GENETIC AND EXPRESSION PROFILING IN SUGARCANE

By

ROSANNE CASU, CHRISTINE DIMMOCK, MERLE THOMAS, NEIL BOWER, DEON KNIGHT, CHRISTOPHER GROF, LYNNE McIntyre, PHILLIP JACKSON, DAVID JORDAN, VICKI WHAN, JANNEKE DRENTH, YUEZHI TAO and JOHN MANNERS

CSIRO Plant Industry, 120 Meiers Rd, Indooroopilly QLD 4068, Australia

Abstract

Commercial varieties of sugarcane accumulate up to 50–60% of the dry weight of stem tissues as sucrose. The accumulation of sucrose occurs in a developmentally programmed tissue-specific manner with a sharp transition from active growth and elongation to sucrose storage functions occurring over a few internodes. We are interested in identifying and characterising the genes that determine the development and function of the sucrose accumulating tissues. Eventually, this may facilitate the manipulation of sugarcane and other plants for enhanced sugar concentration. A dataset of 8504 ESTs has been obtained from two cDNA libraries, one derived from immature stems (YCS-1085 sequences) and one from maturing stems (MCS-7419 sequences) of the sugarcane cultivar Q117. In the YCS library, 59.4% of sequences are homologous to genes of known function, 19.4% are homologous to anonymous genes and 21.0% have no known homologue. In the MCS library, 50.5% are homologous to genes of known function, 26.5% are homologous to anonymous genes, and 23.0% have no known homologue. Expression profiles are being performed on a non-redundant set of ESTs and attempts are being made to correlate expression with the onset of sucrose accumulation and with genotypic variation. ESTs of interest and other relevant markers are currently being mapped in both sugarcane and sorghum populations segregating for levels of sucrose accumulation and other traits of interest.

Introduction

Production of new varieties of cereals and sugarcane using conventional breeding techniques takes many years. New technologies have been developed which significantly enhance our understanding of plant genetics and the efficiency of plant improvement. These include the use of molecular markers to map plant genomes and to facilitate map-based cloning of agriculturally important genes. Mapping of the grass genomes has revealed significant synteny in gene order across the genomes of grass species. This makes grasses a cohesive biological group for integrated genetic analysis. For example, advances in understanding the genome of rice can lead to significant advances in other more difficult grass crops such as wheat and sugarcane. Markers also accelerate conventional selection processes. The application of genetic mapping tools for gene isolation and marker-facilitated breeding has been most successful with simply inherited traits.

The recent development of high throughput genome sequencing has made it possible to obtain the nucleotide sequence of entire organisms and to measure the expression of all of its genes, instead of analysing genes one or a few at a time. While this comprehensive information is only available for a few organisms, it is possible to obtain large amounts of nucleotide sequence information on expressed genes from any plant tissue, including those from major commodity crops such as wheat and sugarcane. The advent of genomic technologies, coupled with advanced bioinformatic analyses of laboratory and field data is providing a paradigm shift in the research approach to key traits of agricultural importance. In particular, it allows complex traits with a multitude of interconnected pathways and processes to be studied at the gene level.

Sugarcane is genetically complex and it is not feasible to sequence its entire genome. With resources available in Australia, we have endeavoured to undertake a focussed EST-based genomics approach to sugarcane research. Our research involves utilising both transcriptome- and genome-based approaches to identify a suite of promising EST candidates from which genes, gene networks and genomic regions with critical roles in targeted agricultural processes, e.g. sucrose accumulation and pest and disease resistance in sugarcane, can be identified and characterised. This paper summarises progress to date in three key areas: construction and characterisation of a focussed sugarcane EST database; expression profiling of sugarcane using microarrays; and genetic identification of genomic regions of sugarcane associated with key agricultural processes.

Materials and methods

Library construction and characterisation

cDNA library preparation

Maturing stem mRNA was isolated from a pool of total RNA isolated from each of internodes 6 to 11 from a maturing cane stem taken from variety Q117. Immature stem mRNA was prepared in a similar
fashion from meristem and internodes 1–2 of another maturing cane stem also taken from variety Q117. Both plants were grown at the CSIRO field station at Samford, Queensland. Both the maturing cane stem (MCS) and immature cane stem (YCS) cDNA libraries were prepared by the directional cloning of cDNA derived from the appropriate mRNA in the bacteriophage vector λZIPLOX (Life Technologies).

