BIOINFORMATICS

Linking genotype to phenotype: the International Rice Information System (IRIS)

Richard M. Bruskiewich[1,*]*, Alexander B. Cosico*[1]*, William Eusebio*[1]*, Arllet M. Portugal*[1]*, Luralyn M. Ramos*[1]*, Ma. Teresa Reyes*[1]*, May Ann B. Sallan*[1]*, Victor Jun M. Ulat*[1]*, Xusheng Wang*[1,3]*, Kenneth L. McNally*[2]*, Ruaraidh Sackville Hamilton*[2] *and Christopher Graham McLaren*[1]

[1]*Biometrics and Bioinformatics Unit,* [2]*Genetic Resources Center, International Rice Research Institute (IRRI), PO Box DAPO 7777, Metro Manila, Philippines and* [3]*Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, People's Republic of China*

ABSTRACT

The International Rice Information System (IRIS, http://www.iris.irri.org) is the rice implementation of the International Crop Information System (ICIS, http://www.icis.cgiar.org), a database system for the management and integration of global information on genetic resources and germplasm improvement for any crop. Building upon the germplasm genealogy and field data components of ICIS, IRIS is being extended to handle diverse rice genomics data including: genetic mapping, genome annotation, genotype, mutant, transcriptome, proteome and metabolomic data. Users can access information in the database through stand-alone programs and WWW interfaces offering specialist views to researchers with different interests.

Availability: http://bioinformatics.org/project/?group_id= 141

Contact: r.bruskiewich@cgiar.org

Keywords: database, crop, rice, germplasm, genotype, phenotype, genomics

INTRODUCTION

International germplasm exchange was the engine of the Green Revolution. In the past, however, much of the important information produced from this exchange was only accessible locally, i.e. in field books or researchers' files. Although major international initiatives for germplasm collection and conservation followed the Green Revolution, much collected material is still not used because it is difficult to access. As a result, the full potential impact upon agriculture has not yet been realized.

In 1995, the international agricultural research centers CIMMYT and IRRI partnered with other CGIAR centers to establish a project to develop an International Crop Information System (ICIS; Fox and Skovmand, 1996) to overcome these deficiencies in crop data management. Since that time, the project has undergone a quiet evolution in design and implementation, and expansion to embrace many different crops, collaborative partners and categories of related information, including newly available molecular and genomic data.

ICIS ARCHITECTURE

A group of international scientists are collaborating to develop ICIS as a generic system that will accommodate all data sources for any crop and breeding system. The goals of ICIS are, firstly, to integrate different data types in both private and public datasets into a single information system and, secondly, to provide specialist views and applications that operate on this integrated platform.

To achieve the first goal, the database is partitioned into a 'Central' (public) and 'Local' (private) installation, with a full user authentication and data ownership model built into the architecture. To achieve the second goal, a middleware layer was designed to provide a high-level data model and application program interface to insulate the details of the database schema from user applications. This has enabled the development of several independent standalone applications for ICIS, and also, provided modularity for the WWW interface.

The system architecture is modular. A 'Genealogy Management System' (GMS) manages information concerning breeding methodology and chronology for germplasm. The GMS is also structured to assign a single unique 'Germplasm Identifier' (or GID) to each

---

*To whom correspondence should be addressed.

germplasm entity, while recording the diverse and often duplicated series of names that may be associated with that germplasm.

A second component of ICIS is the 'Data Management System' (DMS) for capturing experimental data. The DMS embodies a generic data model structured with n-way indexed 'observation units' coupled with experimental variables which are independent 'factors' or dependent 'variates'. Each variable is documented with an ontology-typed 'property', 'scale' and 'method' specification. Sets of observation units are grouped into complete 'studies' that represent a particular round of experiments. This kind of data model with its strong ontology facilitates powerful queries that integrate related data across many independent studies.

After all phases of development are complete, ICIS will support a range of activities such as germplasm conservation and evaluation, through functional and comparative genomics, to allele mining, breeding, testing and varietal release. The database is amenable to distribution on CDROM for PC using specialized tools, or on the WWW.

THE INTERNATIONAL RICE INFORMATION SYSTEM

The extension of ICIS to molecular and genomic data is being driven by elaboration of the rice implementation, the International Rice Information System (IRIS, www.iris.irri.org) which stores information on more than one million varieties, breeding lines and accessions of rice, as well as 5 million data values in more than 500 studies from breeding, screening and international testing trials. Aside from featuring a sophisticated browsing interface for germplasm searches (Fig. 1), the web version of the database also provides access to IRRI's rice mutants, EST clones, genetic maps and related functional genomics data.

