

THE SUCEST-FUN REGULATORY NETWORK DATABASE: DESIGNING AN ENERGY GRASS

By

M.Y. NISHIYAMA-JR, F.F.R. VICENTE, C.G. LEMBKE, P.M. SATO,
M.L. DAL-BIANCO, R.A. FANDIÑO, C.T. HOTTA and G.M. SOUZA
Instituto de Química – Departamento de Bioquímica – Universidade de São Paulo
Av. Prof. Lineu Pretes 748, B9S, sala 954
São Paulo – SP – 05508-900 – BRAZIL
glmsouza@iq.usp.br

KEYWORDS: Sugarcane, SUCEST-FUN Database,
Bioinformatics, Genome Sequence, Transcriptome, Regulation.

Abstract

MODERN sugarcane cultivars are complex hybrids resulting from crosses among several *Saccharum* species. Traditional breeding methods have been employed extensively in different countries over the past decades to develop varieties with increased sucrose yield and resistance to pests and diseases. Conventional variety improvement, however, may be limited by the narrow pool of suitable genes. Thus, molecular genetics is seen as a promising tool to assist in the process of developing improved varieties. The SUCEST-FUN Project (<http://sucest-fun.org>) aims to associate function with sugarcane genes using a variety of tools, in particular those that enable the study of the sugarcane transcriptome. An extensive analysis has been conducted to characterise, phenotypically, sugarcane genotypes with regard to their sucrose content, biomass and drought responses. Through the analysis of different cultivars, genes associated with sucrose content, yield, lignin and drought have been identified. Currently, tools are being developed to determine signalling and regulatory networks in grasses, and to sequence the sugarcane genome, as well as to identify sugarcane promoters. This is being implemented through the SUCEST-FUN (<http://sucest-fun.org>) and GRASSIUS databases (<http://grassius.org>), the cloning of sugarcane promoters, the identification of cis-regulatory elements (CRE) using Chromatin Immunoprecipitation-sequencing (ChIP-Seq) and the generation of a comprehensive Signal Transduction and Transcription gene catalogue (SUCAST Catalogue).

Introduction

The increased importance of sugarcane as a bioenergy feedstock has increased pressure to generate new varieties optimised for energy production. The generation of improved varieties has relied largely on traditional breeding methods, which is limited by the narrow pool of suitable genes and the time-consuming process of selection of plants with desired agronomic traits. In this sense, molecular genetics can assist the process of developing improved varieties by generating molecular markers that can be used in the breeding process or by introducing new genes into the sugarcane genome (Menossi *et al.*, 2008).

In the last few years, the use of many high-throughput methods has exponentially increased the amount of data available to molecular geneticists. As a result, bioinformatics and efficient data management are essential in order to generate useful information. Furthermore, most of the available biological data are complex and stored in dozens of smaller databases (Lacroix and Critchlow, 2003). These databases are frequently not easy to identify and integrate, making the use of information very difficult because of the variety of semantics, interfaces, and data formats used by the underlying data sources (Lacroix and Critchlow, 2003).

The aim of this paper is to describe the SUCEST-FUN Regulatory Network Database (CaneRegNet), which contains data and tools of interest for sugarcane functional genomicists and molecular breeders. This database has been developed in the concept of the mediator approach that incorporates concepts from Data Warehouse and Federation approaches (Lacroix and Critchlow, 2003). It also has flexible data integration to assemble heterogeneous distributed data sources, experimental data, resources, the applications of scientific algorithms and computational analysis.

The database community has extensively investigated architectures and tools for data integration (Lacroix and Critchlow, 2003). A key disadvantage of the warehouse approach is the need for local administrators to maintain the data, while providing control over the contents of the warehoused data. Mediators and heterogeneous Database Management Systems (DBMS), on the other hand, submit the queries directly to the wrappers; program routines that take data from Internet websites or any other data repositories and convert the information into a structured format, and integrate the results locally (Boulcelma *et al.*, 2003).

A geographic data warehouse must provide an expressive query language that may be difficult to maintain and less expressive than those available at remote sources. On the other hand, non-materialised approaches must not only integrate data, but also support all these query capabilities of interest. The purpose of the mediator approach is to integrate both data and source capabilities.

