Critical technologies for bioinformatics

Scientific advances over the last 50 years have provided a basis for parallel revolutions in engineering and in biomedicine. Advances in automation and miniaturization have enabled the development of modern instrumentation that supports large-scale measurement of biological entities. The burgeoning fields of genomics and proteomics, as well as other ‘omics’, keep generating large amounts of molecular expression and interaction data. Advances in cytometry enable quantification of cellular states and related functional properties. The latest scanning and imaging technologies have made it possible to scan and make detailed measurements of whole organisms. Clinical data complement biological data, enabling detailed descriptions of various healthy and diseased states, progression and responses to therapies. The availability of data representing various biological states, processes, and their time dependencies enable the study of biological systems at various levels of organization, from molecule to organism, and even population levels. Multiple sources of data support a rapidly growing body of biomedical knowledge; however our ability to analyze and interpret these data and extraction of new knowledge requires bioinformatics approaches of ever increasing sophistication. This special issue of *Briefings in Bioinformatics* covers eight bioinformatics technologies which have made a significant impact on both the theory and practice of biomedical research. These are representative of important bioinformatics topics that can be described as enabling technologies for generating and sharing of biological knowledge.

Lefranc *et al*. describe IMGT, a system that integrates complex immunogenetic data, ontologies, analysis tools and web resources. This system offers a formal description of objects, processes and relations necessary for conceptualization of knowledge in molecular immunogenetics and provides tools for multi-scale approaches at various levels of hierarchy from molecule to organism and, possibly, population level. In particular, this review offers an insight on merging the conceptual models and ontologies with sequence databases and their combination with related analytical tools. The IMGT technology demonstrates the conversion of precisely defined biological concepts into an integrated bioinformatics environment.

The article by Standley *et al*. describes the latest developments in molecular structure databases, in particular the enhancement of structural information with biochemical, functional and experimental information. A new generation of tools enable querying by keywords/text, sequence similarity and structure similarity. Advanced visualization tools have been developed to assist the analysis and interpretation of molecular structures. The new infrastructure of the 1980s comprised simple molecular databases containing several thousand molecular entries, basic sequence comparison algorithms and simple mathematical models. Current infrastructure comprises a network of databases, web-accessible analytical tools, computer networks and supercomputers that connect users to bioinformatics resources across the world.
Multiple sequence alignment (MSA) methods are among the most important bioinformatics tools whereby biological sequences are compared to identify the regions of similarity that confer evolutionary, structural or functional relatedness. Because the MSA problem is NP-hard, the optimal solution is not possible unless MSA involves a small number of sequences. The rapid growth of biological sequence data, both in length and number of sequences, the need for identification of distant homologs and for web accessibility require increasingly sophisticated algorithms. Katoh and Toh have described the latest version of the MAFFT program for MSA. MAFFT is an example of tools where high-quality MSAs can be performed rapidly over the Internet.

Kumar et al. have described MEGA, a tool for evolutionary analysis of biological sequences. MEGA is designed for comparative analysis of genes and proteins and investigation of their evolutionary relationships. It offers a convenient means for managing data from local files and web repositories, for statistical analysis of data, and visualization of results. In their article, the authors describe the design features that make MEGA a biologist-centric tool, that enable complex analyses where the results are clearly, unambiguously presented to the user through advanced visualization tools.

The analysis of complex biological data brings formidable challenges to the developers of advanced bioinformatics tools. Formal mathematical and statistical approaches are complemented with tools of computational intelligence that employ heuristic algorithms. Computational intelligence methods employ fuzzy logic, search and classification to study systems than learn, evolve and adapt. Fogel provides a primer of computational intelligence methods and their application to selected biological problems.

Systems biology is concerned with the study of complex interactions in biological systems. It uses integrative rather than reductionist approach to discover properties arising from combination of multiple biological entities and their effects on multiple levels of biological organization. Hu et al. have described a visual data mining system VisANT that integrates multiple data types, performs the analysis of complex biological networks, and displays the results through visualization tools.

Transcriptional regulation is a key area of systems biology. Wingender’s article describes the TRANSFAC database and analysis of system whose original aim was to provide a genome-wide map of interaction sites for transcription factors. Over the years, TRANSFAC became the industry standard. However, with the growth of genomic information and genome mapping of model organisms, the original development evolved into a system that provides classifications and ontologies for the domain, and also integrates a catalog of transcription networks and their parts with a set of prediction tools.

Hunter et al. have described the Physiome project that focuses on multi-scale modeling necessary for relating molecular and cellular processes to the events observed at the organ or whole organism. Such multi-scale modeling requires advanced techniques of mechanistic modeling, visualization and numerical techniques.

Advances in bioinformatics are reshaping biomedical research and applications in biotechnology. Integration of large quantities of data, instant access to information, and the ability to perform complex analyses and simulations provide the capacity to amplify the results of biomedical research manyfold. Biomedicine is increasingly becoming an information-rich field. Deciphering qualitative relationships between components of living organisms, and quantification of interactions will depend on further development of bioinformatics technologies. The combination of large-scale screening experiments with large-scale simulations in silico is already used for selection of a limited number of key experiments, resulting in decreased cost and time required for biomedical research and discovery.

Vladimir Brusic1
Shoba Ranganathan2

1Dana-Farber Cancer Institute, Boston, MA 02115, USA
2Macquarie University, Sydney, NSW 2109, Australia