**Sequencing of random clones**

Plasmid DNA, derived from random recombinant colonies from the YCS library, or enzymatically treated PCR products corresponding to inserts contained within random recombinant phages chosen from the MCS library, were sequenced in 96-well plates using an ABI PRISM® BigDye™ Terminator Cycle Sequencing kit (PE Applied Biosystems), and the T7 sequencing primer according to the manufacturer’s directions. Gel separations of reaction products were performed on a PE Applied Biosystems DNA sequencer (at CSIRO or the Australian Genome Research Facility).

**Analysis of random clone sequences and database construction**

**Sequence editing and homology searching**

Sequence editing and analysis for the MCS library sequences were performed using programs accessed at the Australian National Genome Information Service (ANGIS). The YCS library sequences were edited manually using equivalent software on a Macintosh computer (Sequencher). Each sequence was then submitted to the National Center for Biotechnology Information (NCBI) for homology searching against the non-redundant DNA, EST (blastn) and protein (blastx) databases (Altschul et al., 1990). Information relating to the best match from each blast search for a particular sequence was tabulated.

**Database construction, sequence annotation and cluster analysis**

A relational database was constructed using FileMaker Pro 4.0 (Claris Corporation) to organise the data generated from the sequence editing, homology searching, clonebank information and many additional annotations. The significance of the best match from each homology search was assessed by evaluating the score and p(N) and, if necessary, manual inspection. Based on this information, each sequence entry in the database was annotated for possible function and/or role, as well as many additional parameters. Direct access to the sequence and the relevant homology searches, which are all held at ANGIS, is also available from within the database. All the ESTs sequenced were clustered using gcphrap and gap4 (at ANGIS). This indicated the degree of redundancy in the EST collection and aided discovery of multi-allele/gene variants. All singleton ESTs and the most 5’ member of any cluster of ESTs were designated as members of the ‘non-redundant’ set (NR set—5866 ESTs in all).

**Expression profiling in sugarcane**

**Tissue choice and RNA isolation**

For initial technique refinement and stem profile experiments, meristem and internodes 1 to 11 were harvested from 12-month-old field-grown variety Q117. For varietal and species comparisons, meristem and internodes 1–3, internode 4, internode 8 and internode 11 were harvested from 12 month old field grown varieties 74C42 and Q124, and species Saccharum officinarum Badilla and S. spontaneum Mandalay (all grown at Samford, Queensland). For progeny comparisons, meristem and internodes 1–3, internode 4, internode 8 and internode 11 were harvested from 20 progeny from the cross Q117 × MQ77-340 (all grown in North Queensland). All samples were stored separately at −70°C until required for total RNA extraction.

**Preparation of microarrays, hybridisation, scanning and data analysis**

A sub-set of the NR set was chosen for use on the initial set of microarrays to be prepared. The criteria used were: 1. The first 400 ESTs sequenced from the MCS library; 2. All of the non-redundant sugar metabolism-related ESTs not covered by criterion 1; 3. All of the non-redundant signal transduction-related ESTs not covered by criterion 1; 4. A selection of the non-redundant ESTs related to gene expression and RNA metabolism (534 ESTs in all). Six control DNAs were also included. All inserts were amplified by PCR. The microarrays were prepared by spotting the purified PCR products on amino-silane-treated slides (Corning or Telechem Superaldehyde) with 3 replicates of each EST and control DNA on a GMS 417 arrayer (Affymetrix) at the Queensland Institute of Medical Research (QIMR) or on an SDDC-2 MicroArrayer (ESI) at CSIRO Plant Industry, Canberra.

Probes were prepared by the labelling of purified total RNA using an arrayTRACKER™ kit (Display Systems Biotech) and either Cy-3 or Cy-5-labelled dUTP (Amersham) according to the manufacturers' directions. Two reactions (one labelled with Cy-3 and other with Cy-5) were mixed, precipitated and resuspended in hybridisation buffer as recommended by the manufacturers of the slides. Hybridisation and washing procedures were performed as recommended by the manufacturers of the slides. Slides were scanned with a GenePix 4000A scanner (Axon Instruments) and analysed using GenePix Pro 3.0 (Axon Instruments) and Microsoft Excel.

**Genetic analysis of genomic regions in sugarcane**

**Plant material**

A progeny population from the cross (made by Dr Nils Berding, BSES, Gordonvale, Australia) was evaluated with four replications at two sites over two years and evaluated for 8 traits: CCS, brix, pol, fibre, moisture, stalk number, suckering and stem wax, using standard BSES protocols.
Molecular markers

DNA was isolated and RFLP analysis performed using the methods described in the Laboratory Protocols Manual developed by CIMMYT (Holsington, DA. Laboratory Protocols, CIMMYT Applied Molecular Genetics Laboratory, Mexico, D.F.). RFLP probes were obtained from Dr Angelique D’Hont, CIRAD (SSCIR clones) and from Dr Andrew Paterson, UG (sugar metabolism probe set). All techniques relevant to the RAF (radio-labelled amplified fingerprint) were as developed by Dr Bernie Carroll (UQ).