Methods

The databases were developed using MySQL Server (<http://www.mysql.com>) while the interface and search systems were based on the WebServer Apache (<http://www.apache.org>).

The website uses the Joomla platform, as a Content Management System. This platform is developed using PHP (server-side HTML-embedded scripting language). Both Joomla and PHP are free software released under the GNU/GPL Licence. Joomla provides a more interactive website, which allows keyword searches and also manages the access control for groups of users. Moreover, Joomla has a toolkit that provides flexibility in the integration of scripts and programs developed in other languages.

The tools and scripts developed and implemented in the website were produced using CGI (Simple Common Gateway Interface Class), PHP (<http://www.php.net>), PERL (<http://www.perl.org>) and R (<http://www.r-project.org/>) a statistical environment. The data are available to the SUCEST-FUN community through the integration of other platforms such as the Gbrowse (Generic Genome Browse – <http://www.gmod.org>) and BioPerl (Stajich *et al.*, 2002) modules.

Results

The CaneRegNet has been developed following the concept of the mediator approach, which integrates concepts from the Data Warehouse and the Federation approaches. Accessed data are up-to-date, but access can be costly. In addition, wrappers (Boulcema *et al.*, 2003) must be maintained since data providers may change the entry points to the data sources as well as the database organisation. Neither materialised into data warehouses nor non-materialised approaches address the problem to access or maintain sophisticated tools that enhance data manipulation. Indeed, in addition to standard data manipulation such as those performed by SQL, geographic languages usually express spatial selections, metric or topological queries and allocations among other features (Boulcelma *et al.*, 2002).

The SUCEST-FUN Regulatory Network Database (CaneRegNet) assembles different sugarcane databases such as the Sugarcane Expressed Sequence Tags (SUCEST) Genome Project (<http://sucest.lad.ic.unicamp.br/en/>) (Vettore *et al.*, 2003), the SUCAST and the SUCAMET Catalogues, which include expression data (<http://sucest-fun.org>), the GRASSIUS database (Yilmaz *et al.*, 2009) and records of the agronomic, physiological and biochemical characteristics of

sugarcane cultivars. This database is part of the SUCEST-FUN Regulatory Network Project (<http://sucest-fun.org>), which aims to study gene expression regulation through the use of tools that will allow a System Biology approach for the study of sugarcane. This database is based on five main topics: 1) Gene annotation, 2) Gene Expression, 3) Public Resources, 4) Sequencing Projects and 5) Functional Genomics (Figure 1).

