PAG XXI – January 13, 2013
Palm genomics and genetics Workshop

Factors controlling accuracy of genomic selection in oil palm (*Elaeis guineensis*)

Cros D, Denis M, Sanchez L,
Cochard B, Edyana-Suryana H, Omoré A, Durand-Gasselin T & Bouvet JM

The oil palm

- world major oil crop
- one cultivated species, *Elaeis guineensis*
- **allogamous** (monoecious with male and female cycles)

- vegetative multiplication difficult

Selection criteria:

Average bunch weight (ABW), Bunch number (BN)

Fruits to bunch ratio (F/B), Pulp to fruits ratio (P/F), Oil to pulp ratio (O/P)

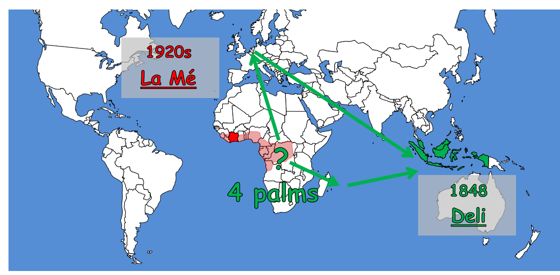
Height increment (INC)

$\underbrace{\text{Average bunch weight (ABW), Bunch number (BN)}}_{\textbf{Bunch production (FFB)}}$

$\underbrace{\text{Fruits to bunch ratio (F/B), Pulp to fruits ratio (P/F), Oil to pulp ratio (O/P)}}_{\textbf{Oil extraction rate (OER)}}$

Breeding populations:

Distant populations with narrow genetic bases
- Deli
- La Mé



Genomic selection

Method of MAS (**Meuwissen et al 2001**):

- Training population phenotyped and genotyped
- Dense genotyping of the whole genome
- All markers effects estimated simultaneously
- No test of significance of marker effects
- Selection on markers alone (GEBV) in test population

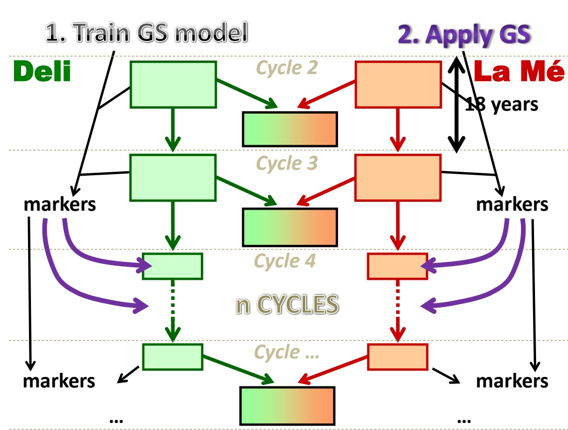
"Selection on genetic values predicted from markers could substantially increase the rate of genetic gain in animals and plants"

Genetics 157: 1819–1829 (April 2001)

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes[†] and M. E. Goddard[†,‡]

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, [†]Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and [‡]Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia



Could we increase the rate of genetic gain in oil palm breeding with genomic selection ?

Hypothesis: Progeny-tested individuals could be used to train a GS model that could be applied to predict breeding values of individuals of the same populations

With:

- Narrow genetic base / Low effective size
- Small training populations
- Small number of markers
- Multiallelic markers

Hypothesis: Progeny-tested individuals could be used to train a GS model that could be applied to predict breeding values of individuals of the same populations

➔ Will be checked by measuring the **accuracy of GS in a cross-validation study with real data**

$$\textbf{Genetic gain per year} = \frac{\textbf{Intensity} * \underline{\textbf{Accuracy}} * \sigma_a}{\textbf{Generation interval}}$$

With accuracy = r(TBV, EBV) and current accuracy ~ 0.8

Materials and methods

Plant material:

- ➢ **Deli**: **131** individuals
- ➢ **La Mé**: **93** individuals

Materials and methods

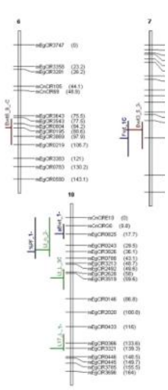
Molecular data:

235 SSR (Billotte et al 2005; Tranbarger et al 2011)

~ 1 SSR / 7.4 cM

Phenotypic data:

1. Progeny tests, **10 quantitative traits**
2. **Estimated breeding values** (BLUP)
3. **Deregressed** and used in a **weighted analysis** to derive genomic estimated breeding values (Garrick et al 2009)



Materials and methods

5-fold cross-validation:

1/ Individuals (genotyped and phenotyped) divided into 5 groups to make training population (4 groups) and test population (5th group)

→ **5 replicates**

2/ Estimation of allelic effects

3/ Calculation of GEBV for test individuals

4/ Calculation of accuracy of genomic selection in test population

Materials and methods

Definition of groups for training and test populations:

2 methods, in order to get a **range of accuracy of GS**:

1. Lower bound: **CLUSTERING** (Saatchi et al 2011)

- Calculate matrix of additive genetic relationships between individuals,
- Use K-means clustering to make 5 groups

