Rice at the Forefront of Plant Genome Informatics

Bal A. Antonio[1] **Katsumi Sakata**[2]
`antonio@staff.or.jp` `ksakata@abr.affrc.go.jp`
Takuji Sasaki[2]
`tsasaki@abr.affrc.go.jp`

[1] Rice Genome Research Program, Institute of the Society for Techno-innovation of Agriculture and Forestry,446-1, Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan
[2] Rice Genome Research Program, National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

Abstract

As rice genomics data continue to accumulate at a rapid rate, databases are becoming more valuable to warehouse and access large and rigorous data sets. This article gives an overview of available resources on rice bioinformatics and their role in elucidating and propagating biological and genomic information in rice. Of particular focus here is the informatics infrastructure developed at the Rice Genome Research Program (RGP) following an extensive rice genome analysis. The database named INE (Integrated Rice Genome Explorer) integrates the genetic and physical mapping information with the genome sequence being generated in collaboration with the International Rice Genome Sequencing Project (IRGSP). Database links are initially evaluated using an interoperable query tool to explore and compare data across the rice and maize genome databases and potential application to multiple crop database querying. A proposed logistics for interlinking these resources is presented to integrate, manipulate and analyze information on the rice genome. One of the biggest challenges of rice bioinformatics lies in the emerging role of rice as a model system among grass crop species. In view of the importance of comparative genetics in the formulation of new knowledge on plant genomes and genes, comparative bioinformatics remains an essential strategy to gain new insights on the needs and expectations on rice genomics.

Keywords: rice, genome sequence, INE, cereal crops, synteny, comparative genomics

1 Introduction

As genomics data from cereal crops continue to accumulate at a rapid rate, specific databases to warehouse and transform these data into information for biologists to access and query are critically needed. Because of the economic importance of these grass crops such as rice, wheat, and maize, analyses of their genomes will greatly contribute to the improvement in their productivity. Of particular importance is the emerging role of rice as a model system among the cereal crops in view of its small genome size (430 Mb) and because it is the major source of food energy for half of the world's population. As an emerging model system, it is necessary to develop an informatics infrastructure in which genomic data will become relevant across the cereal crop species. Using comparative genomic analysis, information derived from genomic maps and sequences of rice may be extrapolated to other cereal crops. Following completion of sequencing the entire rice genome, the next challenge will be directed towards understanding the function of all its genes. Achievements in functional genomics of rice can provide valuable information to the functional analysis of other cereal species. Comparisons between or among the genomes of these related species would provide understanding of their structure and function.

Following the successful launching of the human genome project, a new era of whole genome science has emerged [4] ranging from humans to plants and yeasts. In plant genomics, the nearly completed genome sequence of the model plant, *Arabidopsis thaliana*, which represents the first higher plant utilized for genome sequencing, is anticipated to revolutionize the future direction of plant biology. A similar venture of sequencing the entire rice genome is underway that will directly transform the feasibility and utility of sequencing the large genomes of cereal crops that provide the bulk of the world's food supply. While complete sequencing of the rice genome will uncover valuable genomic information in other cereal crops, we will continue to embrace and benefit from the challenges undergone in whole genome sequencing of other organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*, the metazoan *Caenorhabditis elegans* and even human *Homo sapiens*. Comparisons between distantly related genomes provide insight into the universality of biological mechanisms and identify experimental models for studying complex processes [4].

Informatics resources available to the rice research community are very diverse ranging from genomic to biological and genetic which are represented by independent database projects. For this reason, rice bioinformatics resources that are available to the research community will be presented here in the context of a database by briefly describing their main features encompassing biological and genomic information in rice, services they offer, software used for data analyses, organization or representation and other associated information required to access or retrieve information. Of particular focus of this article is the bioinformatics infrastructure developed at the Rice Genome Research Program (RGP) [13] principally covering integrated information on the structural analysis of the genome as well as the elucidated sequences from the International Rice Genome Sequencing Project (IRGSP) [12]. One of the most formidable limitations facing the rice research community today involves the lack of interlinking among the rice informatics resources. Proposals for integrating these rice database resources are outlined. The current and future prospects of bioinformatics at RGP will be presented including measures designed to help improve the current status of rice bioinformatics resources.

