Bioinformatics: Benefits to Mankind

Himanshu Singh*

Dept. of Biotechnology, Lovely Professional University, Phagwara, India

Abstract: The storage and analysis of biological data using certain algorithms and computer softwares is called bioinformatics. The applications that bioinformatics offer to the civilized world are more than just being a researcher's tool for structural and functional analysis. Development and implementation of computational algorithms and software tools facilitate an understanding of the biological processes with the goal to serve primarily agriculture and pharmaceutical sectors. In this paper, we highlight the role of bioinformatics in health care, drug discovery, forensic analysis, crop improvement, food analysis and biodiversity management. Some ethical issues related to bioinformatics research are also covered.
Keywords : Bioinformatics, Computational Biology, Drug Discovery, Chemiinformatics.

Introduction

Bioinformatics is defined as "a scientific discipline that encompasses all aspects of biological information acquisition, processing, storage, distribution, analysis, and interpretation that combines the tools and techniques of mathematics, computer science, and biology with the aim of understanding the biological significance of a variety of data" [1]. Bioinformatics-tools include any products that store, organize, evaluate, integrate, analyse, and/or distribute biological data [2]. Bioinformatics developed after automated protein and DNA sequencing technologies were introduced around the mid-1970s, and after researchers started to use computers as central sequence repositories that were accessible remotely, a development that occurred in the mid- to late 1980s [3].

Bioinformatics covers a multiple of scientific disciplines and encompasses various areas of genomic analysis, including: *(i)* computational genomics (the storage and analysis of DNA sequence using various software), *(ii)* proteomics (laboratory methods used to determine what proteins are actually expressed and possible cellular location), *(iii)* transcriptional profiling (the analysis of mRNAs at different stages of cell growth and conditions), *(iv)* functional genomics (combined function prediction and experimental biological analysis of gene function), and *(v)* structure function determinations (predicted as well as determined 3-D structures and comparison to known protein structures and functions) [4].

In general, the aims of bioinformatics are three-fold [5]. First, at its simplest bioinformatics organises data in a way that allows researchers to access existing information and to submit new entries as they are produced. The second aim is to develop tools and resources that aid in the analysis of data. Development of such resources dictates expertise in computational theory, as well as a thorough understanding of biology. The third aim is to use these tools to analyse the data and interpret the results in a biologically meaningful manner. Some of the well-known databases include GenBank (Genetic Data Bank), SWISS-PROT, PDB (Protein Data Bank), PIR (Protein Information Resource), SCOP (Familial and Structural Protein Relationships), CATH (Hierarchical Classification of Protein Domain Structures), etc. These databases are available as public domain information and hosted on various Internet servers across the world. Basic research and modelling is done using these databases with the help sequence analysis tools like BLAST(Basic Local Alignment Search Tool),

FASTA(Fast-All), CLUSTALW, etc., and the modelled structures are visualized using visualization tools such as WebLab, MOLMOL, Rasmol, etc.

Bioinformatics is faced with a strong demand for immediate solutions; because the genomic data that are being uncovered encode many biological insights whose deciphering can be the basis for dramatic scientific and economical success. With the pre-genomic era that was characterized by the effort to sequence the human genome just being completed, the post-genomic era that concentrates on harvesting the fruits hidden in the genomic text. Today the researcher investigating humans or any interesting organism, whether microbe, plant or animal, for which the data on the genome sequence are already or will be available in the near future, is offered all the necessary information concerning the complete DNA sequence. It is predicted that bioinformatics research will have a dramatic impact on improving our understanding of diverse areas [6]. Bioinformatics is expected to be an important contributor to the global economy. Research in this field has already made a major impact on the pharmaceutical industry and drug discovery, agriculture, and health care and environment. Bioinformatics tools will increase the potential of curing inherited human diseases and producing new human medicines. Benefits of bioinformatics to mankind are vast; some of these are discussed below.

