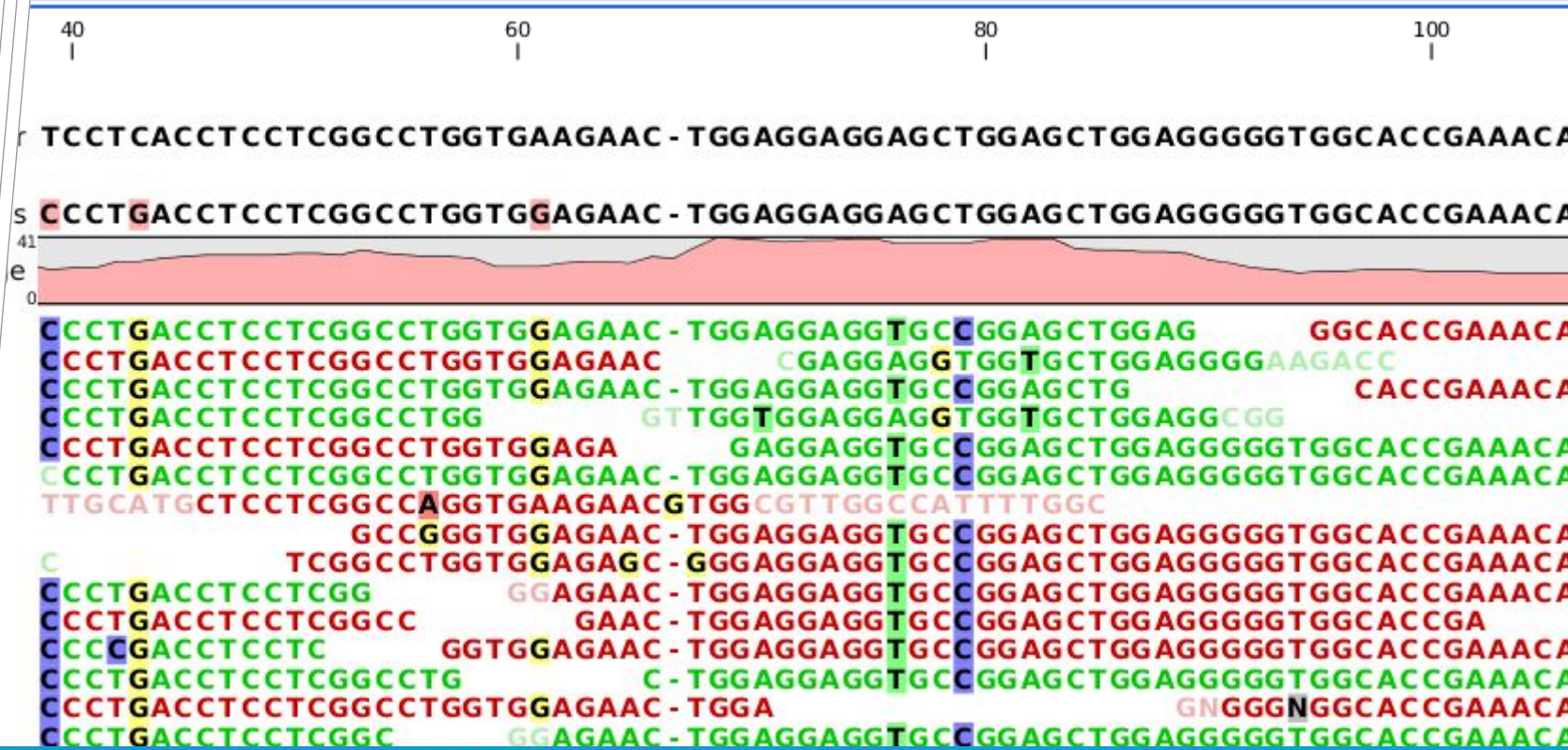


www.csiro.au

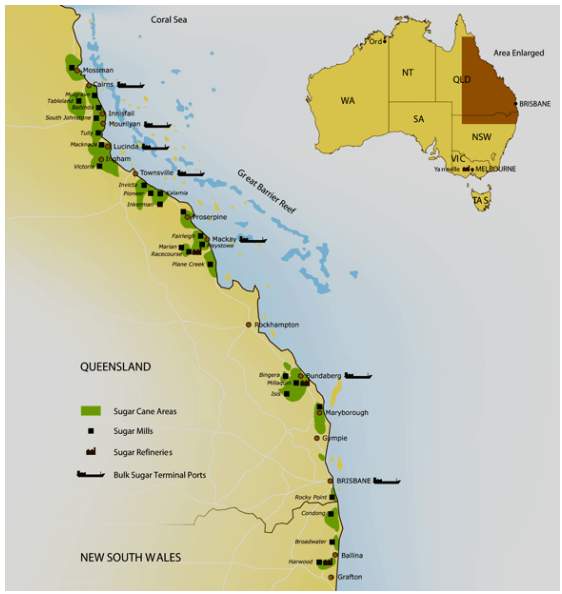
Sequencing the Sugarcane Genome: Sequencing of long insert clones underlying QTL associated with production traits in the Australian Sugar Industry

