web reviews

Bioinformatics and its impact on plant science

Bioinformatics allows us to come to terms with the vast amount of data being generated by the genome sequencing projects. Powerful tools are needed to organize the data and to extend our ability to analyze these complex biological systems. Understandably, the development of bioinformatics has been tightly linked to international collaboration in genome sequencing projects and to the efforts of the pharmaceutical industry in its drive for drug discovery and development.

Although in its infancy, bioinformatics is already well established. For example, in studies on genome regulation and structure, bioinformatics covers many topics including: databases on regulatory sequences; the regulation of gene expression; analysis and recognition of genomic sequences; gene structure prediction; modeling of transcriptional and translational control; and large scale genome analysis. From this list, we need to assess the suitability of these applications to plant science and to assess the impact of bioinformatics on crop improvement.

The integration of bioinformatics will influence plant science and lead to crop improvements in the following areas:

- The identification of important genes through genomics, expression analysis and functional genomics. In conjunction with the design and construction of transgenic plants this will allow new target genes to be identified that will improve quantitative and qualitative traits in commercially important crops.
- The design of agrochemicals based on an analysis of the components of signal perception and transduction pathways to select targets, and with cheminformatics, to identify potential compounds that can be used as herbicides, pesticides, or insecticides.
- The utilization of plant genetic resources to preserve genetic diversity in agricultural species. The need for taxonomic data goes far beyond the field of classical taxonomy, and a catalogue of all species, with phenotypic and genotypic attributes is required. The core taxonomic effort gives stability to the work of regulatory, management and conservation bodies.
- Efficient utilization of biological repositories of clones, cell lines, organisms and seeds. Typically, existing repositories are not linked to each other databases. The many commercial databases and repositories are also part of the bioinformatics infrastructure but operate largely outside of the present day cooperative activities.

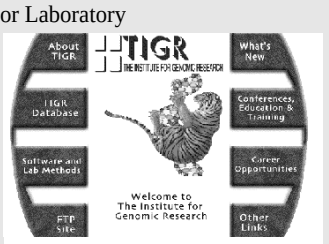
An initiative in plant genomics has emerged[1]. This initiative emphasizes the point that the goals of the Plant Genome Initiative will only be met through appropriate investment in bioinformatics. Along these lines, the National Center for Genome Resources (NCGR) in New Mexico has established a partnership with New Mexico State University to develop a National Biotechnology Information Facility (NBIF), with a budget of $8.5 million over five years. The NBIF is developing a plant-specific metabolic pathways database, and is strengthening or expanding into such fields as agricultural genomics, biological interface technology, computational biology and software development. NBIF will also provide bioinformatics support to leading agricultural research institutions.

Following the lead set by the *Arabidopsis* Genome Initiative (a multinational effort to sequence the entire *Arabidopsis* genome), agricultural biotechnology companies (such as Zeneca Agrochemicals and Ceres Inc.) have recognized the impact of informatics and provided increased resources for bioinformatics as part of their research and development plans. Leading universities and plant science institutions are also incorporating bioinformatics into their research projects. The John Innes Centre (Norwich, UK) is now working on comparative *Brassica* and cereals bioinformatics projects, to provide comparative mapping of monocot and dicot plant genomes. Similarly, the Department of Plant Breeding at Cornell University (Ithaca, NY, USA) is integrating bioinformatics into the study of genetic diversity in agricultural species. Many of these bioinformatics resources are available on the Internet (see Box 1 for selected Internet sites for plant bioinformatics). Plant scientists have an opportunity to use these resources to the full, to ensure that bench work, both in the present and in the future, can be combined with bioinformatics to fully reap the rewards of the genomics revolution.

Box 1. Selected Internet resources for plant bioinformatics

Arabidopsis Genome Initiative (AGI)
http://genome-www.stanford.edu/Arabidopsis/AGI

Arabidopsis Genome Data Analysis, Cold Spring Harbor Laboratory
http://nucleus.cshl.org/protarab



Plant Genome and Information Center, USDA
http://www.nal.usda.gov/pgdic

UK Crop Plant Bioinformatics Network
http://synteny.nott.ac.uk/agr/agr.html

The Institute for Genomic Research (TIGR) Database
http://www.tigr.org

Arabidopsis Genome Center at the University of Pennsylvania
http://genome.bio.upenn.edu/ATGCUP.html

The Genome Sequencing Center of Washington University in St Louis
http://genome.wustl.edu/gsc

Grain Genes Database
http://wheat.pw.usda.gov

Maize Genome Database
http://www.agron.missouri.edu

Arabidopsis Internal Coding Exon Finder
http://clio.cshl.org/genefinder/ARAB/arab.htm

NetPlantGene V2.0 Web Prediction Server
http://www.cbs.dtu.dk/NetPlantGene.html

Ti Xing
Agriculture and Agri-Food Canada, Eastern Cereal and Oilseed Research Centre, Ottawa, Ontario, Canada K1A 0C6
(tel +1 613 759 1814; fax +1 613 759 1701; e-mail xingtx@em.agr.ca)

References

1 Bennetzen, J.L. *et al*. (1998) A plant genome initiative, *Plant Cell* 10, 488–493