Linkage analysis

All progeny were pre-screened to ensure that they were progeny of the Q117 × 74C42 cross. One hundred and eight progeny were screened for the presence or absence of each RFLP or RAF marker band. Markers that segregated 1:1 (single dose restriction fragments—SDRFs) were identified and used for linkage analysis using JoinMap (Stam, 1993). The CP (cross pollination) option with a lod threshold of 4.0 was used. Single factor analysis was also applied to the data set using the Kruskal-Wallis test option.

Results and discussion

cDNA sequencing

8504 sequences (all sequenced from the 5’ end of each insert) were generated with 1085 sequences being from the YCS library and 7419 sequences from the MCS library. The average read-length after vector and quality clipping was 483 bp and 508 bp for the YCS and MCS libraries respectively, indicating that the quality of sequencing was uniform.

Assembly results

A total of 1177 clusters were formed after assembly of all 8504 ESTs. Each cluster consisted of at least two ESTs and was assumed to be derived from the same gene due to the stringent clustering conditions (verified by random manual inspection). The clusters contained a total of 3905 sequences, leaving 4599 ESTs as singletons having no significant homology to other sequences in the collection. This represented a redundancy of 45.9%, indicating that further gene discovery is viable from these libraries.

Similarities

Many ESTs in both cDNA libraries were assigned a keyword relating to possible function or role on the basis of sequence similarity to either proteins or corresponding genes of known function. Each query sequence was submitted for homology searching using the BLAST algorithm. The best match to each query from the NR protein, nucleic acid and EST databases following a BLAST search was assessed in order of significance using the score and p(N) as a guide and the results of cluster analysis. Using these criteria, 59.4% and 50.5% of the ESTs from the YCS and MCS libraries respectively were assigned a putative identity. These percentages were not significantly different from those achieved from other plant tissue libraries e.g. poplar cambial region and developing xylem libraries (63% and 54%—Sterky et al., 1998), loblolly pine developing xylem library (59%—Allona et al., 1998) and Medicago truncatula root hair library (59%—Covitz et al., 1998). 762 proteins or protein families with known function were identified across both libraries. Of these, 188 were found in both libraries, 516 only in the MCS library and 58 only in the YCS library. The remainder of the ESTs had either no significant similarity to any DNA or protein sequence in the public databases (21.0% and 23.0% for the YCS and MCS libraries, respectively) or were significantly similar to ESTs or genomic sequences with no known function (19.4% and 26.5% for the YCS and MCS libraries, respectively). It is possible that many of these unknown ESTs represent genes that are specifically involved in processes unique to sugarcane and, more importantly, unique to the stem. Functional analysis of these genes may disclose novel regulatory controls relevant to the process of sucrose accumulation.

Those ESTs with putative protein or DNA identities were classified into 16 functional groups (Figure 1). In the YCS library, most of the identified genes are associated with categories such as chromatin and DNA metabolism, cell wall structure and metabolism, and protein synthesis and processing. This suggests that the young cane stem is more dynamically involved in growth. In the MCS library, additional categories e.g. fibre biosynthesis and degradation, and defence and stress-related proteins gained greater significance while others e.g. chromatin and DNA metabolism, and protein synthesis and processing diminished in importance. The proportion of sugar metabolism-related transcripts was low for both libraries, but there was a greater abundance of such transcripts in the MCS library. 80 sugar metabolism-related transcripts were found in all, five from the YCS library and 75 from the MCS library. Only one EST type, sucrose synthase, was present in both libraries.

Expression profiling

Early results achieved with the comparison of meristem to internode 8 of sugarcane stem (both derived from Q117) indicate there is a significant difference in expression for a variety of ESTs between these two tissues. For 1602 data points (534 ESTs spotted in triplicate), 1434 points registered significant signal with 413 points exhibiting a greater that 2.5 fold change in expression. Of these, 51 data points were up-regulated and 362 were down-regulated in internode 8 when compared to meristem. Further analysis will reveal a final set of candidate genes from this experiment.

