2006

Methods for SNP identification and analysis in the sugarcane genome

Giovanni M. Cordeiro
Southern Cross University

Frances G. Eliott
Southern Cross University

C Lynne McIntyre
CSIRO Plant Industry

Robert J. Henry
Southern Cross University

Publication details

Cordeiro, GM, Eliott, FG, McIntyre, CL & Henry, RJ 2006, 'Methods for SNP identification and analysis in the sugarcane genome', paper presented to the Tropical Crop Biotechnology Conference, Cairns, Qld., 16-19 August.

Methods for SNP identification and analysis in the sugarcane genome

Giovanni M Cordeiro[1,3]
Frances Eliott[1,3]
Lynne McIntyre[2,3]
Robert Henry[1,3]

1 Centre for Plant Conservation Genetics, Southern Cross University, Lismore 2480, Australia
2 CSIRO Plant Industry, 306 Carmody Road, St Lucia, Queensland 4067, Australia
3 CRC for Sugar Industry Innovation through Biotechnology

Abstract

The sugarcane genome poses the challenge of being highly polyploid and requiring ingenuity to overcome impediments of straightforward genome analyses methods that are often taken for granted in simpler genomes and inbred species. **We have developed and evaluated a number of technologies to accurately locate and score the level of frequency of a SNP at any individual locus.**

1. Ecotilling

The method involves amplification of a desired region with end-labelled primers followed by digestion with the CEL I mismatch-cleavage enzyme on heteroxuplexed DNA strands. We have improved the sensitivity of the original protocol and adapted it for use with sugarcane. This method can be used for SNP discovery and genetic linkage mapping in sugarcane.

2. Mass-spectrometry

The Sequenom platform was used to score SNP frequencies at specific loci. Accuracy is to within a standard deviation of 0.26% to 4.5% (average of 2.00%, n > 1,000). We have used this information to derive possible copy number of genes as well attempt to associate SNPs with phenotypic traits. This method can be used to determine likely copy number, association mapping and genetic linkage mapping.

Reference

Cordeiro G, Eliott FG, Henry RJ (2006). Analytical Biochemistry. *In press*

SNP Discovery and Mapping in Sugarcane ESTs

through ecotilling on the ABI 3730 Capillary Electrophoresis system

The conceptual basis of QTL mapping is relatively simple but requires large numbers of well distributed markers covering the entire genome. The markers used to provide this coverage are often anonymous and their functions unknown. Utilising a candidate gene approach (assuming a gene with known or assumed function controls the trait in question), actual genes are mapped and should they map to a QTL location, then it is likely to be this QTL that directly controls the trait. Mapping ESTs or genes has however been difficult in sugarcane, but through identification of single dose SNPs in candidate genes, it is now possible to map these genes.

Table 1: Number of SNPs, Single Dose (SD) SNPs and the mapped locations for five Sucrose Phosphate Synthase gene family members. Segregation of the SD SNPs is based on 190 progeny of a genetic mapping population.

SPS Gene Family Member	Fragment length examined (bp)	No. of SNPs identified	No. of SD SNPs	SNP used in mapping	Linkage Group (LG) mapped to
I(A)	291	11	2	51	Group 4 LG31
I(B)	311	5	Note [A]		Not mapped
I(C)	307	11	7	204 210, 241	Group 2 LG8 Group 2 LG30
II	389	Note [B]	-		Additional SNPs being identified
III	415	10	2	231	Group 1 LG17
IV	337	10	2	267	Group 6 LG122
V	260	5	1	174	Group 5 LG33

Notes: [A] Unable to determine clear segregation; [B] Unable to ecotill fragment.

With our modified ecotilling protocol (Cordeiro et al., 2006) adapted for use with Capillary Electrophoreses systems, we are able to discover and map with relative ease, SNPs found in candidate genes believed to be associated with traits of agronomic importance.
Table 1 gives an indication on the total number of SNPs identified in the respective fragment lengths and the number of SD SNPs. Where SD SNPs occur on separate homo(eo)logous alleles, these can be mapped to their respective loci.

Unveiling Gene Copy Number in Sugarcane

with mass-spectrometry on the Sequenom system

Whilst base proportions of a single SNP tell little about the copy number of the gene locus it occurs on, frequencies derived from several SNPs at a gene locus are able to indicate the likely copy number of the gene. This information can be utilised to predict the possible haplotypes of the gene present, which is more important for predicting individual phenotypes than are the underlying SNPs.

By determining the possible ratios represented by the frequency scores and by using a combination of multiple SNP loci from a homo(eo)logous locus, it becomes possible to determine the likely copy number of the gene based on the recurrence of a common possible copy number across the assayed loci. For example, the base frequency call of A59.3 : G41.7 for the genotype Mida has two possible ratios, 6:4 or 7:5 that represent possible copy numbers of either 10 or 12. Repeating this deduction across the assayed loci, the possible copy numbers for Mida would range between nine and 12, with 12 being consistent across the six SNPs, indicating this to be the most likely copy number for this locus.

Knowledge of the number of homo(eo)logous loci will assist in the deduction of the allelic composition of the locus in any particular sugarcane genotype.

SD SNPs identified through ecotilling can also be mapped using mass-spectrometry.

Table 2. Variation of SNP base frequencies from six SNP loci in a sugarcane EST contig and the deduced likely copy number of the contig in the genotype Mida.

Marker	Base frequency Score(%:%)	Possible ratios	Possible copy numbers
Genotype: Mida			
crcSNP5938-A1663	A59.3:G41.7	6:4 or 7:5	10 or 12
crcSNP5938-G1776	T58.0:G42.0	6:4 or 7:5	10 or 12
crcSNP5938-T2026	C33.1:T66.9	3:6 or 4:8	9 or 12
crcSNP5938-C2083	C34.8:G66.2	3:6 or 4:8	9 or 12
crcSNP5938-C2338	C57.0:T43.0	8:6 or 7:5	14 or 12
crcSNP5938-C2377	C59.3:T40.7	6:4 or 7:5	10 or 12
	Likely copy number		12