Health Care

Bioinformatics is a rapidly emerging field of biomedical research. As genomic and biomedical data is accumulating, bioinformatics tools are helping in managing and analysing this data for health care [7]. Bioinformatics can be considered a sub-discipline of Biomedical Informatics which addresses the biomedical challenges at molecular level [8]. It ha s a direct impact not only on the development of new diagnostics, therapeutics, and vaccines but also on understanding of infectious disease mechanisms, pathogen-host interactions and transmission cycles [9]. Availability of bioinformatics tools is helping in realizing the potential benefits of human genome project by identification of disease susceptibility genes and the development of many new treatments, the ability to predict those patients at risk for experiencing adverse reactions or patients with a high probability of experiencing improved efficacy. Application of bioinformatics approach for understanding of proteomic data can lead to accurately correlate clinical parameters of patient responsiveness to a particular therapy [10].

Uncovering of the differences in people's genetic makeup will not only yield increased understanding of the genes that cause or contribute to diseases, but ultimately will help health care providers to prescribe the right medicine in the right amount for each person. In addition, this knowledge might prevent the negative side effects of the "one-size-fits-all" method of prescribing drugs that is most often used today [11]. Bioinformatics is also likely to lead to new classifications of diseases. Diseases traditionally characterised through symptoms (phenotype) may be reclassified according to their genetic characteristics (genotype) [12].

Drug Discovery

The drug discovery process is complex, time consuming and very expensive. Typically, about 5 years time is required to develop a candidate drug, while the clinical phases leading, possibly, to the commercial availability of the drug are even longer for a total cost of more than 700 Million dollars [13]. The focus of the pharmaceutical industry has shifted from the trial and error process of drug discovery to a rational, structure based drug design. A successful and reliable drug design process could reduce the time and cost of developing useful pharmacological agents.

Computational methods are used for the prediction of 'drug-likeness' which is nothing but the identification and elimination of candidate molecules that are unlikely to survive the later stages of discovery and development. Drug-likeness could be predicted by genetic algorithms [14]. Genomics has become a major source of drug targets, and bioinformatics is crucial for finding and validating novel targets so as to minimize investment in laboratory resources [15]. With regularly updated public databases, bioinformatics can contribute to these processes by providing functional information of target candidates and correlating this information to the biological pathways [16]. Bioinformatics tools can be used to understand the mode of action and the efficacy of a given compound and to help elucidating the pathophysiology of a disease [13].

Many tools have been developed for drug discovery. A comprehensive bioinformatics software package 'BioSuite' was developed that integrates the functions of macromolecular sequence and structural analysis, chemoinformatics and algorithms for aiding drug discovery [17]. DOVIS: a utility software was developed for

high-throughput virtual screening using AutoDock as molecular-docking-based virtual screening is an important tool in drug discovery that is used to significantly reduce the number of possible chemical compounds to be investigated [18].

Forensic Analysis

Bioinformatics and forensic DNA are inherently interdisciplinary and draw their techniques from statistics and computer science bringing them to bear on problems in biology and law. Personal identification and relatedness to other individuals are the two major subjects of forensic DNA analysis [19]. Mass disaster cases require managing, analysing, and comparing large numbers of biological samples and DNA profiles, making necessary the use of electronic laboratory information management systems for large-scale sample logging and tracking, coupled with bioinformatics tools for DNA database searching [20]. Polymerase chain reaction (PCR) amplification of short tandem repeat (STR) loci is followed by use of bioinformatics tools such as GeneScan^{TM} and GenoTyper^{TM} for determination of the presence or absence of STR alleles associated with a sample [21]. Whenever crime scene investigation needs identification of bacteria, insects, and plants, genomic sequences can be resequenced using microarray and analysed using bioinformatics standard techniques [19].

Bioinformatics tools such as CODIS (Combined DNA Index System), DNA-View, MDKAP (Mass Disaster Kinship Analysis Program), MFISys (Mass Fatality Identification System) were used for forensic analysis of World Trade Centre disaster, New York, USA, September 11, 2001. DNA database searching bioinformatics tools have also been used for the victims of the South Asian Tsunami disaster [20].

