SHORT COMMUNICATION

INTEGRATING CROP MODELLING AND GENOMICS TO ACCELERATE PLANT IMPROVEMENT: PRELIMINARY ASSESSMENT OF PROGRESS FROM A PILOT STUDY IN SUGARCANE

SINGELS A, SMIT MA, BUTTERFIELD M,
VAN HEERDEN PDR and VAN DEN BERG M

South African Sugarcane Research Institute, Private Bag X02, Mount Edgecombe, 4300, South Africa
abraham.singels@sugar.org.za

Abstract

Crop yield is a complex trait governed by dynamic interactions between plant and environment. Integrating crop modelling and genomic technologies provides an opportunity to gain a better understanding of the underlying physiological mechanisms and genetic basis of these interactions and has the potential to enhance plant improvement. This communication reviews progress in an exploratory project to integrate genomics and crop modelling in sugarcane.

Stalk and leaf elongation rate per unit thermal time, fully expanded leaf area, and photochemical light use efficiency were measured in two experiments for 80 clones of a mapping population. These genetically determined traits were chosen because they are important in determining resource capture and yield formation, and are relatively easily measured.

The suitability of crop models for simulating simple trait effects on crop performance is assessed. The heritability of traits and their association with molecular markers are reported. The reliability of regression equations to predict trait values from marker data is assessed.

Keywords: molecular marker, stalk elongation, crop model, genomics, traits, light use efficiency

Introduction

Crop yield is a complex trait governed by numerous dynamic interactions between plant processes, the environment and management. Accelerated improvement in plant performance is only likely when these interactions and the underlying physiological mechanisms and genetic basis are better understood. Integrating crop modelling and genomic technologies provides an opportunity to achieve this.

Credible crop models can simulate the effect of simple traits on crop response to environmental and management factors (gene to phenotype modelling), and therefore have the potential to identify trait sets that are most likely to enhance yields in target environments (Hammer *et al.*, 2006). Genomics on the other hand, allows molecular markers to be detected for these traits for numerous genotypes. These markers could potentially be used to enhance breeding (Hammer and Jordan, 2007).

This communication reviews progress in an exploratory project to integrate genomics and crop modelling in sugarcane.

Methods

Modelling requirements
The suitability of the DSSATv4.5 Canegro sugarcane model (Jones *et al*., 2007) to simulate the effect of simple traits on crop performance was explored by comparing current capabilities with envisaged requirements.

Traits
Stalk elongation rate per unit thermal time (SER), fully expanded leaf area (LA) and leaf appearance rate per unit thermal time (LAR) were obtained for a mapping population in two field experiments. Photochemical light use efficiency for each clone was determined by means of chlorophyll *a* fluorescence measurements. These traits were chosen because they are important in determining resource capture and yield formation, are relatively easily measured and are believed to be genetically determined.

Experiments
Two experiments (exp 1 and 2) were conducted at Mount Edgecombe (29º42'S; 31º03'E) on a sandy clay loam covered with a trash blanket to discourage tillering, and consisted of four replications with five plants per replication. Plants were adequately fertilised and irrigated. In exp 1, single-eyed setts were planted into the field in May 2006. In exp 2, setts were germinated in vermiculite in a germination chamber at 30ºC and transplanted into the field from 29 May to 7 June 2007, when a height of around 10 cm was reached. In both experiments, secondary tillers were removed as soon as they were visible.

On a weekly basis the top visible dewlap (TVD) leaf was dated and numbered, and the leaf length and width recorded together with the TVD collar height from ground level. SER was taken as the slope of the regression of TVD height vs. thermal time, and LAR was taken as the slope of the regression of number of fully expanded leaves vs. thermal time. A distinction was made between the phase between leaf number 8 and 14 (denoted by subscript 14), and the phase between leaf number 15 and 24 (denoted by subscript 24). LA was taken as the average fully expanded area of leaves 15 to 24. Due to limited manpower and changes in measurement protocol, the number of plants and replications that were measured for some traits varied.