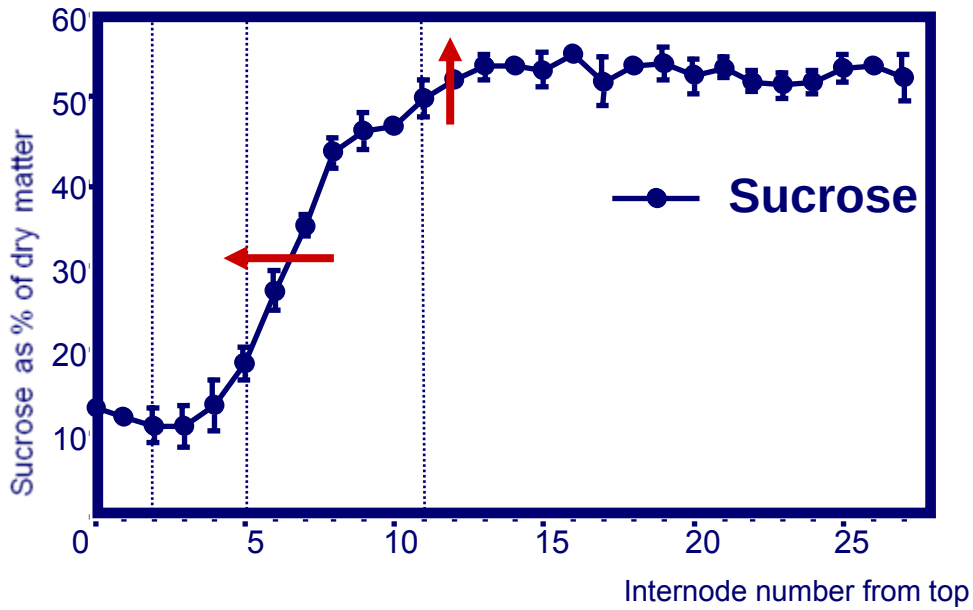
Rosanne Casu
Senior Research Scientist
15 January 2012

CSIRO

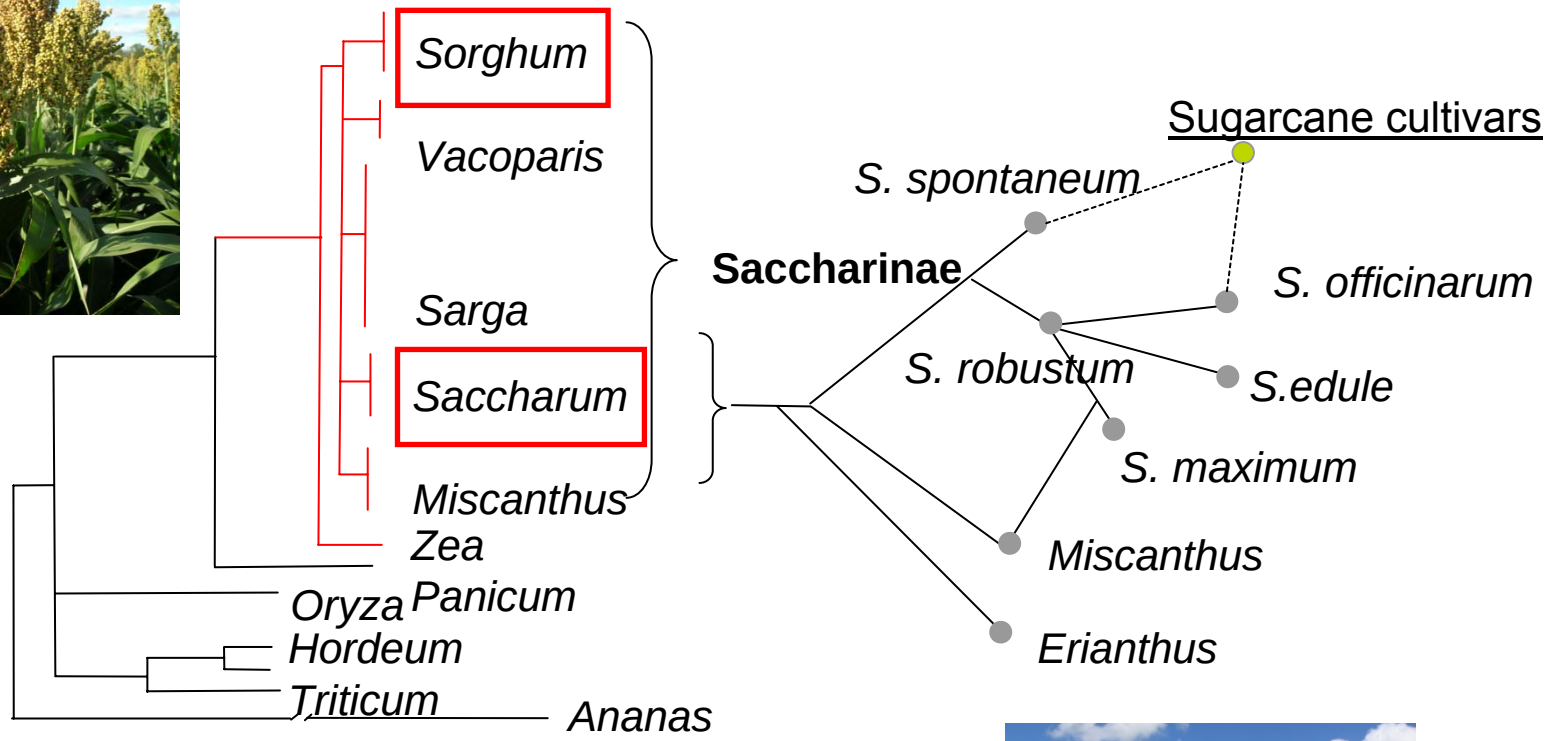
Australian Sugar Industry

- 6,000 cane growers (4,500 cane farming business operations)
- 35 million tonnes sugarcane (4.75 million tonnes raw sugar)
- 80% exported, principally to Asia-Pacific region
- $1.75 billion to the Australian economy



