Next Generation Cassava Breeding: Progress Implementing Genomic Selection

Marnin Wolfe, Chiedozie Egesi, Robert Kawuki, Peter Kulakow, Ismail Y. Rabbi, Jean-Luc Jannink and many others!
Next Generation Cassava Breeding:

Testing Genomic Selection

- International collaboration
  - 3 African breeding institutes
  - Plus Cornell U. et al.

Nigeria (West Africa)  
Uganda (East Africa)

www.nextgencassava.org  
www.cassavabase.org
First Part

Intro to Genomic Selection and Role in Cassava Breeding
Gain per Year = \( \frac{i \times r \times \sigma_a}{T} \)
Traditional Cassava Breeding Cycle

Crossing Blocks Source Pops

Year 1
Seedling nursery (SN)

Year 2
Clonal evaluation (CE)

Year 3
Prelim. yield trial (PYT)

Year 4
Advanced yield trial (AYT)

Years 5 & 6
Multi-location Uniform yield trial (UYT)

NARS and On-Farm Trials

Year 8+
Multiplication and Release

Individuals tested Candidates

Plot size Replications Locations Traits

Courtesy I. Y. Rabbi
Crossing Blocks Source Pops

Year 1
Seedling nursery (SN)

Year 2
Clonal evaluation (CE)

Year 3
Prelim. yield trial (PYT)

Year 4
Advanced yield trial (AYT)

Years 5 & 6
Multi-location Uniform yield trial (UYT)

Two levels of selection
- Parent selection (Population Improvement)
- Variety selection (Product Development)

Same rankings used for both!

Traditional Cassava Breeding Cycle

NARS and On-Farm Trials

Year 8+
Multiplication and Release

Courtesy I. Y. Rabbi
How can we accelerate cassava breeding?
Genomic selection: Enables a rapid breeding cycle with markers

- All plants are genotyped
- Build a reference or “training” population (TP)
  - Genotyped
  - Phenotyped
    - As extensively as possible (many locations / years)
- Statistical model relates genotype to phenotype in the TP
- The TP model can generate genome-estimated breeding value (GEBV) of unevaluated selection candidates
Rapid population improvement by \textit{genomic selection}

Heffner, Sorrells, Jannink 2009. \textit{Crop Science}.
Genomic BLUP Mixed Model Approach

- Effects (levels) of $g$ called BLUPs
- They are estimated breeding values for each individual
- $g$ calculated with covariance matrix = the genomic relationship (kinship) matrix based on SNP markers

\[ Y = X\beta + Zg + e \]

Fixed Effects

Random Effect: $g \sim N(0, K\sigma^2_g)$

Genomic relationship matrix (K)
Genomic BLUP Mixed Model Approach

- Adds new information for selection
- Markers used to disentangle...
  - Environment
  - Non-additive variation
    ... from breeding (additive) value
- Even before phenotypic evaluation

\[ Y = X\beta + Zg + e \]

Fixed Effects

Random Effect: \( g \sim N(0, K\sigma^2_g) \)

Genomic relationship matrix \( (K) \)
One-year GS Cycle in Cassava

Select & Cross

Seedlings

Genotype

Predict Breeding Value

Clonal Evaluation

Training Model Update

Year 1

February 5 (Most of the seedlings germinated)

February 7 (field transplantation)

February 20 (Tagging and DNA sampling)

Year 2

March 9 (Good establishment)

March 29 (Rapid growth)

May 14 (seedlings nearly ready for cloning)

Also testing a two-year cycle
Traditional Cassava Breeding Cycle

Crossing Blocks Source Pops

Year 1
Seedling nursery (SN)

Year 2
Clonal evaluation (CE)

Year 3
Prelim. yield trial (PYT)

Year 4
Advanced yield trial (AYT)

Years 5 & 6
Multi-location Uniform yield trial (UYT)

NARS and On-Farm Trials

Multiplication and Release

Year 8+

Two levels of selection
• Parent selection
• Variety selection

Parent selection based on genome estimated breeding value!

