com](mailto:kamal.rawal@gmail.com)

**Keywords** BIG  
Data;  
Bioinformatics;  
MapReduce; MPI;  
GPU.

**ABSTRACT:** With significant advances in high-throughput sequencing technologies and consequently the exponential expansion of biological data, bioinformatics encounters difficulties in storage and analysis of vast amounts of biological data. The gap between sequencing throughput and computer capabilities in dealing with such big data is growing (Almeida, 2012). Considering the annual growth of data generation, the digital universe - data we generate annually - will reach 44 zettabytes, or 44 trillion gigabytes by the year 2020, which is ten times the size of the digital universe in 2013 (Kashyap,2014). Given the urgency of establishing a new computational framework, high-performance computing has become extremely important for large-scale data analysis. In this paper, we review various technologies such as MapReduce, message passing interface (MPI) and graphics processing unit (GPU) in context of bioinformatics and biological domain and also the work conducted by our lab in the area of network based therapeutics (Jaisri and Rawal et al 2016). © 2016 iGlobal Research and Publishing Foundation. All rights reserved.

Conference Proceedings: International Conference on Advances in Plant and Microbial Biotechnology (PMB-2017); JIIT, Noida: February 02-04, 2017

*Indo Global Journal of Pharmaceutical Sciences( ISSN 2249 1023 ; CODEN- IGJPAI; NLM ID: 101610675) indexed and abstracted in EMBASE(Elsevier), SCIRUS(Elsevier),CABI, CAB Abstracts, Chemical Abstract Services(CAS), American Chemical Society(ACS), Index Copernicus, EBSCO, DOAJ, Google Scholar and many more. For further details, visit <http://iglobaljournal.com>*