Light use efficiency was determined only in exp 2 by recording fast polyphasic fluorescence transients (Strasser and Govindjee, 1992) at night in TVD leaves with a fluorescence meter (PEA, Hansatech Instruments Ltd, UK). Four measurements were taken per leaf on three plants of each clone in each replication (total of 48 measurements per clone). The recorded chlorophyll fluorescence data were used to calculate the Performance Index (PI_{ABS}), which is a sensitive indicator of photosynthetic electron transport efficiency and photosynthetic capacity (Strasser *et al*., 2000).

Daily maximum and minimum temperatures were recorded by a nearby automatic weather station and were used to calculate thermal time using a base temperature of 16ºC.

Mapping population
The population of 80 clones, derived from the South African Sugarcane Research Institute breeding population, were mapped using linkage disequilibrium methods (Butterfield *et al*., 2008). The map consists of 2054 AFLP and DArT markers arranged on 492 haploblocks.

Data analysis
Phenotypic data was analysed using restricted maximum likelihood (RML) methods, with genotype as a fixed effect and plant-within-replication as random effects. Clonal repeatability (ratio of genetic variance to phenotypic variance) was estimated from the derived variance components.

Association between trait value and marker presence or absence was calculated by Pearsons correlation coefficient. Stepwise linear regression was used to select markers ascribing maximum phenotypic variation for each trait. The prediction error was taken as the standard error of regression, expressed as a percentage of the observed range in values of each trait.

Results and Discussion

Modelling requirements
Credible gene-to-phenotype simulation requires a biologically robust simulation of crop development and growth processes and requires appropriate levels of detail and hierarchy (Hammer and Jordan, 2007). Chenu *et al*. (2008) give an excellent example of incorporating a simple trait module into a crop model to accurately predict growth of individual leaves and whole crop response to environmental stresses and management.

The DSSATv4.5 Canegro sugarcane crop model (see Singels *et al*., 2008) provides a good platform, because it:

- Uses 76 parameters to describe the genetics of sugarcane thereby providing ample scope for improved trait parameterisation.
- Simulates canopy development using genetic parameters such as maximum leaf size and thermal time requirements for organ development.
- Uses a source-sink framework for dynamic partitioning of carbon to the different pools as determined by temperature and water stress.

The following improvements to the model will be required:

- Introduce a more mechanistic simulation of the dynamic effect of light environment and temperature on tillering and leaf size.
- Link development and growth processes through the carbon balance.
- Link carbon partitioning to source availability and to sink demand from leaf and internode structural growth.

Recent gains in understanding the mechanisms of sugarcane growth and development (see e.g. Smit and Singels, 2006; Singels and Inman-Bamber, 2008; McCormick *et al*., 2006) could be used to address these shortcomings.

Heritability of traits, association with markers

Clonal repeatability was generally high (above 0.9) for all traits, except for LAR14 and LAR24. This could be ascribed to poor phenotyping accuracy, as it is well established that LAR is strongly genetically driven (Bonnett, 1998).

The number of markers with significant trait associations ($P<0.05$) common to both experiments is shown in Table 1, along with the phenotypic variation accounted for by multiple regressions and the reliability of its predictions. The multiple regressions determined for traits SER14, LA and PI_{ABS} are promising, with r^2 values exceeding 0.5 and prediction errors below 20%.

Table 1. Number of markers with significant correlations with a given trait, the highest correlation achieved for a single marker, the number of markers selected in the stepwise regression to predict the trait value, the coefficient of determination for the regression (r^2) and the prediction error (PE).

Trait	No. of markers	Highest correlation	No. of markers in regression	Regression r^2	PE (%)
LAR14	12	0.43	4	0.29	39
LAR24	18	0.35	3	0.21	44
SER14	59	0.55	6	0.56	19
SER24	13	0.46	4	0.29	47
LA	71	0.44	6	0.62	17
PI_{ABS}	167	0.43	6	0.62	16

The low r^2 values and the high prediction errors for the other traits suggest that these marker associations cannot be used reliably to predict trait values. Possible reasons are:

- Poor phenotyping accuracy due to low number of samples or inappropriate sampling techniques.
- Low repeatability caused by insufficient separation of genetic and environmental aspects in the trait definition. For example, Smit and Singels (2007) found that the genetic control of SER was confounded by site and season effects. Reymond *et al*. (2003) showed that SER in maize was controlled by soil water potential and atmospheric vapour pressure deficit (VPD). In our study, an attempt was made to eliminate soil water status as a source of variation. The impact of VPD on observed SER still needs to be investigated.