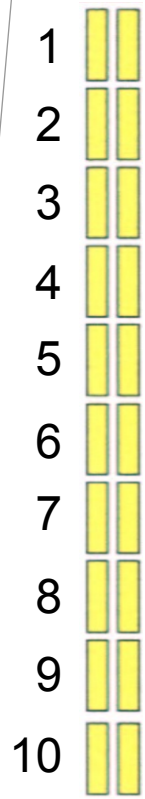


C_4 biomass plants



Adapted from Paterson et al Nature 2009 & Grivet et al 2006 Darwin's Harvest

Sorghum and sugarcane genetics

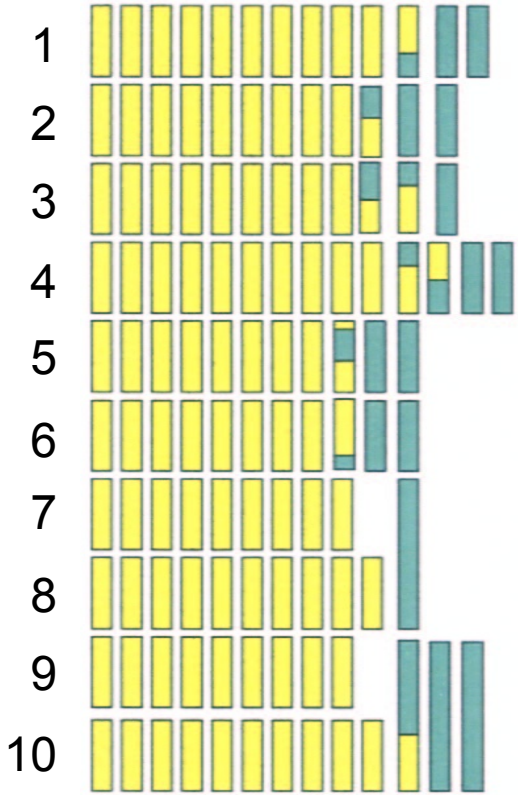


Sorghum

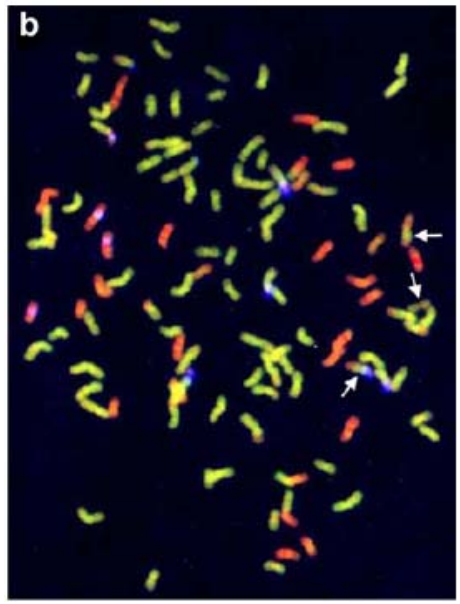
- Diploid
- x=10, 2n=20
- Simple genetics
- Inbred (AA or aa)
- Reference genome

Sugarcane

- Polyploid (10-12x)
- x=8, x=10, 2n=? (100-120)
- Complex genetics
- Heterozygous
- No reference genome

