Available online at www.sciencedirect.com
SCIENCE DIRECT
Field Crops Research 92 (2005) 137–147
Field Crops Research
www.elsevier.com/locate/fcr

Genomics approaches for the identification of genes determining important traits in sugarcane

Rosanne E. Casu [a,d,*], John M. Manners [a,d], Graham D. Bonnett [a,d], Phillip A. Jackson [b,d], C. Lynne McIntyre [a,d], Rob Dunne [c], Scott C. Chapman [a,d], Anne L. Rae [a,d], Christopher P.L. Grof [a,d]

[a] *CSIRO Plant Industry, Queensland Bioscience Precinct, 306 Carmody Road, St. Lucia, Brisbane, Qld 4067, Australia*
[b] *CSIRO Plant Industry, Davies Laboratory, Private Bag, Aitkenvale, Townsville, Qld 4818, Australia*
[c] *CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW 1670, Australia*
[d] *Cooperative Research Centre for Sugar Industry Innovation through Biotechnology, Australia*

Abstract

Sugarcane is a genetically complex polyploid grass, which makes the identification of associations between genes and traits difficult. Genomics science facilitates characterization of entire eukaryote genomes at the DNA sequence level, but for crop plants with complex genomes such as sugarcane, gene characterization is currently best achieved via expressed sequence tag (EST) analysis where sequence information is restricted to genes that are actually functioning in a particular tissue or situation. DNA microarrays allow simultaneous expression analysis of thousands of genes. Current work on EST and array analysis of gene expression in sugarcane is reviewed and insights on stem functions associated with maturation and sucrose accumulation are discussed. A strategy for associating gene expression with a trait is described in which individuals exhibiting particular traits are selected from segregating populations of sugarcane and their gene expression profiles compared. A preliminary experiment to test the feasibility and experimental design for this 'genetical genomics' strategy on a population segregating for sugar content is described. Given the complex genetics of sugarcane, this strategy and refinements of it, represent an attractive pathway to the identification of candidate genes that may control sugar accumulation and other traits in sugarcane.

Keywords: Sugarcane physiology; Sugar metabolism; EST sequencing; DNA markers; Microarrays; Genetical genomics

1. Introduction

The physiology of a plant is determined by the interaction of its genome and its internal and external environment. Plant breeding manipulates the genome to develop new cultivars with desirable attributes. Genetic control of physiological 'traits' is traditionally studied by analysing the segregation of the trait of interest in conjunction with other phenotypic, biochemical and molecular markers in the progeny of crosses. In some diploid plants, such studies have led ultimately to the isolation of the DNA sequences of

* Corresponding author. Tel.: +61 7 3214 2364; fax: +61 7 3214 2950.
E-mail address: rosanne.casu@csiro.au (R.E. Casu).

0378-4290/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.fcr.2005.01.029

genes that control specific plant traits (Fridman et al., 2000). The identification of genes that determine economically important plant traits provides important tools to further manipulate plant function and performance, either through enhanced conventional breeding using the gene's DNA sequence as a 'perfect' marker for trait selection or through manipulation of the gene and trait in transgenic plants.

Rapid advances in techniques for DNA sequence analysis coupled with powerful computer-driven tools for managing this information have allowed sequencing and analyses of the genomes for *Arabidopsis thaliana* and rice. Genomics represents a new paradigm for scientific research. Traditionally, the starting point for studies on genes that encode a particular function has been the demonstration that a particular trait is inherited, with the responsible gene(s) then sought either via its protein product, expression pattern or genetic map position. In genomics science, the starting point is the assembly of information on all the genes in the genome at the level of DNA sequence analysis. This resource can then be further studied either in silico, or by the systematic application of functional tests to identify genes that determine particular physiological processes. The shift in emphasis is that a major resource of information on all plant genes is developed first; this resource can then be used repeatedly to address a range of diverse hypotheses. The advantage of such an approach is a reduction in assumptions that are made when identifying genes responsible for traits, the disadvantage is the substantial costs associated with the up-front development of the resource. Systems biology (Moore, 2005) is an extension of the genomics paradigm where information about the relationships among genes and pathways is integrated across multiple levels from molecule to crop.

One approach to studying genes expressed at the transcript or mRNA level of an organism is to isolate cDNA copies of the mRNA in specified tissues and undertake DNA sequence analysis of these copies. Sequences obtained in this manner are termed expressed sequence tags (ESTs) and represent the genes functioning in the tissues at the time of sampling. However, to capture all expressed genes very large EST collections must be obtained from multiple tissues and treatments. ESTs thus identified can be compared to databases of DNA sequences from other organisms, and putative functions for each EST assigned via its homology with known genes. Substantial technical breakthroughs in the development of DNA microarrays or DNA 'chips' allow the study of the expression of thousands of genes in parallel (Schena et al., 1995). Typically, DNA or oligonucleotide copies of mRNA sequences are spotted or synthesised in high-density arrays on slides. Two or more RNA samples are labelled, each with a specific fluorescent tag, and these are simultaneously hybridised to the slide. The relative intensity of the fluorescence at each cDNA spot is measured to compare the relative expression levels of transcripts for the gene represented by the spot and enables development of a comprehensive catalogue of information on where genes are expressed and their relative levels of expression.

The application of the genomics research approach to sugarcane involves particular challenges (Grivet and Arruda, 2002). Sugarcane has the most complex genome of any crop (D'Hont et al., 1996) and commercial sugarcane plants are the result of a limited series of crosses and backcrosses derived from the domesticated species *Saccharum officinarum* L. ($2n = 80$) and the wild species *S. spontaneum* ($2n = 40$–120). Commercial sugarcane plants are interspecific poly-aneuploid hybrids with chromosome numbers usually in excess of 100. Most traits are polygenic and/or multi-allelic, and quantitatively inherited. Furthermore, the size of the genome in commercial sugarcane is large (~3000 Mbp) compared to that of model plant systems such as rice (430 Mbp) (Tomkins et al., 1999). While it is unlikely that the genome sequence for sugarcane will be determined in the foreseeable future, rapid progress has been achieved through the characterisation of the transcriptome of sugarcane using EST collections and gene expression analysis via DNA arrays.

In this article we briefly review what has been achieved in genomics research in sugarcane, focussing on transcriptome analysis in particular, and how it may contribute to our knowledge of sugarcane physiology and be used to facilitate sugarcane breeding. We describe our own work in the development of EST collections from stem tissue and the bioinformatic analysis of these collections leading to