

SHORT COMMUNICATION

GENE DISCOVERY: APPROACHES, DEVELOPMENTS AND APPLICATIONS TO SUGARCANE IMPROVEMENT

WATT DA^{1,2}, POTIER BAM¹, SWEBY DL¹, JACOB RM¹, ALBERTSE EH¹,
McFARLANE SA¹, NAIDOO PV¹ and RUTHERFORD RS¹

¹*Crop Biology Resource Centre, South African Sugarcane Research Institute,
Private Bag X02, Mount Edgecombe, 4300, South Africa*

²*School of Biological and Conservation Sciences, University of KwaZulu-Natal,
Private Bag X54001, Durban, 4000, South Africa*

*derek.watt@sugar.org.za bernard.potier@sugar.org.za deborah.sweby@sugar.org.za
robyn.jacob@sugar.org.za ewald.albertse@sugar.org.za sharon.macfarlane@sugar.org.za
prabashnie.naidoo@sugar.org.za stuart.rutherford@sugar.org.za*

Abstract

Knowledge of the identity of genes participating in the delivery of important traits or displaying tissue- or organ-specific expression is advantageous in the development of rational molecular breeding strategies. Over the past 15 years, the South African Sugarcane Research Institute (SASRI) has enjoyed considerable success in the discovery of such genes, which have been used within marker-assisted breeding and genetic engineering strategies. For example, identified genes have facilitated the development of genetic markers depicting resistance or susceptibility to the stalk borer, *Eldana saccharina*, as well as in the modification of the activity of enzymes involved in sucrose metabolism. Over the years, several technologies have been used at SASRI for gene discovery, including Expressed Sequence Tag Analysis, cDNA Differential Display, Suppression Subtractive Hybridisation and Affymetrix[®] Sugarcane Genome GeneChip analysis. While the application of these technologies has been successful, the recent advent of the next-generation DNA sequencing and gene expression analysis technologies and the release of the *Sorghum* genome sequence herald a new era for gene discovery in sugarcane; one which presents SASRI scientists with new challenges and opportunities. Presented here is an overview of gene discovery strategies employed at SASRI and the impact that they have had on sugarcane improvement research. Also described is the manner in which recent technological and bioinformatical advances are being embraced within this area of research.

Keywords: gene isolation, gene expression, transgenesis, DNA markers, molecular breeding

SASRI : Gene discovery landscape

Gene discovery research at SASRI has benefitted from advances in DNA sequencing and gene expression technologies, the capacities of which have progressed during the past decade from low, to high and ultra-high throughput. In parallel to these technological advances, the volume and availability of expressed sequence tag collections (Arruda, 2001) and annotated genome sequences, including those of species closely related to sugarcane (Paterson *et al.*, 2009), have increased markedly. The availability of these technologies and resources has

facilitated an evolution in the gene discovery research paradigm at SASRI from descriptive to hypothesis-driven, and which now stands poised to harness the potential of systems biology approaches. Application of the principles of systems biology (Hood, 1998) to sugarcane research may ultimately provide insights into the 'phenome'; that is, the phenotype for all traits of all individuals within populations. As such, the approach holds enormous potential for sugarcane improvement research, particularly in the identification of candidate genes for molecular breeding.

Trajectory: From basic to applied research

Providing initial impetus to gene discovery at SASRI was the acquisition in 1996 of a low-throughput automated DNA sequencer, for which one of the first applications was the analysis of selected genomic regions of local sugarcane pathogens. DNA sequencing of bacteria that caused ratoon stunting disease, leaf scald and gumming disease led to the development of rapid diagnostic tests (van Antwerpen, 1999), while sequence analysis of the coat protein gene of local strains of the sugarcane mosaic virus (SCMV) (Huckett and Botha, 1996) provided essential information for the development of transgenic approaches to engineering SCMV resistance into susceptible varieties (Sooknandan, 2002). This translation of gene sequence information derived from basic research into useful outcomes, and applications continued with the discovery of sugarcane genes participating in the regulation of sucrose accumulation in the stalk (Carson and Botha, 2000; 2002a; 2002b; Watt *et al.*, 2005) and those involved in the response to infection by the fungal pathogen causing sugarcane smut disease (Thokoane and Rutherford, 2001). Greatly aiding this research at SASRI was the availability of the Arabidopsis (The Arabidopsis Genome Initiative, 2000) and rice genome sequences (Goff *et al.*, 2002; Yu *et al.*, 2002), as well as a large collection of expressed sugarcane gene sequences identified by the Brazilian research community (Arruda, 2001). These large international genome sequencing and gene discovery programmes were made possible by the advent of high throughput DNA sequencing technologies.

Applications of gene discovery to sugarcane improvement

Genes identified during early gene discovery research at SASRI have found practical applications in sugarcane improvement, for example as DNA markers for parent selection in breeding (Heinze *et al.*, 2001; Butterfield *et al.*, 2004) and as transgenes for the genetic engineering of sugarcane (e.g. Bekker, 2007; Groenewald and Botha, 2007; Roussouw *et al.*, 2007). Although identified genes have found other applications at SASRI, such as in gene promoter isolation and the investigation of sugarcane biology, this discussion focuses on examples within marker-assisted breeding and genetic engineering research.

