

CPI002

FINAL REPORT

Title: **Functional genomics for enhanced sugar accumulation in sugarcane**

Project Number: CPI002

Research Organisation: CSIRO Plant Industry

Chief Investigators: Dr John Manners
Dr Rosanne Casu

Phone: 07 3214 2304

Address: CSIRO Plant Industry
Queensland Bioscience Precinct
306 Carmody Rd
St Lucia QLD 4067

The research presented in this report has been funded in part by the Sugar Research and Development Corporation.

Disclaimer:

“The Research Organisation is not a partner, joint venturer, employee or agent of SRDC and has no authority to legally bind SRDC, in any publication of substantive details or results of this Project.”

Contents

| | |
|--|------------|
| Executive Summary | 4 |
| Background..... | 6 |
| Objectives..... | 7 |
| Methodology | 8 |
| Results | 11 |
| Publications cited in this report | 21 |
| Outputs..... | 22 |
| Expected Outcomes | 23 |
| Future Research Needs | 24 |
| Recommendations | 25 |
| Publications arising from this Project..... | 26 |
| Appendix 1 | 27 |
| Appendix 2 | 44 |
| Appendix 3 | 71 |
| Appendix 4 | 100 |

Executive Summary

Improvement in CCS of sugarcane would provide considerable benefits to the whole sugar industry by improving profitability via enhanced efficiencies in both sugarcane and raw sugar production. Improvements in CCS bring benefits by increasing sugar input to mills with no new costs in cane growing, harvesting and transport and enhanced sugar output with only moderate changes in the sugar milling process. Despite the economic attractions of the CCS plant trait for plant improvement there has been little progress made in improving CCS in released varieties in the past forty years and new approaches are needed.

One new approach to breeding high CCS sugarcane varieties is to use DNA markers to select for diverse attributes that contribute to CCS and combine these attributes to produce improved varieties. The CCS trait is complex and involves many genes and a range of plant functions. A key contributor to high CCS is sucrose accumulation and the aim of this project was to identify sugarcane genes that are associated with high levels of sucrose accumulation. These genes provide an input to further research where the CCS trait is being mapped on the sugarcane genome and genes identified in CPI002 are tested as markers. Ultimately, benefits to growers will accrue through the use of these markers in the breeding program to select improved varieties.

In a previous SRDC-funded research project (CTA035), we created a sugarcane genebank. This is a collection of clones and a database of ~ 8,000 gene sequences that represent genes expressed in sugarcane stems, the sucrose accumulating tissue. In CPI002, this resource was used to create a new tool, termed cDNA microarrays, which allow measurement of the expression of thousands of genes in plant tissues simultaneously. This permitted the identification of sugarcane genes that were expressed differentially in 1. Maturing stems accumulating sucrose in comparison to immature stems that were not accumulating sucrose in three high sucrose accumulating sugarcane clones and 2. Maturing stems of progeny with high CCS in comparison to maturing stems of progeny with low CCS selected from a segregating population of sugarcane. This allowed the identification of genes with expression patterns that correlate with high sucrose accumulation. A sub-set of the identified genes was examined further for cellular localisation and function of their products. Sugar transporters have been localised to the phloem sieve elements, parenchyma surrounding the vascular bundles and sucrose storage parenchyma. Sugar transporters appear to have a critical role in transporting and accumulating sugar in the sugarcane stem and a model for the transport pathway was developed as a result of the findings in the project.

This project is strategic research and its outputs will feed into further research leading to improved varieties via breeding. The following briefly describes the outputs and their progress along respective delivery pathways to an outcome.

- Sugarcane genes that correlate either developmentally or via genetic association with high sucrose accumulation have been identified. The project has demonstrated the potential of gene expression profile as a marker for high CCS and this represents a new selection tool in sugarcane breeding. Genes associated with high CCS have been passed on to other SRDC projects (e.g. CTA047, CTA049) and an ACIAR project for genetic mapping studies, and to Southern Cross University for sequence-specific marker development. Research will continue on these genes within the new CRC for Sugar Industry Innovation through Biotechnology (CRCSIIB) for assessment whether these genes will correspond to loci for high CCS.

- Sugarcane genes that are expressed in particular cell-types such as parenchyma cells surrounding vascular bundles, lignifying parenchyma within vascular bundles, sieve elements and storage parenchyma were identified. There has been considerable interest in using sugarcane as a biofactory for the production of high-value products. To do this without disrupting plant function it is necessary to target metabolic engineering to specific cells. The genes identified in this project are an essential starting point towards this aim and have provided a platform for SRDC project UQ39, another CSIRO-BSES project and for CRCSIIB projects.
- Sugarcane has unique physiology by accumulating high concentrations of sucrose in stem tissues. Results show that diverse sugar transporters are positioned in specific cell types to form a probable symplastic sugar transport from sieve elements to parenchyma cells. This model represents valuable intellectual property for manipulating sucrose accumulation and will be tested further in CRCSIIB.

Plant breeding in the future will be increasingly influenced by our knowledge of the genes that underpin valuable plant functions and our ability to manipulate these genes via either enhanced selection or genetic modification. Gene technologies are proprietary allowing particular industries to gain competitive advantage over each other. The Brazilian industry has a strong advantage via its large genebank for sugarcane. However, in this project we believe that we have obtained a significant technological advantage for the Australian sugar industry in the identification of genes associated with enhanced sucrose accumulation, a key physiological trait for sugarcane.

Background

The performance of sugarcane has been greatly enhanced through plant breeding. These gains have mostly been made by increased crop yield, but over the past 30-40 years, there have been few gains in CCS. A “Whole of Industry” analysis of crop traits in SRDC project CTA041 has indicated that increased CCS would be the most important trait for economic gains in the sugar industry as a whole. Some regions, e.g. tropical north, have experienced a run of years with extremely low CCS in the late 1990s which threatened the viability of the industry in these regions. Therefore, strategically, gene technologies for the manipulation of CCS in plant improvement programs either as DNA markers or as transgenes, have high value for the Australian industry, making it more profitable and robust in bad years.

In the past decade, genomics has emerged as the most powerful technology for the discovery of genes important in plant function. It is based on knowledge of the DNA sequence of the genome. Sugarcane is genetically complex and DNA sequencing of the sugarcane genome is not feasible at present. Structural genomics approaches in sugarcane have focussed on developing “expressed sequence tags” (ESTs) to identify the expressed gene sequences from particular tissues. In CTA035 and other associated research in CSIRO we developed an EST collection of at least 8324 gene sequences from immature and maturing stem tissue. At the start of this project there was no experimental information on which of these genes have utility for the manipulation of sucrose accumulation.

The other significant program developing sugarcane ESTs is being undertaken by a consortium of university researchers in Sao Paulo State in Brazil. The Brazilian program commenced in July 1999 and aimed to obtain sequence information on 50,000 unique genes of sugarcane. At this time, little sequence from this program has been released other than some ESTs containing microsatellites and some resistance gene analogues.

In the research in Australia our EST collection had focused on sugarcane stem because of the key role this tissue plays in the sucrose accumulation process. There are many genes in the SRDC/CSIRO EST collection that have potential roles in determining sucrose accumulation. Experimental evidence of function and utility on these genes would be required for IP protection.

Objectives

The aims of this research project were:

1. To identify candidate genes that potentially determine sucrose accumulation by correlating gene expression with sucrose levels during development and maturation, and across varieties and progeny populations varying in CCS.
2. To provide experimental evidence of the utility of the candidate genes identified in 1. By, a) testing their use as DNA markers for CCS and early CCS and b) testing whether the manipulation of these candidate genes in sugarcane cells leads to enhanced sucrose accumulation.
3. To undertake, together with SRDC, appropriate protection of the intellectual property developed in 1. and 2. above. This will provide freedom to use this IP in the Australian sugar industry and the potential to leverage other IP that may benefit the sugar industry.
4. To develop and promote strategies for the exploitation of this IP in sugarcane plant improvement programs.

These objectives have been achieved. A number of candidate genes have been identified through the use of bioinformatics and microarray-based transcriptional profiling particularly from the comparison of transcript expression in immature stem and maturing stem tissue. The ESTs developed in this project have been used extensively in joint research (CTA046) between BSES and CSIRO in the development of molecular markers. In addition, many of the ESTs, including those involved in carbohydrate metabolism, have been mapped in sorghum and have been assigned to specific linkage groups. The benefit of this to the sugar industry is that sorghum and sugarcane show great synteny (similarity) in both DNA sequence and gene order on chromosomes. This data will be useful in strategies to develop markers for sucrose accumulation. Transcript and protein localisation information is now available for three sugar transporters and two of these transporters, PST Type 2a and ShSUT1, have been functionally tested. These transporters have been found to transport hexoses and sucrose, respectively.

PST Type 2a was placed under a provisional patent during the project but, in consultation with SRDC, it was determined that there was no advantage in proceeding to a full patent. No other EST has been patented at this time. At present, 197 ESTs and one full-length sequence (PST Type 2a) have been lodged in GenBank and are publicly available. It is anticipated that the remainder will be lodged at GenBank during the coming year. This should not preclude the development of new IP based on these sequences.

The EST collection and associated microarray data has been accessed by other parties within the Australian Sugar Industry research community as part of the development of collaborative projects. It is also a major input in the CRCSIIB and the CSIRO/BSES Joint Venture.

Methodology

The following steps were followed during the course of this project in order to achieve the agreed milestones.

1. Identification of a set of sugarcane ESTs for inclusion on a sugarcane transcript microarray (CaneArray) and production of the CaneArray.
2. Isolation of RNA from selected plant tissues
3. Microarray hybridisation
4. Statistical analysis
5. Candidate gene identification
6. Tissue localisation studies of selected genes
7. Functional analysis of selected genes

Production of the CaneArray and microarray hybridisations

A description of all biological resources and most methods has been presented in Casu et al. (2003 – Appendix 1), Casu et al. (in prep – Appendix 2) and Rae et al. (in prep – Appendix 3). In brief, random clones from immature stem and maturing stem cDNA libraries were subjected to single pass sequencing, with the trace files being edited and the extracted sequences then analysed by homology searching of the non-redundant (NR) DNA, EST (both BLASTN) and protein (BLASTX) databases (Altschul et al., 1990) located at NCBI in June 2002. All ESTs were extensively annotated for possible function and/or role by a combination of automated filtering and manual inspection, and were also clustered into contigs using gcphrap (<http://www.phrap.org/> - deviation from default settings: gap penalty = 15, shatter_greedy, a bandwidth of 30 and a minimum score of 100) with all of the results being captured in a custom database constructed in FileMaker Pro (FileMaker, Inc., Santa Clara, CA). All of the sequencing and the first round of homology searches and concomitant annotations were performed as part of CTA035 and associated CSIRO research. However, in order to maintain the relevance of the annotations, additional periodic homology searching and reannotation, as necessary, is performed (the last being the searches detailed above). Individual sequence analysis was performed using MacVector 7.0 (Accelrys Inc., San Diego, CA, USA), and multiple sequence alignment was performed using the CLUSTALW algorithm as implemented in MacVector 7.0. Microarrays were prepared by spotting purified PCR products corresponding to 4715 ESTs derived from either immature stem or maturing stem in duplicate. Hybridisations, washing, and array visualisation were performed as described in Casu et al. (2003).