Crop Improvement

For centuries, humans have selected plant varieties that best fit their purposes and developed crop plants that have many advantages compared to natural (wild) plants in quality, quantity and farming practises. However, as many traits are involved in resistance and quality, it is extremely difficult to improve them in combination. Genomics revolution enables to achieve a broad as well as detailed genetic understanding of plant performance at multiple levels of aggregation. The complex biological processes that make up the mechanisms of pathogen resistance and provide quality to our crops are now open for a systematic functional analysis using bioinformatics tools [22]. Current genome programs generate a large amount of data that will require processing, storage and distribution to the multinational research community. The data include not only sequence information, but information on mutations, markers, maps, functional discoveries, etc. Key objectives for plant bioinformatics include: to encourage the submission of all sequence data into the public domain, through repositories, to provide rational annotation of genes, proteins and phenotypes, and to elaborate relationships both within the plants' data and between plants and other organisms [23].

Crop bioinformatics is undeniably fundamental in modern plant breeding. In particular, the combination of modeling and simulation with real breeding programs and the standardization of data collection for phenotypic, environmental and genomic information is opening the way to next generation plant breeding [24]. Bioinformatics is used for sequence analysis, transcriptome, proteome, metabolome analysis [25]. The one-dimensional –omics information (e.g. genomics, proteomics, metabolomics) is translated into cellular reality through bioinformatics-based data integration [24]. Bioinformatics will change the ways in which biological research will be conducted in 2050. Each researcher will spend more time on the computer and the Internet to generate and describe data and experiments, to analyse the data and find other people's data relevant for comparison, to find existing knowledge in the field and to relate it to his or her results into the current body of knowledge, and to publish his or her results to the world [25]. Bioinformatics tools and databases must be developed so as to build an in-silico model of the plant, encompassing all the levels of organization that can be used to predict responses of all systems and to rationally engineer plants for agronomic benefit [26].

Food Analysis

Agricultural biotechnology has been used to engineer a variety of agronomically important crops, such as corn, soybean, potato and cotton. Commercialization of such products requires premarket authorization. Extensive food, feed and environmental safety assessments are conducted for each product prior to commercialization; this includes an allergy assessment for each introduced protein in the genetically modified crop. A principal aim is to prevent the introduction of known allergens or clinically cross-reactive incomplete allergens so as to protect the food-allergic individuals from unexpected and unwanted exposure to offending

foods or allergens. The use of bioinformatics algorithms and specialized databases containing known and suspected allergens provides important and effective tools to identify known and potentially cross-reactive allergens that may pose a risk to those with pre-existing allergies. If a significant match is identified, the introduced protein must be evaluated further using sera from individuals who are allergic to the similar allergenic protein to determine if it is likely to cause cross-reactions. *Bacillus thuringiensis* insecticidal proteins, common non-allergenic food proteins and some randomly selected corn *(Zea mays)* proteins were assessed for allergenicity by comparing to sequences in ALLERGEN 3 database using bioinformatics tool FASTA [27]. Structural database of allergenic proteins, SDAP (Structural Database of Allergenic Proteins)-Food, was used to aid in predicting the Immunoglobulin E-binding potential of novel food proteins and cross-reactivities among known allergens [28].

Research has shown that nutrients affect gene expression and formation of various proteins at discrete points in the processes leading to enzymes, structural proteins, and other chemicals on which life depends. Thus, the amount and even the form of nutrients present during gene expression can affect the synthesis of protein, resulting in less of a protein being produced, production of a less than optimally functional form, or no protein at all. Each of those possibilities exists due to the hereditary form of genes present and whether the genes are normal or contain polymorphisms that affect gene expression. With enough data and proper bioinformatics tools, scientists will be able to detect all biological effects of a food component, food or diet [29]. Bioinformatics tools can be used to predict various disorders individually, including the difference of susceptibility or the difference of disease risk, and the effectiveness of drugs by analysing the Single Nucleotide Polymorphisms, which represents personal variation of genomes. Therefore, it is thought in the future that more personalized information on food functions can be provided for disease prevention, in correspondence with individual differences of genome [30].

