Sugar Research Australia

# Genomic selection in sugarcane

**E Deomano, X Wei, P Jackson**

8th ISSCT Molecular Biology Workshop, Reunion 2015

# Background

- DArT
  - 1500 to 2000 “discrete” markers (markers presenting as either present or absent);  15360 “continuous” markers (markers presenting as a variable intensity score: possibly related to allele dosage)
  - Three populations
    - Association mapping population, 480 clones
    - Narrow population
    - Parental population
  - Identified many markers with evidence for association with important traits (TCH, CCS, disease resistance), but fdr of about 30%

# Implementation/testing in practice:

- Simulation suggested that:
  - Marker assisted selection unlikely to be cost effective for cultivar selection for cane yield and CCS, but could be highly effective is for parental selection and improvement
- For pilot implementation marker assisted breeding program:
  1. Developed a DArT plate with 384 “important effect” markers
  2. Screened a population of ~1000 clones from elite cross (S2 stage) for markers + cane yield + CCS
  3. Four groups of clones selected (random, just phenotype, just markers, marker+phenotype index) :
  4. Clones put in crosses to evaluate breeding value - progeny generated and currently (2015 season) being evaluated in S1 trials

# BUT – recent move to genomic selection…

- Limitations with association mapping study
  - Marker by marker, only significant ones included
  - Inherent problem for quantitative (complex) traits
    - Sanna et al, Nature Genetics, Human height (h2 = 0.8), one locus ~ 0.44cm (<1%)
    - J Crossa (CIMMYT), Per comm, <5%
  - Marker effects biased upwardly
  - Most noticeable / significant – few application in commercial program

# Genomic selection

- All markers considered simultaneously
- All QTLs covered by markers for dense markers (SNPs)
- Not significant test – potentially all genetic variance tracked by markers

# Genomic selection

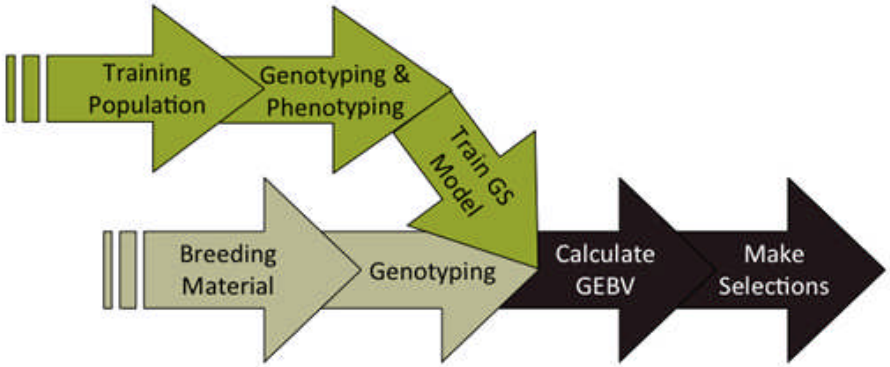



Figure 1: The 'training population' is genotyped and phenotyped to 'train' the genomic selection (GS) prediction model. Genotypic information from the breeding material is then fed into the model to calculate genomic estimated breeding values (GEBV) for these lines. *From Heffner et al. 2009 Crop Sci. 49:1–12*

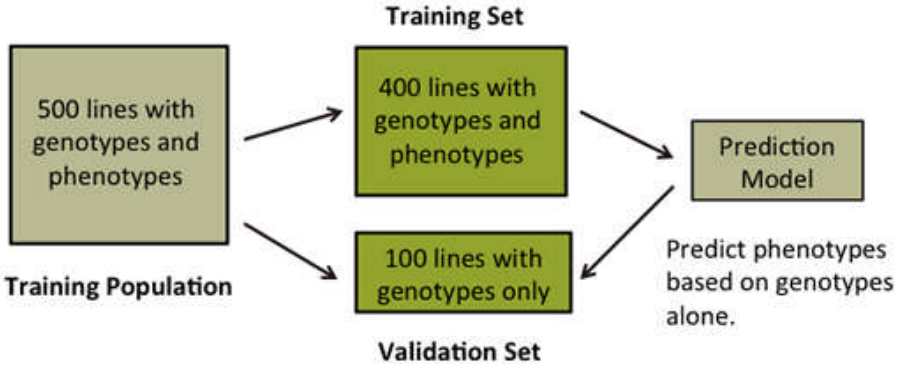



Figure 2: Information from a majority of lines in the breeding population (the training set) is used to create the prediction model. The model is then used to predict the phenotypes of the remaining lines (the validation set), using genotypic information only. The results from the model are compared to the actual data to give the prediction accuracy. *Image courtesy of Martha Hamblin, Cornell University*