Microarray data analysis

Microarray data was analysed using tRMA (Wilson et al., 2003), a set of statistical functions implemented in R code (Ihaka and Gentleman, 1996; <http://www.r-project.org/>). A variation of the procedure described in Klok et al. (2002) was used. In brief, the raw gpr files were imported into tRMA, spots for which data was absent, empty or deemed of insufficient quality were removed, and normalised using the SlideNormalise function (an intensity dependent normalisation function – see Yang et al., 2001). Normalised ratio values for each array were exported to Excel, a mean ratio computed and the ESTs ranked in descending order by the absolute value of the mean ratio (from the most differentially expressed EST to the least differentially expressed EST). Confidence limits (at $\alpha = 0.05$) and standard errors were determined for each EST and an expression trend noted. Further analysis was confined to the top 100 differentially expressed entities that possessed absolute values of normalised ratios that were equal or greater than twice the confidence limit.

Northern analysis

Expression profiles of selected differentially expressed ESTs were confirmed by Northern analysis as described in Casu et al. (2003) and Casu et al. (in prep). The preparation of tissue sections and the methodology for the performance of *in situ* hybridisations in order to localise the transcripts corresponding to various ESTs is described in Casu et al. (2003).

Immunolocalisations

As a first step in the preparation of appropriate antibodies for use in immunolocalisation experiments, a series of synthetic peptides ranging from 15 to 20 residues were designed from the predicted amino acid sequences of the putative sucrose and hexose transporters (ShSUT1, and PST2a and PST3, respectively). The peptides were carefully chosen to regions of the transporter protein not embedded in the membrane and were located either in the central loop region separating the two blocks of six transmembrane helices or the carboxyl terminus of the protein. The peptides were produced commercially by Mimotopes, Melbourne. The peptides were conjugated to diphtheria toxoid and antisera raised in rabbits by the Institute of Medical and Veterinary Science, Adelaide. Antisera raised against the sucrose transporter was immunoaffinity-purified using the synthetic peptide bound to a CNBr-activated Sepharose4B (Amersham Pharmacia) column. The other antibodies were used without further purification.

For localisation of antigens *in situ*, tissue sections are incubated with the specific primary antibody followed by a secondary antibody which recognises the primary antibody and which carries a label that can be visualised microscopically. For antibody labelling, sections were blocked for 1 h at room temperature then incubated overnight in solutions of antibody or pre-immune serum diluted 1 in 1000 in blocking buffer (3% bovine serum albumin, 0.1% v/v Tween-20 in Tris-buffered saline). After washing, sites of antibody binding were detected by incubation with anti-rabbit antibody conjugated to alkaline phosphatase (diluted 1 in 2000), and visualised by incubation with the substrates nitro-blue tetrazolium and 5-bromo 4-chloro 3-indolyl phosphate. Photographs of the labelled sections were taken with a Zeiss Axioskop transmitted light microscope fitted with a Spot digital camera.

Functional analysis

Functional analysis of the sucrose transporter ShSUT1 were performed as detailed in Rae et al. (in prep – Appendix 3). Functional analysis of the hexose transporter PST2a was performed as follows:

The *PST2a* gene from sugarcane was blunt-end cloned into the pGUN vector and was under the control of the maize ubiquitin promoter and NOS terminator. The S65T version of GFP was cloned into the same vector and used as an *in vivo* marker for transformed cells. Sugarcane callus was established from the variety Q117 as previously described (Taylor et al. 1992) and transformation was carried out via microprojectile bombardment as described by Bower et al. (1996) and Elliott et al. (1998). After at least 8 weeks growth on selection media, a small number of transformed cells were introduced into liquid culture. The components of the liquid media were identical to the plate media without the addition of agar. Cells were grown in the dark, maintained at 28°C on a revolving platform and sub-cultured into fresh liquid media every 2 weeks. Cells were grown in a volume of 100mls of media and experiments were performed between 12 and 14 days after introduction into the fresh media. Cells were incubated in sugar free medium 24 h prior to the initiation of the experiment. The cells were harvested by vacuum filtration, washed in 5 mM sodium phosphate buffer (pH 6.0) and allowed to equilibrate for 2 h at 30°C whilst shaking. Cells were resuspended at a density

of either 50 or 100 mg fresh weight per ml and shaken at 30°C during the course of the experiment. Uptake was initiated by the addition of sugar (either glucose or OMG) and measured as the depletion of sugar from the medium. Aliquots of 200 µl were withdrawn from the medium, centrifuged at 20,000 x g (4°C) for 3 min and the supernatant removed and stored at -20°C in readiness for measurement. Sugars were measured using an HPLC-PAD system.

RNA was extracted from a series of transgenic lines transformed with both the PST2a and GFP constructs and the control lines transformed with GFP alone in order to determine expression specificity.

Results

Expression profiling of sugarcane stem transcripts

Much of this work is detailed in Casu et al. (2003 – Appendix 1), Casu et al. (in prep. – Appendix 2) and Rae et al. (in prep – Appendix 3). Three microarray analyses were performed, each addressing a separate question. In brief, microarray experiments contrasting the expression of transcripts derived from sugarcane meristem and those derived from internode 8 of the maturing stem of three sugarcane varieties (Q117, 74C42 and Q124) were analysed in two separate groups. This was followed up by Northern analysis of transcript expression in various stem sections, *in situ* and immunolocalisations, and functional analysis. A separate microarray experiment was performed to examine the expression of transcripts from maturing stem of low and high CCS individuals selected from the progeny of the Q117 x MQ77-340 cross.

Question 1: Which sugarcane Carbohydrate Metabolism Genes, represented by arrayed sugarcane ESTs, were differentially regulated between immature stem (represented by meristem and internodes 1-3) and maturing stem (represented by stem tissue derived from internodes 8 and 11) in high sucrose accumulating varieties of sugarcane?

A sequence survey of the 7242 ESTs derived from the sucrose-accumulating, maturing stem revealed that transcripts for CMGs are relatively rare in this tissue. However, within the CMG group, putative sugar transporter ESTs form one of the most abundant classes observed. A combination of EST analysis and microarray and Northern hybridisation revealed that one of the putative sugar transporter types, designated PST Type 2a, was the most abundant and most strongly differentially expressed CMG in maturing stem tissue. PST Type 2a is homologous to members of the major facilitator super-family of transporters, possessing 12 predicted transmembrane domains and a sugar transport conserved domain, interrupted by a large cytoplasmic loop. Its transcript was localised to phloem companion cells and associated parenchyma in maturing stem, suggesting a role in sugar translocation rather than storage (Casu et al, 2003).

Functional analysis of PST Type 2a

Functional analysis of PST Type 2a by the expression in sugarcane suspension cell cultures was also undertaken. RNA was extracted from a series of transgenic lines transformed with both the PST 2a and GFP constructs and the control lines transformed with GFP alone. The PST 2a gene was not expressed in those lines transformed with GFP alone however in a number of lines the introduced PST 2a gene was expressed strongly (Figure 1). A select number of these lines were used for uptake experiments and it was possible to demonstrate an increased rate of glucose uptake in lines transformed with the PST2a gene when compared to control lines.

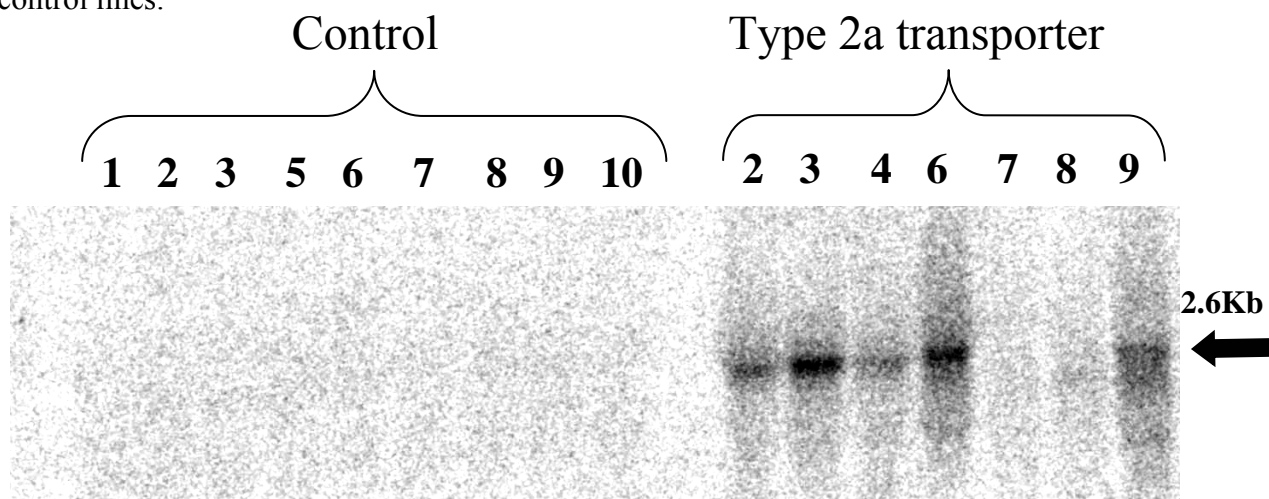
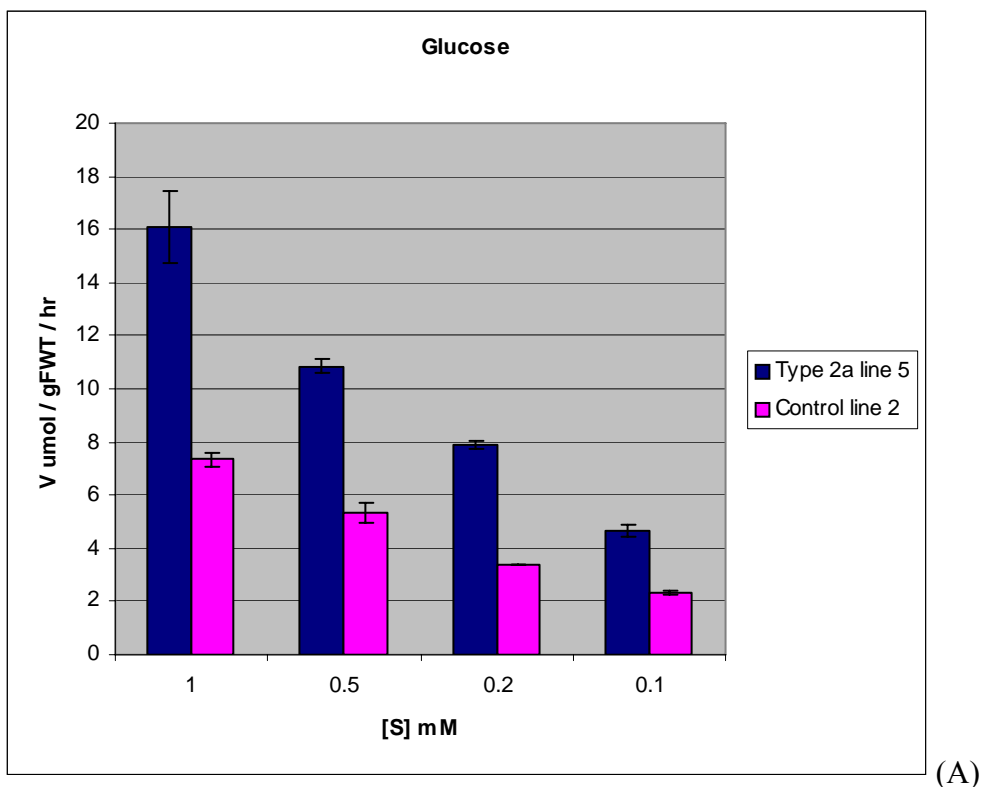


Figure 1. RNA extracted from independent cell lines transformed with GFP alone (control) and lines transformed with both GFP and the PST2a gene (type 2a transporter). The Northern blot was probed with the full-length cDNA of PST2a (C. P. L. Grof).