Biodiversity Management

Scientific information about biodiversity is either scattered in many databases or resides on paper or available through museums, herbaria and repositories. The information about organisms is usually not available and thus not used for making policies and decisions regarding them. Biodiversity management and conservation requires that information about it be available easily [31]. Research activities are being conducted around the world to improve our ability to manage biodiversity information, and the emerging field of biodiversity informatics is attempting to meet the challenges posed by this domain [32]. This new suite of methodologies and tools extends contemporary computer science and informatics principles within the context of biodiversity data. Biological complexity of the organisms, presence of large number of species and ecosystems, interactions of species with multiple abiotic factors makes biodiversity information very complex. Good database designs are required for management and long term stewardship of information related to biological diversity [33]. Biodiversity informatics requires frameworks and approaches that can accommodate the full range of biological information - from molecules to morphological features, to populations, to habitats [34]. Mechanisms are required that enable us to make correlations among new datasets and traditional data and data from even more disparate sources, such as ecosystem-scale global change and carbon cycle data, compile those data in new ways, analyse and synthesize them, and present the results in an understandable and usable manner. [32]. Many nations are now Parties to the Convention on Biological Diversity, and many have joined other treaties related to biodiversity conservation. These treaties will have direct consequences to the way in which biodiversity information is gathered, organised, maintained and disseminated within countries [35].

Technologies such as 'scratchpads' [36] and 'DNA bar code methods' [37] have been developed for managing and sharing biodiversity information. Software packages such as Plant Info, Garden Info, Seed Pack, Biolit and databases such as Fungal Database Meliolales, Sacred Groves of Kerala and Wild Ornamental Plants of the Western Ghats have also been developed for organising, managing and monitoring biodiversity information [38]. Tools can be developed for invasive species risk assessment by creating predictive ecological models for all species not native to a particular region and then using these models to develop strategies for avoiding species invasions [39]. The fledgling field of biodiversity informatics looks set to deliver major advances that could turn the Internet into a giant global biodiversity information system [40].

Other Benefits

Genomics, proteomics and bioinformatics offer the potential to identify genes or loci regulating traits of economic importance in animals. Production traits that are proposed as eligible for transgenic modification include increased growth rate and improved carcass composition, improved feed utilisation, modified milk composition, improved mohair production improved reproductive performance and increased disease resistance [41]. In veterinary research, bioinformatics tools were used in the detection of new castle diseases and to generate novel solutions for the continued improvement and development of molecular diagnostics [42].

Bioinformatics is being used in the building of global databases in microbiology to build an accumulative knowledge repository that captures the reams of experimental data and meta-data about micro-organisms and to develop general data mining tools for knowledge discovery within this data-rich environment, in order to establish dynamically updated and flexible portals upon the observed bacterial diversity and related biotechnological innovations with the ultimate goal of valorising newly discovered insights as new applications or end-products [43].

Bioinformatics is also helping in climate change studies. There are many organisms which use carbon dioxide as their sole carbon source and increasing levels of carbon dioxide emission is one of the major causes of the global climate change. The study of genomes of these microbial organisms, which is possible through bioinformatics, helps in proposing ways to decrease the carbon dioxide content [44].

Ethical Issues

Bioinformatics research creates number of ethical, legal and social, security concerns like fairness in the use of genetic information, privacy and confidentiality of genetic information, psychological impact and stigmatization due to an individual's genetic differences [6]. If personal data become accessible to unauthorized persons, if information on genetic dispositions are passed on to employers and insurers, they can use the genetic information to the unfair disadvantage of some people. It can also promote racism as particular diseases appear in particular groups only. The more exact investigation of the human genome will show specific differences between the DNA sequences of ethnical groups, which are connected with their environment and the physiological particularities of the group [45]. Development of drugs that target specific individuals might lead to stratification [46].

The Ethical, Legal and Social issues research programme (ELSI) is considered the world's largest bioethics program. The programme is jointly funded by the US Department of Energy (DOE) and the National Institutes of Health (NIH). The programme is concerned about safeguarding the privacy of individuals and groups who contribute DNA samples for large-scale sequence variation studies [47].

