Gene discovery as an aid to the understanding of agronomically important metabolic processes of sugarcane

Rosanne Casu
Background
Australian Sugar Industry

- 6,000 cane growers (4,500 cane farming business operations)
- 35 million tonnes sugarcane
- 4.75 million tonnes raw sugar
- 80% exported
  Major export customers include Japan, Korea, Malaysia, Taiwan, Saudi Arabia, New Zealand, Canada and USA
- $1.75 billion to the Australian economy
Historical trends in conventional plant breeding have favoured yield increase rather than sucrose content.

Results from trials in the Herbert region, North Queensland (Jackson, 2005)
Sucrose increases exponentially down a sugarcane stem during growth and development.

Sucrose as % of dry matter

Internode number from top

(Chris Grof)
Historical perspective
Sugarcane and the “genomics” paradigm

- **Functional Genomics**
  The expansion of biological investigation from studying single genes or proteins to **studying all genes or proteins at once in a systematic fashion**

- **30th June 1996**
  23 entries for Saccharum*[Organism] at GenBank
  (17 of these correspond to 28S rRNA sequence)
Genomics and bioinformatics for productivity outcomes in sugarcane in Australia

- A sugarcane genomics project commenced in 1998 at CSIRO. It has received funding from:
  - CSIRO CEO Special Project (Genomics and Gene Discovery for Australia: 1998 – 2000)
  - Cooperative Research Centre for Sugar Industry Innovation through Biotechnology (CRCSIIB: 2003 - present)

- Traits targeted
  - Sucrose accumulation
  - (Wounding of sugarcane roots – minor trait)
Gene Discovery
Australian sugarcane gene discovery project strategy

Immature stem → 1078
Maturing stem → 7234
Roots (MJR) → 829
CSIRO/SRDC sugarcane stem EST collection

- 8324 ESTs
  - Immature stem: 1082
  - Maturing stem: 7242

Bar chart showing the distribution of ESTs across various functional categories:

- ATP synthesis/electron transport
- Carbohydrate metabolism
- Cell division cycle
- Cell wall structure or metabolism
- Chromatin and DNA metabolism
- Cytoskeleton
- Defence/stress related proteins
- Fibre biosynthesis and degradation
- Gene expression and RNA metabolism
- Membrane transport
- Miscellaneous
- No assigned function
- Novel
- Primary metabolism
- Protein sorting and secretion
- Protein synthesis and processing
- Secondary and hormone metabolism
- Signal transduction
Carbohydrate metabolism-related ESTs

- 4-alpha glucanotransferase
- 6-phosphogluconate dehydrogenase
- 6-phosphogluconolactonase
- beta-glucanase
- beta-glucosidase
- invertase
- cellulose synthase
- enolase
- fructokinase/sugar kinase
- fructose-1,6-bisphosphatase
- fructose-2,6-bisphosphatase
- fructose bisphosphate aldolase
- glucose-6-phosphate dehydrogenase
- glucose-6-phosphate isomerase
- GAPDH
- glycogenin
- phosphoglucomutase
- phosphoglycerate kinase
- phosphoglycerate mutase
- PFP
- ribulose-phosphate 3-epimerase
- starch synthase
- sucrose phosphate synthase
- sucrose synthase
- transaldolase
- transketolase
- transporters
- trehalose-6-phosphate synthase
- triose phosphate isomerase
- UDP-glucose dehydrogenase
- UDP-glucose pyrophosphorylase
- UDP-glucuronyltransferase
- UDP-N-acetylglucosamine pyrophosphorylase

*immature stem
maturing stem
Maturing stem transcript bioinformatics

- Sugar transporters are the most up-regulated carbohydrate metabolism genes
- Fibre biosynthesis, defence and oxidative stress transcripts are up-regulated
- Transcripts related to cell division, growth and development are down-regulated