2 Rapid Growth of Genomics Data

Biological data are increasing at a phenomenal rate with the advent of high-throughput genome projects. The growing number of biological data on rice is reflected in the total number of submitted sequences to the International Nucleotide Sequence Database Collaboration where rice ranks 5^{th} in number of accessions among all organisms and second among plants. Informatics infrastructure on rice is evidenced in several databases offering both general and specific information online. Key research centers involved in rice research have also made their resources available on the web including distribution of germplasm. More recently, the establishment of the international rice genome sequencing consortium, IRGSP, has mobilized the participating groups to create individual repositories for their accumulated sequence data.

Much of the information on rice genomics has been brought about by the large-scale rice genome projects in the last decade. RGP has contributed greatly in developing the fundamental tools for rice genome analysis. Genomic information currently available on the Internet includes genetic maps, physical maps, DNA markers, EST sequences and morphological markers. As of July 2000, IRGSP has already submitted a total of about 20 Mb of rice genome sequences to GenBank, EMBL and DDBJ. This collaboration of public sequencing centers from 10 countries is expected to make a major contribution on rice genomics [13]. In addition to the genome sequences, about 50,000 BAC-end sequences that will be useful as sequence tag connectors are also available. Equally important as the molecular tools are the germplasm resources that will provide the basis for direct application of molecular tools. These resources range from tester lines to cultivated varieties and wild species maintained in different institutes involved in rice research. Literatures on different aspects of rice research have also been catalogued for searching on the web for the user community.

Table 1: Summary of rice genomic resources available on the World Wide Web.

Resource[1]	Number
DNA sequence accessions	153,153
Nucleotides (bp)	103,128,543
Genetic maps	9
Physical maps	2
Molecular markers	19,840
Morphological markers	185
QTLs	342
ESTs	51,000
BAC-end sequences	110,438
Worldwide strains	126,369
Tester and mutant lines, etc.	2,515
Landraces/improved varieties	6,293
Wild species	1,609
References	1,500

[1] Accessions and nucleotides are based on DDBJ release 42, July 2000. Other datasets are current August, 2000. Map information including molecular and morphological markers was derived from INE, Oryzabase, and RiceGenes. Other sources of information are BAC-end sequences from the CUGI site; ESTs from the TIGR Rice Gene Index; and germplasm information from NIG, MAFF, NPGS-GRIN, CGIAR-SINGER, and NPGRC Web sites. QTLs = quantitative trait loci, ESTs = expressed sequence tags, BAC = bacterial artificial chromosome, INE = Integrated Rice Genome Explorer, MAFF = Ministry of Agriculture, Forestry, and Fisheries, TIGR = The Institute for Genomic Research, NPGS-GRIN = National Plant Germplasm System-Germplasm Resources Integrated Network, NIG = National Institute of Genetics, CGIAR-SINGER = Consultative Group on International Agricultural Research-Systemwide Information Network for Genetic Resources, NPGRC = National Plant Genetic Resources Center.

3 Informatics for the Rice Community

There are currently more than 20 databases and websites available online for rice genomics. Most of these databases focus on specialized and detailed information covering a broad range of research on rice genomics or rice biology whereas some are designed for ordinary users and offer general information on rice. An overview of these databases is briefly discussed here.

3.1 Rice Genome Databases

One of the earliest databases that have been serving the rice community since 1993 is RiceGenes [15]. Supported by the USDA-ARS Center for Bioinformatics in partnership with Cornell University, the database focuses on comparative genetic maps, molecular markers, morphological markers, quantitative trait loci (QTL) and germplasm data. Newly added features include microsatellite marker data and microsatellite maps of the 12 rice chromosomes. RiceGenes was originally constructed using ACEDB but is currently being converted to a relational database management system to allow for increased data integration and greater flexibility.