**Link between training and test populations**

http://www.nextgencassava.org/genomic_select.html

# Data

- Association mapping population
  - 480 clones
  - TCH, CCS,
  - DArTs (continuous 15,360, discreet 1,531)

# Models

- GS models
  - GBLUP
  - RKHS (Reproducing Kernel Hilbert Space)

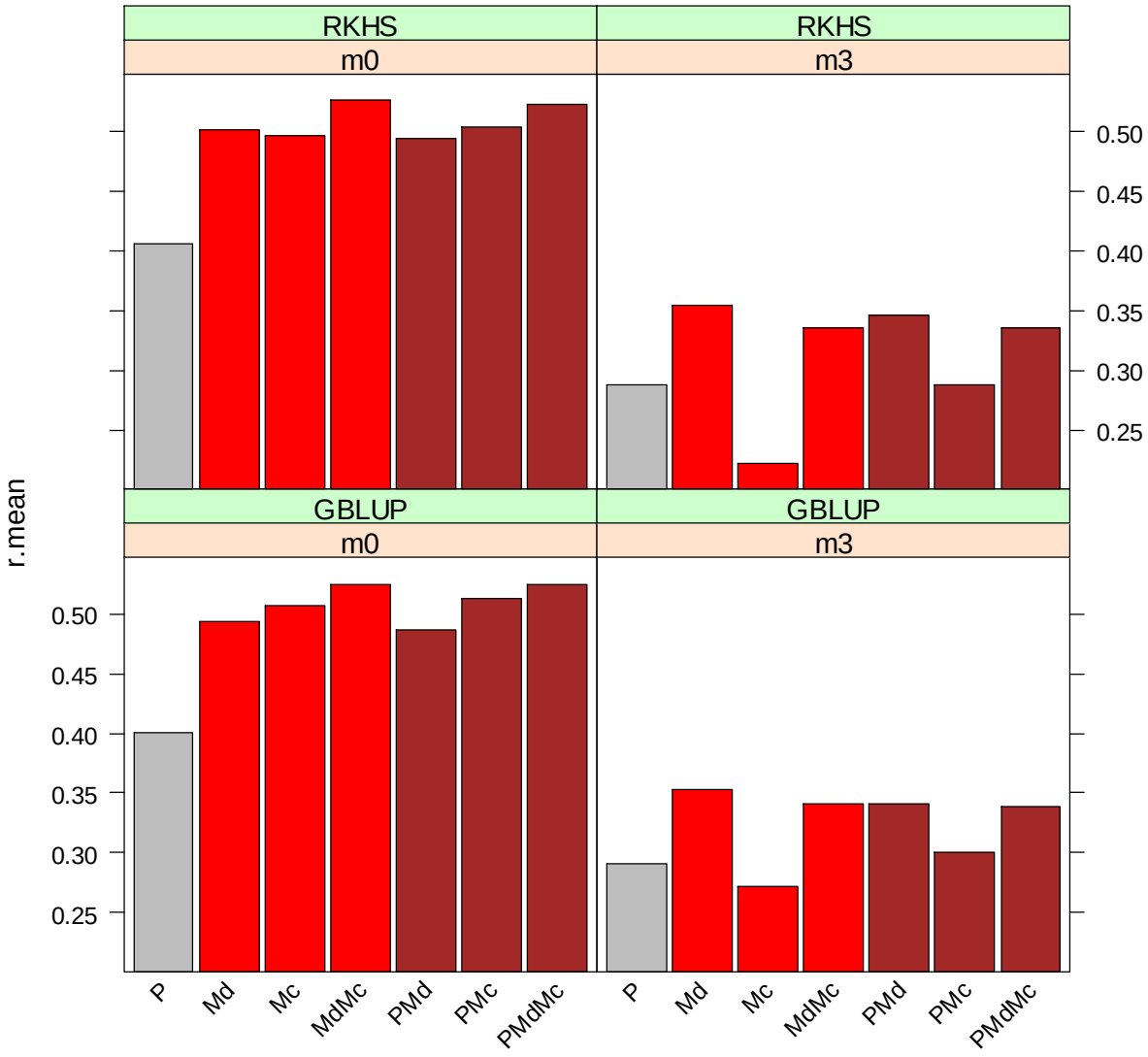
- Sub-models
  - P – pedigree
  - G1 – discrete markers
  - G2 – continuous markers
  - G1G2 – discrete and continuous markers
  - PG – pedigree and discrete markers
  - PG2 – pedigree and continuous markers
  - PG1G2 – pedigree, discrete and continuous markers