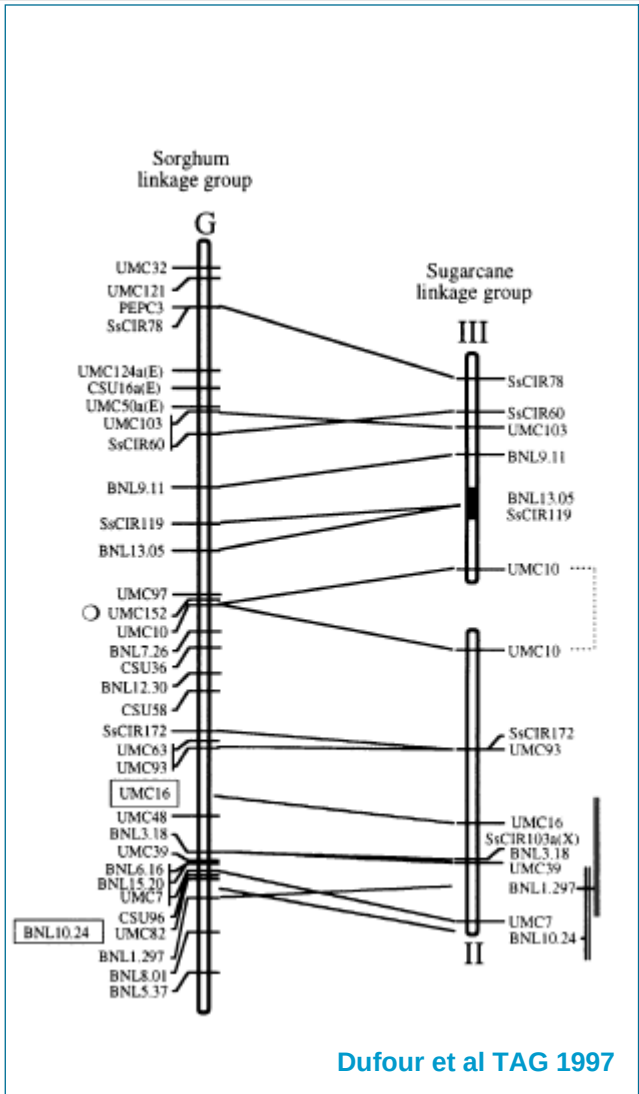


NCo376

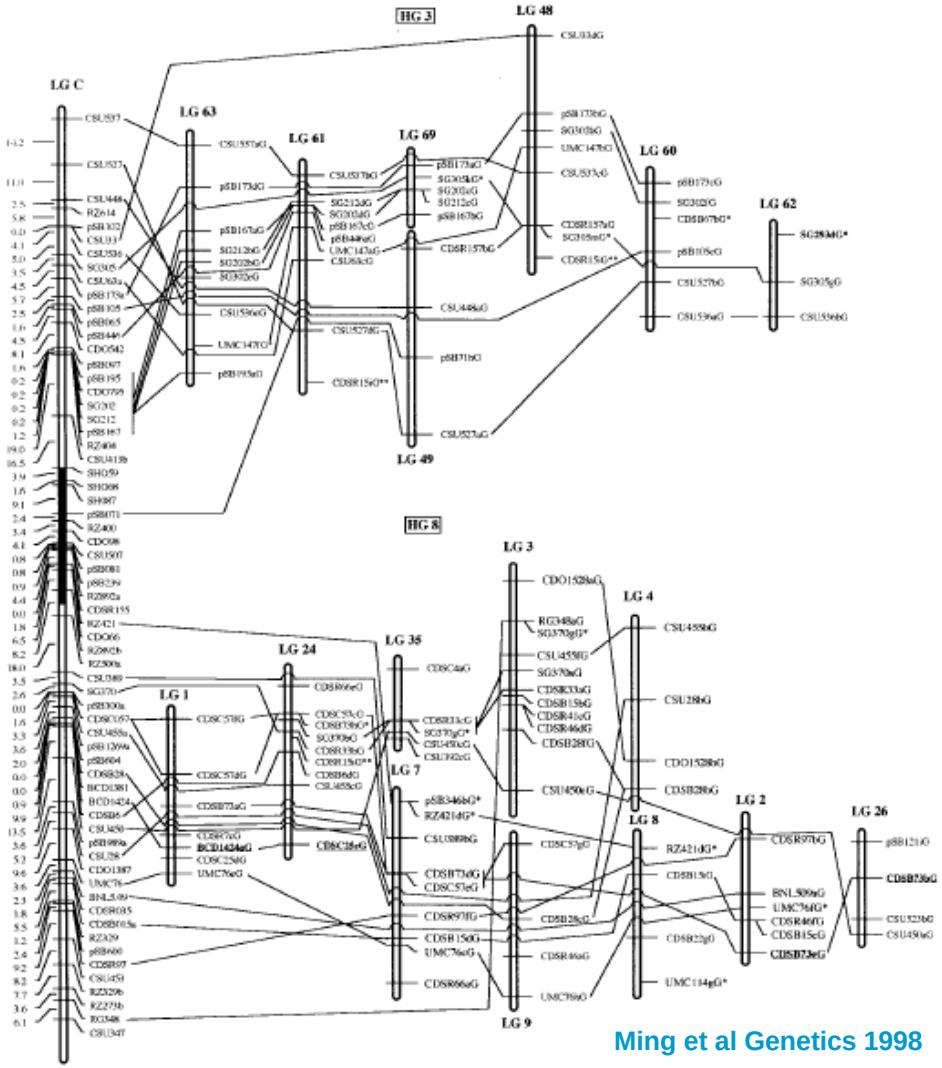


Piperidis et al Mol Genet Genomics, 2010

Synteny between sugarcane and sorghum

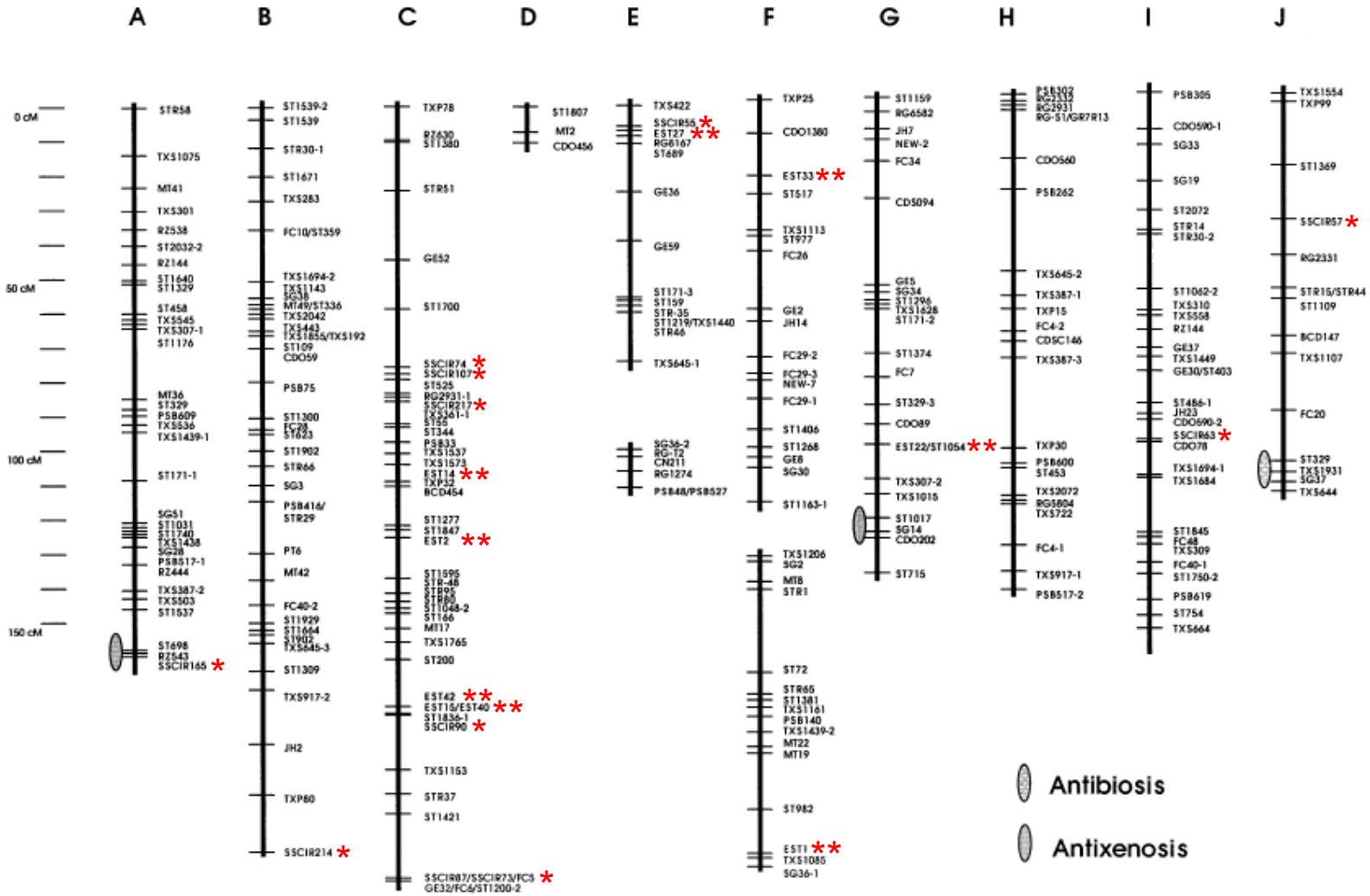


Dufour et al TAG 1997



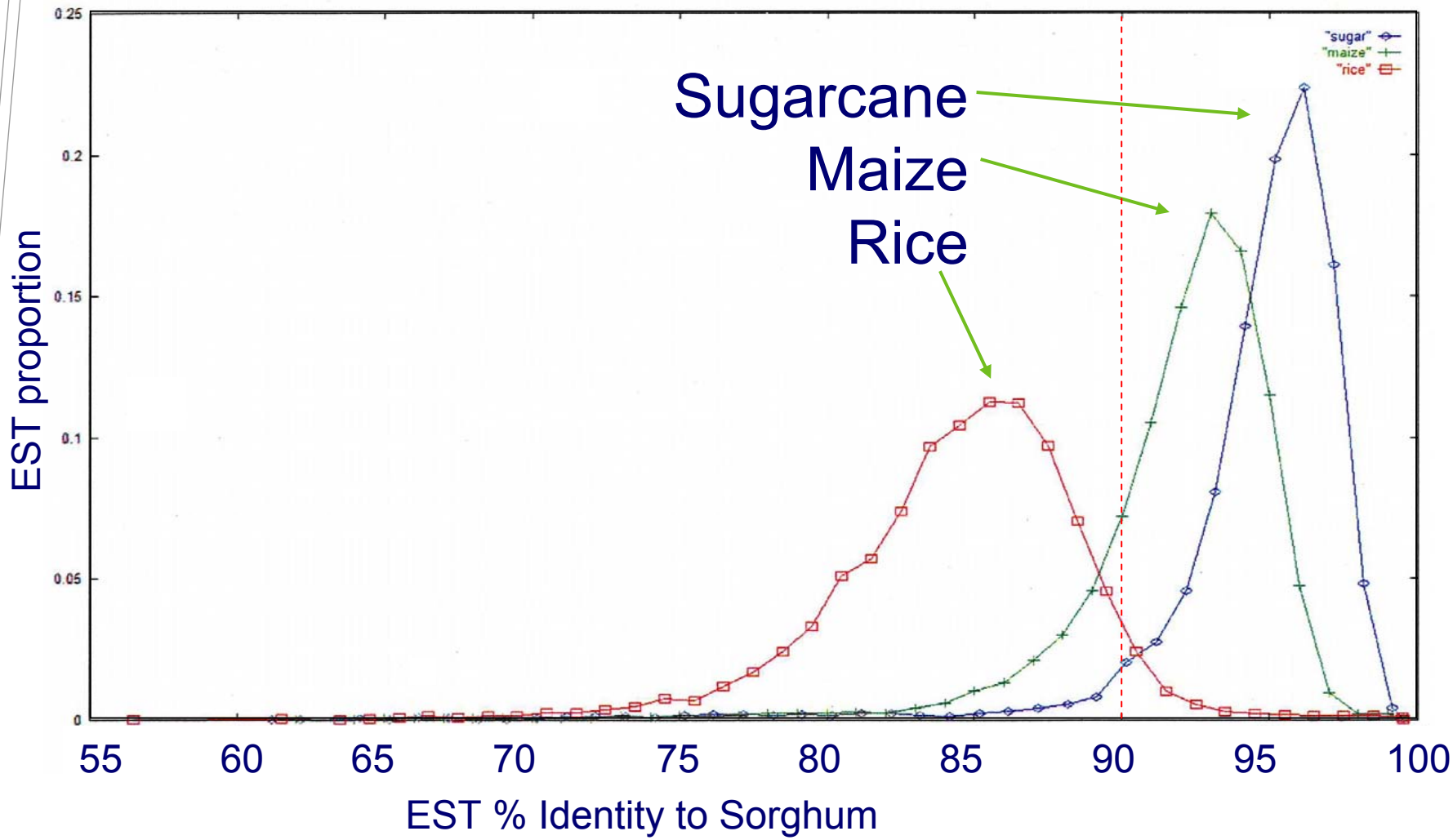
Ming et al Genetics 1998

Synteny between sugarcane and sorghum II



Tao et al (2003) TAG 107:116-122

Similarity of grass ESTs to sorghum



AH Paterson *et al. Nature* **457**, 551-556 (2009)

Aim

To generate sequence data that contributes to a complete sugarcane genome sequence and to identify genes that underpin important agronomic traits in the Australian Sugar Industry

- Generate sugarcane genomic sequences relevant to traits of importance to the Australian sugarcane industry to be contributed to SUGESI (monoploid R570 genome)
- Improve genetic map of Q165
- Develop web-based tools
  - Integrate sugarcane genetic map data with sugarcane genome sequence
  - Add other relevant sugarcane sequence data
  - **Simplify identification of genome regions associated with particular traits**

Generation of sugarcane genomic sequence

- R570 was designated as the reference at the Consortium meeting in August 2009
  - Due to the polyploid genome of sugarcane, whole genome shotgun sequencing is unlikely to result in an assembled genome
  - R570 BAC library already in existence
  - A comprehensive genome map is already available
  - Some members have already commenced sequencing R570 BACs
- Our contribution to SUGESI needs to add value to Australian germplasm
  - Screen the R570 BAC library and select regions that underlie QTL of interest to the Australia sugarcane Industry
  - Generate data that will be the Australia contribution to SUGESI
  - Use the sequence to identify potential genes within Australian germplasm for use in Australian breeding programs

Identification of target genome regions for sequencing

- A number of genetic studies have been carried out over the last decade
  - Biparental mapping
    - IJ76-514 (*S. officinarum*) x Q165
      - 3 years data
    - cultivar x cultivar crosses
      - 2 years data
  - Association mapping
    - 480 clones from core breeding program
      - 1 year of data across three sites
    - 860 parents from the breeding program
      - Up to 10 years data