Genetic analysis of genomic regions in sugarcane

Genetic mapping

To date, approximately 300 markers have been mapped on 108 progeny of the Q117 × 74C42 population. These markers comprised 59 RFLP markers, derived from 26 probes, and more than 200 RAF
markers from 10 RAF primers. Of these, 163 were SDRFs. These markers formed approximately 30 co-segregation groups with approximately 70 markers remaining unlinked. The linkage groups varied in size from 2 to 8 markers. The 30 co-segregation groups could be reduced to 24 homology groups, using common RFLP probes, enabling many of the homology groups to be tentatively aligned with other sugarcane and sorghum maps. The number of linked markers is very small and much more work is required to obtain a useful skeleton map of an Australian sugarcane cross.

**Linkage analysis**

Single factor analysis was undertaken with all of the markers generated. Linkage was found between markers, both RFLP and RAF, for all of the traits measured. Many markers were significantly associated with a trait across both sites and years (plant and ratoon crop), while other markers were associated with a trait at one site for both years or at one site in a single year. This variability is not unexpected, due to the small population size.

Twelve markers were found that were significantly associated with one or more traits across both years and both sites. These ‘robust’ markers explained varying percentages of the variation in a trait in the population and were both positively and negatively associated with each trait.

In general, each marker explained only small amounts of variation in the sugar-related traits (up to 10%), suggesting QTLs of individually small effect. This is consistent with other studies (AH Paterson—ICSB Final Report; L. Grivet—pers. comm.). For Brix, 3 markers with effects of 8–10% were identified. For Pol, 5 markers each explained more than 5% of the variation, while for CCS, 3 markers explained 5–7% of the variation. Markers strongly associated with one trait were usually strongly associated with the other 2 traits.

Five robust markers are associated with moisture content and explained 6–11% of the variation. For fibre, three markers were identified which explained 5–16% of the variation. For stalk number, 4 markers explaining 7–17% of the variation were identified, while for suckering, 3 markers explaining between 8% and 20% of the variation were found. Two markers each explaining 9% of the variation were found for stem wax.

Many markers were significantly associated with more than one trait. Given the obvious correlation between traits, this is not surprising. For example, markers associated with a positive effect on moisture, were usually associated with a negative effect on sugar-related traits and on fibre, and vice-versa. Several markers associated with stalk number were also (but not always) associated with suckering.
We have attempted to position many of these markers onto our skeletal sugarcane map. Although the map is sparse, our preliminary analysis indicates at least 2 different homology groups are associated with each trait.

**Future activities**

The major question to be asked is ‘Can high CCS (commercial cane sugar) or the onset of sucrose accumulation be correlated with a suite of up-regulated or down-regulated ESTs?’ This is being addressed in a variety of profiling experiments using large-scale microarrays. These include: (i) stem profiles on RNA from tissue derived from variety Q117; (ii) comparison of RNA from matching internodes from various commercial varieties and progenitor species; and (iii) comparison of RNA from matching internodes of members of a family displaying extreme segregation for CCS. The results of these experiments will indicate for which EST a change of regulation can be correlated with the onset of sucrose accumulation and in which internodes these occur, thus providing a suite of candidate ESTs implicated in sucrose accumulation.

Future expression profiling activities targeting traits other than sucrose accumulation include suckering analysis in sugarcane, pest and disease resistance mechanisms in sugarcane and use of sugarcane EST microarrays as reagents for investigation in other crops.

The preliminary mapping and linkage analysis are encouraging but require confirmation with additional progeny and markers. The use of additional reference markers will enable comparison of this map to other sugarcane maps. Markers identified in this population require validation in other populations to confirm marker utility and utility of the tagged QTL in different backgrounds. Screening of the ICSB Sugarcane Microsatellite Consortium microsatellites has been completed and their mapping has commenced. This will increase both the number of markers and reference points.

A candidate gene mapping approach has also been initiated. ESTs from categories relevant to traits segregating in this population have been identified from the Sugarcane EST database detailed above and approximately 400 have been screened over the parents of the population. Progeny screening and mapping with these ESTs will commence shortly to see if any map to QTLs of interest. These ESTs will be further investigated as perfect markers in sugarcane breeding, possibly providing additional confirmation of the role of these genes in the sugarcane plant.

Given the excellent synteny between sorghum and sugarcane, the majority of our sugarcane markers are concurrently mapped in sorghum. This provides additional markers for our sorghum analyses and greatly simplifies mapping in sugarcane. More than 100 sugarcane ESTs have already been mapped in sorghum.

**Acknowledgments**

This study is being funded in part by the Sugar Research and Development Corporation. The authors thank Drs Ray Shorter and Chris Lambrides (CSIRO Plant Industry) for helpful comments on this manuscript.

**REFERENCES**