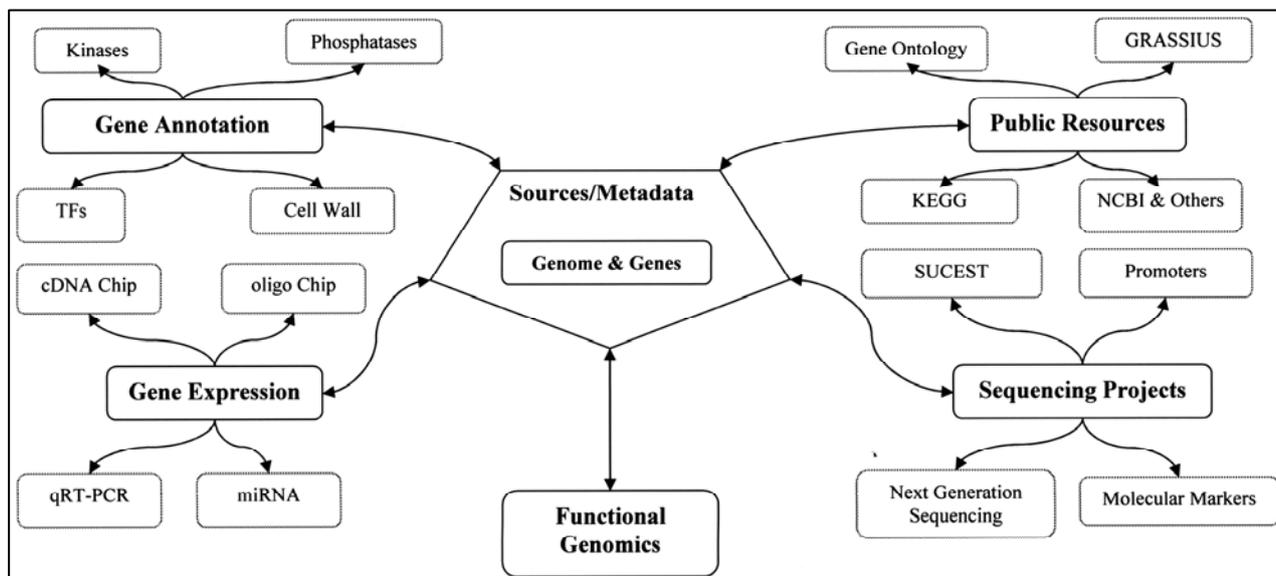


Fig. 1—Diagram of the CaneRegNet database and data abstraction. The CaneRegNet is based on five main topics: Gene Annotation, Gene Expression, Public Resources, Sequencing Projects and Functional Genomics. The topics and source arrows are bidirectional allowing the interaction and relationship between them. TFs stand for Transcription Factors.

1) Gene Annotation

The **Gene Annotation** resource integrates the many catalogues generated from the annotation and categorisation of the putative transcripts sequenced by the SUCEST project (<http://sucest.lbi.ic.unicamp.br/public/>), which is the largest sugarcane EST collection and is represented by 43 141 putative transcripts known as the Sugarcane Assembled Sequences (SAS).

The SUCAST Catalogue (Sugar Cane Signal Transduction—http://sucest-fun.org/cane_regnet/en/database-tools/gene-catalogue/sucast-catalogue) is the first characterisation of the putative transcripts sequenced by the SUCEST project.

It contains more than 3500 components involved in several aspects of signal transduction, transcription, development, cell cycles, stress response and pathogen interaction (Souza *et al.*, 2001).

The **SUCAST Catalogue Tool** allows the search of genes by categories and contains genes involved in signal transduction pathways that integrate with the sugarcane kinome based on a phylogenetic categorisation (Rocha *et al.*, 2007).

The SUCAMET Catalogue (Sugar Cane Metabolism) contains genes involved in metabolic pathways. As in the SUCAST Catalogue, the **SUCAMET Catalogue Tool** was developed to enable the search for genes by categories.

Both the SUCAST and SUCAMET projects were used to generate cDNA-arrays for the study of the sugarcane transcriptome.

More recently, a Cell Wall Catalogue and a Transcription Factor Catalogue were included in the CaneRegNet. The Transcription Factor Catalogue was generated as part of a collaborative effort

to develop the GRASSIUS Platform (Grass Regulatory Information Services) (Yilmaz *et al.*, 2009). GRASSIUS holds information on transcription factors (TFs) from maize, sorghum, sugarcane and rice classified into families.

The TFs for grasses can be accessed by either browsing family members or searching by name or by sequence similarity through BLAST searches or by phylogenetic homology. The GRASSIUS Platform contains 1647 TFs catalogued for sugarcane, which represents about 83% of the number estimated to be present in sugarcane (Yilmaz *et al.*, 2009).

The identification of sugarcane TPs is an essential step in the process of building regulatory networks, which also requires the identification of promoters, CREs and knowledge on their interactions with TFs (Schlitt and Brazma, 2007).

2) *Gene Expression*

The **Gene Expression** resource is focused on the comparison of the expression profile of different sugarcane varieties, different tissues within one individual and individuals subjected to different biotic and abiotic stimuli.

The expression profiles were made from two different custom cDNA microarrays. One cDNA microarray contained 2208 elements from the SUCAST Catalogue and was designed to identify the transcriptome changes in response to drought, phosphate starvation, stress in response to herbivory and N₂-fixing endophytic bacteria.

Also investigated was the response of sugarcane to phytohormones, such as abscisic acid (ABA) and methyl jasmonate. The other cDNA microarray (SUCAMET Catalogue) contained 4594 elements that are involved in cellular process, stress response and basal metabolism.