From additional functional assays using sugar analogues (Figure 2) it is clear that the difference in transport rates between the representative control and transformed lines remains significant even when the non-metabolised analogue of glucose (3-OMG) is the substrate presented to the suspension cells. Furthermore, this indicates that the difference in the rates of glucose uptake is not a direct consequence of differences in downstream metabolic activity between the transformed and control cell lines. This suggests that the PST Type 2a gene encodes a functional hexose transporter which is in accord with the predicted function as determined by sequence homology.



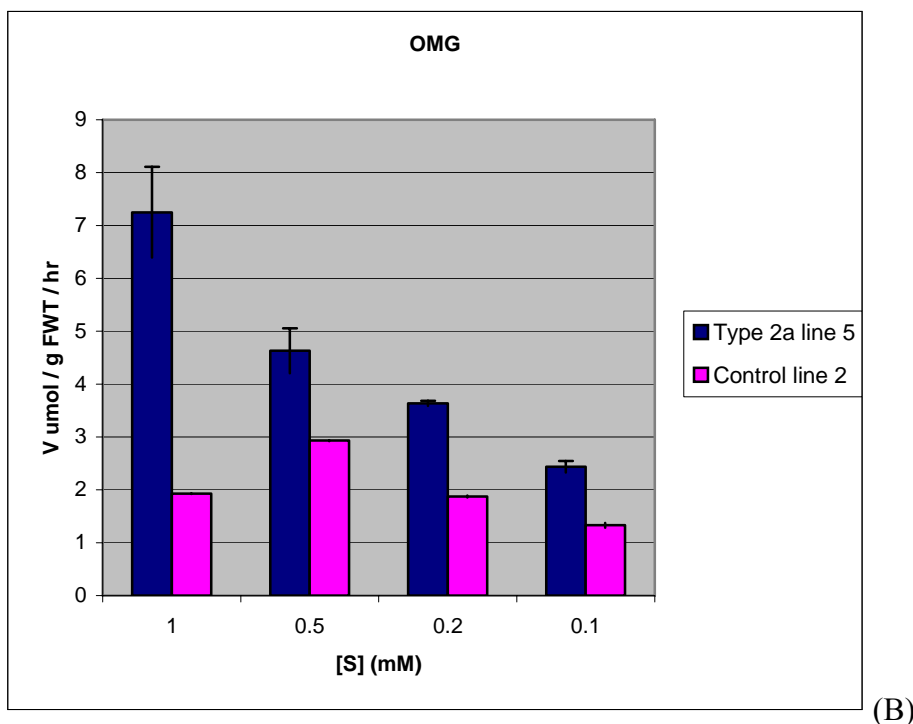


Figure 2. The rate of glucose (A) or 3-O methyl glucose (OMG) (B) uptake from the surrounding medium in a line transformed with GFP alone (control) and a line transformed with both GFP and the PST2a gene (type 2a). The cells were exposed to four different concentrations of the substrate, 0.1, 0.2, 0.5 and 1 mM (C. P. L. Grof).

Functional analysis of ShSUT1

Of the remaining ESTs that corresponded to sugar transporter homologues that were present on the array, only one EST showed a significant change in expression and this corresponded to the up-regulation of the sucrose transporter ShSUT1. ShSUT1 encodes a protein of 519 amino acids including 12 predicted membrane spanning domains and a large central cytoplasmic loop. It was demonstrated to be a functional sucrose transporter by expression in yeast. The estimated K_m for sucrose of the ShSUT1 transporter was 2 mM. *ShSUT1* was expressed predominantly in mature leaves of sugarcane that are exporting sucrose and in stem internodes that are actively accumulating sucrose. Immunolocalisation with a ShSUT1-specific antiserum identified the protein in cells at the periphery of the vascular bundles in the stem. These cells were shown to become lignified and suberised as stem development proceeds, forming a barrier to apoplastic solute movement. However, the movement of the tracer dye, carboxyfluorescein diacetate from phloem to storage parenchyma cells suggested that symplastic connections are present. ShSUT1 may have a role in partitioning of sucrose between the vascular tissue and sites of storage in the parenchyma cells of sugarcane stem internodes (Appendix 3 - Rae et al, in prep).

Isolation and molecular characterisation of additional putative sugar transporters

Three additional putative sugar transporters have also been isolated and sequenced (nucleotide and deduced amino acid sequences are presented in Appendix 4). They all possess a large conserved domain associated with sugar transport (pfam00083). This was located by searching the Conserved Domains Database at NCBI using the Reverse Position Specific BLAST algorithm <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> (Altschul *et al.*, 1997). PST Type 3 is most homologous to a mannitol transporter from celery (AF215837). The cDNA for this putative sugar transporter was 1874 nucleotides in length with an open

reading frame of 1314 nucleotides. This codes for a polypeptide 437 amino acids in length that has a calculated weight of 46,793 Da. The predicted protein has been designated a putative sugar transporter based on homology with other putative sugar transporter sequences from various phyla as well as the previously mentioned mannitol transporter. Figure 3 shows the results of localization of PST3. An antibody raised to specific peptides from the PST3 protein sequence was used to detect the PST3 protein in the tissue, followed by a gold-linked secondary antibody and silver enhancement. The results showed that PST3 is also located in the sieve elements of the phloem. Examination at higher magnifications suggested that PST3 was located in the sieve elements but not in the companion cells, in contrast to the pattern obtained for PST2.

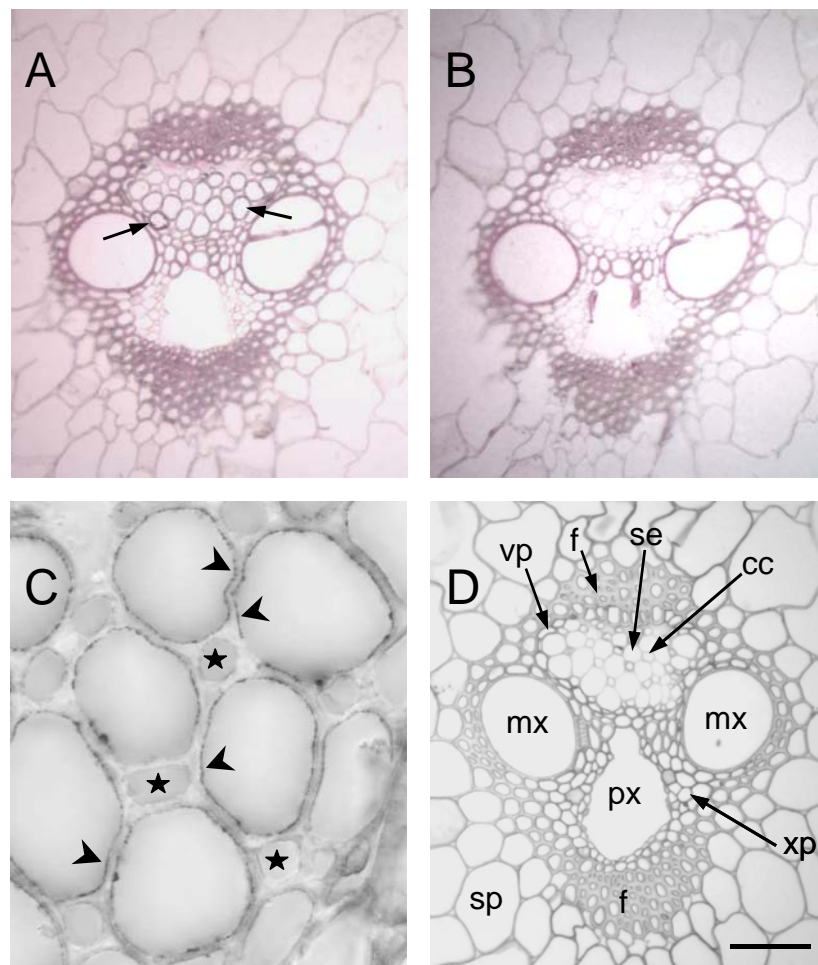


Figure 3. Localisation of PST3 by immunocytochemistry. Immunolabelling of sugarcane stem sections (internode 4) with antibody to PST3 (A, C), or pre-immune serum (B) followed by gold-labelled secondary antibody and silver enhancement. (A) Specific labelling is present in the sieve elements of the phloem (arrows). (B) No labelling is detectable in similar sections treated with pre-immune serum. The pink colour represents post-staining with basic fuchsin. (C) A higher magnification reveals labelling in the plasma membrane of sieve elements (arrowheads) but not companion cells (stars). (D) An unstained section shows the location of cell types within the vascular bundle. (mx, metaxylem; px, protoxylem; se, sieve elements; cc, companion cells; xp, xylem parenchyma; sp, storage parenchyma; f, sclerenchymatous fibres; bar equals 100 μ m). (A. L. Rae)

PST Type 4 is most homologous to sugar alcohol transporters from rice and *Prunus crasus*. The cDNA for this putative sugar transporter was 1970 nucleotides in length with an open reading frame of 1581 nucleotides. This codes for a polypeptide 527 amino acids in length that has a calculated weight of 55,571 Da. The predicted protein has been designated a putative sugar transporter based on homology with other putative sugar transporter sequences from various phyla as well as the previously mentioned sugar alcohol transporters. The cDNA for PST Type 5 was 1990 nucleotides in length with an open reading frame of 1512 nucleotides. This codes for a polypeptide 504 amino acids in length that has a calculated weight of 53,646 Da. The predicted protein has been designated a putative sugar transporter based on homology with other putative sugar transporter sequences from *Arabidopsis thaliana* and *Lycopersicon esculentum*. A multiple sequence alignment of the deduced amino acid sequences for PSTs Types 2a, 2b, 3, 4 and 5 revealed numerous regions of sequence homology (Appendix 4). The most obvious of these (marked with a double-headed arrow) is PESPRWL, which is present in all five sequences and is also present in many sugar, hexose and other metabolite transporters. This alignment also highlights the large central loop structure present in PSTs Type 2a and 2b.

Other categories of CMGs also showed evidence of coordinated expression e.g. transcripts for enzymes involved in sucrose synthesis and catalysis were significantly down-regulated in maturing stem, as were the transcripts for three enzymes belonging to central hexose phosphate metabolism, including key regulatory enzymes such as fructose-6-phosphate 2-kinase/fructose-2-6 biphosphatase, PFP and phosphoglucomutase. Transcripts for enzymes involved in polysaccharide synthesis and catalysis showed no particular coordination in either up- or down-regulation in maturing stem. Transcripts of different isogenes of glucosidase were differentially regulated while a strong and consistent down-regulation of UDP-glucose dehydrogenase, an enzyme involved in pectin synthesis, was observed.

