Incorporating genotype-by-environment and genomic selection information into oat breeding programs

Lucía Gutiérrez
Quantitative Genetics and Cereals Breeding
Agronomy Department
University of Wisconsin - Madison
gutierrezcha@wisc.edu
July 2016
International Oats Conference, St. Petersburg, Russia
Selection Response

\[ \Delta Y = b(\Delta X) \]

\[ R = b_{x,y}S \quad b_{x,y} = \frac{R}{S} = h^2 \]

\[ \Delta G = \frac{C_i V_A}{y \sigma_P} \]

\( \mu_0 = \) mean of the initial Random Mating population

\( \mu_s = \) mean of a selected individual s from R.M. population

\( \mu_1 = \) mean of a progeny of selected individual s

\( c = \) truncation point

\( S = \) selection differential \( (\Delta X) \)

\( R = \) response to selection \( (\Delta Y) \)

Holland et al. 2010

Hull, 1945
Plant Breeding

TRADITIONAL PLANT BREEDING

Relies mainly on plant or population phenotypic evaluations and pedigree information

MARKER ASSISTED SELECTION

Uses some genotypic information to improve selection

GENOMIC SELECTION

Uses a lot of genotypic information to improve selection
What is the Phenotype?

\[ P = G + E + GE \]
GxE Interaction

\[ P_{ijk} = \mu + G_i + E_j + GE_{ij} + \epsilon_{ijk} \]
Dealing with GxE

1. Ignore
2. Avoid
3. Exploit

Bernardo (2010)
Genomewide scan with QTLxE

\[ P_{ij} = \mu + E_j + G_i + GE_{ij} \]

\[ P_{ij} = \mu + E_j + x_i \alpha + G_i^* + x_i \alpha_j^* + GE_{ij}^* \]

\[ P_{ij} = \mu + x_i \alpha_j + \epsilon_{ij} \]

\( \alpha = \) QTL main effect

\( \alpha_j^* = \) QTLxE = deviation of QTL main effect

\( \alpha_j = \) Environment - specific QTL effect

Gutierrez et al., 2015 (TAG 128: 506-521)
Handling GxE for QTL Mapping

QTLxE in a RIL population for saline stress in *Lotus*

*Quero et al., 2014 (Crop Pasture Sci. 65(2): 139-149)*
Handling GxE for QTL Mapping

QTLxE on a GWAS population for disease resistance traits in Barley

Global QTL as well as Mega-environment specific QTL were detected. Different strategies should be used for each type.
QTL for Marker Assisted Selection


Spot Blotch

Leaf Rust

Winterhardiness

Malting Quality

Zitzewitz et al., 2011 (The Plant Genome 4(1): 76-91)

Gutierrez et al., 2011 (The Plant Genome 4:256 – 272)
Some Limitations of MAS

LIMITATIONS OF MAS (BEAVIS EFFECT)

1. Underestimation of the number of QTL
2. Over-estimation of effects

Bernardo 2010
# Molecular Breeding

## SELECTION RESPONSE VS. UNDERLING CAUSES

<table>
<thead>
<tr>
<th>Trait</th>
<th>Genetic control</th>
<th>Population used for selection</th>
<th>Accuracy of gene/QTL position</th>
<th>Molecular Breeding (MB) method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>Few major genes</td>
<td>Elite x Elite</td>
<td>High accuracy - Ideally marker = gene - Often, marker close to gene (stable association)</td>
<td>MAS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elite x Non-elite</td>
<td></td>
<td>MABC</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Few QTL of large effects</td>
<td>Elite x Non-elite</td>
<td>Medium to low accuracy (CI &gt;10cM) - Markers linked to QTL - Associations marker-QTL depend on the genetic background</td>
<td>MARS (or GWS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Elite x Elite</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Numerous QTL of small effects</td>
<td>Elite x Elite</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Building blocks

Black box
Genomic Selection

STEPS IN GENOMIC PREDICTION

1. **Create/choose a Training Population**: this is a group of individuals that will be used to train the model (i.e. obtain marker predictors).
   Things to take into account:
   a) Population size
   b) Trait: $h^2$, number of QTL, and trait *per se*
   c) Population structure
   d) Relationship to Testing Population
   e) Markers: Number and platform
   f) Genotype by Environment Interaction

2. **Genotype and Phenotype individuals in the Training Population**.