Oryzabase is one of the most recently established rice genome database developed at the National Institute of Genetics in Mishima, Japan as part of a huge collection of biological germplasm databases called SHIGEN (SHared Information of GENetic Resources) [16]. It encompasses a wide range of information from general facts about rice and classical genetics to recent advances on rice genomics. The rice maps consist of the classical genetic map with phenotypic genes, molecular maps and an integrated map with RFLP markers and phenotypic markers for cross-referencing of marker positions in each map. The database also features a compilation of rice strain information, a dictionary of all identified genes in rice with corresponding references, basic information on rice classification, morphology, rice cultivation and mutant collection with images.

The rice genome analysis projects of various research institutions are also documented in several databases. The genome initiative of the Japanese Ministry of Agriculture, Forestry and Fisheries (MAFF) are represented at the RGP website (described below) [17] and the MAFF DNA Bank [18] affiliated with the National Institute of Agrobiological Resources. The latter features the sequence analysis of all rice cDNA clones generated from RGP. The Korea Rice Genome Database [19] is maintained by Myongji University as part of the Korea Rice Genome Research Program. It features rice

EST database, genetic maps and chloroplast genome sequences. The Thai Rice Genome Project website [20] focuses on rice genome projects in Thailand including genome sequencing and gene discovery.

3.2 Genetic Resources Databases

The genetic resources databases principally cover important resources on the source of materials for rice researches. These databases also offer essential information on the specific attributes of these resources, evaluation data and their availability for distribution. The MAFF Gene Bank [21] features a total of 8671 genetic stocks including evaluation data such as morphological characters, resistance to stress, yield, quality and ingredients. The National Institute of Genetics (Mishima, Japan) provides information on about 11,000 genetic stocks developed in Japanese institutes and universities [16]. These resources include marker gene testers, mutant lines, isogenic lines, autotetraploid lines, primary trisomics, reciprocal translocation homozygote lines, cytoplasm substitution lines and cell cultured lines. The System-wide Information Network for Genetic Resources (SINGER) is an information exchange network of research centers affiliated with the Consultative Group on International Agricultural Research (CGIAR) [22]. It contains key data on the identity, source, characteristics and transfers to users of more than 106,000 accessions of rice genetic resources. The National Plant Genetics Resources Center in Taiwan [23] has a total of 3980 strains. Rice germplasm data are also available via GRIN (Germplasm Resources Integrated Network), the USDA germplasm database comprising pedigrees and trait data [24].

3.3 Specialized Databases

In contrast to generalized databases, some specialized databases for rice offer in-depth information surrounding a specific biological function or aspect of research. Participating members of IRGSP present the sequencing progress of their assigned chromosome(s) of the rice genome via independent databases. These databases incorporate the sequence of PAC or BAC clones for individual chromosomes including the annotation of the sequence. The National Center for Gene Research Chinese Academy of Sciences features the sequencing progress of chromosome 4 for the indica rice variety [25]. The Plant Genome Center Academia Sinica provides the sequencing progress of chromosome 5 [26]. The US Rice Genome Sequencing integrates the sequencing efforts of the four participating US groups: The Institute for Genomic Research (TIGR), Clemson University / Cold Spring Harbor Laboratory / Washington University consortium (CCW), Plant Genome Initiative at Rutger (PGIR), and the University of Wisconsin [27]. Each group also has independent sequencing databases covering specific regions of chromosomes 3, 10 and 11. The sequencing efforts of other participating groups such as Korea (chr. 1 region), UK (chr. 2), Canada (chr. 2), Thailand (chr. 9), India (chr. 11) and France (chr. 12) are expected to be integrated into their respective databases in the near future.