DNA sequences as markers: Marker-assisted breeding

Together with DNA sequences of unknown identity, genes expressed in response to *Ustilago scitaminea* (smut) infection and *Eldana saccharina* (stalk borer) predation have been used to develop DNA markers for resistance or susceptibility within an elite parent breeding population at SASRI (Butterfield *et al.*, 2007). These markers, both singly and in combination, have been used to select parents for crosses that have been specifically designed to increase *E. saccharina* and smut resistance amongst offspring. To date, 434 crosses of this type have been made at SASRI and the results obtained indicate that parent selection based on DNA marker information is more reliable in predicting offspring resistance to smut and *E.*

saccharina than the resistance rating of the parents, which is the conventional method (Butterfield, 2006). Although the results of these validation studies are promising, implementation into routine breeding activities and expansion to encompass other traits of interest, including those relating to yield, will necessitate the use of high-throughput DNA sequencing technologies. Within that context, subsequent data analysis will be dependent on rigorous comparison of sugarcane sequence information to that of published genome sequences of close relatives of sugarcane, such as sorghum (Paterson *et al.*, 2009).

DNA sequences as transgenes: Engineering increased sucrose accumulation

Together with the comprehensive analysis and modelling of sugarcane sucrose metabolism (e.g. Whittaker and Botha 1997; Rohwer and Botha, 2001; Botha, 2007; Uys *et al.*, 2007), the discovery of genes involved in internode maturation and, by association, sucrose accumulation, has provided access to several promising gene candidates for genetic manipulation. The discovery of such gene targets has relied on existing technologies at SASRI that permit DNA sequence analysis, albeit at low throughput, and analysis of gene expression during development of the sugarcane plant. SASRI scientists, in conjunction with the Institute of Plant Biotechnology at Stellenbosch University, have studied the effects of manipulating the activity of several of these genes as a means to increase sucrose accumulation. One promising target is the gene responsible for producing the enzyme UDP-glucose dehydrogenase, which reportedly plays a central role in controlling the flow of sugars between cell wall biosynthesis and sucrose synthesis (van der Merwe, 2005). Consequently, the activity of this gene was deliberately reduced to test whether decreasing the flow of sugars through the biochemical reaction mediated by the UDP-glucose dehydrogenase enzyme would result in increased availability of substrate for sucrose synthesis and accumulation (Bekker, 2007). Results from glasshouse studies yielded some startling insights into the effects of this genetic manipulation, in that both stalk sucrose content and cellulose content were significantly increased in the genetically engineered lines when compared to the non-transgenic NCo310 control plants. Further investigation revealed that the down-regulation of UDP-glucose dehydrogenase enzyme activity stimulated another pathway for cell wall biosynthesis, namely that catalysed by the enzyme *myo*-inositol oxygenase (Bekker, 2007). A field trial will be conducted at SASRI between 2009 and 2010 to confirm these results. The outcomes of this research have highlighted the necessity of not only examining the effects of a genetic modification on the disrupted metabolic pathway, but also on surrounding metabolism (Coleman *et al.*, 2008). To examine such collateral effects, the availability of next-generation, high-throughput technologies are required, including those for the analysis of DNA sequences (Droege and Hill, 2008), global gene expression (Eveland *et al.*, 2008) and metabolite profiling (Sato *et al.*, 2004).

Opportunities and challenges:

Era of ultra-high throughput technologies and poaceae genome sequences

The development of next-generation instruments for DNA sequencing, such as those based on 454-technologies, has paved the way for rapid sequence analysis of plant genomes (Droege and Hill, 2008) and gene expression patterns (Eveland *et al.*, 2008). For monocotyledonous crops such as sugarcane, a further vital link in facilitating such comprehensive genome and gene expression analyses has been the recent publication of the *Sorghum bicolor* genome sequence (Paterson *et al.*, 2009). The opportunities presented to SASRI by the availability of these technological and bioinformatical resources are profound, and include the potential for expansion of marker-assisted breeding to encompass yield traits, as well as for rational

genetic engineering strategies to modify central carbon metabolism, including that associated with sucrose synthesis, transport and storage. The major challenge facing SASRI lies in the development of the bioinformatical expertise and infrastructure that are required to process, assemble and interpret the vast volumes of data generated by ultra-high throughput technologies; a challenge that is not unique to SASRI, sugarcane research or South Africa. To this end, SASRI is currently engaging with national and international specialists to uncover ways to increase bioinformatics capacity. When such capacities are in place, focus may turn to the sequencing of the genomes of local varieties, which will provide the ultimate tool for sugarcane improvement.