References

1. D. Benton, "Bioinformatics—Principles and Potential of a New Multidisciplinary Tool," *Trends in Biotechnology*, vol. 4, pp. 261-272, 1996.
2. C. Vorndran and R.L. Florence, "Bioinformatics: Patenting the Bridge Between Information Technology and the Life Sciences," *IDEA - The Journal of Law and Technology*, vol. 42- 1, pp. 93-142.
3. A. Persidis, "Data mining in biotechnology," *Nature Biotechnology*, vol. 17, pp. 828-830, 1999.
4. R.J. Zagursky and D. Russel, "Bioinformatics: Use in Bacterial Vaccine Discovery," *Biotechniques*, vol. 31, pp. 636-659, September 2001.
5. N.M. Luscombe, D. Greenbaum and M. Gerstein, "What is Bioinformatics? A Proposed Definition and Overview of the Field," *Method Inform Med*, vol. 40, pp. 346-358, 2001.
6. P. Nagarajan, "An Over View of Bioinformatics," *Trends in Biomaterial and Artificial Organs,* vol. 17(2), pp. 4-8, 2004.
7. J. H. Kim, "Transforming Healthcare through Translational Bioinformatics."
8. R. A. Greenes and E.H. Shortliffe, "Commentary: Informatics in Biomedicine and Health Care," *Academic Medicine*, vol. 84-7, July 2009.

9. W. Degrave , C. Huynh, D. Roos, A. Oduola and C.M. Morel, *"Bioinformatics for Disease Endemic Countries: Opportunities and Challenges in Science and Technology Development for Health," IUPAC Journal, 2002.*

10. V.R. Srinivasa, K.S. Das, K.R. Nageswara and E.K. Kusuma, "Bioinformatics is a Key Life Science R & D Activity," *International Journal of Bioinformatics Research,* Volume 1:2, pp. 81-84, 2009.

11. "Personalized Medicine*," Issue Brief, About Science,* October 2006.

12. N. Brown, A. Nelis, B. Rappert, A. Webster and G.J.B.V. Ommen, "Bioinformatics: A technology Assessment of Recent Developments in Bioinformatics and Related Areas of Research and Development Including Highthroughput Screening and Combinatorial Chemistry," Final Report for the Science and Technological Options Assessment (STOA) Unit, European Parliament, May 1999.

13. A.A. Impiombato and D.D. Bernardo, "Computational Biology and Drug Discovery: From Single-Target to Network Drugs," *Current Bioinformatics*, vol. 1, pp. 3-13, 2006.

14. M. M. Babu, "Integrating Bioinformatics, Medical Sciences and Drug Discovery."

15. S. Pollock and H. M. Safer, "Bioinformatics in the Drug Discovery Process," *Annual Reports in Medicinal Chemistry*, 2001.

16. Y. Wang, J.F. Chiu and Q.Y. He, "Bioinformatic Application in Proteomic Research on Biomarker Discovery and Drug Target Validation," *Current Bioinformatics*, vol. 2, pp. 11-20, 2007.

17. NMITLI-BioSuite Team, "BioSuite: A comprehensive bioinformatics software package," *Current Science*, vol. 92:1, January 2007.

18. S. Zhang, K. Kumar, X. Jiang, A. Wallqvist and J. Reifman, "DOVIS: an implementation for high-throughput virtual screening using AutoDock," *BMC Bioinformatics* 2008, vol. **9**:126, February 2008.

19. L. Bianchi and P. Lio, "Forensic DNA and Bioinformatics," *Briefings in Bioinformatics*, vol. 8:2, pp. 117-128, March 2007.

20. A. Alonso et al., "Challenges of DNA Profiling in Mass Disaster Investigations," *Croat Med J*, vol. 46:4, pp. 540-548, 2005.

21. D.R. Paoletti, D.E. Krane, M.L. Raymer, and T.E. Doom, "Inferring the Number of Contributors to Mixed DNA Profiles," *Transactions on Computational Biology and Bioinformatics*, May 2010.

22. D. Vassilev, J. Leunissen, A. Atanassov, A. Nenov and G. Dimov, "Application of Bioinformatics in Plant Breeding," *Biotechnol. & Biotechnol. Eq.*, vol. 19, 2005.

23. D. Vassilev, J. Leunissen, A. Atanassov, A. Nenov, G. Dimov and L. Getov, "Application of Bioinformatics in Fruit Plant Breeding," *Journal of Fruit and Ornamental Plant Research*, vol. 14:1, 2006

24. P.Faccioli, A.M. Stanca, C.Morcia, V.Terzi, "From DNA Sequence to Plant Phenotype: Bioinformatics Meets Crop Science," *Current Bioinformatics*, vol.4, pp. 173-176, 2009.