Other specialized databases on rice include the Rice Gene Index [28] and the Rice Repeat Database [29] both of which are maintained by TIGR. The Rice Gene Index is part of the TIGR Gene Indices integrating ESTs from international sequencing projects and consists of 51,000 ESTs and ETs (expressed transcript sequences) from rice. These sequences are clustered and assembled using stringent overlap criteria to produce a high quality consensus sequence. The advantage of this approach is that it provides a mechanism to link to candidate orthologues in other species [9]. The Rice Repeat Database features a compilation of repeats and mobile elements comprising 50% of the rice genome. It provides a search engine using BLASTN and an anonymous FTP link to download rice repeat sequences. Rice blast DB [30] developed by the RiceGenes group features various types of information on the rice blast fungus, *Magnaporthe grisea*.

4 RGP Informatics

The RGP website encompasses the activities of RGP, the sequencing progress of IRGSP and an integrated database of the rice genome. The informatics effort of RGP centers on the database aptly named INE (INtegrated Rice Genome Explorer) which literally means “rice plant” in the Japanese language [11]. This database was developed primarily to serve as repository for all the information accumulated from the first phase of the rice genome project and to integrate the genetic and physical mapping data with the sequence of the rice genome that will be elucidated as part of the second phase of the program. In addition, it also functions as a repository of the rice genome sequences from the international sequencing collaboration.

4.1 Database Structure

A characteristic feature of INE is the integrated genetic and physical map of rice. The high-density linkage map currently contains 2275 DNA markers [5] and an updated version with about 3000 markers will be released soon. Information on the markers include the probe size, the RFLP pattern as a result of parental screening, the sequence data as well as the results of similarity search. The markers on the genetic map are traceable to the ordered YAC clones covering 70% of the whole genome [10]. Information on the insert size of the clone is also provided. At present, approximately 4500 ESTs (expressed sequence tags) are being assembled in the database [14]. These ESTs were mapped by PCR screening of our YAC library and positioned using the YAC physical map as base. Together with the genetic markers, these ESTs are subsequently used for ordering the PAC (P1-derived artificial chromosome) clone contigs to construct a sequence-ready physical map, which serve as the template for genome sequencing. Integrating the structural information of the rice genome allows efficient utilization of available data particularly in the functional and applied aspects of genomics.

As part of the sequencing effort of RGP, the sequence data for chromosomes 1 and 6 are incorporated in INE. The ordered PAC clones in these chromosomes are displayed and linked to the sequence of the clone and the annotation of the sequence using various similarity search and gene prediction programs. The annotation page consists of the annotation map for each clone and tables summarizing the predicted genes including the results of the similarity searches and gene prediction programs. Furthermore, a list of low quality sequence data is provided in a separate link as a requirement of the IRGSP to maintain an accuracy of less than one error in 10,000 base pairs (greater than 99.99%).

4.2 Navigation and Visualization

The integrated maps for each chromosome facilitates a general overview of the genomic information in a particular chromosome. The maps can be manipulated by zoom in/out to facilitate browsing at detail-oriented levels. Another distinct feature of INE is the rapid display of integrated maps. This has been achieved by programming the viewer in Java language using an application GIOT (Genome Information-display Orderly Tool) developed by Mitsubishi Space Software Co. Ltd. This attribute contributes to a smooth navigation of specific information associated with each dataset. Clicking on a particular marker opens a Java applet containing details such as screening data, image sets, sequence etc. These features allow getting all specific information together with an overall view of the distribution of the markers and clones in a short time. In addition, different chromosomes can be displayed simultaneously for direct comparisons between or among chromosomes.

The integrated map allows the user to correlate various genomics information even when sequencing is progressing at the initial stages. The preliminary genome sequence can be directly traced to the physical and genetic map. This will undoubtedly provide valuable information for positional cloning of target genes and for analysis of more complicated traits such as QTLs (quantitative trait loci). For instance, a richly informative map can be established for a particular trait by integrating the genetic and EST map. With the information from the YAC-based physical map and PAC contigs, specific clones carrying markers tightly linked to the gene of interest can provide the template either for developing more markers or directly cloning the gene. Linking these data with the annotation of the sequence may provide valuable information on the genes involved in a particular trait of interest.