An outliers searching method resulted in the identification of 179 genes that were considered differentially expressed in at least one of the treatments analysed (Rocha *et al.*, 2007). qRT-PCR was used to validate the expression profiles obtained using the cDNA arrays.

Expression data were stored in different stages and formats in flat files and Relational Databases, which can be visualised and grouped by correlation, Venn diagrams or experiments using different cut-offs and levels of confidence. The analysis of both cDNA arrays resulted in the development of several tools that allow the search for differentially expressed genes. For example, a virtual Matrix of expression profile by tissues was constructed using data from the SUCAST microarray, which allowed the identification of sugarcane genes that are tissue specific.

Furthermore, a Self Organising Map (SOM) tool was implemented to select differentially expressed candidate genes to make hierarchical clustering analysis (Rocha *et al.*, 2007). The Gene Expression database in CaneRegNet allows the identification of genes that are co-regulated by different treatments.

For example, a comparison between the transcripts from sugarcane treated with 100 µM ABA and those present in drought tolerant and drought sensitive sugarcane varieties subjected to water deficit stress showed that 14 of the 24 genes differentially regulated by ABA also had different expression levels in one of the sugarcane varieties.

In addition, 115 of the 184 genes regulated in tolerant varieties submitted to drought were also differentially expressed in sensitive varieties (Figure 2). Among the ABA treatment and the drought tolerant varieties submitted to drought and between ABA treatment and drought sensitive varieties, the four gene categories that were more represented included stress, carbohydrate metabolism, hormone biosynthesis and no matches.

Eleven genes were differentially expressed in the three samples. These genes belonged to stress, carbohydrate metabolism, hormone biosynthesis, no matches, protein phosphatase and unknown protein functional categories.

Another example of co-regulation has recently been described for genes associated with sucrose content and drought responses (Papini-Terzi *et al.*, 2009).

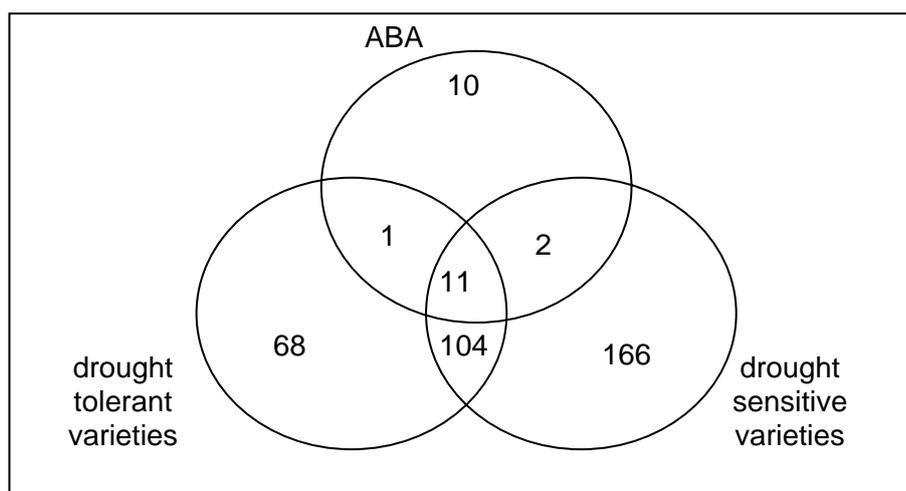


Fig. 2—Comparison of differential gene expression associated with ABA treatment and drought stress in sugarcane. Genes were identified as associated with ABA if they were differentially expressed when treated (for 0.5, 1, 6 and 12 h) and untreated plants were compared. Genes regulated during drought stress were identified by comparing plants with irrigation and without irrigation after 24, 72 or 120 h of water deficit. The figure represents a Venn diagram of the three differential expression data sets. The credibility level used to define outliers was 0.96 in all three data sets.

