

Mapping sugarcane

The Genome Project paved the way for the development of new varieties of the plant

Fabício Marques



The race to find new varieties of sugarcane adapted to Brazil's variety of climates and soils is reaping the benefits from a major collaboration that brought together 240 researchers representing 22 institutions from 1999 to 2002. The Sugarcane Genome Project, tasked with mapping 238,000 functional sugarcane gene fragments, paved the way for the use of molecular markers in crop improvement. The identification of the fragments, called expressed sequence tags (ESTs), was followed by data mining related to sugarcane metabolism so as to obtain varieties that are more productive and resistant to drought or poor soil. "We obtained 238,000 transcript fragments, set out to identify the function of the genes, studied the associated agronomic characteristics and analyzed the transcriptoma to help produce more-efficient transgenic plants," says Glauca Souza, a professor at the Chemistry Institute of the University of São Paulo who took part in the Sugarcane Genome collaboration.



Souza now coordinates FAPESP's Research on Bioenergy Program, or Bioen (*see report on page 140*), aimed in part at improving the productivity of Brazilian ethanol and making advances in basic science and technology related to biomass energy production. One of its divisions incorporates research based on the sugarcane genome. The information obtained through the functional genome of sugarcane also engendered other studies, including one to identify canecystatin, a protein with anti-fungal properties that was studied as a possible inhibitor of pathogens that attack plants. The research group that conducted the study was headed by the geneticist Flávio Henrique da Silva of the Center for Structural Biotechnology (CBME) at the Federal University of São Carlos (UFSCar).

Known officially as the FAPESP Sucest (Sugarcane EST) Project, the sugarcane genome was the focus of one of the projects connected to the Onsa (Organization for Nucleotide Sequencing and Analysis) Network, an infrastructure of laboratories located in several cities and outfitted with new sequencers and other equipment. The network—a kind of virtual research institute—was created in 1997, and its first undertaking was the

240
 researchers
 from 22
 institutions
 collaborated
 on the project

Greenhouse with seedlings at the Sugarcane Technology Center: industry interest spurred the Sugarcane Genome Project

sequencing of the genome for the *Xylella fastidiosa* bacterium in 2000 (*see report on page 44*). Soon thereafter, the researchers undertook other projects under FAPESP's Genome Program. In 1998, the network became immersed in issues of great social and economic interest, including the mapping of genes associated with cancer and the expressed genome of sugarcane.

The Sugarcane Genome Project was characterized by a robust interaction between universities and the private sector that has distinguished bioenergy research efforts thus far. Paulo Arruda, a professor at the State University of Campinas (Unicamp) and the project's coordinator, recalls being invited to head the project after the Sugar and Alcohol Producers' Cooperative of the State of São Paulo (Copersucar) and its research arm, the Sugarcane Technology Center (CTC), called on the scientific directors of FAPESP and proposed a partnership between universities and industry with the aim of sequencing the sugarcane genome. "Professor José Fernando Perez, then one of the scientific directors, asked me what I thought. I observed that sugarcane has a very complex genome, and I recommended mapping the functional fragments of the genome," says Arruda, who today is one of the coordinators of FAPESP's Research for Innovation department. Sugarcane is truly a complex organism. Its genome is three times larger than that of humans, and this is exacerbated by the fact that, instead of two copies of each chromosome, there are eight to twelve copies that are not always identical. Owing to this peculiarity, the idea of full sequencing of the genome was abandoned because the process would be too lengthy and tedious.

The project received around US\$4 million in funding from FAPESP, and another US\$400,000 from Copersucar. For the first time in a joint enterprise, it brought together laboratories in the states of São Paulo, Pernambuco, Bahia, Rio de Janeiro, Paraná, Rio Grande do Norte, Minas Gerais and Alagoas. "All the laboratories had access to the data bank and were able to study the genes identified," Arruda says. "It was really innovative. With its focus on very young people, who were more at ease working with technology than the more-experienced researchers, the Sugarcane Genome Project showed that it is possible to identify major challenges and bring together the talent to solve them," he adds. "The project had enormous repercussion. We compiled the research findings in a special edition of the Brazilian Genetics Society's journal that became the most heavily cited in the publication's history," Arruda notes. In September 2003, an article in the scientific journal *Genome Research* presented the principal result of the

program: a detailed description of the genetic composition of sugarcane, Brazil's longest-cultivated plant on a large scale. The article showed that the sugarcane genome consists of 33,620 possible genes, of which about 2,000 appear to be associated with sugar production.

