BIOINFORMATICS AND ITS APPLICATIONS IN AGRICULTURE

Jian Xue [1], Shoujing Zhao [1,*], Yanlong Liang [1], Chunxi Hou [1], Jianhua Wang [1]

[1] *College of Biological and Agricultural Engineering, Jilin University, Changchun, China, 130022*

[*] *Corresponding author, Address: College of Biological and Agricultural Engineering, Jiling University, 5988 Renmin Street, Changchun, Jilin, 130022, P.R.China, Tel: +86-431-85095253, Fax: +86-431-85095253, Email: swgc@jlu.edu.cn*

Abstract: The field of bioinformatics emerged as a tool to facilitate biological discoveries more than 10 years ago. With the development of Human Genome Project (HGP), the data of biology increased fabulously and marvelously. The ability to capture, manage, process, analyze and interpret data became more important than ever. Bioinformatics and computers can help scientists to solve it. Here are introduced roles of bioinformatics, meanwhile Web tools and resources of bioinformatics were reviewed. And its applications in agriculture were also discussed.

Keywords: bioinformatics, WorldWideWeb, agriculture.

1. INDUCTION

Bioinformatics which is coming with HGP brings together the fields of life science, computer science and statistics and strives to understand medical and biological systems by the creative application of statistics and computer analysis.

Bioinformatics is the use of computer technology to help scientists keep track of the genetic information they find. Using computers, researchers can gather, store, analyze and compare biological data with great speed and accuracy.

Imagine studying gene structures without the help of a computer. It would take many years to compare the 15,000 genes of Arabidopsis to the genes of a similar plant. And keeping track of the 100,000 genes of a human being would be inconceivable. With computers, the process of comparison is automated. By storing information as it is discovered, computers ease the immense job of genome mapping. But computers can analyze as well as store information. They can be used to construct models that reduce the need for experimentation.

In this way, biotechnology has become more efficient. Scientists are able to use fairly reliable computer-assisted predictions of test results on genetic modifications. This complements the time-consuming process involved in growing out every modified plant in the laboratory or greenhouse to test for the desired modification.

2. ROLES OF BIOINFORMATICS

Bioinformatics today has entered every major discipline in biology. In genomics, Bioinformatics has aided in genome sequencing, and has shown its success in locating the genes, in phylogenetic comparison and in the detection of transcription factor binding sites of the genes (Liu et al.,1995; Thijs G et al.,2002), just to name a few. Microarray technology has opened the world of transcript me in front ‘biologists (Spellman et al., 1998; Eisen et al., 1998). Bioinformatics provides analytical tools for microarray data. These tools range from image the processing techniques that read out the data, to the visualization tools that provide a first-sight hint to the biologists; from preprocessing techniques (Durbin et al., 2002) that remove the systematic noise in the data to the clustering methods (Eisen et al., 1998; Sheng et al., 2003) that reveal genes that behave similarly under different experimental conditions. In proteomics, bioinformatics helps in the study of protein structures and the discovery of sequence sites where protein-protein interactions take place. To help understanding biology at the system level, bioinformatics begins to show promise in unraveling genetic networks (Segal et al., 2003). Finally, in the study of metabolome, bioinformatics is used to study the dynamics in a cell, and thus to simulate the cellular interactions.

3. WEB TOOLS AND RESOURCES OF BIOINFORMATICS

The WorldWideWeb provides a mechanism for unprecedented information sharing among researchers. Today, scientists can easily post their research findings on the Web or compare their discoveries with previous results, often spurring innovation and further discovery. The value of accessing data from other institutions and the relative ease of disseminating this data has increased the opportunity for multi-institution collaborations, which produce dramatically larger data sets than were previously available and require advanced data management techniques for full utilization.

As a side effect of these types of collaborations, some tools become de facto standards in the communities as they are shared among a large number of institutions. For instance, consider the BLAST (Altschul et al., 1990) family of applications, which allow biologists to find homologs of an input sequence in DNA and protein sequence libraries. BLAST is an example application that has been enhanced as a Web source, which provides dynamic access to large data sets. Many genomics laboratories provide a Web-based BLAST interface (http://blast.wustl.edu/) to their sequence databases that allow scientists to easily identify homologs of an input sequence of interest. This capability enhances the genomics research environment by allowing scientists to compare new sequences with every known sequence and to have their work validated by other members of the community. The addition of new sequences at an increasingly frequent rate (NIAS DNA Bank, http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.htm) further increases the value of this capability.

There are a number of common bioinformatics analyses one can perform at other sites, such as European Bioinformatics Institute (EBI), BioWeb Pasteur and Canadian Bioinformatics Resource, including BLAST and sequence analyses, primer tools and phylogenetics tree construction. EMBOSS sequence analysis package and SRS bio-database access are among the widely useful web tools available at these and other resource sites.

There are numerous web lists of bioinformatics resources, with many aimed at the biologist looking for software. Some of these, such as Bioinformatics.net, include discussion forums on the use of biology software. These are useful for biologists, as well as bioinformatics engineers looking for tools related to their work, or to be used at service centers. Many of these share a similar organization by functional categories, with many of the same links. It is useful to compare these for their different editorial perspectives, e.g. genomics/molecular biology or proteomics/biochemistry,

as well as effort to update and remove obsolete links. General resources such as Google, Amazon's Alexa and Open Directory Project at Mozilla.org include biology and bioinformatics categories in their directories. These directories are populated by robots or from submissions; they tend to lack the comprehensiveness of biologist-maintained lists.