Genome-enabled Cassava Breeding Cycle

Courtesy I. Y. Rabbi
Progress so far

GS Cycle 0
Training Pop.
N = 969
83 parents selected
Phenotyped
• 2000-2015

GS Cycle 1
N = 2322
106 parents selected
Phenotyped
• 2013-2014
• 2014-2015

GS Cycle 2
N = 2216
135 parents selected
Phenotyped
• 2014-2015

Training Population Size Increase!!!
N = 969 → N = 5507
Does GS improve selection accuracy?

Gain per Year = \( \frac{i \times r \times \sigma_a}{T} \)

- Selection Accuracy
- Time per Cycle
Second Part

- Markers vs. Pedigree
- Predicting breeding value (selecting parents)
- Distinguishing within-families
- Preliminary hints of genetic progress
- Predicting total gen. value (selecting varieties)
Selection Accuracy: Markers vs. Pedigree

• Pedigrees can also be used for selection
• Also disentangle
  – environment,
  – non-additive genetics
  – From breeding value genetics
• Pedigrees are cheaper
Predicting GS Cycle 1

- Predict C1 with C0 (Training Pop.)
- Using 2013-2015 trial data
- Compare Pedigree vs. Marker Prediction

Accuracy = \text{cor}(\text{Breeding Value}, \text{Genotypic Value})
Predicting GS Cycle 1 Markers vs. Pedigree

Prediction Accuracy

\[ \text{cor}(\text{Prediction}, \text{Observation}) \]

- Fresh Root Yield
- Dry Matter
- Mosaic Dis. Severity
Markers > Pedigree because...

- Pedigree may be mis-recorded
- Pedigree is shallow
- Pedigree: *Expected* relationships among families
- Markers: *Realized* relationship among families
- Markers: *Realized* relationships *within* families
Predicting *Breeding Value*

- **Markers** and *pedigree* predict *breeding value*
- NOT the genotypic value
- **Genotypic value** = *Additive* + *Non-additive genetics*
  - Not reliably passed to offspring

Accuracy = cor( Prediction Breeding Value, Observed Breeding Value )
• Estimate breeding value of C1
  – With C1 + C2 (progeny of C1)
  – Using markers / genomic relationships
• Predict C1 with C0
• Compare correlation of prediction with:
  – Obs. Genotypic Value
  – Obs. Breeding Value
Accuracy Predicting Breeding vs. Genotypic Value

Prediction Accuracy

cor(Prediction, Observation)

- Fresh Root Yield
- Dry Matter
- Mosaic Dis. Severity

Genotypic Val
Within-Family Prediction

• Distinguishing among families: Easy
  – Pedigrees do this.

• Distinguishing within families: Hard
  – Markers hopefully do this.

• Two ways to quantify:
  – Prediction with parent means vs. GEBV
  – Accuracy within each family
Prediction Using Mean of Parents

Overall Prediction
Predict C1 GEBV with all C0 (as before)

Prediction with Mean of Parents

\[
\text{C1 Prediction} = \frac{\text{Parent 1 GEBV} + \text{Parent 2 GEBV}}{2}
\]

- If mean of parent’s \( \approx \) Overall pred.
  - Markers *do not* distinguish within families well
- If mean of parent’s \( < \) Overall pred.
  - Markers *do* distinguish within families!
Prediction Using Mean of Parents

Family mean prediction: means of parent’s breeding values

Prediction Accuracy
\[ \text{cor(Prediction, Observation)} \]

- Fresh Root Yield
- Dry Matter
- Mosaic Dis. Severity

Bar chart showing prediction accuracy for different traits.
Accuracy Predicting Each Family

- 17 C1 families have more than 25 progeny
- Calculate accuracy within each family

Yield is a challenge.

Non-additive effects?

Accounting for major gene resistance (CMD2) fixes this
Quick sum-up: Within-family prediction

• All evidence indicates markers capture additive value
  – Even for root yield
• This is also true within families
Is genomic selection working?
Preliminary evidence of gain: Dry Matter Content

- 2013-2015 data
- Distribution of genome estimated breeding values (GEBV) for C0, C1 & C2
Preliminary evidence of gain: CMD Severity
Preliminary evidence of gain: Fresh Root Yield
Summary so far….