Question 2: Which sugarcane genes with no previously documented role in carbohydrate metabolism, as represented by arrayed sugarcane ESTs, were differentially regulated between immature stem (represented by meristem and internodes 1-3) and maturing stem (represented by stem tissue derived from internodes 8 and 11) in high sucrose accumulating varieties of sugarcane

A sequence survey of stem tissue identified a number of classes of sequence that are correlated with growth and development in immature stem, and with fibre biosynthesis and degradation in maturing stem. A combination of EST analysis and microarray hybridisation revealed that homologues of the dirigent protein, a protein that assists in the stereospecificity of the bimolecular phenoxy radical coupling reaction, were the most abundant and most strongly differentially expressed transcripts in maturing stem tissue. This protein is particularly important in lignin and lignan biosynthesis, indicating a major role in both fibre biosynthesis and defence (Casu et al., in prep – Appendix 2).

Figure 4 shows the results of localization of dirigent. *In situ* hybridization to sugarcane stem tissue with an antisense probe showed that transcripts are located in the thick-walled cells surrounding the vascular bundles and in some adjoining parenchyma cells. Labelling was only detected in bundles towards the periphery of the stem which have extensive thick-walled sheaths. This localization is consistent with the predicted function of the dirigent protein as a scaffold for lignin biosynthesis.

Other categories of fibre biosynthesis and defence-related transcripts also show evidence of coordinated expression. In addition, a putative plasma membrane Na^+/H^+ antiporter has been identified that potentially has a role in osmotic regulation which may aid in the accumulation of sucrose to very high levels within cells (Casu et al., in prep – Appendix 2).

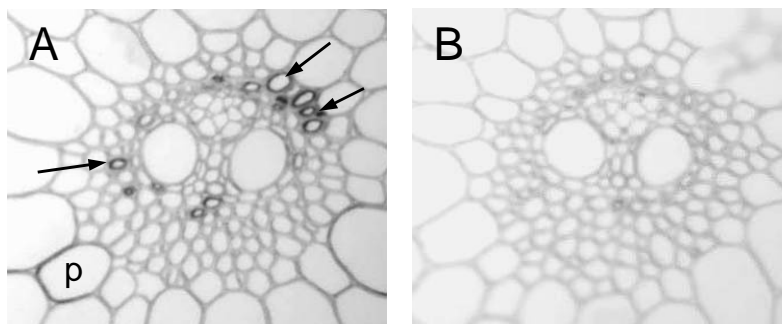


Figure 4. Localisation of dirigent by *in situ* hybridisation. *In situ* hybridisation of sugarcane stem sections (internode 5) with antisense RNA probes for dirigent (A) shows that transcripts are present in the thick-walled cells (arrows) surrounding the vascular bundle and in some parenchyma cells (p). Hybridisation with the corresponding sense control probe (B) shows no labelling (A. L. Rae).

A model for sucrose transport to the stem parenchyma

The location of sugar transporter genes demonstrated in this project taken in combination with some physiological analysis carried out at CSIRO has led to a model for the pathway of sucrose transport from sieve elements to sucrose parenchyma. It would appear that high sugar concentrations are maintained in sieve elements by hexose-like transporters (PST2 and 3) absorbing hexoses from leaked sucrose that would be hydrolysed by apoplastic invertase and probably by a sucrose transporter detected using antibodies to a rice sucrose transporter. Tracer dye and histological staining have demonstrated that a symplastic, but not an apoplastic, pathway exists for sucrose to move from the sieve elements through the vascular parenchyma to the storage parenchyma through plasmodesmata. Apoplastic transport is blocked by cell wall suberin and lignin. The storage parenchyma cells that surround the vascular bundles have a sucrose transporter that maintains the symplastic pathway so that sucrose can continue to move symplastically deeper into the parenchyma, any apoplastic hexose would be returned to the symplastic compartment via the PST2 transporter. One important question remains unanswered and this is “what drives this process?”. Presumably, there is a strong concentration gradient of sucrose from the intracellular sieve element compartment to the storage parenchyma, but continuous transport must be maintained by either removal of sucrose from the cytoplasmic compartment by efflux to the apoplast or vacuole in storage parenchyma or by gating of one-way sucrose traffic through the plasmodesmata at the parenchyma/vascular interface. These are hypotheses that can now be tested experimentally and would lead to identification of the critical controlling processes.

Question 3: Can the patterns of gene expression in maturing stem tissue in high CCS and low CCS progeny from the Q117 x MQ77-340 cross be correlated with their CCS level?

A third set of analyses involved the transcript profiling of maturing stem from selected individuals from the progeny of the Q117 x MQ77-340 cross (Q3). This population is segregating for high and low CCS as judged by CCS measurements at early season samplings. Progeny with either high or low CCS ratings were selected. Other research (Phil Jackson personal communication) has indicated a high correlation of the early CCS ratings with eventual mature crop measurements. Using the results of these array hybridisations on these high-CCS and low-CCS genotypes, it should be possible to identify genes with expression that is either up-regulated or down-regulated in a segregation-association with high CCS. This may permit the identification of candidate genes that correlate with sucrose accumulation. Fluorescent probes were prepared using RNA isolated from internode 8 of 5 low CCS individuals and 5 high CCS individuals, as well as from a pool of RNA derived from internode 8 of the 5 low CCS individuals. Ten hybridisations were performed in 2 main groups: (i) low CCS pool (labeled with Cy-3) v. each low CCS individual (labeled with Cy-5); (ii) low CCS pool (labeled with Cy-3) v. each high CCS individual (labeled with Cy-5). Technical replication was achieved by duplicate arrays within each slide. Each of the hybridisations within each group also functioned as a replicate. The ratios achieved for each EST from each array were normalised as detailed in the methods and were further analysed using GeneSpring 5.1. The expression ratios of a group of 62 ESTs were found to display statistical significant difference at $p = 0.05$ for the ratios achieved in the Low pool v. Low CCS individuals and Low pool v. High CCS individuals. These ESTs were further separated into two clusters based on whether they were up-regulated or down-regulated (Figure 4). The identities of the 59 ESTs that display differential regulation between low CCS and high CCS progeny from the Q117 x MQ77-340 cross (Q3) are presented in Table 1.

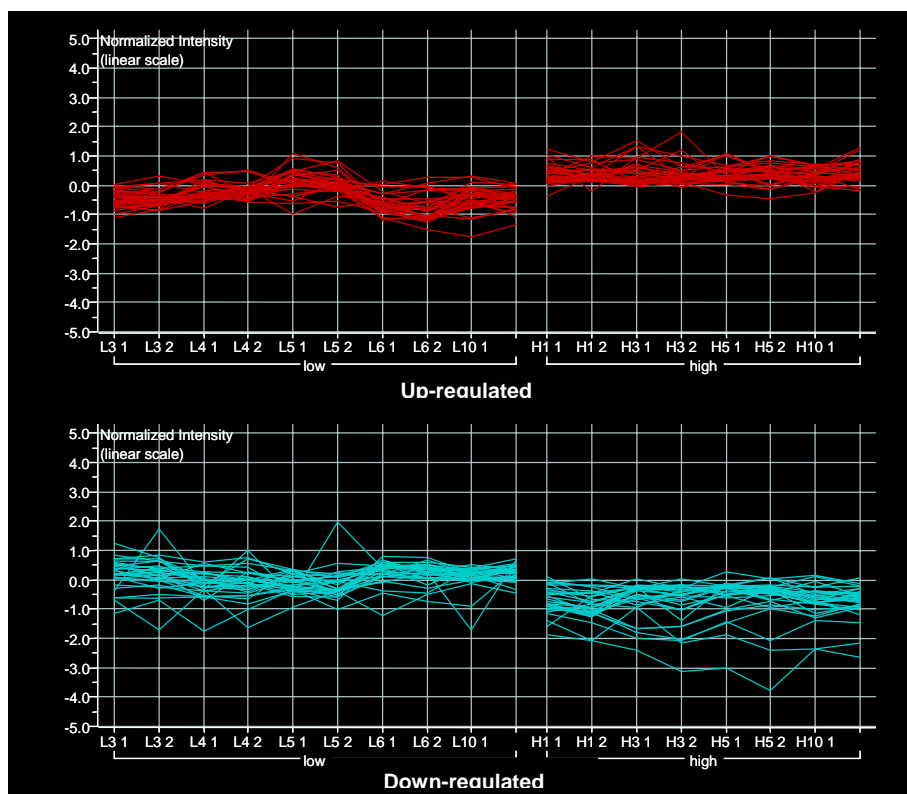


Figure 4. Regulation status of ESTs that are discriminatory between individuals with low CCS and high CCS.

| Regulation | Classification | Product | EST | |
|--|------------------------------------|---|---|------------|
| up | ATP synthesis/electron transport | ATP synthase | MCS017E05 | |
| | Carbohydrate metabolism | pyrophosphate-fructose-6-phosphate 1-phosphotransferase | MCS002C01 | |
| | Gene expression and RNA metabolism | RNA helicase | MCS016C05 | |
| | Membrane transport | ABC transporter | MCSA061A07 | |
| | Miscellaneous | | apospory-associated protein | MCSA179A11 |
| | | | rTSbeta | YCS23.043 |
| | | | tub protein (putative transcription factor) | MCS004G04 |
| | No assigned function | | no function assigned | MCSA038D07 |
| | | | | MCSA061E12 |
| | | | | MCSA066D10 |
| | | | | MCSA069E09 |
| | | | | MCSA075B06 |
| | | | | MCSA075F06 |
| | | | | MCSA117F05 |
| | | | | MCSA118A12 |
| | | | | MCSA180B11 |
| | | | | YCS43.019 |
| | Novel | | novel | MCS007H05 |
| | | | | MCSA031D06 |
| | | | | MCSA142H02 |
| | | | | MCSA147A11 |
| | | | | YCS42.013 |
| | Primary metabolism | | C-14 sterol reductase | YCS16.002 |
| | | | dTDP-glucose 4-6-dehydratase | MCSA067D03 |
| | | | glutamate-/aspartate-binding peptide | MCSA115G05 |
| | | | glycinamide ribonucleotide synthetase | MCS003A04 |
| | Protein synthesis and processing | | DnaJ | MCSA118H05 |
| | Signal transduction | | calmodulin | MCSA064F11 |
| | | | GTP-binding protein | MCSA175H02 |
| | | | phosphatidylinositol-4-phosphate 5-kinase | MCS015G06 |
| | down | Carbohydrate metabolism | sucrose phosphate synthase | MCSA172D09 |
| | | Chromatin and DNA metabolism | nucleic acid binding protein | YCS04.063 |
| Cytoskeleton | | tubulin | MCSA062F08 | |
| Miscellaneous | | | auxin induced protein | YCS13.035 |
| | | | IAA25 | MCSA037B02 |
| No assigned function | | | no function assigned | MCS017B03 |
| | | | | MCSA032B03 |
| | | | | MCSA035G10 |
| | | | | MCSA062E05 |
| | | | | MCSA075F02 |
| | | | | MCSA111B01 |
| | | | | MCSA174E08 |
| | | | | MCSA174F12 |
| | | | | MCSA202H06 |
| | | | | MCSA221B10 |
| YCS10.059 | | | | |
| Novel | | | novel | MCS011C12 |
| | | | | MCSA037C06 |
| | | | | MCSA063G06 |
| | | | | MCSA071G06 |
| Protein synthesis and processing | | | protein methyltransferase | MCSA063B08 |
| | | | ribosomal protein | MCSA180H10 |
| | | | ubiquitin-conjugating enzyme | YCS16.042 |
| Signal transduction | | | calcium-dependent protein kinase | MCSA111H11 |
| | | | GTP-binding protein | MCSA063B12 |
| | | | MAP kinase | MCSA180D12 |
| | | | nucleoside diphosphate kinase | MCSA074E04 |
| | | | sucrose export defective 1 (tocopherol cyclase) | YCS38.012 |
| Vesicular trafficking, protein sorting and secretion | | | shrunken seed protein | MCSA177D02 |

Table 1. Possible role and product identity of genes (ESTs) differentially expressed between low and high CCS individuals from Q3.