- Traits selected
  - Sucrose and yield
    - Sucrose content
    - Biomass
    - Sucrose yield
    - Other yield components - stalk weight, stalk number, stalk diameter, stalk height
  - Disease resistance
    - Smut
    - Pachymetra
    - Fiji disease

Identification of targets for sequencing

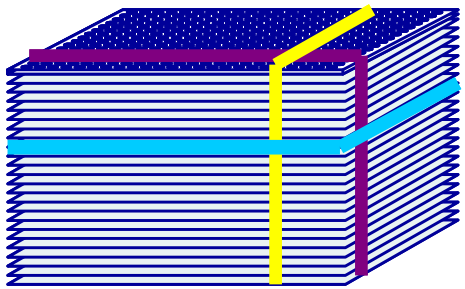
- Identification of markers and quantitative trait loci (QTL)
  - QTL were significant at P≤0.001
  - More than one marker was identified as associated with the trait
  - QTL effects were identified across the homology group
  - Ideally, the QTL region had been identified in more than one population
- Most selected markers were SNP or DArT markers with a known sequence
- Marker sequences were aligned to the sorghum genome to determine the abundance of the sequence
  - Only those aligning to a single chromosomal location in sorghum were selected to screen the BAC clones
  - Preference was given to those that were homologous to members of the sugarcane EST collection
- 200 sequences were selected for primer design with predefined criteria
  - Primer length 20-24bp, %GC ~50%, Tm = 60°C, no primer duplexes, amplicon length under 500bp, single amplicon achieved using R570 genomic DNA
- 112 primer pairs met all of these criteria and were used for screening

3D-pooling strategy of the R570 BAC library

- R570 library
  - Constructed by Tomkins et al. (1999) at Clemson University, USA
  - Total of 103,296 clones
  - Average size of insert of 130 kb
  - Since sugarcane genome size of 10 Gb, library coverage is about 1.3x
  - This suggests that recovery of all alleles would not be always effective

- The 3D-pooling strategy was developed at INRA in Toulouse, France
  - Allows the whole library to be screened with the minimum number of quantitative PCR reactions

  - 269 384-well plates are pooled to obtain ~0.12x coverage per block
    - These form 11 super pools of 24 plates (final super pool – 30 plates)

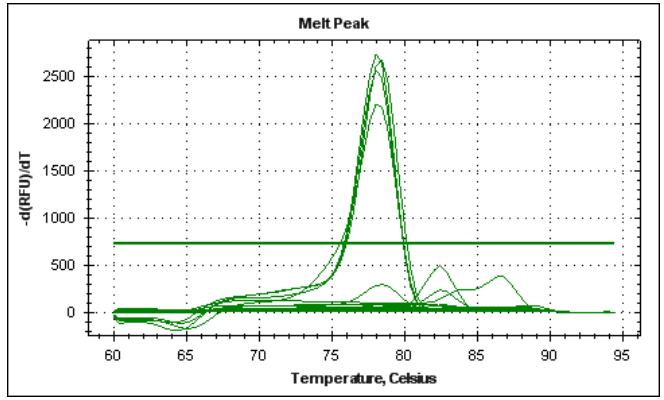


11 super pools of

- 16 row pools
- 24 column pools
- 24 plate pools

Screening super pools and 3D-pools

- 112 primer pairs were screened across the 11 super pools of the sugarcane BAC library
  - 32 primer pairs amplified products in 1 – 4 super pools



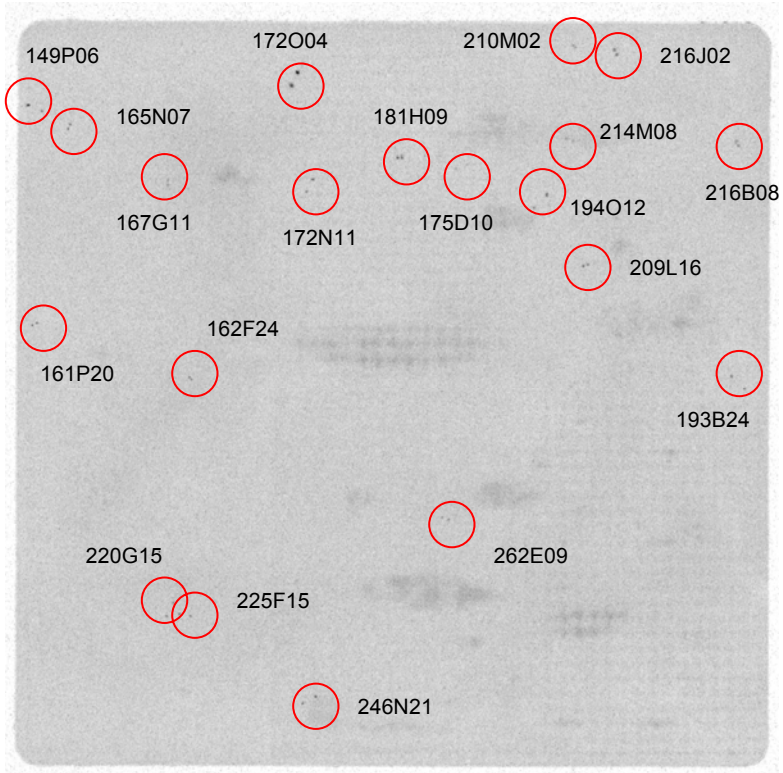
- Re-screened across the 3D-pools to identify positive clones
  - Up to 22 BAC clones were identified with each of these primer pairs
  - 8 primer pairs identified single BAC clones
  - 377 BAC clones in total were identified

A	1	17	P9
B	2	18	P10
C	3	19	P11
D	4	20	P12
E	5	21	P13
F	6	22	P14
G	7	23	P15
H	8	24	P16
I	9	P1	P17
J	10	P2	P18
K	11	P3	P19
L	12	P4	P20
M	13	P5	P21
N	14	P6	P22
O	15	P7	P23
P	16	P8	P24
Block 1			