Bioinformatics.ca provides a curated list of links that are well organized in categories, with main sections that include human genome and model organisms, sequences, gene expression, education and computer-related resources. Most or all of these include useful editorial comments on the content and value of the linked resources, making this list especially useful in learning about resources.

The Genome Web at MRC, UK, offers a similar very useful catalogue of links with editorial abstracts. An interesting function at Bioinformatics.ca is provided by an XML standard for web news called RSS, for sharing bioinformatics links. This allows customers and other web sites to have computable access to this catalogue. For instance, you can use an RSS program to notify you of additions and changes to this catalogue.

The Bionetwork project at Pasteur Institute provides an example of resource lists that are searchable by several bioinformatics criteria: Biological Domain, e.g. sequence analysis or structural biology, Resource type, e.g. database or online analysis tools, and Organism. This biology-focused search engine proves especially useful in finding that tool or resource most relevant to one's research. This project also has implemented link maintenance by using semi-automatic scanning of internet news and resources (robot-like) to update the catalogue. A similar project is BioHunt, which uses internet robot technology to search and update molecular biology resources.

BioHunt maintains current entries (it shows update times of this review month for several searches), making it especially useful to find new or updated tools that one has heard of, but lacks crated cataloguing of these to make it easy to find by subject matter.

Bioinformatics.net is a catalogue of online biology resources, specializing in bioinformatics tools. Its focus is towards he needs of molecular biologists and life science professionals, more than for bioinformaticians, and includes discussion and help forums on the use of software and bioscience topics. Jonathan Rees, who developed this resource, also curates biology lists in the Open Directory project. This service is supported in part y advertising, as are others reviewed here, one of the limited options available o maintain such services.

Bioinformatik.de offers a similar directory style collection of crated bioinformatics and biology resource links. The CMS molecular biology resource is an extensive catalogue of biology resources, including software

tools. The Southwest Biotechnology Center also maintains a useful catalogue covering a broad range of biology resources.

Bioinformatics.org and SourceForge.net are resources that support software developers and bioinformatics engineers, but are also useful to biologists looking for tools. Open-source software development in bioinformatics and other fields is being invigorated through agencies such as these. The number of active, widely used and valuable bioinformatics projects at these services is growing, including Generic Model Organism Database, Gene Ontology, GeneX Gene Expression Database and Staden Package for sequence analysis. These agencies allow for software archiving, but the primary attractions to software developers are infrastructure and tools that enable collaborative software development. A historical archive or catalogue service of bioinformatics software is limited, and maintenance of software releases is left to developers using this service.

4. APPLICATIONS OF BIOINFORMATICS IN AGRICULTURE

Plant life plays important and diverse roles in our society, our economy, and our global environment. Especially crop is the most important plants to us. Feeding the increasing world population is a challenge for modern plant biotechnology. Crop yields have increased during the last century and will continue to improve as agronomy re-assorting the enhanced breeding and develop new biotechnological-engineered strategies. The onset of genomics is providing massive information to improve crop phenotypes. The accumulation of sequence data allows detailed genome analysis by using-friendly database access and information retrieval. Genetic and molecular genome co linearity allows efficient transfer of data revealing extensive conservation of genome organization between species. The goals of genome research are the identification of the sequenced genes and the deduction of their functions by metabolic analysis and reverses genetic screens of gene knockouts. Over 20% of the predicted genes occur as cluster of related genes generating a considerable proportion of gene families. Multiple alignment provides a method to estimate the number of genes in gene families allowing the identification of previously undescribed genes. This information enables new strategies to study gene expression patterns in plants. Available information from news technologies, as the database stored DNA microarray expression data, will help plant biology functional genomics. Expressed sequence tags (ESTs) also give the opportunity to perform “digital northern” comparison of gene expression levels providing initial clues toward unknown regulatory phenomena. Crop plant networks collections of databases and bioinformatics resources for crop plants genomics have been

built to harness the extensive work in genome mapping. This resource facilitates the identification of ergonomically important genes, by comparative analysis between crop plants and model species, allowing the genetic engineering of crop plants selected by the quality of the resulting products. Bioinformatics resources have evolved beyond expectation, developing new nutritional genomics biotechnology tools to genetically modify and improve food supply, for an ever-increasing world population. So bioinformatics can now be leveraged to accelerate the translation of basic discovery to agriculture. The predictive manipulation of plant growth will affect agriculture at a time when food security, diminution of lands available for agricultural use, stewardship of the environment, and climate change are all issues of growing public concern.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., Lipman, D.J. 1990. Basic local alignment search tool, Mol. Biol., 215: 403.

Durbin, B.P., Hardin, J.S., Hawkins, D.M., Rocke, D.M. 2002. A variance-stabilizing transformation for gene-expression microarray data, Bioinformatics, 18(Suppl. 1): s105.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. 1998. Cluster analysis and display of genome-wide expression paterns, Proc Natl Acad Sci (USA), 95: 14 863.

Liu, J.S., Neuwald, A.F., Lawrence, C.E. 1995. Bayesian models for multiple local sequence alignment and Gibbs sampling strategies, J Amer Stat, 90: 1156.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., Friedman, N. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Gen, 34(2):166.

Sheng, Q., Moreau, Y., De Moor.B. 2003. Biclustering microarray data by Gibbs sampling, Bioinformatics, 19(Suppl. 2): ii 196.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B. 1998. Comprehesive identification of cell cycle-regulated genes of the yeast saccaromyces cerevisiae by microarray hybridization, Molecular Biology of the Cell, 9: 3 273.

Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor.B., Rouze, P., Moreau, Y. 2002. A Gibbs Sampling method to detect over-represented motifs in upstream regions of coexpressed genes, Journal of Computational Biology, 9(2): 447.