Only two of the candidate genes appear to be known carbohydrate metabolism genes based on DNA sequence homology to known genes. In fact, 30 of these ESTs are either sequences for which there is no known function or are novel sequences. To determine whether these sequences can act as expression markers for high CCS would require validation in progeny from another population segregating for CCS. Additional validation by Northern analysis or Real-Time PCR studies will also be required.

All candidate ESTs were homology-searched against the non-redundant DNA, protein and EST database at NCBI in order to update any putative product and function assignments. This was performed as part of curation of the CaneDB which was created to organise all data pertaining to the sugarcane stem ESTs that were developed during the preceding project CTA035.

Implications of these analyses

The use of expressed sequence tag and DNA microarray analysis is very powerful for providing correlative associations between particular gene sequences and important sugarcane functions such as sucrose accumulation. There are four principle ways that the materials and information developed in this project can be used to deliver outcomes for the industry:

1. The sugarcane EST collection has been extensively used as a source of DNA markers for marker identification projects in both sugarcane and sorghum. In particular, ESTs of putative function have been sourced as part of studies aimed at:
 - identifying perfect markers for marker-assisted breeding projects;
 - understanding the physiological basis of specific QTLs; and
 - determining the genetic location of specific genes.

ESTs from the sugarcane EST collection including those identified within CPI002 have been used in the following projects.

- Candidate genes for sucrose accumulation and fibre content in sugarcane. Within CTA046, a part of the PhD project of Ms Nathalie Reffay was to investigate the usefulness of candidate genes as markers in sugarcane. She screened across the R570 sugarcane mapping population with seven sugar-related and six fibre-related ESTs from the sugarcane EST collection. Four markers were consistently associated with a trait over the two years for which data was collected. One sugar accumulation-related candidate gene (UDP-glucose dehydrogenase) was identified as associated with Brix (BRX95 and 96), while another (cell wall invertase) was identified as associated with stalk diameter (DIA95 and 96). One fibre-related candidate genes were identified as associated with height (HT95 and 96 – cinnamyl alcohol dehydrogenase). No markers were consistently identified with sucker number. Two other sugar-related candidate gene markers were associated with height and Brix in one year only. These associations are of interest and will be followed in other populations.
- Genetic location of sugarcane ESTs in sorghum. The sugarcane genome is very complex, due to the large number of chromosomes and its interspecific origins. There are 8 homology groups, each containing an estimated 10-12 linkage groups. By contrast, sorghum, a close relative of sugarcane, contains 10

pairs of chromosomes. Mapping in sorghum provides us with a useful genetic reference map for sugarcane.

Approximately 400 ESTs from the CSIRO/SRDC sugarcane stem EST collection have been screened for polymorphism in sorghum and more than 100 have been mapped in two sorghum populations. Particular emphasis has been placed on mapping sugar metabolism and stress-related ESTs. Six of the seven sugar transporter families have been mapped in sorghum and are widely distributed throughout the sorghum genome. PST Type 3 and ShSUT1 cluster together while PST Type 2a and PST Type 4 cluster with other sugar metabolism genes. Numerous ESTs co-locate to QTLs for a range of sorghum traits, including yield. These associations are of interest as they may provide information on the role of specific QTLs in the expression of the trait, as well as providing perfect markers.

- Identification of new SSRs from the sugarcane EST collection. Simple sequence repeats (SSRs) are the most commonly used and most useful of markers in plant mapping activities as they are easy to use and very informative. However, they are difficult and expensive to identify. In collaboration with Drs Giovanni Cordeiro and Robert Henry at SCU, we therefore decided to see if any were contained within ESTs in the sugarcane EST collection, as a new source of SSRs. Approximately 250 SSR sequences were identified. Primer sets were designed to 35 sequences and tested on 5 sugarcane genotypes. Twenty-one gave an amplification product of which 17 were polymorphic. However, the level of polymorphism amongst the 5 genotypes was lower (PIC value of 0.23) than that seen when using genomic DNA-derived SSRs (PIC value of 0.72). By contrast, the EST-derived SSRs were more effective in related genera, such as *Erianthus* and sorghum. These results suggest that the EST-derived SSRs are more useful across greater evolutionary distances such as inter-generic comparisons, while genomic DNA-derived SSRs are better for within cultivar variation (Cordeiro et al, 2000, *Plant Science* 160: 1115-23).
2. The identification of a small suite of genes that, in one cross, show a genetic association in their expression with high CCS provides an expression fingerprint for the high CCS state and potentially new tools that could be added to selection programs. The use of such expression fingerprints will need to be validated on other crosses and at earlier stages of development when efficient selection would be more beneficial. However the principle of gene expression fingerprints appears to have substance from results so far.
 3. Characterisation of the function of the genes associated with high CCS may open up options for genetically manipulating gene expression in transgenic plants. Although this may be achieved with current candidates it is more likely that sugar accumulation is a complex process this will be achieved by understanding its regulation and the manipulation of key master genes that control carbon partitioning processes, the identification of the candidate genes is an important first step. We have successfully localised both the PST Type 2a transcript and protein to the companion cells and the sieve elements, respectively. In addition, the PST Type 2a protein was also localised at lower density to the parenchyma cells. These localizations, the fact that this transporter successfully facilitates the import of glucose into the cells in sugarcane suspension cultures and that the transcript is up regulated in maturing and mature stem tissue to the bottom of the stem suggest that PST Type 2a is involved in the transport

of hexoses in all maturing and mature internodes. The specificity of the type of localisation techniques used is illustrated by the immunolocalization of PST Type 3 to the sieve elements of the phloem alone. Dirigent protein, the most up-regulated transcript in maturing stem, has been located to a tissue that is consistent with its role in lignin biosynthesis. These examples illustrate the power of continued bioinformatic analyses in combination with *in situ* hybridisation and immunolocalization techniques, as well as functional testing in sugarcane suspension cultures. The continued identification of additional putative sugar transporter sequences indicates the complexity of sucrose and hexose transport processes and is leading to an integrated picture of sugar transport pathways. Given their important role in sugar accumulation experiments investigating the expression of sucrose transporters in germplasm that varies in sucrose accumulation will provide important clues on whether these genes are functionally related to increased CCS.

4. Finally, the project has developed tools that can be used to develop tissue-specific promoters, and sub-cellular targeting signals to direct other metabolic processes for the production of new products. Further applications of this will be developed in alliances with other research providers, particularly within the CRC for Sugar Industry Innovation through Biotechnology and the CSIRO/BSES Joint Venture.

Publications cited in this report

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389-3402.
- Bower, R., Elliott, A. R., Potier, B. A. M. and Birch R. G. (1996) High-efficiency, microprojectile-mediated cotransformation of sugarcane using visible or selectable markers. *Mol. Breeding*, 2, 239-249.
- Casu, R. E., Grof, C. P. L., Rae, A. L., McIntyre, C. L., Dimmock, C. M. and Manners, J. M. (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol. Biol.*, 52, 371-386.
- Elliott, A. R., Campbell, J. A., Brettell, R. I. S. and Grof, C. P. L. (1998) Agrobacterium-mediated transformation of sugarcane using GFP as a screenable marker. *Aust. J. Plant Physiol.*, 25, 739-743.
- Ihaka, R. and Gentleman, R. (1996) R: A language for data analysis and graphics. *J. Comp. Graph. Stat.*, 5, 299-314.
- Jackson, P. A. and Morgan, T. E. (2003) Early stage selection for commercial cane sugar (CCS) in sugarcane clones: effects of time of sampling and irrigation. *Aust. J. Ag. Res.*,
- Klok, E. J., Wilson, I. W., Wilson, D., Chapman, S. C., Ewing, R. M., Somerville, S. C., Peacock, W. J., Dolferus, R. and Dennis, E. S. (2002) Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. *Plant Cell*, 14, 2481-2494.
- Taylor, P. W. J., Ko H. – L., Adkins, S. W., Rathus, C. and Birch R. G. (1992) Establishment of embryogenic callus and high protoplast yielding suspension cultures of sugarcane (*Saccharum* spp. hybrids). *Plant Cell, Tissue and Organ Culture*, 28, 69-78.
- Wilson, D.L., Buckley, M.J., Helliwell, C.A. and Wilson, I.W. (2003) New normalization methods for cDNA microarray data. *Bioinformatics*, 19, 1325-1332.

Outputs

1. Identification of a suite of genes whose expression pattern can discriminate between high and low CCS individuals from one cross.
2. Identification of ESTs with potential as DNA markers for traits relevant to Australian Sugar Industry
3. Methodologies for evaluating transporter function in sugarcane cell and yeast suspension cultures.
4. Identification by bioinformatic analyses and transcriptional profiling of a novel putative sugar transporter that is the most up-regulated carbohydrate metabolism gene in maturing stem of sugarcane.
5. Two functionally characterised sugar transporters, one transporting sucrose (ShSUT1) and the other transporting hexoses (PST Type 2a).
6. Methodologies for the routine localisation of transcript and corresponding proteins in stem sections.
7. Updated version of the CaneDB holding all information pertaining to the CSIRO/SRDC EST collection (a progression on the original database developed as part of CTA035).

Expected Outcomes

This project is strategic research and its outputs will feed into further research leading to industry outcomes via improved varieties produced via breeding. Expected outcomes after this subsequent research process include:

1. Varieties with improved CCS produced via the application of DNA markers or expression fingerprints.
2. A more efficient breeding program using new tools (DNA markers or expression fingerprints) for varietal selection for the production of high CCS varieties.
3. The production of novel compounds of greater value than sucrose generated using gene technologies for the cell and tissue specific of transgenes developed using genes identified in this project.
4. Improved profitability for the industry via improved efficiency of sugar production and varietal development and in the longer term via product diversification from engineered sugarcane.
5. A competitive and leading position for the Australian industry in sugarcane functional genomics and sucrose accumulation research, leading to the generation of proprietary intellectual property.

Future Research Needs

Because the strategic research conducted in this project provides outputs for further research in order to deliver industry benefit these research steps are needed. They are briefly outlined below under the heading of the desired outcome.