4.3 An Interoperable Query System for Rice and Maize

In collaboration with the Missouri Maize Project, an interoperable query system will be incorporated to INE to explore multiple crop databases [1]. MaizeDB [31] is a comprehensive public database of phenotypic, genetic and genomic information on maize such as genetic maps, loci, sequences, phenotypes, QTLs and relevant references. Cross-database querying and display of text objects between INE and MaizeDB is being implemented using a web-based object-oriented query system called the OPM (Object-Protocol Model) data management tools of Gene Logic Inc. These tools are unique in their capacity to impose a uniform object-oriented data model on existing relational database framework where users can explore and assemble biological information from heterogeneous databases. This query system promotes direct analysis of colinearity at the nucleotide level in rice and maize species that may also be applied for exploring multiple crop databases.

4.4 Present and Future Prospects

Bioinformatics support is essential for the implementation of the rice genome research project. As INE will be used as the central repository of the international sequencing collaboration, a system that will allow pulling sequence data from the various sequencing centers is being incorporated. This will facilitate easy monitoring of the sequencing progress as well as maintenance of quality standard. A standardized rice annotation approach will be developed as well. This may require an automated analysis suite of tools and a platform that will allow re-annotation on a regular basis.

Functional genomics at RGP is currently progressing with accumulated data derived from gene expression monitoring, insertional mutagenesis and map-based cloning. The ESTs used in the microarray analyses are incorporated in the genetic and EST maps where direct linking of gene profiles can be made. A database of gene expression information will provide insights into the function and specificity of all genes in rice as well as extensive information on the profiles of genes related to growth, development and in response to various environmental changes. Eventually integrating information on rice structural and functional genomics will provide an overall view of the network of genes involved in complex biological responses.

We are also developing methods to further improve the interface of our database to facilitate efficient data mining and improve facile access to the database of all types of users. An annotation database is also being constructed as we proceed with the annotation of the rice genome to categorize all the rice genes that have been identified so far. As the rice genome data will increase exponentially adequate tools for input, integration and query will become necessary.

5 Rice Informatics in the Next Decade

Rice genome databases that evolve from rigorous and systematic sequencing efforts should not merely function as storehouses for thousands of bases or amino acids. Of particular importance is the ability to attach substantial genomic information to the sequence. Studies on identifying genes and predicting the proteins they encode, determining when and where the gene proteins are expressed and how they interact, and how these expression and interaction profiles are modified in response to environmental signals will follow. However, meeting the demands and challenges of an ideal crop improvement strategy remains a matter of combining traditional breeding concepts and genomic tools through rigorous phases of experimentation. Therefore, emphasis on the underlying value of genotypic and genomic elements must be balanced with a phenocentric approach [8]. Data representations for rice genomics should be able to cope with these innovations. One way to address this need is to interlink the resources of various types of information such as genomic data, phenotypic or expression data and genetic resources as shown in Fig. 1. Logical connections to other information will enhance the intrinsic value of the raw sequence data to facilitate the formation of new biological discoveries. For a given gene, the database would horizontally link sequence, structure, and map position and would link related elements of the same type pertaining to the expression profile, proteins and phenotypes. All these information should be correlated with the genetic resources available for rice.

Rice, being a model system for other grass crops, should establish an informatics infrastructure designed to interlink database resources on rice genomics to better serve a more focused research community utilizing this system in contrast to a larger user community. The major cereal species evolved from a common ancestor some 60 million years ago and considerable differences have arisen in genome size and haploid chromosome number [2]. However, these gross comparisons conceal a high level of conservation in both gene structure and order. Thus even if maize genome is 25-fold larger than rice, conserved linkage groups can be identified across much of both genomes. This striking result has been extended to a wide range of cereal species. The predictive aspect of genome colinearity is potentially valuable and offers a framework in which genetic data become relevant across species boundaries. Ultimately the rice genome sequence that will be elucidated will be potentially valuable in understanding the genome structure and function of other cereal crops as well.