A biologist can approach the database with simple text searches (with wildcards permitted) directed to a specified 'entry point' and subject to entry point specific constraints. Some of the available entry points are:

- Germplasm
- Trait
- Mutant
- Clone
- Map

Typical 'search by' constraints that may be applied to these entry points are:

- Search by name
- Search by accession identifier
- Search by ontology term

The outcome of a given query is a list of hits to that string. Each hit accesses a specialized 'View' to the entity type (e.g. Germplasm View) summarizing information about that entity. Germplasm entries are cross referenced to associated field evaluation data and, increasingly, to associated molecular data.

The 'Mutant' entry point provides a graphical and descriptive term-directed browser organized by controlled vocabularies and ontology for traits.

FUTURE DIRECTIONS

IRRI is undertaking a project for high-throughput molecular characterization of the 'in trust' International Rice Genebank Collection. Use of this data for analysis of biodiversity and allele mining will facilitate utilization of the germplasm in future plant breeding efforts. Currently, IRRI maintains distinct legacy databases for breeding, genetic resources and international evaluation. Efforts are underway to consolidate these databases and their specialized views into IRIS.

IRIS is extensively cross-linked to other international plant databases such as the USDA Gramene database (Ware *et al.*, 2002) using common plant ontologies (Jaiswal *et al.*, 2002). In the near future, IRIS interoperability with such systems may be enhanced using emergent web data protocols like BioMOBY. IRIS potentially plays a combined role in BioMOBY as a WWW client for its users, server for its data and rice-specific registry for the community (Wilkinson and Links, 2002).

As time progresses, additional functional genomic, genetic resource and crop improvement data will be captured and cross integrated in IRIS, providing increasingly seamless linking of genotype (germplasm alleles at candidate gene loci) to phenotype (expression of important agronomic traits) information.

PROJECT PHILOSOPHY AND SOFTWARE AVAILABILITY

IRIS is being developed under the open-source ICIS project. Code is freely available to anyone wishing to enhance or deploy it at their sites, whether for rice or for other crops. Although the oldest components are based upon Microsoft Windows and ODBC, newer components are being developed using the methodology, languages and code base of other public, open-source bioinformatics projects such as the Perl community and the Generic Model Organism Database (http://www.gmod.org) project. To promote the open, collaborative spirit of the project and software exchange, a community project development site for ICIS is located at http://bioinformatics.org. Community participation is also promoted by open, annual ICIS workshops. The latest information about the ICIS project can be accessed via http://www.icis.cgiar.org.
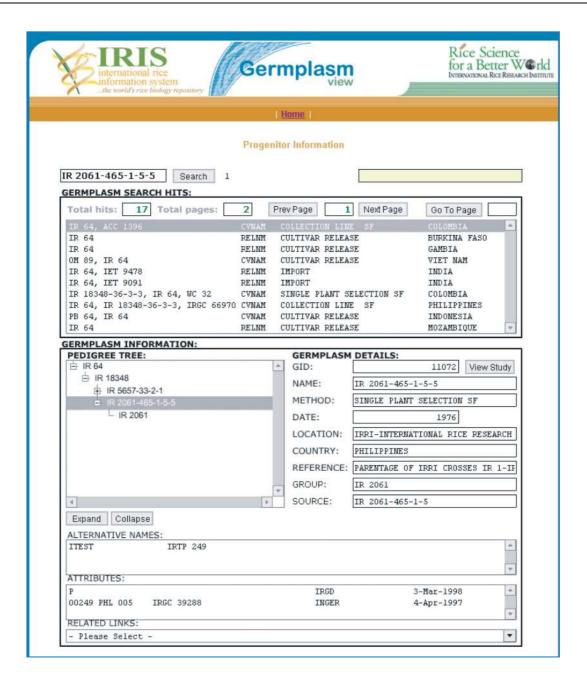
Fig. 1. IRIS Germplasm view pedigree browser.

REFERENCES

Fox,P.N. and Skovmand,B. (1996) *The International Crop Information System (ICIS)—Connects Genebank to Breeder to Farmer's Field*, In Plant Adaptation and Crop Improvement, Cooper,M. and Hammer,G.L. (eds), CAB International.

Jaiswal,P., Ware,D., Ni,J., Chang,K., Zhao,W., Schmidt,S., Pan,X., Clark,K., Teytelman,L., Cartinhour,S. *et al.* (2002) Gramene: development and integration of trait and gene ontologies for rice. *Comp. Funct. Genomics*, **3**, 132–136.

Ware,D., Jaiswal,P., Ni,J., Pan,X., Chang,K., Clark,K., Teytelman,L., Schmidt,S., Zhao,W., Cartinhour,S. *et al.* (2002) Gramene: a resource for comparative grass genomics. *Nucleic Acids Res.*, **30**, 103–105.

Wilkinson,M and Links,M. (2002) BioMOBY: an open-source biological web services proposal. *Brief Bioinformatics*, **3**, 331–341.