- Development of high CCS varieties through more efficient marker selected breeding. This project has produced outputs that may develop two marker types for enhanced selection of high CCS varieties. Research needs to develop and utilise the following are as follows:
 - i. *DNA markers.* Candidate genes identified in this project will need to be tested as markers for genetic segregation in multiple populations segregating for CCS. This is underway in projects CTA47 and an ACIAR project and will be further developed in CRCSIIB. One essential aspect of testing these genes as markers will be the development of all probes for all alleles of the gene as markers. A project in CRCSIIB aimed at developing single nucleotide polymorphisms will test the feasibility of this and R. Casu is participating in this project.
 - ii. *Expression fingerprints as markers.* An exciting possibility that has been experimentally tested in this project is that transcriptional profiles that are specific to high CCS can be identified in segregating populations. This now needs to be confirmed in the current Q3 population in another season and in other populations. Ultimately, this may lead to a DNA chip able to detect high CCS types and, perhaps more importantly, it may be able to differentiate between mechanisms leading to high CCS.
 - iii. The above markers will need to be used in the sugarcane breeding and selection program and be used to develop high CCS varieties. The key use of these markers would be for efficient combining of multiple independent mechanisms that can lead to CCS, leading to maximal expression of the trait.
- An important objective for the sugar industry is the generation of sugarcane ‘biofactory’ plants that produce alternative products of greater value than sugarcane. To do this it will be necessary to redirect carbon from sucrose production to other molecules. Sucrose is a vital metabolite in plants and perturbation of sucrose metabolism leads to impaired plant growth and development. Therefore, there is a requirement to isolate the enzymatic conversions of sucrose to cell types and sub-cellular compartments that are not likely to affect plant development. The most appropriate cell type for this in sugarcane is the mature stem sucrose storage parenchyma. In this project we have identified many genes that are highly expressed in mature cells as well as methods for localising transcripts to cell types in the stem. These techniques can now be applied to identify genes that are specifically expressed in parenchyma cells. These experiments are already underway in SRDC project UQ39 and in CRCSIIB.

Finally, the project has contributed to the development of a model for sucrose accumulation in the stem. The principle tenet of this model is the maintenance of a symplastic pathway through to parenchyma cells remote to vascular bundles with symplastic concentrations maintained by efflux from cytoplasm of parenchyma cells to the apoplast and/or vacuole and one way flow through the plasmadesmata of the vascular/parenchyma interface. This model is a hypothesis that can now be tested experimentally and it may lead to the identification of key controlling genes for sucrose accumulation in sugarcane. This would be extremely valuable intellectual property for the manipulation of sucrose accumulation via breeding.

Recommendations

1. Further assessment of whether the sugarcane genes that correlate either developmentally or via genetic association with high sucrose accumulation that have been isolated so far can be developed for use as DNA markers corresponding to loci for high CCS for application in breeding high CCS varieties.
2. Analysis of genes expressed in specific cell types such as storage parenchyma and vascular bundles to identify genes that can provide tools to target metabolic engineering products to specific cell types to avoid disruption of plant function.
3. Experimental testing of the symplastic sugar transport model and identify the key rating limiting steps as a first step to further manipulation.
4. Transfer of the outputs of this project to a range of research providers for the Australian sugar industry most probably via the new CRC.

Publications arising from this Project

1. Casu, R. E., Grof, C. P. L., Rae, A. L., McIntyre, C. L., Dimmock, C. M. and Manners, J. M. (2003) Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol. Biol.*, 52, 371-386 [Appendix 1]
2. Casu, R. E., Dimmock, C. M., Chapman, S. C., Grof, C. P. L., McIntyre, C. L., Bonnett, G. D. and Manners, J. M. Identification of transcripts differentially regulated in maturing stem of sugarcane by *in silico* characterisation and expression profiling of stem ESTs (in preparation) [Appendix 2].
3. Rae, A. L., Perroux, J. M. and Grof, C. P. L. Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the ShSUT1 sucrose transporter (in preparation) [Appendix 3].

Appendix 4

Nucleotide and deduced amino acid sequences of sugarcane putative sugar transporters – PST Type 3, PST Type 4 and PST Type 5. Underlined sequence is homologous to pfam00083 (sugar_tr), a sugar transport domain.

PST Type 3

```

GCCTCACGTA GCTAGCACTA GCAGCGAGGA GTGAGGAGTG AGAGGACACG GTAACTCCGG CGGCATTTCG TGCTGGTACA      80
GTACACCAAT AATCGAGCAG CGTGCGCGAG AGAGAACCGA GCCAGAGGCC GGCCGGCACG CCGGGACACG GCCAGTACAG      160
AGAACGAAAAG CTAGTGCCAA AATAATCGGT AGATCTACAC CACCGACACT TGGCC ATG AGC ACG GGC ACG GAC GCC      236
                               M S T G T D A>

GTC GCG GTC GCC GTG GCG CCG GCG AAG CGC CCC CCC ATC AAC GAG TAC GCC TTC GCC TGC GCC CTG      302
V A V A V A P A K R P P I N E Y A F A C A L>

CTC GCC TCC ATG AAC TCC GTC CTC CTC GGC TAT GAC ATC TCG GTG ATG AGT GGC GCG CAG CTG TTC      368
L A S M N S V L L G Y D I S V M S G A Q L F>

ATG AAG GAG GAC CTC AAG ATC ACG GAC ACG CAG ATC GAG ATC CTC GCG GGC GTC ATC AAC ATC TAC      434
M K E D L K I T D T Q I E I L A G V I N I Y>

TCT CTG TTC GGC TCC CTC GCC GCC GGA TTC ACC TCC GAC TGG CTC GGC CGG CGC TAC ACC ATG GTA      500
S L F G S L A A G F T S D W L G R R Y T M V>

CTG GCG GCC GCC ATC TTC TTC ACG GGC GCG CTC CTC ATG GGC TTG GCT CCG GAC TAC GGG CTC CTC      566
L A A A I F F T G A L L M G L A P D Y G L L>

ATG GTG GGC CGC TTC GTG GCC GGC ATC GGC GTC GGC TTC GCG CTC ATG ATC GCC CCC GTG TAC ACC      632
M V G R F V A G I G V G F A L M I A P V Y T>

GCC GAG GTG GCG CCC ACC TCG GCG CGC GGC TTC CTC ACC TCC TTC CCG GAG GTG TTC AAC AAC TTC      698
A E V A P T S A R G F L T S F P E V F N N F>

GGC ATC CTC CTG GGC TAC GTC TCC AAC TTC GCC TTC GCG CGC CTG CCC GTG CAC CTC AGC TGG CGC      764
G I L L G Y V S N F A F A R L P V H L S W R>

GCC ATG TTC CTG GTC GGC GCC GTG CCG CCC GTC TTC CTG GGC GTC GCC GTC CTC GCC ATG CCG GAG      830
A M F L V G A V P P V F L G V A V L A M P E>

TCG CCA CGC TGG CTC GTC ATG CCG GGC CGC ATC GAC GAC GCG CGC CGC GTG CTG CAG AAG ACC TCC      896
S P R W L V M R G R I D D A R R V L Q K T S>

GAC TCC CAC GCC GAG GCT GAG GAG CGC CTG CTC GAC ATC AAG AAG GTG GTC GGC ATC CCC GAG GGC      962
D S H A E A E E R L L D I K K V V G I P E G>

GTG TCC GAC GCC GAC GAC GTG GCC GCC ATC GTC CGC GCC AAC AAG GGC TCC TCG CGC CAC GAC GGG      1028
V S D A D D V A A I V R A N K G S S R H D G>

GTG TGG AAG GAA CTG CTC ATC AAC CCT TCC CGC CCC GTG CGC CGC ATG CTC GTG GCT GGG CTG GGG      1094
V W K E L L I N P S R P V R R M L V A G L G>

CTC ATG TTC ATC CAG CAG GCC ACG GGC GTG GAC TGC GTG GTC ATG TAC AGC CCG CGG GTG TTC GAG      1160
L M F I Q Q A T G V D C V V M Y S P R V F E>

CGG GCG GGC ATC AAG TCC AAG ACC AAC TCG CTG GGC GCG TCC ATG GCG GTG GGC GCG TGC AAG ACC      1226
R A G I K S K T N S L G A S M A V G A C K T>

TTC TTC ATA CCC ATA TCG ACG CTG CTC CTG GAC CGC ATC GGC CGG CGG CCG CTG CTG CTG GCG AGC      1292
F F I P I S T L L L D R I G R R P L L L A S>

GGC GGC GGG ATG GCC ATC TTC CTC TTC ACG CTG GCC ACG TCC CTG CAC ATG ATG GAC CGG CGC CCC      1358
G G G M A I F L F T L A T S L H M M D R R P>

GAG GGG GAG GCC GCG GCG CTG GCG CGG TCA GCA TCG CCG CCA TGC TGT CGT TCG TGG CGT CGT TCG      1424
E G E A A A L A R S A S P P C C R S W R R S>

CGT CCG GCC TGG GCC CCG TCG CCT GGG TCC ACT GCT CCG AGA TCT ACC CGC TTC GGC TGC GCG CGC      1490
R P A W A P S P G S T A P R S T R F G C A R>

AGG CCG CCG CCA TCG GCA CGG GGC TCA ACC GGA TCA TGA G CGGCGCCACC ACCATGTCCT TCCTCTCGCT      1560
R P P P S A R G S T G S *>

CTCCAACACC ATCACCATCG CCGGAAGCTT CTACCTCTAC GCCTGCATCG CCGCCGCAGG GTGGGTGTTT ATGFACTTCT      1640
TCCTGCCGGA GACGATGGGC AAGAGCCTGG AGGACACCGT GAAGCTCTTC GGAAAGGACG CGGACGACGA GGACGTCCGG      1720
GACAGCCGTC GTCACGTGCC CAGTCAGAAG CCCTCCACTG AGCTGAGTGC TCAGCAGTGA GTAATATGCC ATGCAACACG      1800
TGTCCTCCCG GGTGTGGCCG AAGCTTATGT TCCGTAAGC TTCCGATGATA TGATTACAGC ATATACACTA GTACTGTATT      1880
TTGTGTGCTC GTGCGTGGTG AATCTTTGGG CCGATGTAAG AAATTTGGCA AAATGGATAC CGTAGTATTT TCGTTGTAAA      1960
AAAAAAAAAA AAAA      1974

