

TropGENE-DB, a multi-tropical crop information system

Manuel Ruiz*, Mathieu Rouard, Louis Marie Raboin¹, Marc Lartaud, Pierre Lagoda and Brigitte Courtois

CIRAD, Biotrop, TA 40/03, Avenue Agropolis, 34398 Montpellier Cedex 5, France and ¹CIRAD, CA, Station de Ligne-Paradis, 7 chemin de l'IRAT, 97410, Saint-Pierre, Réunion, France

Received August 13, 2003; Revised September 17, 2003; Accepted October 13, 2003

ABSTRACT

TropGENE-DB, is a crop information system created to store genetic, molecular and phenotypic data of the numerous yet poorly documented tropical crop species. The most common data stored in TropGENE-DB are information on genetic resources (agro-morphological data, parentages, allelic diversity), molecular markers, genetic maps, results of quantitative trait loci analyses, data from physical mapping, sequences, genes, as well as the corresponding references. TropGENE-DB is organized on a crop basis with currently three running modules (sugarcane, cocoa and banana), with plans to create additional modules for rice, cotton, oil palm, coconut, rubber tree, pineapple, taro, yam and sorghum. The TropGENE-DB information system is accessible for consultation via the internet at <http://tropgenedb.cirad.fr>. Specific web consultation interfaces have been designed to allow quick consultations as well as complex queries.

INTRODUCTION

At Centre de Coopération Internationale en Recherche Agronomique pour le Développement, France (CIRAD), plant genomics is aimed at assisting tropical crops breeding programs. The TropGENE-DB information system has been conceived to manage genetic and genomic information on these yet poorly documented crops. The present data include information on genetic resources (agro-morphological data, parentages, allelic diversity), molecular markers, genetic maps, results of Quantitative Trait Loci (QTL) analyses, data from physical mapping, sequences, genes, as well as the corresponding published paper references. These data are highly useful for rational use of genetic diversity available from germplasm collections, refinement of gene incorporation or introgression methodologies through accurate genome mapping, marker-assisted selection and genetic transformation.

TropGENE-DB is organized on a crop basis. The first genetic maps for cocoa, banana and a sugarcane cultivar were developed and published by scientists from CIRAD (1–3). The corresponding three TropGENE-DB crop modules (cocoa, banana and sugarcane) have been implemented and are now accessible for consultation via the internet. Database browsing is available at <http://tropgenedb.cirad.fr>. For each crop database, web consultation interfaces have been created to allow quick and complex queries, and user-friendly results representation.

DESIGN AND IMPLEMENTATION

TropGENE-DB development is based mainly on the AceDB database management system (J. Thierry Mieg and R. Durbin, 1996, <http://www.acedb.org>) version 4_91, running on Linux Redhat 7.1. AceDB is an object-oriented system capable of storing and retrieving complex biological information, and is currently used by many databases: WormBase (4), crop-related databases available from the UK Crop Plant Bioinformatics Network WWW site (5), MagnaportheDB (6), ESTHER (7), ParaDB (8), etc. AceDB provides an intuitive object-oriented view of biological data, and a graphical user interface with many specialized data visualization tools, such as a genetic map viewer and a sequence annotation display.

The first step in the development of the TropGENE-DB has been to design a generic database model with standardized class and tag names. The same object classes were created for all the species allowing easy comparison and interoperability between the different modules. All object types have interactive links with each other's.

The web consultation interface is implemented with Perl/CGI scripts, using modules of the AcePerl Application Programming Interface (API) and the AceBrowser generic web interface (9).

TropGENE-DB DATABASE CONTENTS

Currently, TropGENE-DB contains around 8200 clone entries (6300 for sugarcane, 800 for cocoa and 1100 for banana). These objects are linked to much phenotypic detailed information: geographic origin, collection description,

*To whom correspondence should be addressed. Tel: +33 4 67 61 56 31; Fax: +33 4 67 61 56 05; Email: manuel.ruiz@cirad.fr
Present address:

Pierre Lagoda, International Atomic Energy Agency, Plant Breeding and Genetics Section, Wagramer Strasse 5, PO Box 100, A-1400 Vienna, Austria

Figure 1. QTL query interface for the sugarcane module. This search allows QTL, trait, linkage group names and numeric phenotypic R2 values to be combined. The logical operators available are AND (corresponding to the intersection between the web forms), OR (corresponding to the union between the web forms) and [OR]AND (corresponding to the union between each element of one web form and the intersection between the web forms).

agro-morphological data, disease, pest and abiotic reactions, parentage, fertility, flowering, breeding data, etc. Associated molecular and genetic data comprise the genotypes at various markers (RFLP, AFLP, microsatellites, isozymes, etc.), genetic maps, information on the markers themselves (TropGENE-DB contains ~2200 marker entries linked to probes, primers, sequences data, etc.), QTL, genes, etc. AceDB allowed us to easily include images in TropGENE-DB: gel profiles, photos of fruits or plant disease reactions, etc.