A new platform 60-mer Agilent oligoarray technology with a custom-designed array (44 k = 44 000 elements) has been developed recently to improve the investigation of the sugarcane transcriptome. Approximately 40% of the unique genes in the SUCEST database are represented in this array, with a total of 21 902 specific probes replicated twice and designed in the sense and antisense direction with high specificity. This oligoarray will allow the identification of antisense expressed sequences and the sugarcane regulatory networks associated with sucrose content and drought stress.

3) Public Resources

A great number of databases and resources have been developed for general genome analysis, such as the National Center for Biotechnology Information (NCBI), Gene Ontology (GO) (Ashburner and Lewis, 2002) and Kyoto Encyclopedia of Gene and Genomes (KEGG) (Kanehisa and Goto, 2000), and for the study of grass genomes, such as MaizeGDB (Lawrence *et al.*, 2004) Sorghum Phytozome (Paterson *et al.*, 2009) and Rice Genome Annotation (Yuan *et al.*, 2003). These databases allow the integration of a variety of data types and analyses in a single framework. The flexibility of these generic data abstractions will allow the integration of all data and joint analysis of sugarcane and other grasses, and will provide a simple and fast searching platform.

4) Sequencing Projects

The CaneRegNet database contains the sequences produced by the SUCEST Project (<http://sucest.lad.ic.unicamp.br/en/>; (Vettore *et al.*, 2003). This project produced 237 954 high quality ESTs, obtained from 27 cDNA libraries from different sugarcane tissues of different developmental stages and stress conditions. This collection of ESTs was assembled into 43 141 putative genes referenced as Sugarcane Assembled Sequences (SAS) that can be analysed with the following tools:

- **Tissue and Library Search**—gives access to the names, descriptions and EST sequences by library or tissue specific
- **Cluster alignment viewer**—shows the cluster assembly of an EST
- **Object Search**—permits direct access to any object, such as ESTs or a SAS name

- **Generic SQL Query**—allows users to create generic queries typed in a web form and the results are exhibited in a tabular format
- **SAS Information Viewer**—integrates and exhibits all SAS information available in the CaneRegNet database.

Sugarcane and sorghum are syntenic organisms that show co-linearity between the genes and striking homology in most of the coding parts of the genome (Jannoo *et al.*, 2007). When SUCEST was compared with the sorghum CDS sequences using the pairwise alignments, 4212 orthologue candidates were identified (Table 1). These putative orthologues were identified using stringent parameters of e-value, coverage and identity. Because SAS are expected to contain UTRs and some of them are incomplete transcripts, a more rigorous coverage threshold was applied to sorghum. Less stringent parameters may result in a larger number of candidates, at the cost of increase in false positives.

Table 1—Features and numbers from pairwise alignment between sugarcane and sorghum genes.

Feature	Lowest value	Median	Highest value
SAS length (nt)	262	1304	6194
Sorghum CDS length	153	933	3417
Sorghum coverage	90%	100%	100%
SAS coverage	50%	70%	100%
% identity	79%	96%	99%

A number of SAS (19 418) were mapped onto the sorghum genome using tools that allowed the identification of intron/exon boundaries spanning 10 chromosomes and 3294 supercontigs. The best match of each SAS between all chromosomes and supercontigs was selected. Next, a threshold of 90% of coverage and 90% of identity was used to select the orthologue candidates. Further work is needed in order to extend the SAS and identify more orthologue candidates.

The EST sequences from the SUCEST database are essential for the Gene Expression and the Gene Annotation databases. However, they lack information about the promoters, and other regulatory elements, that are essential for the understanding of gene regulatory networks. A major sequencing effort has recently started in order to generate more information about regulatory elements of sugarcane genes. The Sugarcane Genome Sequencing Initiative (<http://bioenfapesp.org>) is focused in sequencing the sugarcane gene space using next generation sequencing tools. One approach is the use of massively parallel pyrosequencing in association with gene-enrichment methods. The CaneRegNet database is currently being prepared to integrate these sequences to the SUCEST database and the sorghum genome using GBrowser. Sequencing strategies are expected to include BAC sequencing, surveys of several cultivars using the shot-gun approach, ChIP-Seq (Chromatin Immunoprecipitation followed by sequencing), those targeting gene promoters and full length cDNA sequencing.