- The major processes in maturing stem relate to sugar accumulation and fibre biosynthesis
Identification of an N-terminal vacuolar signaling peptide
Endopeptidase MCSA201C03 is up-regulated in the maturing stem of sugarcane varieties

(Casu et al. 2004)
Sugarcane legumain homologue

- Cysteine protease (family C13)
- Vacuolar Processing Enzymes (VPEs)
- Most reside in vacuole

Also present in 4 other sugarcane vacuolar proteins

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugarcane</td>
<td>MVTA&lt;sup&gt;L&lt;/sup&gt;RRLALLLL&lt;sup&gt;L&lt;/sup&gt;SVFLC-----SAWARPRLEPTIRLPSERAAAAGDE&lt;sup&gt;T&lt;/sup&gt;TD&lt;sup&gt;AV&lt;/sup&gt;</td>
</tr>
<tr>
<td>Zea mays</td>
<td>MVADRLRLALLLS-&lt;sup&gt;A&lt;/sup&gt;CLC-----SAWARPRLEPTIRLPSERAAAADETD&lt;sup&gt;T&lt;/sup&gt;DAV</td>
</tr>
<tr>
<td>N.tabacum</td>
<td>MIVRYVV&lt;sup&gt;S&lt;/sup&gt;AAILLII-&lt;sup&gt;G&lt;/sup&gt;LSS-----VVAADVGR-DVLKLPS&lt;sup&gt;E&lt;/sup&gt;ASTFFSGNYDD&lt;sup&gt;D&lt;/sup&gt;S&lt;sup&gt;I&lt;/sup&gt;</td>
</tr>
<tr>
<td>A.thaliana γ</td>
<td>TRVSVG&lt;sup&gt;V&lt;/sup&gt;VLFVLLVS-&lt;sup&gt;-&lt;/sup&gt;LVA-----VSAARSGPDDVIDKLPSQA&lt;sup&gt;S&lt;/sup&gt;RFRPA&lt;sup&gt;V&lt;/sup&gt;END&lt;sup&gt;D&lt;/sup&gt;D&lt;sup&gt;S&lt;/sup&gt;</td>
</tr>
<tr>
<td>A.thaliana α</td>
<td>MTTVVSFLALFLFLV-----&lt;sup&gt;-&lt;/sup&gt;AAVSG--DV&lt;sup&gt;I&lt;/sup&gt;KLPSLASKFRPTEND&lt;sup&gt;D&lt;/sup&gt;D&lt;sup&gt;S&lt;/sup&gt;</td>
</tr>
<tr>
<td>A.thaliana β</td>
<td>KSCYFRPALLLLVVLLVHAE---------SRGRFEPK&lt;sup&gt;I&lt;/sup&gt;LMPTEE--ANPADQDEDG</td>
</tr>
<tr>
<td>R.communis</td>
<td>METHKSSLLFF&lt;sup&gt;T&lt;/sup&gt;NYYVLFLVF&lt;sup&gt;T&lt;/sup&gt;LSPGLLLAS&lt;sup&gt;R&lt;/sup&gt;LNPF&lt;sup&gt;E&lt;/sup&gt;PG&lt;sup&gt;I&lt;/sup&gt;LMPTEE---AEPVQV&lt;sup&gt;V&lt;/sup&gt;DDDDD</td>
</tr>
<tr>
<td>A.thaliana δ</td>
<td>MSSPLGHFQILVFLHAL&lt;sup&gt;L&lt;/sup&gt;IFS--------AESR---KTQ&lt;sup&gt;L&lt;/sup&gt;LLLND----NDVE&lt;sup&gt;S&lt;/sup&gt;SS&lt;sup&gt;D&lt;/sup&gt;K&lt;sup&gt;S&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Targeting proteins to the vacuole

- 15-20 hydrophobic amino acids
- Controls transit to E.R.

- Vacuole targeting signals
- Cleaved in vacuole
- NTPP for lytic vacuole
- CTPP for protein storage vacuole
Analysis of GFP fusions

See poster MON-079 by Jackson et al.