WEB CONSULTATION INTERFACE

Specific TropGENE-DB queries

Each crop module has several interfaces to carry out specific requests. Phenotypic, Marker, QTL or Parentage query sections have been created to be used for all crop modules. Some additional ones are more specific to a given crop module, like Fertility for sugarcane for example. These interfaces make it possible to carry out complex requests combining various fields of the database (in the QTL query for example, searching for a QTL for a specific trait and a specific linkage group; Fig. 1). Complex queries can combine different data and relational (less than, equal to, etc.) or logical (and, or, etc.) operators. For the Phenotypic, Marker, QTL, Fertility and Parentage queries the result is, respectively, the list of corresponding clones (Germplasm class), locus (Locus class), QTL (QTL class), fertility data (Pollen_fertility class) and parent clones (Germplasm class).

Results of a query are listed with HTML checkboxes, allowing users to select the object's contents to be displayed in details. Moreover the request results can be visualized according to two different object representations: the standard representation (according to the classical hierarchical AceDB tree view) and the comparative representation (allowing easy comparison between tag values for different objects; Fig. 2).

Each AceDB object displayed in the web interface has a hypertext link. When users click this link, a new window appears that corresponds to the generic AceBrowser system allowing navigation through classes of the AceDB model. Links to the graphical AceDB displays are also added (Fig. 3).

Classical AceDB queries

AceDB data queries (Class, Text and AceDB queries) are also allowed. Class query corresponds to the useful object-browsing mode, which allows the user to retrieve objects by class, with the possibility of restricting the search to names that match a pattern. Text query is a keyword-based search on all the data. AceDB query uses the Ace Query Language (AQL), which was created to formulate complex queries based on several criteria. However, to realize these requests, end-users need to learn a specific syntax and to know the structure of each object model.

The Tropgene package

The development of new web interfaces, containing sophisticated request forms and results displays reformatted into nice HTML pages is a time-consuming task. We have developed a Perl object package, named Tropgene, which can be used, for all the AceDB databases, to create or modify customized web interfaces, quickly and easily. This package uses modules of the AcePerl API and the AceBrowser generic web interface (9). Attributes and methods of the Tropgene modules are used independently of the AceDB database model. In our package, only one configuration file has to be modified (by changing a few variables values) to completely create or modify web interfaces for a given AceDB database. This file defines the connected databases and the characteristics of the request forms and results displays. Our package allows the representation design to be easily modified and different output formats (XML, text, HTML tables, etc.) to be proposed.

	Generic_name	Type	Allele	Map	RFLP_Type	Probe	Primer_pair
PGMA	PGMA	Isozyme	PGMA_1 PGMA_1.2 PGMA_1.3 PGMA_2	Linkage_Group2 Position 74	-	-	-
gTcCIR161	gTcCIR161	RFLP	-	Linkage_Group2 Position 23.6	Genomic	gTcCIR161	-
mTcCIR003	mTcCIR003	Microsatellite	-	Linkage_Group2 Position 19.5	-	GEN473	AGTC21/22
rOPRB.0.5	-	RAPD	-	Linkage_Group2 Position 95.6	-	-	OperonRB

Figure 2. Comparative results display of a cocoa marker query. The table rows correspond to the selected marker objects and the columns to the tag values. Several tag values have AceBrowser links. The Graphic Display tag links to the associated genetic map graphical representation.

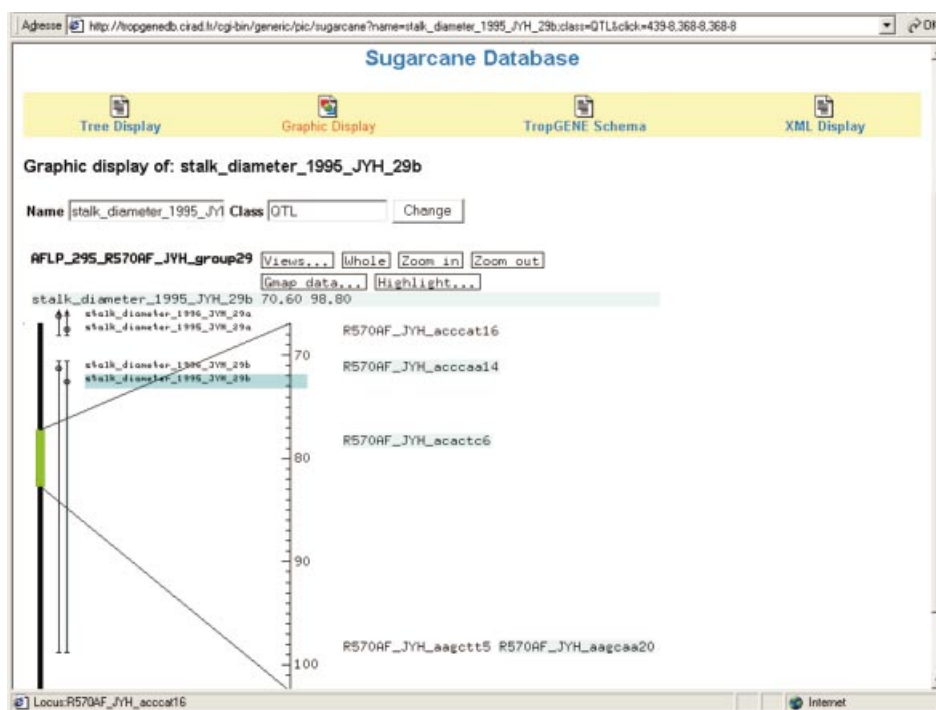


Figure 3. Graphical display of a sugarcane QTL. The selected QTL and the associated most significant locus are highlighted.

