

## Bioinformatics: how it helps to boost modern biological research

Suvakanta Barik, Nilesh Rai, Pradeep Mishra, Santosh K. Singh and Vibhav Gautam

The trending of bioinformatics as an interdisciplinary branch of science in recent decades has enormous impact on biological research, as it enables *in silico* analysis of collected, archived and retrieved biological data<sup>1,2</sup>. The primary focus of bioinformatics during its nascent stage of development was just handling biological data, i.e. DNA, RNA and protein sequences through use of the basic computational tools. However, with increasing popularity of *in situ* and microarray gene expression and high-throughput gene sequencing studies involving massive data, it has immersed to play a bigger role in biological research output and applications<sup>3</sup>. Currently, every field of biological research, including biotechnology, molecular biology, medicine, pharmaceuticals, agriculture, environment and climate change uses bioinformatics and computational biology approaches for better planning of research and output thereof. Thereby, it has become an essential part of biological research. Today, a number of popular tools, software and database are available for a better understanding of biological complexity<sup>4</sup>, and analysis and application of the result output through user-friendly interface and web resources. This remarkable development in utilizing massive and complex biological data in solving biological problems through use of bioinformatics tools has revolutionized the method of research<sup>4-6</sup>. The availability of user-friendly interactive bioinformatics workflows to (re)analyse the publicly available genomics, transcriptomics, proteomics, and metabolomics datasets (commonly called omics data) has contributed immensely to solving biological problems<sup>7,8</sup>. The *in silico* analysis of omics-based high-throughput data and precision prediction has reduced the cost and time for experimental validations.

Biological sequence analysis, which is the hallmark of bioinformatics applications, uses bioinformatics tools to determine the different features of a nucleotide sequence or protein molecule. The sequence analysis tools uncover the underlying features of DNA/RNA sequences, i.e. gene structures, site of post-translational modification, active

sites, regulatory elements and distributions of introns, exons, and are also used in the analysis of genetic markers through SNP (single nucleotide polymorphism) structure–function prediction, comparative genomics and evolutionary relationships, etc.<sup>9</sup>. Besides the large number of currently available web-based tools for sequence analysis, the basic bioinformatics tools available for the purpose are BLAST, FASTA, ClustalW, PROBCONS, MUSCLE, MAFFT, T-Coffee, etc. However, the recent use of high-throughput sequencing methods to generate massive sequencing data has led to the development of advanced bioinformatics tool that use complex software programs and algorithms to analyse the sequences on a large scale<sup>10-14</sup>. This has improved the scope of data analysis in shorter span of time than if analysed individually.

Bioinformatics is no longer limited to mapping and sequencing of genomes, as it had been during the early phase of its development. The evolution of genomics to ‘functional genomics’ assigns functional relevance to genomic information, i.e. gene, gene product and function, interactions, and regulation at genome-wide scale. This uses the high-throughput sequencing, microarray and transcriptome analysis, gene annotation, etc. Unlike the earlier used gene-by-gene approach which produces limited data, functional genomics produces massive and complex biological data which can be analysed through efficient bioinformatics tool. Functional genomics analyses through bioinformatics tools have enabled scientists to perform events of genome analysis on a global scale for candidate gene discovery, gene phenotype association, i.e. improved productivity, abiotic stress tolerance, root system architecture, disease classification, plant–pathogen interactions, gene ontology (GO, <http://www.geneontology.org/>) analysis<sup>15,16</sup>, gene module analysis, and genome-scale miRNA expression profiling<sup>17</sup>.