The proposed interlinking of database resources as outlined above hopes to address this need for the rice research community. This should be extended to databases of other cereal crops as well. One of the more serious challenges of specialized or expert domain databases, best represented by model organism databases, is to balance the needs between the broader scientific community and the specialized focused groups. As comparative genetics is now viewed as key to extending existing knowledge on plant genomes and genes [3], comparative bioinformatics remains an essential strategy to pursue the needs and expectations on rice genomics. Comparative bioinformatics offers the possibilities to link various cereal crops through their genomes and will provide keys to understanding how genes and genomes are structured and how they evolved. Genetic mapping as well as some preliminary sequence data show the extent of synteny among cereal crops [2]. Through identification of synteny, it will be possible to isolate genes from crop plants with large genomes using information of homologous genes in related crops with smaller genomes. As the creation of links between different databases may foster interoperability [6, 7], linkages and interactions should also be promoted between databases of rice and other non-plant species.

An era of biological revolution has begun during which a tremendous amount of information on plant genetics will be accumulated over the next ten years. The need to transform these data into information that the rice community and biologists in general can query and use properly should be given high priority. As in any informatics effort the essential factors necessary to achieve the desired success in rice bioinformatics involve developing the necessary tools for interlinking information or whole databases. As we move from the pre-genomic to the post-genomic era, rice bioinformatics should also be able to adapt to the next challenge of analyzing biological and agricultural problems in multiple dimensions.

Acknowledgments

The authors wish to thank I. Ohta, Y. Mukai, H. Nagasaki, K. Waki, A. Idonuma, M. Masukawa, M. Negishi and K. Arikawa for extensive work on the RGP database. This work is supported by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project MA-1001).

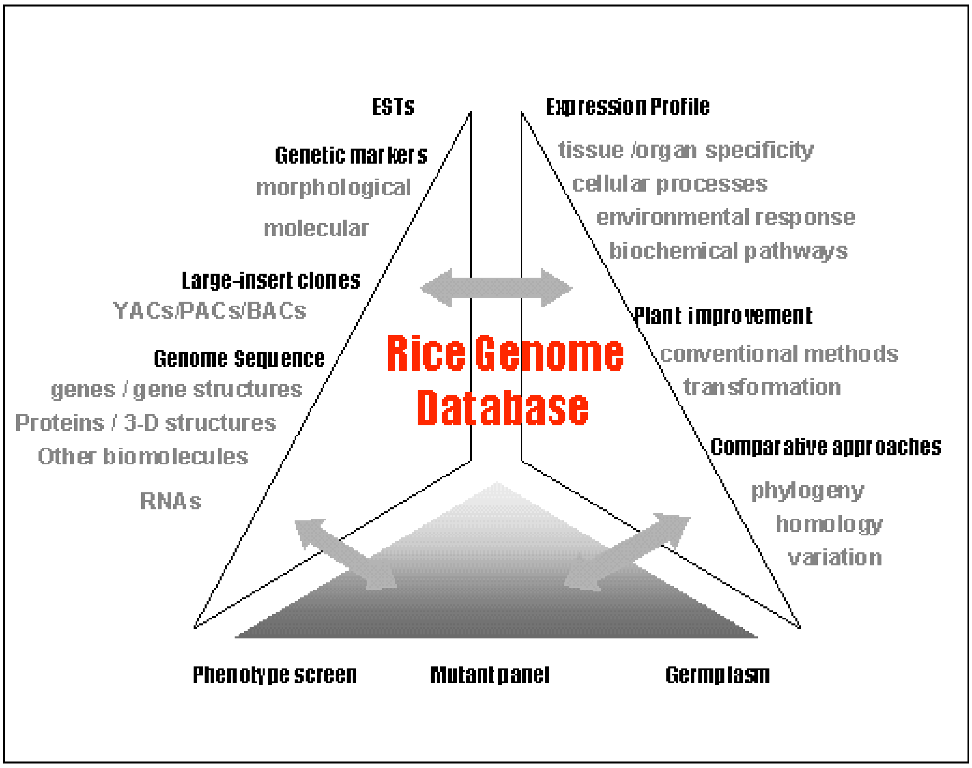

Figure 1: Interlinking of genomics information in rice.