DATA SUBMISSION

Current TropGENE-DB data have been submitted by different CIRAD teams and by scientists from other institutions working on tropical crops (a data origin web page is available for each crop module). Potential submitters can contact us at the following address tropgene@cirad.fr.

Standard Excel submission files corresponding to the various types of data that can be submitted have been created to allow easy data submission. These files and a web form to deposit the standard data files for their incorporation in TropGENE-DB are currently available on our intranet

TropGENE-DB website. The internet version will soon be available. Submitted data quality and integrity are checked by biologist experts of each tropical crop.

FURTHER DEVELOPMENTS

It is planned for the TropGENE-DB content to grow quickly. Indeed, new data for a cocoa module including sequence, map and primer data on 150 recently developed microsatellites, 15 resistance genes, a genetic map including 424 markers, new QTL data (yield and yield components, bean traits, resistance to various species of *Phytophthora*) should be soon added to TropGENE-DB.

TropGENE-DB model flexibility makes it simple to add new types of data as they become available: expressed sequence tag (EST), bacterial artificial chromosome (BAC) data, etc.

Concerning the map representations, new graphical capabilities will also be developed to allow presentation of synteny relationships, building bridges between linkage groups.

Future enhancements of TropGENE-DB include the development of other tropical crops modules for rice, cotton, oil palm, coconut, rubber tree, pineapple, taro, yam and sorghum. Their interface development will be facilitated by the Tropgene package.

AVAILABILITY AND CITATION

Authors who use TropGENE-DB are encouraged to cite this article and to quote the TropGENE-DB home page URL, <http://tropgenedb.cirad.fr>.

ACKNOWLEDGEMENTS

We are deeply grateful to Thierry Erwin and Christine Nouaille for the TropGENE-DB home page design. We would like to thank all the biologist experts involved in the TropGENE-DB model design and data curation. TropGENE-DB is supported by the CIRAD and the Région Languedoc-Roussillon.

REFERENCES

1. Lanaud,C., Risterucci,A.M., N'Goran,J., Clément,D., Flament,M.H., Laurent,V. and Falque,M. (1995) A genetic map of *Theobroma cacao* L. *Theor. Appl. Genet.*, **87**, 987–993.
2. Faure,S., Noyer,J.L., Horry,J.P., Bakry,F., Lanaud,C. and Gonzalez De Léon,D. (1993) A molecular marker-based linkage map of diploid bananas. *Theor. Appl. Genet.*, **87**, 517–526.
3. Grivet,L., D'Hont,A., Roques,D., Feldmann,P., Lanaud,C. and Glaszmann,J.C. (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics*, **142**, 987–1000.
4. Harris,T.W., Lee,R., Schwarz,E., Bradnam,K., Lawson,D., Chen,W., Blasier,D., Kenny,E., Cunningham,F., Kishore,R. *et al.* (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.*, **31**, 133–137.
5. Dicks,J., Anderson,M., Cardle,L., Cartinhour,S., Couchman,M., Davenport,G., Dickson,J., Gale,M., Marshall,D., May,S. *et al.* (2000) UK CropNet: a collection of databases and bioinformatics resources for crop plant genomics. *Nucleic Acids Res.*, **28**, 104–107.
6. Martin,S.L., Blackmon,B.P., Rajagopalan,R., Houfek,T.D., Sceeles,R.G., Denn,S.O., Mitchell,T.K., Brown,D.E., Wing,R.A. and Dean,R.A. (2002) MagnaportheDB: a federated solution for integrating physical and genetic map data with BAC end derived sequences for the rice blast fungus *Magnaporthe grisea*. *Nucleic Acids Res.*, **30**, 121–124.
7. Cousin,X., Hotelier,T., Giles,K., Toutant,J.P. and Chatonnet,A. (1998) aCHEDb: the database system for ESTHER, the α/β fold family of proteins and the cholinesterase gene server. *Nucleic Acids Res.*, **26**, 226–228.
8. Leveugle,M., Prat,K., Perrier,N., Birnbaum,D. and Coulier,F. (2003) ParaDB: a tool for paralogy mapping in vertebrate genomes. *Nucleic Acids Res.*, **31**, 63–67.
9. Stein,L.D. and Thierry-Mieg,J. (1998) Scriptable access to the *Caenorhabditis elegans* genome sequence and other ACEDB databases. *Genome Res.*, **8**, 1308–1315.