Cytosolic GFP

Secreted GFP

Vacuole localised GFP

5 amino acid motif -IRLPS- diverts GFP from the default secretion pathway to the lytic vacuole
Confirming vacuolar localisation

Vacuole Targeted GFP  DAPI Nuclear Stain

Vacuole Targeted GFP  Fluorescent protease substrate
Localisation in parenchyma cells

Vacuole Targeted GFP appears around nuclear envelope and can label the ER.

Does the acidity quench GFP fluorescence?

Inhibition of H^+ATPase by ConcanamycinA
The sugarcane transcriptome
Custom microarray expression profiling

- **EST clustering and data-mining**
  - Database development for efficient data-mining
  - 9,149 ESTs derived from three tissues (7,242 from maturing stem) clustered
  - Identification of the most 5’ EST i.e. longest EST in each cluster to form the “non-redundant EST set”

- **Microarray design & preparation: CaneArray 1 (> 5,500 features derived from NR EST set)**
  - Bioinformatics and microarray analysis of transcripts in maturing stem involved in
    - Carbohydrate metabolism (Casu et al., 2003, PMB, 52, 371-386)
    - General metabolism associated with sucrose accumulation (Casu et al., 2004, PMB, 54, 503-517)
  - Transcript analysis of maturing stem from high and low CCS individuals from a segregating population (Casu et al., 2005, Field Crops Res., 92, 137-147)
Analysis of gene expression as a strategy for sucrose accumulation gene discovery

- **Bioinformatics**
  - Develop sequence and clone bank
  - Integrate with world collection

- **Gene expression analysis**
  - Function
  - Role
  - Utility

- **DNA markers**
  - Cell localisation studies
  - Transgenic tests
The sugarcane “transcriptome” @ GenBank

  Leaf roll and maturing stem

- “Rossi” RGA ESTs: 54 25/4/2003

  - Young cane stem (YCS): 1,078
  - Maturing cane stem (MCS): 7,242
  - Methyl jasmonate-treated roots: 829

- “Nogueira” cold response ESTs: 1,219 16/7/2003

- USA sugarcane ESTs: 8,125 6/8/2003
  - Apex: 3,329
  - Stem: 2,268
  - Leaves: 2,396
  - Misc: 132

- SUCEST (40 cDNA libraries): 236,916 24/9/2003
8,320 ESTs (CSIRO) \(\rightarrow\) 255,135 ESTs (World)
# Sugarcane Gene Index @ DFCI (ex TIGR)

## Table 1 ---- Summary of total unique sequences

<table>
<thead>
<tr>
<th>Release</th>
<th>Number of unique sequences</th>
<th>TCs</th>
<th>Singleton ESTs</th>
<th>Singleton ETs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release 2.2 July 29, 2008</td>
<td></td>
<td>40016</td>
<td>76529</td>
<td>43</td>
<td>116588</td>
</tr>
<tr>
<td>Release 2.1 October 13, 2004</td>
<td></td>
<td>24711</td>
<td>53795</td>
<td>41</td>
<td>78547</td>
</tr>
<tr>
<td>Release 2.0 September 20, 2004</td>
<td></td>
<td>24356</td>
<td>53482</td>
<td>41</td>
<td>77879</td>
</tr>
<tr>
<td>Release 1.0 January 6, 2004</td>
<td></td>
<td>23596</td>
<td>72281</td>
<td>7</td>
<td>95884</td>
</tr>
</tbody>
</table>

**Scale:**

---

**Key:**
- Red = number of TCs
- Green = number of singleton ESTs and ETs

---

[CRC Sugar Industry Innovation through Biotechnology]
DFCI Sugarcane Gene Index

Saccharum officinarum in Wikipedia

About SoGl Gene Index

Development and Goals: Background information about SoGl
Release Summary: Display a statistical summary of all SoGl releases
Category Comparison: Display estimated number of genes among all plant releases