= plate 15, G2

Macroarray hybridisation

- The R570 BAC library has been gridded in duplicate in a 7*7 pattern onto two filters
- 30 amplicons were screened across the macroarrays
- 140 BAC clones were identified
- Total of **517 BAC clones** identified using both methods



Membrane n3 145-269

BAC sequencing

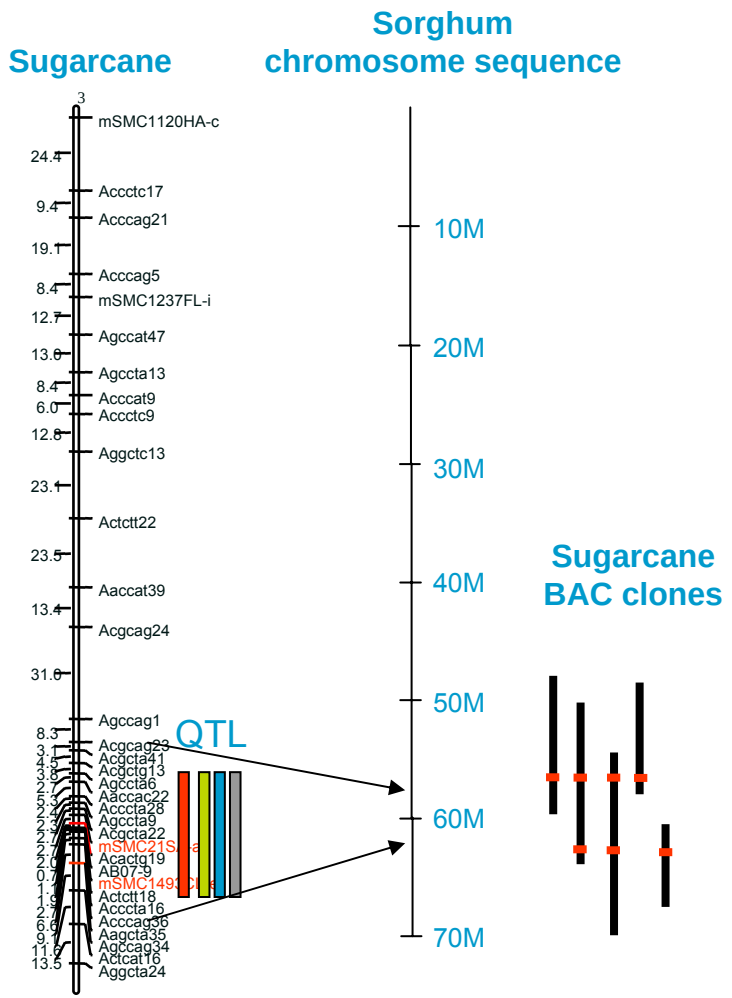
- Using the Illumina HiSeq
  - Sequencing individually tagged BAC clones in pools
    - Paired end sequencing of ~200bp yields 100bp from each end
    - 24 – 96 plex kits now available
    - 187 million reads per lane

Current activities

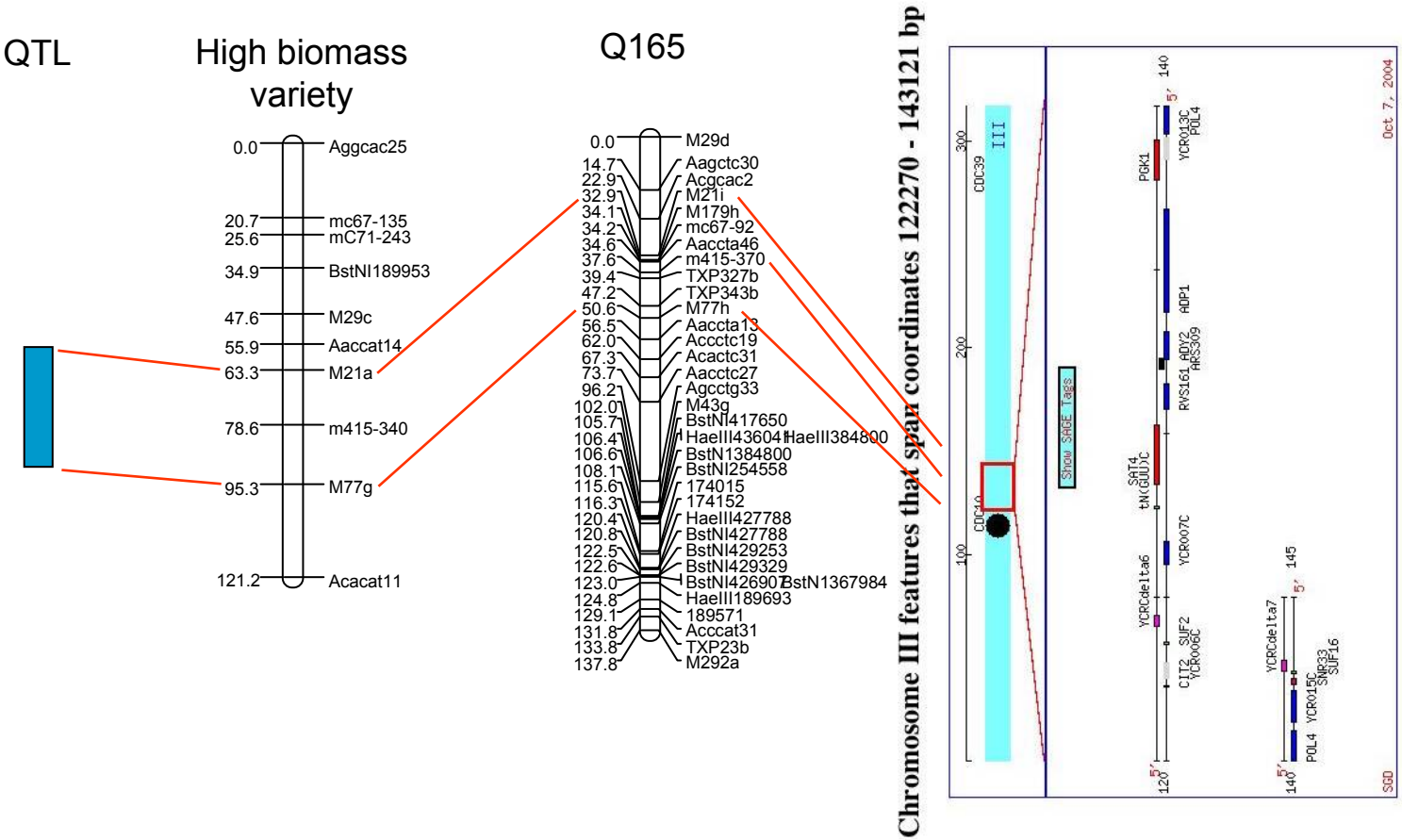
- Sequencing of 72 R570 BAC clones is underway with data delivery expected imminently
- Methodologies for sequence assembly are being developed and trialled with test data
- Additional WGS shotgun sequence will also be obtained to assist in assembly