# Accuracy

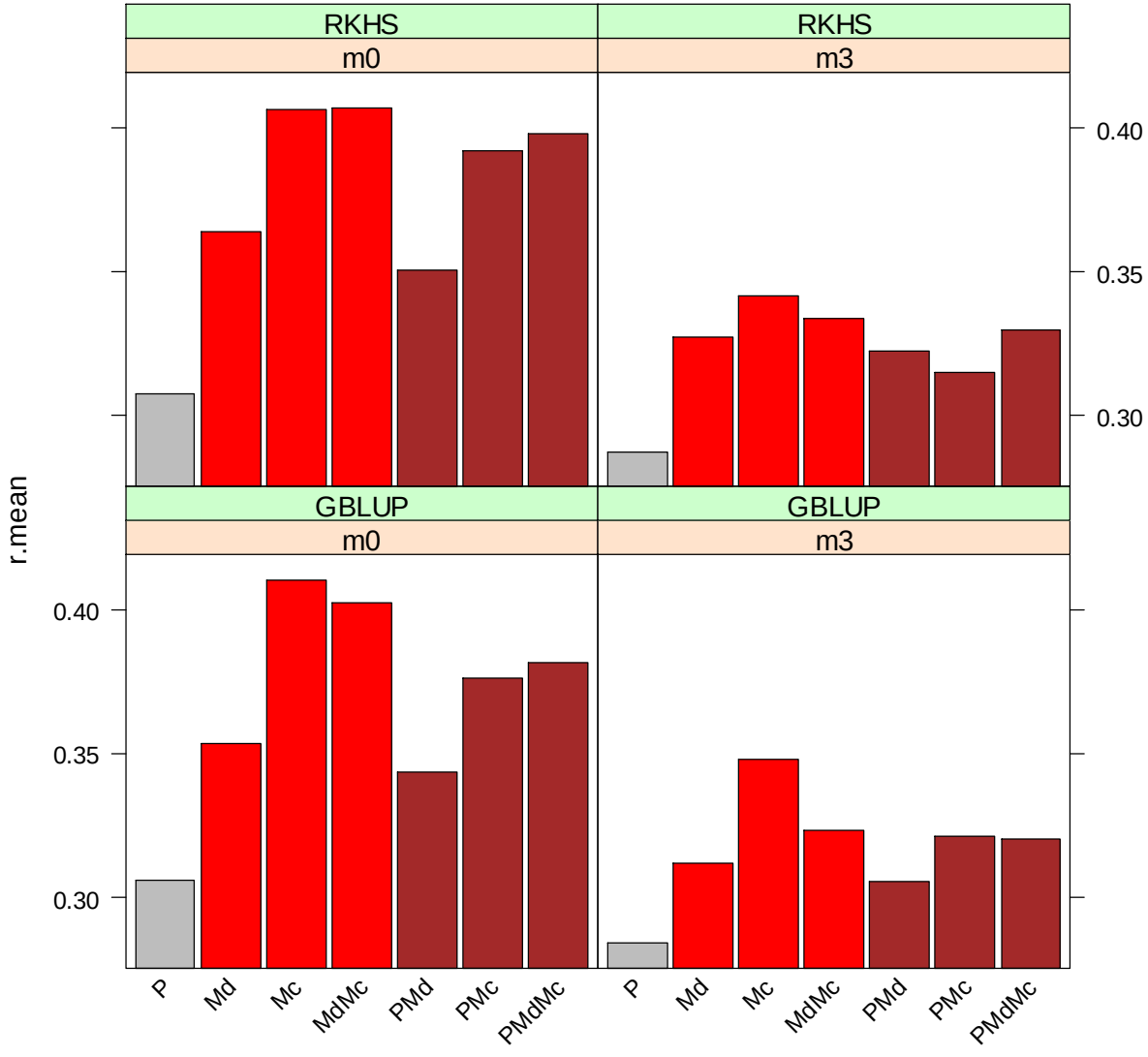
- Pearson's product-moment correlation coefficient
  - observed phenotype vs predicted phenotype
  - Averaged over 50 sets