Felipe Rodrigues da Silva, a biologist, and Guilherme Pimentel Telles, trained in computer science, know how difficult it was to arrive at the final numbers that completed the venture begun in April 1999. To make the initial determination of the number of genes—basic information on any genome—Silva, then a 29-year-old doctoral student, and Telles, 27, had to work out what had not yet been solved in any other laboratory in the world. They had to find out how to eliminate repetitions and find the best possible way to use the information contained in nearly 300,000 gene fragments. The sugarcane genome was one of the first plant projects in the world to adopt that gene identification technique. Until they hit upon the solution, they worked at least twelve hours a day for four months on programs according to which the number of sugarcane genes varied from 9,000 genes to over 100,000, and sometimes an intermediate number of genes that varied in accordance with the differing criteria employed as to what constitutes a gene. At one crucial point, they discovered that sections of genes that could have been used were being discarded.

The Sugarcane Genome Project basically launched the ongoing effort to deepen our understanding of sugarcane metabolism so that varieties that are more productive and resistant to drought or poor soil can be obtained more quickly. Under the current genetic-improvement techniques, a new variety takes ten years of work, from initial testing to approval for use in the field. “We’re still in the very early stages of understanding the sugarcane genome,” Arruda says. “The plant has a potential productivity three times greater, but we still don’t know to what extent the genome presents an obstacle to the exploitation of that potential. In any event, the project showed that it is possible to rise to the challenge,” he maintains.

The completion of the project did not dampen interest on the part of researchers and industry in their continued quest for understanding of the plant. After 2003, Souza became coordinator of Sucest and launched the Sucest-FUN project, a network of researchers focused on analyzing sugarcane genes. The genes associated with sugar content were identified in a joint project of the

Sugarcane Technology Center, the Lucélia Central Alcohol Distillery and researchers from the University of São Paulo (USP) and Unicamp. The project was funded by FAPESP and headed by Glaucia. Another important project involved the identification of molecular markers from the Sucest sequences, under the leadership of researcher Anete Pereira de Souza of the Biology Institute at Unicamp. The markers can be used in identifying a specific gene—for example, one that is linked to sucrose production. “Souza’s and Pereira da Souza’s projects were both landmarks, because they demonstrated that there is a community of researchers prepared to invest time and effort in this issue.

The article in *Genome Research* showed that the sugarcane genome has about 2,000 genes associated with sugar production

Their advances made it possible to actually map the sugarcane genome, which was not feasible in the Sucest era,” said geneticist Marie-Anne Van Sluys, a professor at USP’s Biosciences Institute. ■

PROJECTS

1. Proposal for DNA coordinator of the Sugarcane EST Project (SucEST) – No. 1998/12250-0 (1998-2004)
2. Bioinformatics for the sugarcane EST project – No. 1999/02837-6 (1999-2002)
3. SucEST – data mining – No. 1999/02840-7 (1999-2002)

GRANT MECHANISM

1. 2. 3. Line of Research Project Award – Genome Program

COORDINATORS

1. Paulo Arruda (CBMEG-Unicamp)
2. João Meidanis (IC-Unicamp)
3. Antonio Vargas de Oliveira Figueira (Cena-USP)

INVESTMENT

1. R\$2,324,381.81
2. R\$576,439.24
3. R\$52,496.22

SCIENTIFIC ARTICLE

VETTORE, A. L. *et al.* Analysis and Functional Annotation of an Expressed Sequence Tag Collection for Tropical Crop Sugarcane. *Genome Research*. v. 13, p. 2725-35. 2003.

FROM OUR ARCHIVES

The architects of the new sugarcane
Issue No. 59 – November 2000

Abundant harvest
Issue No. 91 – September 2003