

A monoploid reference sequence for the highly complex genome of sugarcane

Mis en forme : Interligne : simple

Olivier Garsmeur¹, Gaetan Droc¹, Karen Aitken², Bernard Potier³, Marie-Anne Van Sluys⁴, Catherine Hervouet¹, Edwin van der Vossen⁵, Robert Henry⁶, Jeremy Schmutz⁷ and Angélique D'Hont¹

¹ CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), UMR AGAP, F- 34398 Montpellier, France.

³ CSIRO (Commonwealth Scientific and Industrial Research Organisation), Agriculture, St Lucia, QLD 4067, Australia.

² SASRI (South African Sugarcane Research Institute), Mount Edgecombe, 4300, South Africa.

⁴ Universidade de Sao Paulo, Sao Paulo 05508-090, SP, Brazil.

⁵ KEYGENE N.V., Agro Business Park 90, 6708 PW, Wageningen, the Netherlands.

⁶ QAAFI (Queensland Alliance for Agriculture and Food Innovation), University of Queensland, St Lucia, QLD 4072, Australia.

⁷ DOE JGI (Joint Genome Institute), Walnut Creek, CA and HudsonAlpha Institute for Biotechnology, Huntsville, AL 35801

Sugarcane (*Saccharum* spp.) is a major crop for sugar and bioenergy production. Its genome poses challenges that have not been addressed in any prior sequencing project due to its highly polyploid, aneuploid, heterozygous and interspecific genome structure ($2n \sim 12x \sim 120$, $\sim 10Gb$). We exploited the colinearity with sorghum to produce a monoploid genome sequence of sugarcane. We applied whole genome profiling (WGPTM) to 20,736 sugarcane BAC clones and anchored 11,732 of them onto the sorghum genome. A minimum tiling path of 4,660 sugarcane BAC clones that best cover the gene-rich part of the sorghum genome was selected, sequenced and assembled in a 382 Mb single tiling path (STP) of high quality sequence. A total of 25,316 protein-coding gene models were predicted on the STP, 17% of which displayed no colinearity with their sorghum orthologs. We showed that the two species, *S. officinarum* and *S. spontaneum*, involved in modern cultivars differed by their content in transposable elements (TE) explaining their distinct genome size. We also showed that they differ by a few large chromosomal rearrangements, explaining their distinct basic chromosome numbers while also suggesting that polyploidisation arose in both lineages after their divergence. This BAC-based monoploid sugarcane reference sequence represents an essential resource for genetic and genomic studies and for future whole genome sequence assembly programmes.

Mis en forme : Interligne : simple

BAC and STP sequences as well as gene annotations are available on the sugarcane genome hub (<http://sugarcane-genome.cirad.fr>).

We acknowledge members of the International Consortium for Sugarcane Biotechnology (ICSB) and Joint BioEnergy Institute (JBEI) for financial support.