Sequence Similarity Search

BLAST: Search TC sequences based on sequence similarity

Sequence Reports

Identifiers or Keywords: Search TC reports using TC identifiers, GB accessions or keywords
TC Annotator: List all TC annotation
EST Annotator: List all EST annotation
Libraries: Search EST libraries by keywords or tissue origins
CAT# Download: Download EST and TC sequences originating from one library

Functional Annotation and Analysis

Alternative Splice Forms: Prediction of alternative splice variants
EST Expression: Compare EST expression between different libraries or tissues
Gene Ontology: Classification of TCs by GO vocabularies
Metabolic Pathways: Association of TCs with metabolic and signaling pathways
Oligomer Prediction: List all 70-mer oligo predictions

Release 2.2 (July 29, 2008)

Input Sequences
- ESTs: 25635
- ETs: 499

Output Sequences
- TC sequences: 40016
- Singleton ESTs: 76529
- Singleton ETs: 43

Total unique: 116588

Attributions
Photo Courtesy: Vascular Plant Image Library, Digital Flora of Texas
<table>
<thead>
<tr>
<th>DFCI Index 1.1 config #</th>
<th>DFCI Index 1.2 config #</th>
<th>NCBI Index 1.1 config #</th>
<th>Genbank accession number</th>
<th>EST</th>
<th>Identify</th>
<th>Common name</th>
<th>GO terms</th>
<th>EC number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC14017</td>
<td>TC14018</td>
<td>35105168</td>
<td>CA173567</td>
<td>SCULST1025802</td>
<td>glucone synthase</td>
<td>M6E1100</td>
<td></td>
<td>2.4.1.13</td>
</tr>
<tr>
<td>TC14006</td>
<td>TC14008</td>
<td>35295264</td>
<td>CA226712</td>
<td>SCSVFL3026321</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.13</td>
</tr>
<tr>
<td>TC14006</td>
<td>TC14008</td>
<td>35092490</td>
<td>CA226886</td>
<td>SCSL1067302</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.13</td>
</tr>
<tr>
<td>TC14021</td>
<td>TC14024</td>
<td>35256111</td>
<td>CA211105</td>
<td>SCBF02110504</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35273506</td>
<td>CA220359</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35074767</td>
<td>CA150604</td>
<td>SCWSR1100400</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35138075</td>
<td>CA152154</td>
<td>SCCEST1002411</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35069010</td>
<td>CA166820</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35030570</td>
<td>CA166511</td>
<td>SCCEST1002411</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35071166</td>
<td>CA163017</td>
<td>SCDFB0210062</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35010815</td>
<td>CA174321</td>
<td>SCFST1000411</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35050010</td>
<td>CA148750</td>
<td>SCJLST1005102</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>34933918</td>
<td>CA985238</td>
<td>SCWAM3004008</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>34093010</td>
<td>CA098612</td>
<td>SCWAM3004008</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35074006</td>
<td>CA150097</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35722934</td>
<td>CA217238</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>34593708</td>
<td>CA015603</td>
<td>SCWAM3004008</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35040006</td>
<td>CA252765</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35261266</td>
<td>CA211105</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35062574</td>
<td>CA211345</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35062574</td>
<td>CA211345</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
<tr>
<td>TC14008</td>
<td>TC14010</td>
<td>35240397</td>
<td>CA231965</td>
<td>SCBF3042208</td>
<td>glucone synthase</td>
<td></td>
<td></td>
<td>2.4.1.10</td>
</tr>
</tbody>
</table>
Expression profiling – the advent of the commercial array
Large-scale expression profiling

- Access to the world collection of sugarcane ESTs fostered the development of the first commercial expression profiling tool for sugarcane
- Association with Affymetrix to develop the GeneChip® Sugarcane Genome Array
  Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane (Casu et al., 2007, Funct. Integr. Genomics, 7, 153-167)
  Further identification of targets for tissue-enriched expression and manipulation of sucrose accumulation
Expression differences between closely-related transcripts can be distinguished