The next-generation sequencing (NGS) and microarray analysis techniques have opened the flood gate for high-throughput sequencing of many plant/animal genomes of diverse impor-

tance. The massive amount of raw data produced in NGS through different platforms, i.e. Roche/454, Illumina/Solexa, Ion torrent, PacBio RS sequencer, GridION, MinION sequencers, etc. poses a great challenge for downstream analysis<sup>18</sup>. The bioinformatics tools have been developed for analysis of these data. Some of the web-based tools are Galaxy (<http://galaxyproject.org>), Artemis (<https://www.sanger.ac.uk/science/tools/artemis>), GSAP (<http://www.broadinstitute.org/>), etc. Similarly, sequence data through NGS are analysed using NCBI (<https://www.ncbi.nlm.nih.gov/>), GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), EMBL (<https://www.embl.org/>), Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>), PlantGDB (<http://www.plantgdb.org/>), TAIR (<https://www.arabidopsis.org/>), KEGG (<https://www.genome.jp/kegg/>), Gramene (<http://www.gramene.org/>), etc. These web resources are used in plant genome research for developing improved plants with agronomic and horticultural importance.

Comparative genomics uses bioinformatics tools for the analysis of gene regulatory networks between various biological species through structural and functional relationship of the genomes. It reconstructs the evolutionary relationship of living or extinct organisms which are related to each other by fluctuating degrees of evolutionary divergence from a common ancestor by comparing DNA/RNA/protein sequence signature. The phylogenetic studies are carried out utilizing software and computer programs, i.e. Molecular Evolutionary Genetics Analysis (MEGA) and ClustalW/Clustal Omega respectively<sup>19-22</sup>.

For human healthcare in the field of medicine, bioinformatics plays a crucial role in drug discovery and targeting by establishing a strong association with the disease of interest. The identification of the drug target against which the candidate drug acts, is one of the most research-intensive parts of the drug discovery process. The increasing pressure to manufacture more cost-effective low-risk drugs within a limited span of time has generated incredible interest in bioinformatics<sup>23</sup>. The unprecedented growth of the ‘omics’ technologies in the last

decade (transcriptomics, proteomics, metabolomics and pharmacogenomics) has facilitated in monitoring the molecular level of disease processes through systemic characterization of the drug target physiology and has reduced the high rejection rate in the drug target recognition. These allow easy screening of large libraries and databases against the growing list of targets in identifying the drug leads<sup>24</sup>.

With the rapidly expanding field of research in bioinformatics-driven genomics, the focus has shifted from increasing fundamental understanding of biology to its application in areas of agriculture. This has been fuelled by the continued growth and reduced cost of DNA sequencing, NGS and genotyping<sup>25</sup>. The massive information coming from these advanced studies on plant sciences is managed and utilized in plant breeding and transgenic research. The ultimate goal is to develop abiotic stress-tolerant, insect- and disease-resistant plants with improved crop productivity. The tools and database systems for mining data for crop plants are (<http://triticeaetool-box.org>, <http://wheatis.org/>), the International Rice Information Consortium (<http://iric.irri.org>), etc. Similarly, next generation sequencing (NGS) and microarray data are being used to address the climate change driven abiotic changes. Similarly, the analysis of omics data from microorganisms having a role in bioremediation sheds light on the mechanism of genes action, global metabolic and regulatory networks of complex gene and proteins. The best example of microbial bioremediation using bioinformatics approach is ionization resisting properties of the bacterium *Deinococcus radiodurans*<sup>26</sup>.

On the basis of facts that have been summarized earlier, bioinformatics plays an efficient role to influence almost all aspects of human health and plant science. With the massive data being generated everyday through different computational tools, the greatest chal-

lenge bioinformatics faces today is to filter out meaningful productive data. Many tools that are developed to handle enormous datasets are not always done so with the specific needs of biological sciences. The future challenge requires discipline-specific novel and enhanced algorithms for developing powerful bioinformatics tools and skilled manpower for efficient analysis of these data.