```

PST Type 4

| | |
|---|------|
| CACAACACAC ATACCTTGCC TCTAATCAGA GAGCAAAGAA GTAGCCGTTA AGTTAACG ATG GCT TCC GCC GAG CTC | 76 |
| M A S A E L | 6 |
| GCT CAG GCC ATC GAG CCC AGG AAG AAG AGC AAC GTC AAG TAT GCC TCC ATC TGT GCC ATC CTC GCA | 142 |
| A Q A I E P R K K S N V K Y A S I C A I L A | 28 |
| TCC ATG GCC TCC GTC ATC CTT GGC TAT GAC ATT GGG GTG ATG AGT GGA GCG GCG ATG TAC ATC AAG | 208 |
| S M A S V I L G Y D I G V M S G A A M Y I K | 50 |
| AAG GAC CTG AAG ATC ACG GAC GTG CAG TTG GAG ATC ATG ATA GGC ATA CTC AGT GTC TAC TCG CTG | 274 |
| K D L K I T D V Q L E I M I G I L S V Y S L | 72 |
| ATC GGT TCC TTC GCT GGC GCG CGG ACG TCC GCC AGG ATC GGT CGC CGC TTG ACC GTC GTC TTC GCT | 340 |
| I G S F A G A R T S A R I G R R L T V V F A | 94 |
| GCT CTC ATC TTC TTC GTG GGC GCG TTG CTC ATG GGT TTC GCC GTC AAC TAC GGC ATG CTC ATG GTG | 406 |
| A L I F F V G A L L M G F A V N Y G M L M V | 116 |
| GGA CGC TTC GTG GCC GGA GTA GGT GTG GGC TAC GGG GGT ATG ATC GCG CCC GTG TAC ACC GCC GAG | 472 |
| G R F V A G V G V G Y G G M I A P V Y T A E | 138 |
| ATC TCG CCG GCG GCA TCC CGC GGC TTC CTC ACG TCC TTC CCG GAA GTG TTC ATC AAC ATC GGT ATC | 538 |
| I S P A A S R G F L T S F P E V F I N I G I | 160 |
| CTG CTC GGC TAC GTG TCG AAC TTC GCG TTC GCT CGC CTC CCG CTC CAC CTC GGC TGG CGT GTT ATG | 604 |
| L L G Y V S N F A F A R L P L H L G W R V M | 182 |
| CTC GGC ATC GGC GCA GCG CCG TCC GCC CTG CTC GCG CTC ATG GTG TTT GTC ATG CCC GAG TCG CCT | 670 |
| L G I G A A P S A L L A L M V F V M P E S P | 204 |
| CGG TGG CTT GTC ATG AAG GGC GCG CTC GCG GAC GCC AGG GCT GTG CTA GAG AAG ACC TCT GCC TCG | 736 |
| R W L V M K G R L A D A R A V L E K T S A S | 226 |
| CCA GAG GAG GCC GCA GAG CGG CTG GCT GAC ATC AAG GTT GCC GCG GGG ATT CCC AAG GAC CTC GAC | 802 |
| P E E A A E R L A D I K V A A G I P K D L D | 248 |
| GGA GAC GTG GTC ACC GTA CCC AAG GAG CGG AAC GGT GGT GAG ATG CAG GTG TGG AAG GCG CTC ATC | 868 |
| G D V V T V P K E R N G G E M Q V W K A L I | 270 |
| TTT TCA CCA ACC CCG GTG ATC CGA CGC ATA CTG CTA TCG GGC GTC GGC CTG CAC TTC TTC CAG CAG | 934 |
| F S P T P V I R R I L L S G V G L H F F Q Q | 292 |
| GCT TCT GGC AGC GAT TCC GTC GTC CTG TAC AGC CCG AGC GTG TTC AAG AGC GCG GGA ATC ACC GGC | 1000 |
| A S G S D S V V L Y S P S V F K S A G I T G | 314 |
| GGC AAC AAG CTC CTG GGC GTC ACC TGC GCG GTG GGC GTC ACC AAG ACG CTT TTC ATC CTG GTG GCC | 1066 |
| G N K L L G V T C A V G V T K T L F I L V A | 336 |
| ACG TTC CTG CTG GAC CGC GTC GGG CGT CGG CCT CTG CTG CTG ACC AGC GCG GGT GGA ATG ATC ATC | 1132 |
| T F L L D R V G R R P L L L T S A G G M I I | 358 |
| TCT CTC ATC GGC CTC GGG ACG GGC CTC ACC GTC GTG GGG CAT CAC TCG GAC GCC AAG ATC CCG TCG | 1198 |
| S L I G L G T G L T V V G H H S D A K I P S | 380 |
| GCC GTC GCT CTG TGC ATC GCG TCC ACC TTG GCC TAC GTC GCC TTC TTC TCT ATC GGC CTC GGC CCC | 1264 |
| A V A L C I A S T L A Y V A F F S I G L G P | 402 |
| ATC ACG GGT GTG TAC AAC TCG GAG ATA TTC CCG CTG CAG GTG CGC GCG CTG GGC TAC GCG GTC GGC | 1330 |
| I T G V Y N S E I F P L Q V R A L G Y A V G | 424 |
| GGG GCG TGC AAC CGC GTT ACC AGC GCC GTC ATC TCC ATG ACC TTC CTG TCC CTC TCG AAG GGC ATC | 1396 |
| G A C N R V T S A V I S M T F L S L S K G I | 446 |
| ACC ATC GGC GGC AGC TTC TTC CTC TAC TCC GGC ATC GCC GTG GTC GGG TGG GTT TTC TTC TTC ACG | 1462 |
| T I G G S F F L Y S G I A V V G W V F F T T | 468 |
| TGC CTC CCG GAG ACC CGC GGT CGT ACG CTG GAG GAG ATG GGC AAG CTG TTC GGC ATG CCA GAC ACG | 1528 |
| C L P E T R G R T L E M G K L F G M P D T | 490 |
| GAC ATG GCT GAA GCA GAC AAC GCC GCT GCC AAG GAG AAG GTG GTG GAA ATG ATT GAT GCA GCG TCT | 1594 |
| D M A E A D N A A A K E K V V E M I D A A S | 512 |
| TAC CAG AGC ACA GCT CAA CTG AAG CAG CGT GGC AAG ATT GGC CTC TGA GGTGTCGT GGATCGTGGT | 1660 |
| Y Q S T A Q L K Q R G K I G L * | 527 |
| GCTGCAGGAC AGTGACGGCC TGGTCTAAC AGAGACTCTT CTGTAGGACA CCTGTGAATT ATTGGCGTCG GCGTCCAGTG | 1740 |
| TGCGAGCAAG TTGCTCAAGC TGTGCGTTTC TTCTGAGTGC CAGAGAGTGA ACCCAAAATG CAAGCAGTAG GTACTTGCAA | 1820 |
| GGGGCACGTG CTTCCTGATT ACTCCCTTCA TCCTAAATTA TAAGGTATTT CAAGAATCCT AGAGAGTTAA AAGATTTCAA | 1900 |
| GTTTGACCAA ATTTATATGA TAAGATAATA ATATTTATAA CATTAAATTAA GTATCAAAAA AAAAAAAAAA | 1970 |

PST Type 5

| | |
|---|------|
| GCCGGAGCGG GCACCACTGC GTCTGCCCG GAGTGTCCCG AACTCCCGAT CCACCCATCT CCATCTCCAT CTTGGCCAGA | 80 |
| GGCAGGCAGC CTCATCTCAT CTCACGTTGG TAGGAACAAG AACAAGAACA CA ATG GGT ATA GTG CGG GCC GCG | 153 |
| M G I V R A A> | 7 |
| GAC GAT GTG GAG GCC GGG GTG GTC GGG GCC GCC ACG ACG ACG GAG CCG CTT CTC CTC CGC CAC TGC | 219 |
| D D V E A G V V G A A T T T E P L L L R H C> | 29 |
| CAC GGC AGC AAG GAG GAG GAC CAC GAG GAG TCT AAG ATC CAG GGC AGC CCG GAG GAG GCG GCG TCC | 285 |
| H G S K E E D H Q E S K I Q F S P E E A A S> | 51 |
| GAC TGC GGG AGC GAG GGG CGG CCG GAC GCA GCA GGC TCG CTG CGG ATG GTC CTT CTC TCC ACC GCC | 351 |
| D C G S E G R P D A A G S L R M V L L S T A> | 73 |
| GTC GCC GTC TGC GGC TCC TTC GAG TTC GGC ACC TGC GTC GGC TAT TCT GCA CCC ACG CAG TCA GGG | 417 |
| V A V G G S F E F G T C V G G Y S A P T Q S G> | 95 |
| ATA GTA GAT GAA GTC GGG CTA TCT ATC TCC GAG TTT GCC ATT TTT GGA TCT ATA TTG ACA ATC GGA | 483 |
| I V D E A V G G L S I S E F G I F G S I L T I C G> | 117 |
| GCA ATG GTT GGT GCT GTT ACT AGT GGG CGC TTG GCT GAT TTT CTT GGA CGA AAA ATG ACC ATG CGG | 549 |
| A M V G A V T S G G R L A D F L G R K M T M R> | 139 |
| ATA TCA GCG ATT ATT TGC ATC TTT GGT TGG CTT TCT ATA CAT CTG GCC AAG AGC GCT ATC ATG CTC | 615 |
| I S A I I C I F G W L S I H L A K S A I M L> | 161 |
| TAC TTT GGG AGA ATC TTA CTG GGT TTC AGT ACT GGA GTT CTA TCT TAT GTG GTT CCA GTA TTC ATA | 681 |
| Y F G R I L L G F S T G V L S Y V V P V F I> | 183 |
| GCT GAA ATA GCA CCA AAG AAT CTC CGT GGG GGG CTC GCA ACA TCA AAC CAG CTA TTG ATC TGT TCT | 747 |
| A E I A P K N L R G G L A T S N Q L L I C S> | 205 |
| GGA AGC TCA GCT ACA TAC ATC ATA GGA GCC CTG GTT GCA TGG CGC AAT TTG GTG CTT GTT GGT TTA | 813 |
| G S S A T Y I I G A L V A W R N L V L V G L> | 227 |
| TTG CCT TGT GTC CTC CTG CTA GCC GGG CTT TTC TTC ATT CCA GAG TCG CCC AGA TGG CTG GCC AAT | 879 |
| L P C V L L L A G L F F I P E S P R W L A N> | 249 |
| GTC GGA AGG GAA AAG GAA TTC CAC ACC TCA TTG CAA AAG CTG CGT GGA GAG GAT GCT GAT GTA TCA | 945 |
| V G R E K E F H T S L Q K L R G E D A D V S> | 271 |
| GAA GAG GCA TTG AGA TTA AAG AGT ACA TTG AAT CAC TCT GTA GCT TCC AAA GGC CAG GCT TCA AGA | 1011 |
| E E A L R L K S T L N H S V A S K G Q A S R> | 293 |
| CTT ATT TCT AAG CAA AAA TAT ATA TGC AGT CAT TGT GGG TGT TGG CTT GAT GGT TTT CAG CAA CTC | 1077 |
| L I S K Q K Y I C S H C G C W L D G F Q Q L> | 315 |
| GGA GGG ATA AAT GGT GTT GGA TTC TAT GCA AGC TAT ATC TTC AGT TCT GCC GGG TTT TCT GGA AAA | 1143 |
| G G I N G V G F Y A S Y I F S S A G F S G K> | 337 |
| CTT GGA ACC ATC TTG ATT GGC ATT ATT CAG ATT CCG ATT ACA TTG TTT GGG GCC ATC CTC ATG GAT | 1209 |
| L G T I L I G I I Q I P I T L F G A I L M D> | 359 |
| AGG AGT GGA AGA AGG GTT CTC CTG ATG GTC TCT GCA TCT GGA ACA TTT TTG GGC TGC TTC CTC ACT | 1275 |
| R S G R R V L L M V S A S G T F L G C F L T> | 381 |
| GGA GTA TCA TTC TAT CTA AAG GCA CAA GGA CTT CTC CCA GAA TGG GTT CCC ACA TTG GCT CTT TCT | 1341 |
| G V S F Y L K A Q G L L P E W V P T L A L S> | 403 |
| GGC ATA CTG GTT TAC ATC GGG GCA TAC TCA ATT GGA ATG GGT CCT GTC CCT TGG GTT GTC ATG TCT | 1407 |
| G I L V Y I G A Y S I G M G P V P W V V M S> | 425 |
| GAG ATA TTC TCA ATT AAC ATG AAA GCA ATT GGT GGA AGC TTG GTG ACC CTT GTT AGT TGG TTG GGT | 1473 |
| E I F S I N M K A I G G S L V T L V S W L G> | 447 |
| TCC TTT GCG ATT TCT TAC TCA TTT AGC TTC CTT ATG GAC TGG AGC TCT GCA GGT ACC TTC TTC ATG | 1539 |
| S F A I S Y S F S F L M D W S S A G T F F M> | 469 |
| TTC TCT GCA GCC AGC TTG GTC ACT GTA CTC TTC GTG GCG AAG CTA GTG CCT GAA ACC AAA GGA AGG | 1605 |
| F S A A S L V T V L F V A K L V P E T K G R> | 491 |
| ACA CTA GAG GAG ATC CAA GAC TCT CTA AAC ACT AGA AGA TAA TAA TGAATTTACT GCACACTACA | 1670 |
| T L E E I Q D S L N T R R *> | 504 |
| GAAAAATTCC AAGATGGAAA ATGGGAATTT ACAGGACACA ACAGAATAAT TCCAAGATGG AAAATGATAT ACACGTTTAC | 1750 |
| CTTTCCACC TTCACCATGT GTTCCACCTC CTCCCGCTTC CCTCCCTT CTCTCCTGCC CACCTCAGTG TAATATGTAG | 1830 |
| TGTACCCCTA TGTAGGTTAA GGAATGCAGA TTGAGATTAT GTAGTATTAT TATGTACCAT GTCCTCCAA ATTGTTTGAG | 1910 |
| CTTGAGCAAC AACTGAACCT CCCTCTAGAT CAATCAGCAA TGCATTTTTA TATTCATGAT CAGATAAAAA AAAAAAAAAA | 1990 |

