Plant Biotechnology 26, 443–444 (2009)

Preface

Bioinformatics from MOURA to SHIKKAI

Shigehiko Kanaya

Laboratory of Comparative Genomics, School of Information Science, NARA Institute of Science and Technology

Fundamental roles of information science are collection and storage of data in databases and subsequent analysis (classification, manipulation, retrieval, calculation and simulation) corresponding to application fields. Now a day, biologists need to perform systematization of several types of data concerning genome, transcriptome, proteome, interactome, and metabolome to understand organisms as systems. To attain this, close collaboration of information scientists and biologists is necessary to thoroughly understand biology with efficient analysis of huge and versatile data. But the cultures and strategies of research in biology and information science are very different. Information scientists initially define elements and program the idea to understand the behavior of them, while biologists analyze unknown elements and finally clarify them, that is, information scientists and biologists follow opposite strategies.

Fortunately, after determining the genome, we have a set of finite number of genes and thus biology has become a data rich science. This is the most important point in information science because we can make programs taking a finite number of elements corresponding to genes into consideration. In post genome projects related to transcriptome, proteome and metabolome the involvement of information science has steadily become apparent by facilitating the examination of all possible genes/proteins, RNAs and metabolites and interrelation among them.

This special issue on bioinformatics bears the proof of successful researches through collaboration between biologists and information scientists and shows that such collaboration is important for understanding the cellular system. There is no master-slave relation between biologists and information scientists. Biologists and information scientists have different strategies toward unraveling the truth and their mutual cooperation is vital to success in bioinformatics research. Understanding plants as a life system is also a challenging problem for information scientists as well as for biologists, because plants tend to have a very large size of genomes. This special issue on bioinformatics provides some novel strategies to understand the whole cell systems in plants.

There are a total of ten research articles in this special issue. Of them, two are concerning spectral analysis. Ara *et al*. developed a methodology of peak detection for gas chromatography-mass spectrometry data of biological samples. Ikeda *et al*. developed a data mining software tool for Fourier transform near-infrared reflectance spectroscopy to distinguish complicated composition of biological samples.

Five other articles try to systematize biological elements in genome and post genome levels. Tanaka *et al*. proposed a method for prediction of metabolomic pathway based on the logic of inclusive relation between cyclic structures of metabolites by analyzing more than thirty-thousand metabolites together. Abe *et al*. try to systematize the genes in eukaryotic genomes with special reference to plant genomes using Batch-Learning Self-Organizing Map (BLSOM). Matsuda *et al*. proposed compound identifier network for plant metabolomics. Fukushima *et al*. tried to understand the complicated system of gene expression on the basis of a network where nodes correspond to genes and edges correspond to coexpresison between genes. Nishikata *et al*. proposed a statistical method to determine significant domain-domain interactions based on a whole set of protein-protein interaction (PPI) data which is helpful to assess true positive interactions in high throughput PPI data and to predict the conformation of automatically generated protein complexes.

The remaining three articles are concerning biological databases. Ogata *et al*. developed gene co-expression database for understanding biological processes in soybean. Takahashi *et al*. developed gene classification system for *A. thaliana* from about 3000 scientific literatures. This also means that we have some clues for function of half of the genes in Arabidopsis genomes. Nakamura *et al*. organized a database by collecting information on metal-ion-binding sites in proteins.

Bioinformatics concerning 'omics' studies should be changed from comprehensive understanding ("MOURA" in Japanese) to complete understanding ("SHIKKAI"). Each of the papers will be helpful to understand strategies for this challenging theme. I acknowledge the contributors of this special issue.

The editing of this special issue was supported by the following invited editors. The chief editor deeply appreciates their assistance.

Takashi Abe	Nagahama Institute of Bio-Science and Technology
Atsushi Fukushima	RIKEN Plant Science Center
Masami Yokota Hirai	RIKEN Plant Science Center
Toshimichi Ikemura	Nagahama Institute of Bio-Science and Technology
Shigehiko Kanaya	Nara Institute of Science and Technology
Daisaku Ohta	Osaka Prefecture University
Hiroki Takahashi	Nara Institute of Science and Technology
Kazuki Saito	Chiba University, RIKEN Plant Science Center
Hideyuki Suzuki	Kazusa DNA Research Institute