Future directions

- Sequencing of the remaining identified BAC clones (~ 500 clones in total)
- BAC end sequencing
- Finalisation of BAC assembly strategy and implementation
- Identification and annotation of genes and regulatory elements
- Integration of annotated sequence data with genetic data
- Develop markers tightly linked to causative gene and integrate use into the Australian breeding program
- Contribute R570 BAC sequences to SUGESI for integration into R570 monoploid genome sequence

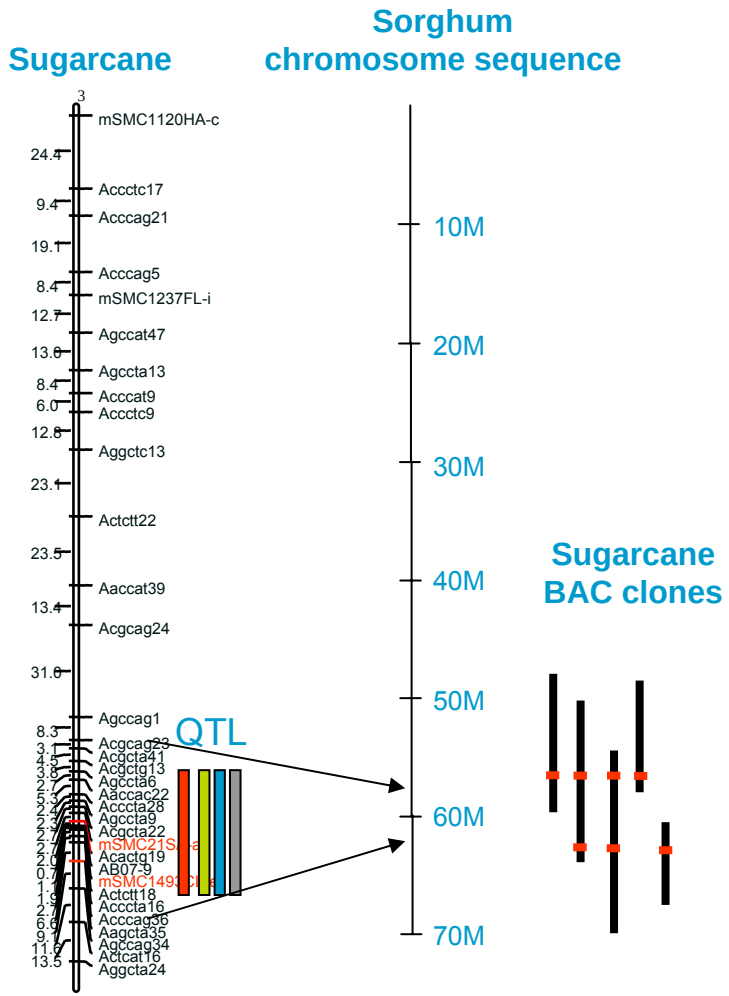


Web-based platform to integrate the genomic sequence and genetic information



Future directions

- Sequencing of the remaining identified BAC clones (>500 clones in total)
- Finalisation of BAC assembly strategy and implementation
- Identification and annotation of genes and regulatory elements
- Integration of annotated sequence data with genetic data
- Develop markers tightly linked to causative gene and integrate use into the Australian breeding program
- Contribute R570 BAC sequences to SUGESI for integration into R570 monoploid genome sequence



Summary

- >500 BAC clones were identified using either RT-qPCR of 3D-pools or macroarray hybridisation of the R570 BAC library
  - These clones were identified with primer pairs corresponding to markers linked to smut, pachymetra and Fiji disease resistance, stalk weight, biomass, stalk diameter, fibre, sucrose content
- Sequencing of the first 72 BAC clones is underway with data delivery imminent
- Methodologies for sequence assembly are being developed

Acknowledgements

- Team members
  - Karen Aitken (Project Leader)
  - Rosanne Casu
  - Paul Berkman
- CSIRO and BSES Ltd sugarcane marker and breeding groups
- Hélène Bergès and her team at INRA, Toulouse, France
- DArT Pty Ltd Team
- Australian Genome Research Facility (AGRF)

- External funding
  - Cooperative Research Centre for Sugarcane Industry Innovation through Biotechnology (CRCSIIB)
  - Australian Sugar Research and Development Corporation
  - Queensland Government Smart Futures Fund (Queensland International Fellowship)

CSIRO Plant Industry
Rosanne Casu
Senior Research Scientist

Phone: +61 7 3214 2364
Email: Rosanne.Casu@csiro.au
Web: www.csiro.au/org/pi.html

www.csiro.au

Thank you

Contact Us
Phone: 1300 363 400 or +61 3 9545 2176
Email: enquiries@csiro.au Web: www.csiro.au