3. **Train the model**: use one of the methods described previously to obtain marker predictors.

4. **Create/choose a Testing Population**: this is a set of individuals whose merit we wish to determine.

5. **Genotype individuals in the Testing Population**.

6. **Predict Breeding Values for the Testing Population**: use the model created with the training population and the genotypic information of the testing population to determine their merit (i.e. to get the GEBV).
Handling GxE in GS

Genomic Selection with GxE

\[ y = 1 \mu + Zg + \varepsilon \]

**GBLUP\(_{(M)}\):**

- \( y_{(Nx1)} \): vector of mean performance
- \( g \sim N(0, A\_{(NxN)} \sigma^2_g) \)
- \( N \): number of genotypes

**GBLUP\(_{(gxe)}\):**

- \( y_{(nx1)} \): vector of mean performance
- \( g \sim N(0, A\_{(NxN)} \otimes \rho\_{(kxk)} \sigma^*^2_g) \)
- \( n \): number of genotypes \((N)\) by number of environments \((k)\)

*Lado et al., 2015 (Crop Sci.)*
Handling GxE for GS

Lado et al., 2015 (Crop Sci.)

- Year
- Location
- Mega-Env.
Handling GxE for GS

<table>
<thead>
<tr>
<th>Set</th>
<th>Training set</th>
<th>Testing Year</th>
<th>LE 1</th>
<th>LE 2</th>
<th>YOU 1</th>
<th>YOU 2</th>
<th>DOL 1</th>
<th>DOL 2</th>
<th>R2 1</th>
<th>R2 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>By year†</td>
<td>All locations in testing year but one</td>
<td>2010</td>
<td>0.31</td>
<td>0.31</td>
<td>0.12</td>
<td>0.17</td>
<td>0.71</td>
<td>0.71</td>
<td>0.57</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>0.42</td>
<td>0.39</td>
<td>0.39</td>
<td>0.31</td>
<td>0.52</td>
<td>0.37</td>
<td>0.44</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>0.74</td>
<td>0.79</td>
<td>0.44</td>
<td>0.54</td>
<td>0.78</td>
<td>0.81</td>
<td>0.57</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013</td>
<td>0.03</td>
<td>0.05</td>
<td>0.40</td>
<td>0.74</td>
<td>0.36</td>
<td>0.59</td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>0.23</td>
<td>−0.14</td>
<td>0.30</td>
<td>0.19</td>
<td>0.58</td>
<td>0.71</td>
<td>0.30</td>
<td>0.19</td>
</tr>
<tr>
<td>By location‡</td>
<td>All years for a location but one</td>
<td>2010</td>
<td>0.16</td>
<td>0.17</td>
<td>0.01</td>
<td>−0.02</td>
<td>−0.06</td>
<td>0.04</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>0.23</td>
<td>0.18</td>
<td>0.27</td>
<td>0.34</td>
<td>0.08</td>
<td>0.14</td>
<td>0.19</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>0.50</td>
<td>0.31</td>
<td>0.10</td>
<td>0.04</td>
<td>0.33</td>
<td>0.29</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013</td>
<td>0.23</td>
<td>0.10</td>
<td>−0.05</td>
<td>0.08</td>
<td>0.31</td>
<td>0.56</td>
<td>0</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>0.21</td>
<td>0.20</td>
<td>0.06</td>
<td>−0.05</td>
<td>0.08</td>
<td>0.36</td>
<td>−0.05</td>
<td>0.33</td>
</tr>
<tr>
<td>By mega-environment§</td>
<td>All environments for a specific mega-environment but one</td>
<td>2010</td>
<td>0.10 (1)</td>
<td>0.20 (1)</td>
<td>0.11 (1)</td>
<td>0.18 (1)</td>
<td>0.72 (1)</td>
<td>0.72 (1)</td>
<td>0.32 (3)</td>
<td>0.35 (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2011</td>
<td>0.36 (2)</td>
<td>0.35 (2)</td>
<td>0.36 (2)</td>
<td>0.44 (2)</td>
<td>0.39 (1)</td>
<td>0.28 (1)</td>
<td>0.38 (1)</td>
<td>0.39 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2012</td>
<td>0.63 (3)</td>
<td>0.45 (3)</td>
<td>0.46 (2)</td>
<td>0.54 (2)</td>
<td>0.75 (3)</td>
<td>0.80 (3)</td>
<td>0.63 (3)</td>
<td>0.30 (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013</td>
<td>0.28 (1)</td>
<td>0.18 (1)</td>
<td>0.46 (3)</td>
<td>0.43 (3)</td>
<td>0.43 (2)</td>
<td>0.50 (2)</td>
<td>0.44 (1)</td>
<td>0.32 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2014</td>
<td>0.24 (2)</td>
<td>0.25 (2)</td>
<td>−0.01 (2)</td>
<td>0.11 (2)</td>
<td>0.63 (3)</td>
<td>0.78 (3)</td>
<td>−0.03 (3)</td>
<td>0.20 (3)</td>
</tr>
</tbody>
</table>