5) *Functional Genomics*

The vast amount of data contained in Gene Annotation, Gene Expression, Public Resources and Sequencing Projects databases must be organised in the context of the plant in order to provide new understanding of gene function and gene interaction, which are instrumental for the discovery of genes that regulate agronomic traits of interest. These genes can be used as molecular markers in breeding programs or to generate transgenic sugarcane plants.

Physiology studies and transgenic sugarcane plants will be used to test hypotheses regarding gene regulatory networks and for assigning function to the genes identified. The CaneRegNet database will store information on all plant samples, including agronomic traits such as growth

rates, water use efficiency, photosynthesis parameters, and sugar and cell wall content of sugarcane genotypes. Information on the protocols, experiments, clones, vectors and primers as well as the results generated will be stored and visualised by a web interface and integrated into the Relational database with the other resources.

Discussion and future directions

An evolving integrated platform has been developed to integrate and store a large amount of data from diverse sources and experimental conditions. The SUCEST-FUN db provides the infrastructure for storage, retrieval and integration especially for deep sequencing genomics, DNA methylation filtration experiments, Molecular Markers, ChIP-Seq data, expression data, gene catalogues and information on sugarcane physiology and from the analysis of transgenic plants. The mediator approach will allow the running of parallel queries, do parallel loads into multiple SUCEST-FUN db instances, and provide unlimited data storage and management. The SUCEST-FUN db will be an important tool for scalable storage and interpretation of these data. New datasets will be available as soon as the accompanying papers are released.

The search and analysis tools developed in the project have allowed correlations to be made among gene expression experiments as well as among groups of experiments treating different conditions and treatments. The tools have also permitted the consolidation of the data with manually curated genes, gene catalogues and aggregated public resources such as GO and KEGG.

The extent of genetic and genomic information available on several species, such as barley, wheat, maize, rice, and sorghum, which belong to different tribes in the grass family, has enabled the most comprehensive comparative genomic studies in plants (Eckardt, 2008). The SUCEST-FUN db will be fundamental for the development of new tools and analysis of the sugarcane genome and functional genomic studies.

In the future, this platform will be composed of new algorithms and analysis tools, which are to be based on mathematical and statistical methodologies and draw on numerous biological resources. These resources will exploit the whole sugarcane genome sequence, the transcriptome and functional genomic data. This will allow the application of a Systems Biology approach to sugarcane in order to identify regulatory gene networks, discover new genes and develop improved 'energy' cultivars.

REFERENCES

- Ashburner, M. and Lewis, S. (2002). On ontologies for biologists: the Gene Ontology—untangling the web. *Novartis Foundation symposium*, 247: 66–80; discussion 80–63, 84–90, 244–252.
- Boucelma, O., Essid, M. and Lacroix, Z. (2002). A WFS-based mediation system for GIS interoperability, in 10th ACM international symposium on Advances in geographic information systems, McLean, Virginia, USA.
- Eckart, N.A. (2008). Grass Genome Evolution. *Plant Cell*, 20: 3–4.
- Jannoo, N., Grivet, L., Chantret, N., Garsmeur, O., Glaszmann, J.C., Arruda, P. and D'Hont, A. (2007). Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant J.*, 50:574–585.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28: 27–30.
- Lacroix, Z. and Critchlow, T. (2003) *Bioinformatics: Managing Scientific Data*: Elsevier Science.
- Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E. and Brendel, V. (2004). MaizeGDB, the community database for maize genetics and genomics. *Nucleic acids research*, 32: D393–397.
- Menossi, M., Silva-Filho, M.C., Vincentz, M. Van-Sluys, M.A., Souza, G. M. (2008). Sugarcane Functional Genomics: gene discovery for agronomic trait development. *Int J Plant Genomics*.