REFERENCES

- Arruda P (Ed) (2001). Sugarcane transcriptome: A landmark in plant genomics in the Tropics. *Genet Molec Biol* 24(1-4): 1-296.
- Bekker J (2007). Genetic manipulation of the cell wall composition of sugarcane. PhD Thesis, University of Stellenbosch, South Africa. 66 pp.
- Botha FC (2007). Precision breeding to improve the usefulness of sugarcane. *Proc Int Soc Sug Cane Technol* 26: 35-43.
- Butterfield MK (2006). Marker assisted breeding in sugarcane: A complex polyploid. PhD Thesis, University of Stellenbosch, South Africa. 91 pp.
- Butterfield MK, Keeping MG and Sewpersad C (2007). Estimating genetic parameters and the efficiency of molecular breeding for resistance to the stalk borer *Eldana saccharina*. *Proc Int Soc Sug Cane Technol* 26: 691-698.
- Butterfield MK, Rutherford RS, Carson DL and Hockett BI (2004). Application of gene discovery to varietal improvement in sugarcane. *S Afr J Bot* 70: 167-170.
- Carson DL and Botha FC (2000). Preliminary analysis of expressed sequence tags for sugarcane. *Crop Sci* 40: 1769-1779.
- Carson DL and Botha FC (2002a). Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Reports* 20: 1075-1081.
- Carson DL and Botha FC (2002b). Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. *Plant Sci* 162: 289-300.
- Coleman HD, Park J-Y, Nair R, Chapple C and Mansfield SD (2008). RNA-mediated suppression of ρ -coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proc Natl Acad Sci USA* 105: 4501-4506.
- Droege M and Hill B (2008). The Genome Sequencer FLX™ System – Longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol* 136: 3-10.
- Eveland AL, McCarty DR and Koch KE (2008). Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiol* 146: 32-44.
- Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W-L, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T,

- Oliphant A and Briggs S (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*: 296: 92-100.
- Groenewald J-H and Botha FC (2007). Down-regulation of pyrophosphate: fructose-6-phosphate 1-phosphotransferase (PFP) activity in sugarcane enhances sucrose accumulation in immature internodes. *Transgenic Res* 17: 85-92.
- Heinze BS, Thokoane LN, Williams NJ, Barnes JM and Rutherford RS (2001). The smut-sugarcane interaction as a model system for the integration of marker discovery and gene discovery. *Proc S Afr Sug Technol Ass* 75: 88-93.
- Hood L (1998). Systems biology: new opportunities arising from genomics, proteomics and beyond. *Exp Hematology* 26: 681.
- Huckett BI and Botha FC (1996). Progress towards a definitive diagnostic test for sugarcane mosaic virus infection. *Proc S Afr Sug Technol Ass* 70: 11-13.
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J and Rokhsar DS (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551-556.
- Rohwer JM and Botha FC (2001). Analysis of sucrose accumulation in the sugarcane culm on the basis of *in vitro* kinetic data. *Biochem J* 358: 437-445.
- Rossouw D, Bosch S, Kossmann JM, Botha FC and Groenewald J-H (2007). Down-regulation of neutral invertase activity in sugarcane cell suspension cultures leads to increased sucrose accumulation. *Func Plant Biol* 34: 490-498.
- Sato S, Soga T, Nishioka T and Tomita M (2004). Simultaneous determination of the main metabolites in rice leaves using capillary electrophoresis mass spectrometry and capillary electrophoresis diode array detection. *Plant J* 40: 151-163.
- Sooknandan S (2002). Generation of transgenic sugarcane plants resistant to sugarcane mosaic virus by induction of post-transcriptional gene silencing. *BTech Report*, Durban Institute of Technology, Durban. 105 pp.
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- Thokoane LN and Rutherford RS (2001). cDNA-AFLP differential display of sugarcane (*Saccharum* spp. hybrids) genes induced by challenge with the fungal pathogen *Ustilago scitaminae* (H&P Sydow) (sugarcane smut). *Proc S Afr Sug Technol Ass* 75: 104-107.
- Uys L, Botha FC, Hofmyer J-H and Rohwer JM (2007). Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. *Phytochemistry* 68: 2375-2392.
- van Antwerpen T (1999). Development of a DNA-based diagnostic method to detect sugarcane bacterial pathogens with emphasis on *Clavibacter xyli* subsp. *xyli*. MSc Dissertation, University of Stellenbosch, Stellenbosch. 90 pp.
- van der Merwe J (2005). The characterization of the UDP-glucose dehydrogenase gene and regulatory sequences of sugarcane. PhD Thesis, University of Stellenbosch, Stellenbosch.
- Watt DA, McCormick AJ, Govender C, Carson DL, Cramer MD, Huckett BI and Botha FC (2005). Increasing the utility in unravelling sucrose accumulation. *Field Crops Res* 92: 149-158.
- Whittaker A and Botha FC (1997). Carbon partitioning during sucrose accumulation in sugarcane intermodal tissue. *Plant Physiol* 115: 1651-1659.

Yu J, Hu S, Wang J, Ka-Shu Wong G, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L and Yang H (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296: 79-92.