25. S.Y. Rhee, J. Dickerson, and D. Xu, "Bioinformatics and its Applications in Plant Biology," *The Annual Review of Plant Biology,* vol.57, pp.335-60, 2006.

26. G. Coruzzi, "Achieving the Virtual Plant," *US-EC Task Force on Biotechnology*, June 2005.

27. R.E. Hileman et al., "Bioinformatic Methods for Allergenicity Assessment Using a Comprehensive Allergen Database," *International Archives of Allergy and Immunology,* vol. 128, pp. 280-291, 2002.

28. O. Ivanciuc et al., "Detecting Potential IgE-Reactive Sites on Food Proteins Using a Sequence and Structure Database SDAP-Food," *Journal of Agricultural and Food Chemistry,"* vol. 51, pp. 4830-4837, 2003.

29. D. Ghosh, "Fututre Perspectives of Nutrigenomics Foods: Benefits *vs*. Risks," *Indian Journal Of Biochemistry and Biophysics*, vol. 46, pp. 31-36, February 2009.

30. Y. Sumi, "Research and Technology Trends of Nutraceuticals," *Science & Technology Trends,* vol. 28, pp.10-22, July 2008.

31. J. L. Edwards, M.A. Lane and E. S. Nielsen, "Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop," *Science*, vol. 289, September 2000.

32. J.L. Schnase, "Research Directions in Biodiversity Informatics," *Proceedings of the 26th International Conference on Very Large Databases,* Cairo, Egypt, 2000.

33. P.J. Morris, "Relational Database Design and Implementation for Biodiversity Informatics," *PhyloInformatics, vol.* 7, pp. 1-66, 2005.

34. I.N. Sarkar, "Biodiversity Informatics: the emergence of a field," *BMC Bioinformatics*, vol. 10:14:S1, November 2009.

35. I. K. Crain and N. M. Collins, "Bioinformatics Implications of the International Biodiversity Conventions," International Conference on Biodiversity and Bioresources: Conservation and Utilization, Phuket, Thailand, November 1997.

36. V. S Smith, S. D Rycroft, K. T. Harman, B. Scott and David Roberts, "Scratchpads: A Data-Publishing Framework to Build, Share and Manage Information on the Diversity of Life," *BMC Bioinformatics* 2009, vol. 10:14:S6, November 2009.

37. P. Bertolazzi, G. Felici and E. Weitschek, "Learning to Classify Species with Barcodes," *BMC Bioinformatics* 2009, vol. 10:14:S6, November 2009.

38. "Workshop on biodiversity informatics and interlinking of databases," *Current Science*, vol. 90: 3, February 2006.

39. A.T. Peterson and D. A. Vieglais, "Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem," *Bioscience*, vol. 51:5, pp. 363-371, May 2001.

40. F.A. Bisby, "The Quiet Revolution: Biodiversity Informatics and the Internet," *Science*, vol. 289, PP. 2309-2312, September 2000.

41. H. Baldassarre,, B. Wang, C. L. Keefer, A. Lazaris and C. N. Karatzas*, "*State of the art in the production of transgenic goats," *CSIRO*, 2004.

42. K.O. Soetan and M.O. Abatan, "Biotechnology a key tool to breakthrough in medical and veterinary research," *Biotechnology and Molecular Biology Review*, vol. 3:4, pp. 88-94, August 2008.

43. P. Dawyndt, T. Dedeurwaerdere and J. Swings, "Exploring and exploiting microbiological commons: contributions of bioinformatics and intellectual property rights in sharing biological information."

44. B. Jayaram and P. Dhingra, "Bioinformatics for a Better Tomorrow."

45. K.F. Kittowski and H.A. Rosenthal, "Biosemiotics, Bioinformatics, and Responsibility: Ambivalent Consequences of the Deciphering of the Human Genome for Society and Science," Triple C, vol. 5:2, pp. 87-100, 2007.

46. S. Hongladarom, "Ethics of Bioinformatics: A Convergence between Bioethics and Computer Ethics," Asian Biotechnology and Development Review, vol. 9:1, pp. 37-44, 2006.

47. D. Jawdat, "The Era of Bioinformatics," IEEE, pp. 1860-1865, 2006.

＊＊＊＊＊