- Papini-Terzi, F.S., Rocha, F.R., Vencio, R.Z., Felix, J.M., Branco, D.S., Waclawovsky, A.J., Del Bem, L.E., Lembke, C.G., Costa, M.D., Nishiyama, M.Y., Jr., Vicentini, R., Vincentz, M.G., Ulian, E.C., Menossi, M. and Souza, G.M. (2009) Sugarcane genes associated with sucrose content. *BMC Genomics*, 10: 120.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberler, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otiillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboobur, R., Ware, D., Westhoff, P., Mayer, K.F., Messing, J. and Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, 457: 551–556.
- Rocha, F.R., Papini-Terzi, F.S., Nishiyama, M.Y., Jr., Vencio, R.Z., Vicentini, R., Duarte, R.D., de Rosa, V.E., Jr., Vinagre, F., Barsalobres, C., Medeiros, A.H., Rodrigues, F.A., Ulian, E.C., Zingaretti, S.M., Galbiatti, J.A., Almeida, R.S., Figueira, A.V., Hemerly, A.S., Silva-Filho, M.C., Menossi, M. and Souza, G.M. (2007). Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. *BMC Genomics*, 8:71.
- Schlitt, T. and Brazma, A. (2007) Current approaches to gene regulatory network modelling. *BMC Bioinformatics*. 8 Suppl 6: S9.
- Souza, G.M., Simoes, A.C. and Oliveira, K.C. (2001) The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. *Genetics and Mol. Biol.*, 24:25–34.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D. and Birney, E. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome research*, 12:1611–1618.
- Vettore, A.L., da Silva, F.R., Kemper, E.L., Souza, G.M., da Silva, A.M., Ferro, M.I., Henrique-Silva, F., Gigliotti, E.A., Lemos, M.V., Coutinho, L.L., Nobrega, M.P., Carrer, H., Franca, S.C., Bacci Junior, M., Goldman, M.H., Gomes, S.L., Nunes, L.R., Camargo, L.E., Siqueira, W.J., Van Sluys, M.A., Thiemann, O.H., Kuramae, E.E., Santelli, R.V., Marino, C.L., Targon, M.L., Ferro, J.A., Silveira, H.C., Marini, D.C., Lemos, E.G., Monteiro-Vitorello, C.B., Tambor, J.H., Carraro, D.M., Roberto, P.G., Martins, V.G., Goldman, G.H., de Oliveira, R.C., Truffi, D., Colombo, C.A., Rossi, M., de Araujo, P.G., Sculaccio, S.A., Angella, A., Lima, M.M., de Rosa Junior, V.E., Siviero, F., Coscrato, V.E., Machado, M.A., Grivet, L., Di Mauro, S.M., Nobrega, F.G., Menck, C.F., Braga, M.D., Telles, G.P., Cara, F.A., Pedrosa, G., Meidanis, J. and Arruda, P. (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome research*, 13:2725–2735.
- Yilmaz, A., Nishiyama, M.Y., Jr., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J. and Grotewold, E. (2009). GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant physiology*, 149:171–180.
- Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J. and Buell, C.R. (2003). The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic acids research*, 31:229–233.

LE RÉSEAU DE LA BASE DE DONNÉES SUCEST-FUN: LA CONCEPTION D'UNE CANNE ÉNERGIE

Par

M.Y. NISHIYAMA-JR., F.F.R. VICENTE, C.G. LEMBKE, P. M. SATO, M.L. DAL-BIANCO, R. A. FANDIÑO, C.T. HOTTA ET G.M. SOUZA

*Instituto de Química – Departamento de Bioquímica – Universidade de São Paulo
Av. Prof. Lineu Pretes 748, B9S, sala 954 São Paulo – SP – 05508-900 – BRÉSIL
gmsouza@iq.usp.br*

MOTS CLÉS: Canne à Sucre, Base de Données SUCEST-FUN, Bioinformatiques, Séquence Du Génome, Transcriptome, Régulation du Gène.