Conclusions

Progress thus far suggests that:

- The Canegro model provides a suitable platform for gene-to-phenotype modelling; however, further refinement such as linking carbon flux to structural growth at phytomer level is required.
- Photosynthetic light use efficiency (as indicated by PI_{ABS}), was predicted reliably from marker data, while marker predictions of stalk elongation rate per unit thermal time and leaf size also showed promise.

- Further validation of marker associations in other germplasm and/or environments are needed before including marker-based trait prediction in the Canegro model.

Although much work remains before reaching the ideal of using gene to phenotype modelling to identify desirable sugarcane sub-traits for target environments, no adverse factors emerged from this study to suggest that the target is not possible or that it will prove too difficult to attain.

REFERENCES

Bonnett GD (1998). Rate of leaf appearance in sugarcane, including a comparison of a range of varieties. *Aust J Plant Physiol* 25: 829-834.

Butterfield M, D'Hont A and Kilian A (2008). International Consortium for Sugarcane Biotechnology Project 26: Mapping the reference cultivar R570 and a population of commercial sugarcane genotypes using diversity arrays technology (DArT), and development of bioinformatics tools to use this information. Final report submitted to the International Consortium for Sugarcane Biotechnology.

Chenu K, Chapman SC, Hammer G, McLean G, Ben Haj Salah H and Tardieu F (2008). Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: An integrated modelling approach in maize. *Plant, Cell and Environment* 31: 378-391.

Hammer GL and Jordan DR (2007). An integrated system approach to crop improvement. pp 45-61 In: JHJ Spierts, PC Struik and HH van Laar (Eds), *Scale and Complexity in Plant Systems Research: Gene-plant-crop Relations*. UR Fortier Series 21, Springer, Wageningen.

Hammer G, Cooper M, Tardieu F, Welch S, Walsh B, van Eeuwijk F, Chapman S and Podlich D (2006). Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science* 11: 587-593.

Jones M, Porter C, Jones JW, Hoogenboom G, Singels A, Shine J, Nayamuth R, Kingston G, Chinorumba M and van den Berg M (2007). Incorporating the Canegro sugarcane model into the DSSAT v4 crop modelling system. *Proc Int Soc Sug Cane Technol* 26: 438-443.

McCormick AJ, Cramer MD and Watt DA (2006). Sink strength regulates photosynthesis in sugarcane. *New Phytologist* 171: 759-770.

Reymond M, Muller B, Leonardi A, Charcosset A and Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology* 131: 664-675.

Singels A and Inman-Bamber NG (2008). A source-sink model to simulate sucrose accumulation at the internode level. Report to the South African Sugar Research Institute and the Council for Scientific and Industrial Research Organisation. 49 pp.

Singels A, Jones M and van den Berg M (2008). DSSAT v4.5 Canegro Sugarcane Plant Module: Scientific documentation. South African Sugarcane Research Institute, Mount Edgecombe, South Africa. pp 34.

Smit MA and Singels A (2006). The effect of row spacing on inter-row competition in sugarcane. *Proc S Afr Sug Technol Ass* 80: 139-142.

Smit MA and Singels A (2007). Quantifying the effects of environment and genotype on stalk elongation rate in sugarcane. *Proc Int Soc Sug Cane Technol* 26: 568-572.

Strasser RJ and Govindjee (1992). The F_0 and the O-J-I-P fluorescence rise in higher plants and algae. pp 423-426 In: JH Argyroudi-Akoyunoglou (Ed), *Regulation of Chloroplast Biogenesis*. Plenum Press, New York, USA.

Strasser RJ, Srivastava A and Tsimilli-Michael M (2000). The fluorescent transient as a tool to characterise and screen photosynthesic samples. pp 445-483 In: M Yunus, U Pathre and P Mohanty (Eds), *Probing Photosynthesis: Mechanisms, Regulation and Adaptation*. Taylor and Francis, London, UK.