→ **Increases within-group relationships / Decreases between-group relationships** (groups represent subpopulations)

2. Upper bound: **ACROSS FAMILIES**

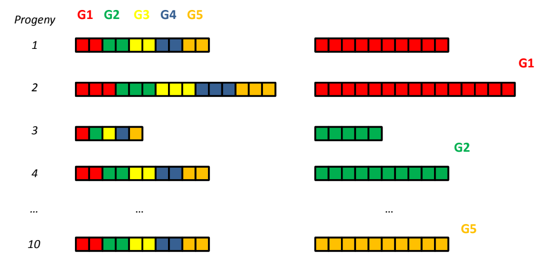
- Each family is randomly divided into 5 groups

→ **Maximizes relationships between training and test populations**

Materials and methods

Definition of groups for training and test populations:

ACROSS FAMILIES **K-MEANS CLUSTERING**



Materials and methods

Statistical methods to estimate GEBV:

Group of method	Method	Marker effects	Comments	Reference
Mixed Model	ABLUP	no	Control	Henderson 1975
Mixed Model	**BLUP**	yes	gi ~ N(0, Vm)	Meuwissen et al 2001
Mixed Model	**GBLUP**	no	Estimate GEBV	Henderson 1975, Eding and Meuwissen 2001
Bayes	**BRR**	yes	gi ~ N(0, σ^2_{BR})	Perez et al 2010
Bayes	**BL**	yes	gi ~ N(0, τ_i * σ^2_e)	Perez et al 2010
Semi-parametric	**RKHS**	no	Estimate GEBV	Gianola et al 2006, Heslot et al 2012

→ Some methods better suited for traits with many small effect genes, others for traits with major genes + small effect genes

Accuracy: $$r_{GEBV,TBV} = \frac{\hat{\sigma}_{DEBV,GEBV}}{\sqrt{\sigma^2_a \hat{\sigma}^2_{GEBV}}}$$ (Saatchi et al 2011)

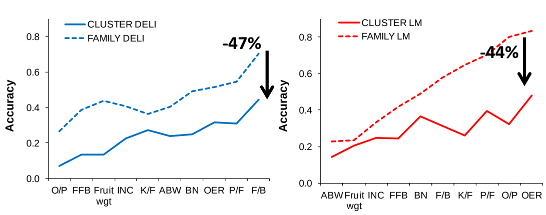
Results

- Factors with the strongest effect on accuracy:
  - **(1) TRAINING MODE,**
  - **(2) TRAIT, POPULATION**
  - **(3) TRAIT * POPULATION INTERACTION**
- No effects of statistical method, no statistical method * trait interaction

Results

Effect of training mode on accuracy:

➔ Range of accuracy for GS



... Effect of training mode related to a_{max}:

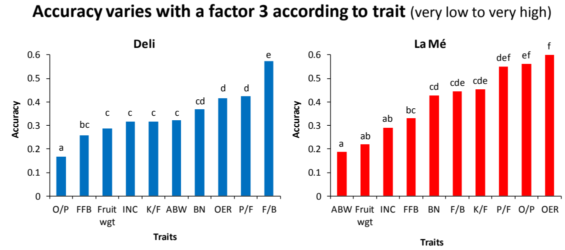
family → cluster, **-22%** family → cluster, **-18%**

1.1 0.86 0.93 0.77

Results

Effect of trait on accuracy:

Accuracy varies with a factor 3 according to trait (very low to very high)

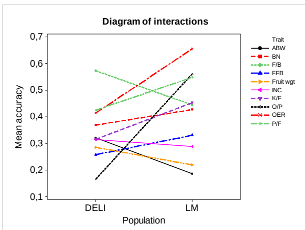


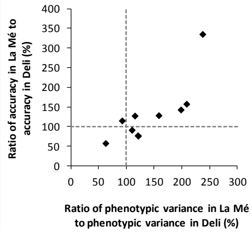
Due to genetic architecture of each trait ?

(number of QTLs, distribution of QTL effects, distribution of QTL along genome versus distribution of SSRs, LD between markers and QTLs)

Results

Effect of trait * population interaction on accuracy:





Due to differences in genetic architecture between populations and traits ?

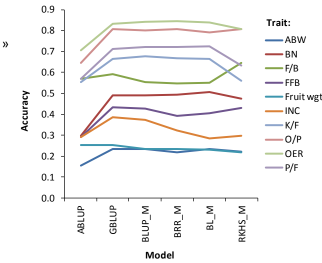
Related to differences in phenotypic variance

Results

Effect of statistical method on accuracy:

No effect, no interaction with trait

Example:
La Mé, training population « family »



… contradictory with trait effect and trait * population interaction
→ **Too small number of phenotypic records ?**

Conclusions

Some traits with very low accuracy
→ **bigger training populations / more markers**

More **studies required** before implementing GS in our oil palm breeding program…

- Effect of increasing training population size ?
- Rate of decrease of accuracy over generations ?
- Accuracy between experimental designs ?
- Genetic architecture of traits in each population ?
- …

Some answers in 2013 (**simulations**)
and 2014 (**more real data**: 2 experimental designs, 2 generations + **GBS genotyping**)