

Success is sweet: Researchers unlock the mysteries of the sugarcane genome

9 October 2018



Harvesting sugarcane or Sugarcane field: Modern sugarcane cultivars are polyploid interspecific hybrids, combining high sugar content from *S. officinarum* with hardiness, disease resistance and ratooning of *S. spontaneum*. The sequenced genome is a haploid accession AP85-441 generated by anther culture from octoploid *S. spontaneum* SES208. Credit: Ray Ming

For centuries, sugarcane has supplied human societies with alcohol, biofuel, building and weaving materials, and the world's most relied-upon source of sugar. Now, researchers have extracted a sweet scientific prize from sugarcane: its massive and complex genome sequence, which may lead to the development of hardier and more productive cultivars.

Producing the comprehensive sequence required a concerted effort by over 100 scientists from 16 institutions; the work took five years and culminated in a publication in *Nature Genetics*. But the motivation to tackle the project arose long before.

"Personally, I waited for 20 years to get this genome sequenced," said Ray Ming, a University of Illinois plant biology professor who instigated and led the sequencing effort. "I dreamed about having a reference genome for sugarcane when I

worked on sugarcane genome mapping in the late 1990s." Ming is a member of the Carl R. Woese Institute for Genomic Biology, one of a group of researchers interested in developing sugarcane and related crops to boost food and biofuel production.

The [complete genome sequence](#) was well worth the wait and the effort because of its potential to aid the effort to improve sugarcane. The sugarcane grown by most farmers is a hybrid of two species: *Saccharum officinarum*, which grows large plants with high sugar content, and *Saccharum spontaneum*, whose lesser size and sweetness is offset by increased disease resistance and tolerance of environmental stress. Lacking a complete genome sequence, plant breeders have made high-yielding, robust strains through generations of crossing and selection, but this is an arduous process relying on time and luck.

"Sugarcane is the fifth most valuable crop, and the lack of a reference genome hindered genomic research and molecular breeding for sugarcane improvement," Ming said. ". . . Sequencing technology was not ready to handle large autopolyploid genomes until 2015 when the throughput, read length, and cost of third generation sequencing technology [e.g. that developed by biotechnology company Pacific Biosciences] became competitive enough."

Why was sequencing the sugarcane genome so difficult? A naturally occurring phenomenon common in plants created a significant technical barrier. Sometime during the evolutionary history of sugarcane, its genome had been duplicated twice, resulting in four slightly different versions of each pair of chromosomes all crammed into the same nucleus together.

These events not only quadrupled the size of the genome (and therefore the sheer volume of DNA sequence), they also made highly similar

sequences from the genome wide duplication much more difficult to assemble into distinct chromosomes. Genomic DNA is typically sequenced, or read, in small, overlapping fragments, and the sequence data from those fragments become overlapping pieces of an enormous linear puzzle. As the sugarcane genome size doubled, then doubled again, this puzzle didn't just get larger; it took on repeated but not-quite-identical elements into which those many tiny pieces were difficult to correctly fit.

To conquer this challenge, the sequencing team used a technique called high-throughput chromatin conformation capture or Hi-C. This method allows researchers to discover what parts of the long, tangled strands of chromosomal DNA lie in contact with one another inside the cell. When analyzed using a customized algorithm called ALLHIC developed by the team, the resulting data served the purpose of the picture on the lid of a jigsaw puzzle box, providing a rough map of which sections of sequence most likely belonged to which chromosome.

"The biggest surprise was that by combining long sequence reads and the Hi-C physical map, we assembled an autotetraploid [quadrupled] genome into 32 chromosomes and realized our goal of allele-specific annotation among homologous chromosomes," Ming said. In other words, the researchers now knew which gene sequences belonged to each of the four variations on the original, pre-duplications genome—a much higher level of detail than they expected to attain.

With this information, the researchers could form better hypotheses about the mysteries of the sugarcane genome's evolutionary history.

Through comparison with the genomes of related species, researchers knew that at some point the number of unique chromosomes had dropped from 10 to eight. To the team's surprise, the new sequence data revealed that two different chromosomes had split apart, and all four halves had then fused to different existing chromosomes, a more complex set of events than the one they hypothesized.

How does understanding these physical changes help? Along with these large physical rearrangements within the genome come changes to the genes in the affected regions. For example, Ming and his colleagues found that the large chunks of chromosome that had been moved to new locations contained many more genes that help plants resist disease than were found in other locations.

"It resolved a mystery why *S. spontaneum* is such a superior source of disease resistance and stress tolerance genes," Ming said. "The chromosomal rearrangements are likely the cause, not the consequence of this enrichment, although the underlining mechanism of this enrichment remains to be investigated. This discovery will accelerate mining effective alleles of disease resistance genes that have incorporated into elite modern sugarcane hybrid cultivars, and subsequently the implement of molecular breeding [of sugarcane]."

The high quality of the [genome sequence](#) also allowed researchers to identify possible origins of modern sugarcane's incredible sweetness: even in the less sweet *S. spontaneum*, mutations that produced multiple copies of genes for sugar-transporting proteins have accumulated. They were also able to observe that in the hybridization between *S. officinarum* and *S. spontaneum*, the *S. spontaneum*-derived DNA sequence is scattered randomly throughout the hybrid genome.

"The ALLHIC method has already proven to be effective for the construction of the autopolyploid sugarcane genome," Ming said. He anticipates that the techniques used successfully for the [sugarcane genome](#) will also assist researchers in sequencing other complex genomes.

More information: Jisen Zhang et al, Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L., *Nature Genetics* (2018). [DOI: 10.1038/s41588-018-0237-2](https://doi.org/10.1038/s41588-018-0237-2)

Provided by University of Illinois at Urbana-Champaign

APA citation: Success is sweet: Researchers unlock the mysteries of the sugarcane genome (2018, October 9) retrieved 17 June 2021 from <https://phys.org/news/2018-10-success-sweet-mysteries-sugarcane-genome.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.