

# Most complete commercial sugarcane genome sequence assembled

3 December 2019, by André Julião



Credit: CC0 Public Domain

An international group of researchers led by Brazilian scientists has assembled the most complete genome sequence of commercial sugarcane. They mapped 373,869 genes or 99.1 percent of the total genome.

This feat is the result of almost 20 years of research, and will serve as a basis for the genetic improvement of the world's largest crop in tonnage according to the U.N. Food & Agriculture Organization (FAO).

An article describing the study is published in *GigaScience*. Its lead authors are Glaucia Mendes Souza, a full professor at the University of São Paulo's Chemistry Institute (IQ-USP) and a member of the steering committee for the FAPESP Bioenergy Research Program (BIOEN- FAPESP), and Marie-Anne Van Sluys, a full professor at the same university's Bioscience Institute (IB-USP) and a member of FAPESP's Life Sciences Adjunct Panel.

"It's the first time all the [genes](#) of the sugarcane plant, or the vast majority, have been seen. In previous projects by various research groups, the

sequences had to be collapsed for lack of a proper assembly tool, so they were only an approximation," said Souza.

"This knowledge opens up many possibilities, from applications in biotechnology to genetic improvement and gene editing [substitution or elimination of genes with specific functions]," said Van Sluys.

## Challenges

As the researchers explained, today's commercial sugarcane hybrids have been bred over thousands of years by crossing different varieties of two species (*Saccharum officinarum* and *S. spontaneum*) and have a highly complex [genome](#) comprising 10 billion base pairs in 100-130 chromosomes. Sequencing the genome is no easy task, requiring substantial computing power to assemble the DNA fragments while keeping homologous chromosomes separate.

For comparison, the wheat genome contains 17 billion base pairs but only 46 chromosomes, while the human genome has a mere 3.2 billion base pairs, also organized into 46 chromosomes.

Although the technology available at the start of the project was capable of producing long sequences, these long sequences had to be built from smaller fragments. Assembling the genome with these sequences required significant computing power, which was supplied by Microsoft.

The idea for the whole-genome sequencing of sugarcane dates to the onset of the BIOEN Program in 2008. A presentation by Souza at a conference held by Microsoft and FAPESP in 2014 left David Hackerman, a researcher at Microsoft Research Institute in Los Angeles, fascinated with the computational challenges posed by the initiative. He proposed a collaboration with FAPESP, which took the form of the project

"Development of an algorithm for the assembly of the sugarcane polyploid genome," with Souza as the principal investigator funded by FAPESP's program Research Partnership for Technological Innovation (PITE). The project was a collaboration with other partners, such as Bob Davidson, then a researcher with Microsoft at its Seattle unit and now with Amazon.

The sequence published has made it possible for the first time to identify the diversity in genome segments called gene promoters—DNA regions that control gene expression.

"Although in some cases the genes are 99.9 percent identical, we can detect differences in their promoters, and these help us determine which ancestor the copies derive from, *S. officinarum* or *S. spontaneum*," Souza said. The achievement permits studies, for example, of how different copies contribute to increased sugar and fiber yields and which copies may be advantageous to the different genotypes selected by programs to breed sugarcane varieties for sugar and for energy.

"The result confirms Brazil's and São Paulo state's leadership in research on sugarcane which is such an important plant for our country. It also reflects foresight on the part of the São Paulo research community and of FAPESP, regarding the challenge of learning about the sugarcane genome to extract knowledge leading to increased efficiency and productivity. We should always recall that research on sugarcane is one of the factors that enabled Brazil to achieve something no other country of a similar size has achieved to date, namely, producing 40 percent of its total energy from renewables and with low carbon emissions," said Carlos Henrique de Brito Cruz, FAPESP's Scientific Director.

## Background

The variety chosen for sequencing was SP80-3280 because more data are available about this variety in scientific literature than about any other variety. During Project Sugarcane Genome (known as FAPESP SucEST, 1999-2002), 238,000 functional gene fragments from this variety were partially sequenced.

Today, SP80-3280 ranks among the top 20 sugarcane varieties grown in São Paulo state. It is also part of the genealogy of several commercial varieties, since it is used in new crossings. Its agricultural yield is high, and it is easily regrown by the sett method (setts are stem cuttings taken from old plants containing one or more buds), thus making it an option for late harvesting at the end of the crop year in São Paulo state.

"The knowledge obtained for this variety can be applied in studies of other genotypes, particularly for the discovery of genes that control biomass accumulation," explained Augusto Lima Diniz, a coauthor of the study and currently on a research internship abroad at Cold Spring Harbor Laboratory (CSHL) in the United States as part of his postdoctoral research for IQ-USP.

Souza and Van Sluys recently participated in an international team that sequenced the genome of *S. spontaneum*, the ancestor species corresponding to 10-15 percent of the commercial sugarcane genome. *S. officinarum* contributes 80-85 percent, and 5 percent is recombinant chromosomes of these two progenitor species. The study is [published](#) in *Nature Genetics*.

In 2018, Van Sluys was one of the authors of an article on the results of a study that mapped about half of the sugarcane monoploid genome (only one chromosome in each pair).

Based on the information obtained from this latest whole-genome sequencing effort, researchers at the University of São Paulo (USP) are developing tools for the genetic improvement of sugarcane and testing several candidate genes in Genetically Modified (GM) plants. They are also conducting comparative genomics studies on large gene families with the aim of understanding their contributions to the sugarcane varieties used in Brazilian genetic improvement programs. They hope to find genes that can increase yields, enhance drought resistance and contribute to the development of novel compounds from sugarcane.

"We're also offering the community a Genome Browser that can be used to search for specific genes and analyze sequences in comparison with

previous sequencing exercises. This will be valuable to biotech projects not just relating to [sugarcane](#) but also to other crops and plants," Souza said.

**More information:** Glaucia Mendes Souza et al, Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop, *GigaScience* (2019). [DOI: 10.1093/gigascience/giz129](#)

Provided by FAPESP

APA citation: Most complete commercial sugarcane genome sequence assembled (2019, December 3) retrieved 16 May 2021 from <https://phys.org/news/2019-12-commercial-sugarcane-genome-sequence.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*