Immature stem (meristem, I1-3)
Low sucrose

Maturing stem (I8)
Accumulating sucrose

Mature stem (I20)
Accumulated sucrose

Leaf susy2
Root susy3
YCS susy2/1
GeneChip-derived expression data correlates consistently with that achieved using RT-qPCR

204 data points (612 RT-qPCR reactions) = 3 tissues x 4 biological replicates x 17 transcripts
CesA and Csl transcripts are differentially expressed in stem tissue
Expression profiling to aid promoter identification
**Parenchyma, Vascular bundles and Rind (PVR) expression profiling**

Field-grown Q117 plants (11 months old, 3 biological replicates)  
Tissue dissected from internode 8

(Anne Rae)
Native sugarcane promoters can stably drive GFP transcript expression in maturing sugarcane stem internodes
Transgenic testing of candidate genes

Can the relevance of our candidate genes to the process of sucrose accumulation in maturing stem be tested using RNAi-mediated gene silencing?
RNAi-mediated gene silencing of genes involved in core metabolism in sugarcane

- In an effort to understand which genes are important for sucrose accumulation in sugarcane, we are interested in interfering with the expression of several key genes

- Desired outcomes
  Production of viable transgenic sugarcane carrying a variety of RNAi constructs relating to a core metabolic function i.e. sucrose accumulation
  Perturbation of transcript levels for each target
  Correlation with sucrose, glucose or fructose levels (if possible) or other phenotype
Candidate genes in core metabolism have been down-regulated in gene X-RNAi lines.
Gene discovery to aid sugarcane genetics and genome sequencing
Gene discovery and sugarcane genetics

- Assessment of ESTs as a source of microsatellite markers
  Microsatellite markers derived from ESTs have lower polymorphism within sugarcane varieties but are cross transferable to *Erianthus* and *Sorghum* (Cordeiro et al., 2001, Plant Sci, 160, 1115-1123)

- RFLP mapping
  EST-derived RGAs and other RGAs – association with resistance to brown rust (McIntyre, C.L. et al., 2005, Genome 48: 391 – 400).

- Allele diversity in sucrose phosphate synthase (SPS) gene family III
  SPS gene families initially identified through data-mining

- SNP marker development from candidate genes derived from large-scale expression profiling studies
  SNP marker assays have been developed and are being mapped
Similarity of grass ESTs to sorghum

Summary

- Gene discovery was a necessary first step for commencing genomics research in sugarcane
- Greatest value from the various sugarcane EST programs was achieved when all of the data became available for analysis
- Commercial arrays for large-scale expression profiling hold promise for the development of a databank of expression profiling for further data-mining
- Candidate genes from large-scale expression profiling as well as other studies have been a source for
  - Targeting peptide discovery
  - Promoter discovery
  - Transgenic testing for trait modification, including use of RNAi
  - DNA marker development
- The world sugarcane EST collection is a valuable resource to assist in the production of the sugarcane genome
Acknowledgements

- **CSIRO**
  Karen Aitken  
  Christine Dimmock  
  Hayati Iskandar  
  Janine Nielsen  
  Jai Perroux  
  Merle Thomas  
  Graham Bonnett  
  Donna Glassop  
  Mark Jackson  
  John Manners  
  Karl Pioch  
  Ben Trevaskis  
  Neil Bower  
  Chris Grof  
  Deon Knight  
  Lynne McIntyre  
  Anne Rae

- **BSES**
  Priya Joyce  
  Li Wang

- **Southern Cross University**
  Peter Bundock  
  Frances Eliott  
  Giovanni Cordeiro  
  Robert Henry

- **Funding**
  **CSIRO CEO Special Project**  
  **SDRC**  
  **CRCSIIB**