- Markers >> Pedigree
- Breeding value predicted >> total genetic value
- Markers can accurately distinguish among- and within-families
  - For most *but not all* families and traits
- GS can accelerate gain for some traits
  - e.g. CMD and Dry Matter Content
  - By reducing cycle time and increasing selection accuracy
Move pop. mean: $\sigma^2_{\text{Additive}}$

Best varieties in tail of distribution:

$\sigma^2_A + \sigma^2_D + \sigma^2_{AA} + \sigma^2_{AD} + \sigma^2_{DD} \cdots$

Adapted from Ceballos et al. 2015. TAG.
Diallel studies indicate significant non-additive variance.

Ratio GCA to SCA

Lower value = relatively higher SCA / non-additive effects

Adapted from Table 3. Ceballos et al. 2015. TAG
Are non-additive genetics strong in our breeding populations?

Can we capture and predict it with markers?
GBLUP Model with Multiple Random Effects

Models Tested

- Additive *only*
- Additive *plus* Dominance
- Additive *plus* Dominance *plus* Add.-by-Dom. Epistasis

\[ Y = X\beta + Zg_{\text{add}} + Zg_{\text{dom}} + Zg_{\text{AxDepi}} + e \]
Non-additive variance in GS Cycle 0 and Cycle 1

Heritability
(Prop. Total Variance Explained)

GS Cycle 0 (Genetic Gain)

Dry Matter

A  A+D  AxD

Dry Matter is *mostly* additive
Non-additive variance in GS Cycle 0 and Cycle 1

Heritability (Prop. Total Variance)

- **Dry Matter**: mostly additive
- **Yield components**: significantly non-additive
- Non-additive effects in C1 > C0
Predicting total genetic value

Models Tested
- Additive (A)
- Additive + Dominance (A+D)
- Additive + Dominance + AxD Epi. (AxD)

Dry Matter

\[
cor(\text{Sum of Genetic BLUPs}, \text{Genetic Value})
\]
# Predicting total genetic value

<table>
<thead>
<tr>
<th>Dry Matter</th>
<th>Harvest Index</th>
<th>Root Weight</th>
<th>Mean CMD Severity</th>
</tr>
</thead>
</table>

Non-additive models predict genotypic value $>_{\text{>}}$ than additive.
Partition additive and non-additive effects to select parents and rank potential varieties

Two levels of selection
- Parent selection
- Variety selection

Parent selection based on genome estimated breeding value!

Variety selection based on sum of additive and non-additive values!

**Genome-enabled Cassava Breeding Cycle**

- **Year 1**: Seedling nursery (SN)
- **Year 2**: Clonal evaluation (CE)
- **Year 3**: Prelim. yield trial (PYT)
- **Year 4**: Advanced yield trial (AYT)
- **Years 5 & 6**: Multi-location Uniform yield trial (UYT)
- **Year 8+**: Multiplication and Release

**Training Data Updates**
- **Crossing Blocks**
- **Source Pops**

**Fast Cycles**
Summary

• Genomic methods confirm non-additive variance, for yield

• Non-additive models predict yield more accurately

• Partition additive from non-additive variance
  – Select parents based on additive prediction
  – Select potential varieties based on total genetic prediction
Thank you!

Cornell University
• Martha Hamblin
• Hale Ann Tufan
• Uche Okeke (IITA)
• Ugochukwu Ikeogu (NRCRI)
• Alfred Ozimati (NaCRRI)
• Roberto Lozano
• Ariel Chan
• Dunia Pino Del Carpio
• Punna Ramu

NaCRRI, Uganda
• Robert Kawuki
• Ismail Kayondo

NRCRI, Nigeria
• Chiedozie Egesi
• Lydia Ezenwaka

IITA, Nigeria
• Ismail Rabbi
• Ipkan Smith Andrew
• Peter Kulakow
• Olumide Alabi

AND MANY OTHERS!!!!

BTI
• Lukas Mueller
• Guillaume Bauchet
• Isaak Tecle

JGI
• Simon Prochnik
• Jessen Bredeson