- Bayesian approach
  - 12000 samples with 200 samples burn-in.
  - BGLR, R package

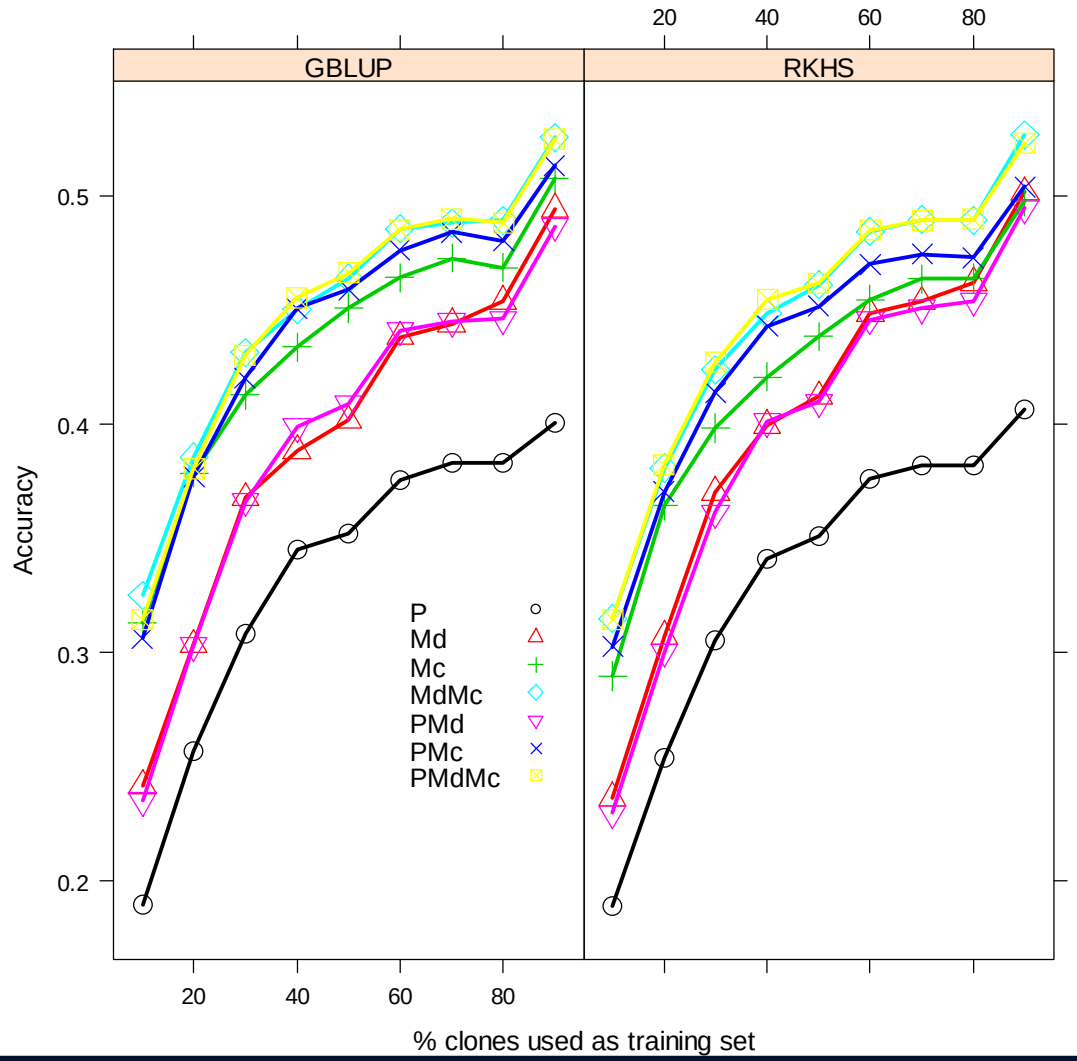
## Selection accuracy for TCH



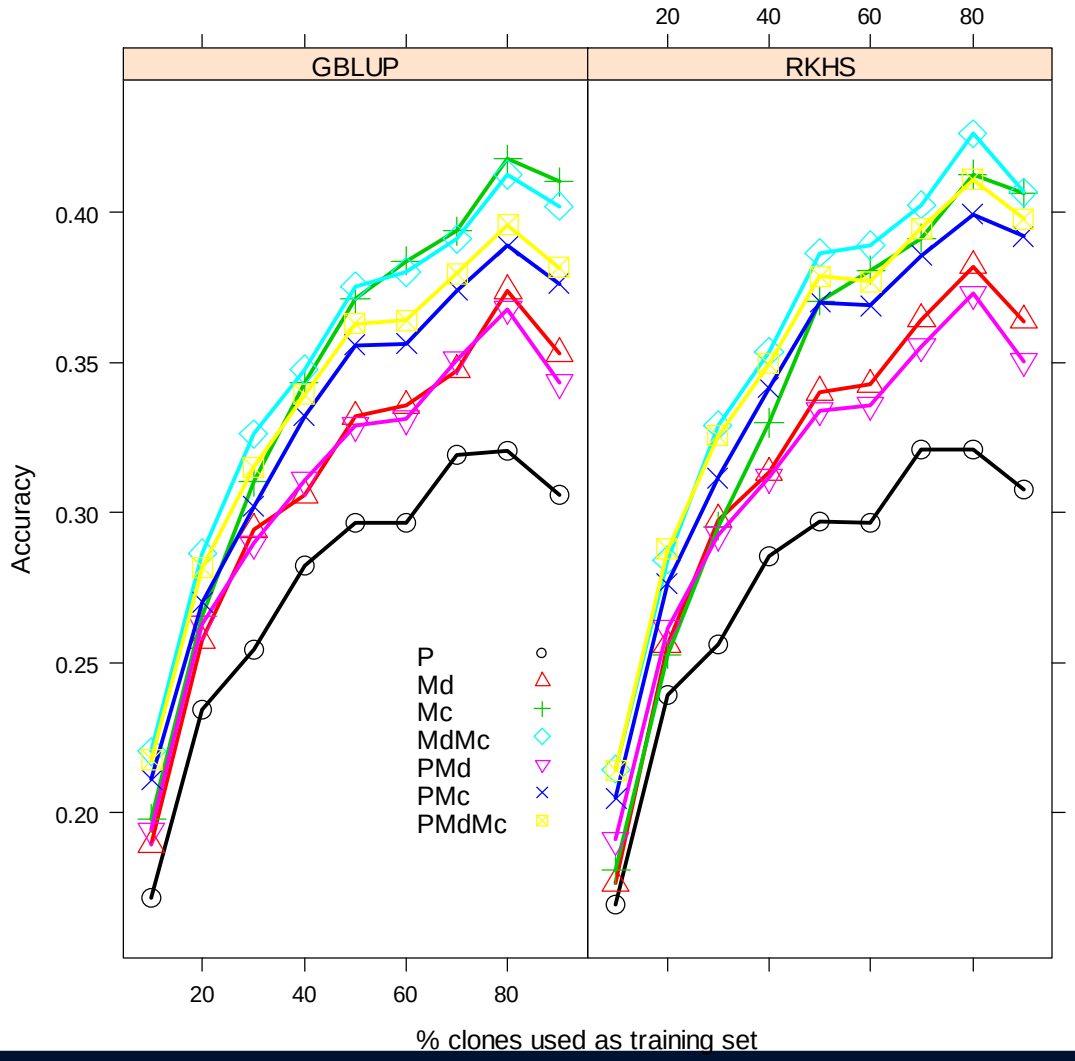
## Selection accuracy for CCS



# Impact of training size – TCH



# Impact of training size – CCS



# $h^2$ (r)

- CAT
  - Random selected clones
  - Better understanding of genetic architecture
    - Additive vs non-additive
  - Estimates of narrow-sense heritability
    - TCH 0.15 (0.39)
    - CCS 0.23 (0.48)

# Remark

- TCH has higher accuracy than CCS
- Training size may be the limiting factor
- GBLUP produced almost the same results as RKHS
- Accuracy is limited by proportion of total genetic variance which is caused by additive genetic effects? Accuracy for breeding value may be greater?

- Recent large set of SNP data (50K array x 480 clones) at first look appears far superior to DArT markers: if yes, may make recent work partly redundant?

# Acknowledgements


- Marker data
  - K Aitken, Scott Hermann, Andrzej Kilian, Katarzyna Heller-Uszynska
  - K Aitken, Raja Kota et al
- Phenotypic data
  - John Foreman, Michael Hewitt, Fengduo Hu
- Genomic selection
  - Jose Crossa, Paulino Rodriguez, Diego Jarquin
- Funding
  - SRA