**Conflict of interest:** The authors declare no potential conflict of interest.

1. Koch, I. and Fuellen, G., *Brief. Bioinformat.*, 2008, **9**, 232–242.
2. Asthana, V., Bioinformatics at present: Description and disciplines in bioinformatics.
3. Sharma, R. K., *Int. J. Biomed. Data Min.*, 2016, **5**, 122.
4. Edwards, D., Stajich, J. and Hansen, D., *Bioinformatics: Tools and Applications*, Springer Science & Business Media, 2009, 1–564.
5. Mount, D. W., *Bioinformatics: Sequence and Genome Analysis*, Cold Spring Harbor Laboratory Press, New York, USA, 2001, 1–564.
6. Rao, V. S., Das, S. K., Rao, V. J. and Srinubabu, G., *Afr. J. Biotechnol.*, 2008, **7**, 495–503.
7. Rai, A. and Saito, K., *Curr. Opin. Biotechnol.*, 2016, **37**, 127–134.
8. Manzoni, C., Kia, D. A., Vandrovцова, J., Hardy, J., Wood, N. W., Lewis, P. A. and Ferrari, R., *Brief. Bioinformat.*, 2016, **19**, 286–302.
9. Raza, K., arXiv preprint arXiv: 12051125, 2012.
10. Hsiao, E. S., Chang, Y. W., Lin, P. and Liao, P.-C., *J. Proteomics Bioinformat.*, 2012, **5**, 15–23.
11. Naghizadeh, S., Rezaeitabar, V., Pezeshk, H. and Matthews, D., *J. Proteomics Bioinformat.*, 2012, **5**, 24–30.
12. Simpson, R. J. and Mathivanan, S., *J. Proteom. Bioinf.*, 2012, 05.
13. Hashemi, M., Behrangi, N., Borna, H. and Akbarzadeh, A., *J. Proteomics Bioinformat.*, 2012, **5**, 50–53.
14. Kanehisa, M. and Bork, P., *Nature Genet.*, 2003, **33**, 305.

15. Cheng, L., Lin, H., Hu, Y., Wang, J. and Yang, Z., *PLoS ONE*, 2014, **9**, e107187.
16. Attrill, H. *et al.*, *Database*, 2019, **2019**.
17. Kong, W., Zhao, J. J., He, L. and Cheng, J. Q., *J. Cell. Physiol.*, 2009, **218**, 22–25.
18. Schuster, S. C., *Nature Methods*, 2007, **5**, 16.
19. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
20. Yang, Z., Zhao, T., Ma, Q., Liang, L. and Wang, G., *Front. Plant Sci.*, 2018, **9**, 927.
21. Barik, S. *et al.*, *Sci. Rep.*, 2015, **5**, 14611.
22. Barik, S., Sarkar Das, S., Singh, A., Gautam, V., Kumar, P., Majee, M. and Sarkar, A. K., *Genomics*, 2014, **103**, 114–121.
23. Ortega, S. S., Cara, L. C. L. and Salvador, M. K., *Drug Metab. Drug Interact.*, 2012, **27**, 199–207.
24. Katara, P., *Netw. Model. Anal. Health Informat. Bioinformat.*, 2013, **2**, 225–230.
25. Edwards, D. and Batley, J., *Plant Biotechnol. J.*, 2010, **8**, 2–9.
26. Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. and Daly, M. J., *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 44–79.

**ACKNOWLEDGEMENTS.** N.R., S.K.S. and V.G. thank the Centre of Experimental Medicine and Surgery, Institute of Medical Science, Banaras Hindu University, Varanasi for funding and internal grants. V.G. also acknowledges Science and Engineering Research Board (SERB) for funding (EME/Q/2019/000025).

*Suvakanta Barik is in the Chemical Engineering Discipline, Indian Institute of Technology, Gandhinagar 382 355, India; Nilesh Rai, Santosh K. Singh and Vibhav Gautam\* are in the Centre of Experimental Medicine and Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221 005, India; Pradeep Mishra is in the Division of Biochemistry, Department of Medical Biochemistry and Biophysics Karolinska Institute, Stockholm, Sweden.*

\*e-mail: vibhav.gautam4@bhu.ac.in