References

[1] Antonio, B.A., Fang, Z., Sanchez, H., Polacco, M., Chen, S.S., Sakata, K., Ohta, I. Mukai, Y., Coe Jr., E., and Sasaki, T., An interoperable query system for INE and MaizeDB: Expanding genomic resources for cereal crops, *Proceedings of the 6th International Congress of Plant Molecular Biology*, Supplement to Reporter 18:2, 2000.

[2] Bennetzen, J.L, Plant genomics takes root, branches out, *Trends Genet.*, 15:85–87, 1999.

[3] Bevan, M. and Murphy, G., The small, the large and the wild - the value of comparison in plant genomics, *Trends Genet.*, 15:211–214, 1999.

[4] Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A, Gesteland, R., and Walters, L., New goals for the U.S. human genome project: 1998–2003, *Science*, 282:682–689, 1998.

[5] Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S., and Sasaki T., A high-density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics*, 148:1–16, 1998.

[6] Karp, P.D., Database links are a foundation to interoperability, *Trends Biotech.*, 14:273–279, 1996.

[7] Macauley, J., Wang, H., and Goodman, N., A model system for studying the integration of molecular biology databases, *Bioinformatics*, 14:575–582, 1998.

[8] Miflin, B., Crop improvement in the 21st century, *J. Exp. Bot.*, 51:1–8, 2000.

[9] Quackenbush, J., Liang, F., Holt, I., Pertea, G., and Upton, J., The TIGR Gene Indeces: reconstruction and representation of expressed gene sequences, *Nucleic Acids Res.*, 28:141–145, 2000.

[10] Saji, S., Umehara, Y., Antonio, B.A., Yamane, H., Tanoue,H., Baba, T., Aoki, H., Ishige, N., Wu, J., Koike, K., Matsumoto, T., and Sasaki, T., A physical map with yeast artificial chromosome (YAC) clones covering 63 percent of the 12 rice chromosomes, *Genome*, (in press), 2000.

[11] Sakata, K., Antonio, B.A., Mukai, Y., Nagasaki, H., Sakai, Y., Makino, K., and Sasaki, T., INE: a rice genome database with an integrated map view, *Nucleic Acids Res.*, 28:97–101, 2000.

[12] Sasaki, T. and Burr, B., International rice genome sequencing project: the effort to completely sequence the rice genome, *Current Opinion Plant Biol.*, 3: 138–141, 2000.

[13] Sasaki, T., Rice genome research in Japan as a frontrunner in crop genomics, *Science Tech. Japan*, 18:18–21, 1999.

[14] Wu, J., Yamamoto, S., Maehara, T., Harada, C., Shimokawa, T., Ono, N., Mukai, Y., Takazaki, Y., Yazaki J, Koike, K., Shomura, A., Fujii, F., Yano, M., Ando, T., Kono, I., Sasaki, T., and Matsumoto, T., A comprehensive map of rice expressed sequence tags (ESTs): It's characteristics and application to genome analysis, *Proceedings of the 6th International Congress of Plant Molecular Biology*, Supplement to Reporter, 18:2, 2000.

[15] http://genome.cornell.edu/rice/

[16] http://www.shigen.nig.ac.jp/

[17] http://rgp.dna.affrc.go.jp

[18] http://bank.dna.affrc.go.jp/

[19] http://bioserver.myongji.ac.kr

[20] http://jasmine.kps.ku.ac.th/

[21] http://www.gene.affrc.go.jp/

[22] http://www.cgiar.org

[23] http://www.npgrc.tari.gov.tw/

[24] http://www.ars-grin.gov/npgs

[25] http://www.ncgr.ac.cn

[26] http://genome.sinica.edu.tw/

[27] http://usricegenome.org/

[28] http://www.tigr.org/tdb/ogi/

[29] http://www.tigr.org/tdb/rice/blastsearch.html

[30] http://genome.cornell.edu/riceblast/

[31] http://www.agron.missouri.edu/