Résumé

LES CULTIVARS modernes de la canne à sucre sont des hybrides complexes résultant de croisements entre plusieurs espèces de *Saccharum*. Au cours des dernières décennies, des méthodes d'amélioration génétique traditionnelles ont été surtout employées dans différents pays pour développer des variétés à fort rendement en saccharose et résistantes aux ravageurs et aux maladies. Cependant, l'amélioration des variétés par ces méthodes, est limitée par la gamme restreinte de gènes appropriés. Ainsi, la génétique moléculaire est considérée comme un outil prometteur pour aider au processus d'amélioration variétale. Le projet SUCEST-FUN (<http://sucest-fun.org>) vise à associer des fonctions à des gènes de la canne à sucre en utilisant une variété d'outils, en particulier ceux permettant l'étude du transcriptome. Une analyse approfondie a été menée afin de caractériser, les génotypes de canne à sucre basés sur les phénotypes suivants: la teneur en saccharose, la biomasse et la réponse à la sécheresse. Grâce à l'analyse de différents cultivars, les gènes responsables pour la teneur en saccharose et la lignine, ainsi que pour une réponse à la sécheresse ont été identifiés. Actuellement, des outils sont en cours de développement pour déterminer les réseaux de signalisation et de régulation parmi les graminées, ainsi que pour le séquençage du génome de la canne à sucre, et pour identifier des promoteurs de la canne. Ces projets sont réalisés par l'intermédiaire du programme SUCEST-FUN (<http://sucest-fun.org>) et les bases de données GRASSIUS (<http://grassius.org>), avec le support du clonage des promoteurs de la canne à sucre, l'identification des éléments cis-régulateurs (ECR) à l'aide de la méthode CHIP (Chromatin Immunoprecipitation-sequencing), et la génération d'un catalogue SUCAST détaillé (Signal Transduction and Transcriptome Gene catalogue).

**LA BASE DE DATOS DE LA RED REGULATORIA SUCEST-FUN:
DISEÑANDO UNA GRAMÍNEA ENERGÉTICA**

Por

M.Y. NISHIYAMA-JR, F.F.R. VICENTE, C.G. LEMBKE, P.M. SATO,
M.L. DAL-BIANCO, R.A. FANDIÑO, C.T. HOTTA and G.M. SOUZA

*Instituto de Química – Departamento de Bioquímica – Universidad de São Paulo
Av. Prof. Lineu Pretes 748, B9S, sala 954 São Paulo – SP – 05508-900 – BRASIL
glmsouza@iq.usp.br*

**PALABRAS CLAVE: Caña de Azúcar, Base de Datos SUCEST-FUN,
Bioinformática, Secuencia del Genomio, Transcriptomo, Regulación.****Resumen**

LOS CULTIVARES de la caña de azúcar son híbridos complejos resultantes de las cruces interespecíficas de *Saccharum*. Métodos convencionales de fitomejoramiento han sido extensivamente usados en diferentes países a lo largo de las pasadas décadas para desarrollar variedades con un rendimiento mayor de sacarosa y resistencia a plagas y enfermedades. Sin embargo, el fitomejoramiento convencional puede limitarse por una reducción en la poza de genes adecuados. Por tanto, la genética molecular es vista como una herramienta prometedora para apoyar el proceso de desarrollo de variedades mejoradas. El proyecto SUCEST-FUN (<http://sucest-fun.org>) tiene como objetivo asociar la función de genes de la caña de azúcar usando una gama de herramientas, en particular aquellas que permiten el estudio del transcriptomo de la caña de azúcar. Un análisis extenso ha sido conducido para caracterizar, fenotípicamente, genotipos de caña de azúcar con respecto de su contenido de sacarosa, biomasa y respuesta a la sequía. A través del análisis de diferentes cultivares, han sido identificados los genes asociados con el contenido de sacarosa, el rendimiento, la lignina y la respuesta a la sequía. Actualmente, se están desarrollando herramientas para determinar las redes de señalización y regulación en gramíneas, y para secuenciar el genomio de la caña de azúcar, así como para identificar promotores de caña de azúcar. Esto ha sido implementado a través de las bases de datos SUCEST-FUN (<http://sucest-fun.org>) y GRASSIUS (<http://grassius.org>), la clonación de promotores de caña de azúcar, la identificación de elementos *cis* de regulación (CRE, por sus siglas en Inglés) usando la Inmunoprecipitación–Secuenciación de Cromatina (ChIP-Seq, por sus siglas en Inglés) y la generación de un catálogo exhaustivo de Señalización de Transducción y Transcripción (Catálogo SUCAST).