Figure 7. Nucleotide sequence and deduced amino acid sequence of

```

aaPST2a      1
aaPST2more   1
aaPST3 corrected 1
aaPST4E1     1
aaPST5_F1    1

```

0
0
- 19
- 16
60

M S T G T D A V A V A V A F A K R P E
M A S A M L R Q A I E P R K X 3 - -

K I G I V R A A D D V E A G V V G A A T T T E P L L L R H C H G S K E E D H Q E S X I Q G S P E R A A S D C G S E G R P D

```

aaPST2a      1
aaPST2more   1
aaPST3 corrected 20
aaPST4E1     17
aaPST5_F1    61

```

56
51
74
71
116

M S G A A V A I A A S I G M L L Q G W D M A T I S R A L L Y I K K E F K L E S E P T V E G L I V A M S L I G A
- - - I N E Y R F A C A L L A S H N S V L L G Y D I S V M S G A Q L F M E D L K I - T D T Q I E I L A G V I R H Y - 3
- - - M V K Y A S I C R I L A S H A S V I L G Y D I G V M S G A A M Y I K K D L K I - T D V Q L E I H I G I L S V Y - 3

A A G S L R M V L S T A V A V C G S F E F G T V G Y S A F T Q S G L V D E V G L - S I T E F A I F G S I L T I G - R

```

aaPST2a      57
aaPST2more   52
aaPST3 corrected 75
aaPST4E1     72
aaPST5_F1    119

```

116
111
134
131
176

T I I T I F S G P V S D W I G K L P M L I L S S V L Y F V S S L I H L W S F M V V L L L A R L I N G E F G V G L A V T L
T V I T T F S G A V S G S V G R R P M L I A S A V L Y F V S G L V H L W R P S V Y V L L L A R L I D G F G I G L A V T L
L F G S L A A G F T S D W L G R E Y T M V L A R A I F F T G A L L H G L A P D Y G L L M V G R F V A G T G V G F A L M I
L I G S F A G A R T S A R I G R E L L T V V F A L I F E V G A L L H G F A V N V G L M V G R F V A G T G V G Y G G M I
M V G A V S G R L A D E L G R K M T H E I S A I I C I F S G L S I H L A K S R I M L Y F G R I L L G E S T G V L S Y V

```

aaPST2a      117
aaPST2more   112
aaPST3 corrected 135
aaPST4E1     132
aaPST5_F1    179

```

175
170
193
190
230

V P L Y I S E T A P F E R G L L N T L F Q F S G S G G M F L S Y C H V E G M S L F P D - W R I H L G V L A L P S L
V P Y I S E T A P T D I R G L L N T L F Q F S G S G G K P L S Y C H V E G M S H M P T P D - W R L H L G G L I P I P S L
A P V Y T R E V A P T S A R G F L T S F E D V T N M F G I L L G Y V S R E A F A R L P V H L S W R A M T L V G A V P - P
A P V Y T R E I S P A A S R G F L T S F E D V T N M F G I L L G Y V S R E A F A R L P L H L G W R V H L G I G A A P - S
V P V F T A E I A P K N L R G G L A T S N Q L L I C S G S S A T Y I I G - - - - - A L V A - - - - - W R N L V L V G L L P - C

```

aaPST2a      176
aaPST2more   171
aaPST3 corrected 194
aaPST4E1     191
aaPST5_F1    231

```

235
230
229
226
266

F F F G L T I F Y L P E S P R W L V S K G R M A E A K K V L Q R L E G X D V T G E H A L L V E G L E V G K D T F I E E
I Y F G L T V E Y L P E S P R W L V S K G R M A E A K R V L Q R L E G R E D V S G E H A L P V E G L E V G K D T R I E E
V E L G V A V L A B E E S P R W L V M R G R I D D A R R V L Q K T S D S -
A L L A L M V F V H P E S P R W L V M K G R L R A D A R A V L E K T S A S -
V L L L A G L E F F I P E S P R W L A N V G R E K E F H T S L Q K L E G E -

```

aaPST2a      236
aaPST2more   231
aaPST3 corrected 230
aaPST4E1     227
aaPST5_F1    267

```

295
289
235
232
267

Y I I G F A I P R P D D H V R D G D P R I I L F P F E E G Q S W I R E P S K G S M L G S V L S I G S R H G S H V N Q
Y I I G P D D E L A D E G L A P - D P M X I K L Y G P E E L G L S W V A R P V R G Q S A L G S A L G L I S R H G S K A S Q
-
-
- -

```

aaPST2a      296
aaPST2more   290
aaPST3 corrected 236
aaPST4E1     233
aaPST5_F1    268

```

355
348
254
249
279

M L P L T D P I V T L F G S V H E M I S L A G G S H E R I V P S K F D S M I S I T D Q H P K T E Q W D E M L H R D M E
G K P L V D P V V T L F G S V H E K H P E I H G S H E S T L F P M F G S M F S V A E Q Q V K A D W D A M - 3 Q R E G D
R - - L L D -
R - - L A D -
-
- -

```

aaPST2a      356
aaPST2more   349
aaPST3 corrected 255
aaPST4E1     250
aaPST5_F1    280

```

415
406
262
253
287

E Y A T D G A S D V E N N V H S P L L S R Q M T S A E G K C I A H H H R G S A L G L R R E S S L S D E G G E A T S T
D Y A S H G G D D E D M L Q S P L I S E R Q A T S V E G K C I A A P - H G S I M G A V R E S S L Q G G E A V S S M
-
-
- -

```

aaPST2a      416
aaPST2more   407
aaPST3 corrected 263
aaPST4E1     258
aaPST5_F1    268

```

475
464
279
272
296

G I G G G Q L A Q X W S E R D G E D G K K E G S E K R I Y L H Q E G V A G S R E G S V V S L P G G G D A S E G G K F I
G I G G G Q L A Q X W T E R E G E D G K K E G S E K R I Y L H E E G V Q G M R - G S I L S L P G G - D V P P G G E F V
- N K G S S R H D G V W X M L L I M -
- R M G S E - M Q V W X A L I F S -
- -

```

aaPST2a      476
aaPST2more   465
aaPST3 corrected 280
aaPST4E1     273
aaPST5_F1    297

```

535
524
297
290
312

H A A A L V S Q S R L Y P R D I T E Q R M A G P A T H H P S E A A R K V P S W R D L F E P G V R E A L L V G I G I Q I L
Q A A A V S Q P A L Y S X E L L E Q H A A G P A H H P S E A V A K P R W A D L F E P G V K K A L F V G I G I Q I L
-
-
- -

```

aaPST2a      536
aaPST2more   525
aaPST3 corrected 296
aaPST4E1     291
aaPST5_F1    313

```

594
583
357
350
366

Q Q F A G I N G V L Y Y T P Q I M E Q A G V A V L I S M L G L S A S - A S I L I S S V T R L L M L P S I G L A M R L H
Q Q F A G I N G V L Y Y T P Q I L E Q A G V G V L S M S L G L S A S S - A S I L I S A L T L L M L P S I G I A M B L M
Q Q A T G V D C V V M Y S P R V F E R A G I K X K T M S L G A S M A V G A C K T F I F I S T L L D E I G R E P L L L
Q Q A S G S D S V V L Y S P V F E X S A G I T G S M K X L G V T C A V G V T K T F I L V R T F L L D R V G R P L L L
Q Q L G G I N G V G F Y A S Y I E S S A G F S G - - - - - K L G T I L I G - I I Q I F I T L F G A I L M D R S G R R V L L H

```

aaPST2a      595
aaPST2more   564
aaPST3 corrected 356
aaPST4E1     351
aaPST5_F1    369

```

651
643
409
402
416

D V S G R - - - E F L L L S T I P V L I A S L I V L V V S M V I E L G T V V H A V L S T I S V I T Y L C E F X M G F G P
D M S G R E F L R L L L A T I P I L I V A L A I L V L V M I Q D V G T M V H A A L S T I S V I V Y F C F E V M G F G P
A S G G G - - - - - M A I F L F T A T S L H M M D R E F E G E A A L G A V S I A M L S E F V S F E A S G L G P
T S A G G - - - - - M I I S L I G L G T S L T V V G H H S D A K I P S A V A L C I A S T L A Y V A F F S I G L G P
V S A S G - - - - - T L G C F L T G V S F Y L K A Q G - - L L P E Q V P T L A L S G L V Y I G A Y S I G M G P

```

aaPST2a      652
aaPST2more   644
aaPST3 corrected 410
aaPST4E1     403
aaPST5_F1    419

```

711
703
469
462
477

I P M I L C A F P F T R V R G I C I A I C A L I P V V G D I V T Y S L P V H L M A I G L G V F G F Y A V A C A I A
I P M I L C A E T F T T V R G I C I A I C A L T F W I G D I V T Y T L P V H L M A I G L R G V F G F Y A I V C I L A
V A Q V H Q S E I Y P L R L R A Q A A I G T G L N R I B S G A T T B S F L S E S M T I T I A G S E Y L Y A C I A A A G
I T G V Y H S E I E P L Q V R A L G V A V G G A C N R V T S A V I S M T F L S L S K G I T H G S E F L Y S G I A V V G
V F Q V H S E I E S I M N K A I G S L V T L V S W L G S F A I S Y S F S F L H D D Q S - S A G T F E R E S A A S L V T

```

aaPST2a      712
aaPST2more   704
aaPST3 corrected 470
aaPST4E1     463
aaPST5_F1    476

```

745
739
522
522
505

F V F V Y L K V P E T K G M P L E V I T E F F A V G R K Q A V A K A
L V F V Y I K V P E T K G M P L E V I T E F F S V G A K Q A E A R E D
V V F M F F L P E T H G K S L E D T V K L E F G K D A D E D V G D S R R H V P S Q K P S T E L S A Q M
V V F F T C L P E T E G R T L E E M G K L E G M P D T D M A E A D N A A A K E X V E M I D A A S Y Q S T A Q L K Q R
V L F V A K L V P E T K G R T L E E I Q M S L R T R R N

```

aaPST2a      746
aaPST2more   740
aaPST3 corrected 523
aaPST4E1     523
aaPST5_F1    506

```

745
739
522
520
505

746 745
740 739
523 522
523 520
506 505

Figure 8. CLUSTALW Alignment of the amino acid sequences for PSTs 2a, 2b, 3, 4, and 5. Shaded and boxed areas indicate regions of homology. Double-headed arrow indicates conserved motif.