Lado et al., 2015 (Crop Sci.)
Handling GxE for GS

USING ENVIRONMENTAL COVARIATES

<table>
<thead>
<tr>
<th>Year</th>
<th>GBLUP(M)</th>
<th>GBLUP(gxe)</th>
<th>GBLUP(gxe) + Cov</th>
<th>GBLUP(cov)</th>
<th>GBLUP(M) + Cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0.204</td>
<td>0.669</td>
<td>0.742</td>
<td>0.243</td>
<td>0.324</td>
</tr>
<tr>
<td>2011</td>
<td>0.469</td>
<td>0.652</td>
<td>0.688</td>
<td>0.236</td>
<td>0.237</td>
</tr>
<tr>
<td>2012</td>
<td>0.475</td>
<td>0.372</td>
<td>0.484</td>
<td>0.176</td>
<td>0.185</td>
</tr>
<tr>
<td>2013</td>
<td>0.551</td>
<td>0.622</td>
<td>0.496</td>
<td>0.292</td>
<td>0.286</td>
</tr>
<tr>
<td>2014</td>
<td>0.234</td>
<td>0.646</td>
<td>0.679</td>
<td>0.243</td>
<td>0.269</td>
</tr>
</tbody>
</table>

Gonzalez-Barrios et al., (in prep)
Right off my plane

Crosses

Increases

Increases

Planting

Planting

Quite a sight!

2-leave

Advanced

Elite

Uniform
Oat Breeding at UW-Madison

- Madison
- Marshfield
- Spooner
- Arlington
- Sturgeon Bay
Oat ideas

Use the Information provided on T3:

- CORE
- POGI
- BREEDING PROGRAMS

GBS Data (Nick Tinker)
Historical Phenotypic data (highly unbalanced)
Good GS predictions (Jean-Luc Jannink)

GxE characterization
Strategic Phenotyping
Strategic Phenotyping

Devise a training GLOBAL Oats data-set
Use GxE to enhance prediction ability
Funding and Acknowledgement

**Statistical Genetics:**
Gaston Quero
Sofia Brandariz
Bettina Lado
Ines Berro
Pablo Gonzalez
Agustín Gonzalez
Alejandra Borges
Pablo Sandro
Natalia Berberian
Juan Rosas

**Rice Breeding:**
Alvaro Roel
Gonzalo Zorrilla
Marco Dalla Rizza
Victoria Bonnecarrere
Pedro Blanco
Fernando Pérez de Vida
Schubert Fernández
Silvia Garaycochea

**Wheat Breeding:**
Martin Quincke
Paula Silva
Silvia German

**Animal Breeding:**
Ignacio Aguiar

**Barley Breeding:**
Ariel Castro
Silvia German

Funding:

-**OSU**
-**Oregon State University**

-**CSIC**

-**Wageningen University**

-**Generation Challenge Programme**

-**INIA**

-**K-State**
-**Kansas State University**

-**The University of Alabama**

-**Jesse Poland**

-**Monica Balzarini**

-**Cecilia Bruno**

-**Andrea Pena**

-**Gustavo de los Campos**

-**Ana Ines Vazquez**

-**Fred van Eeuwijk**

-**Marcos Malosetti**
Genomic Selection

A) POPULATION SIZE AND B) TRAIT $h^2$:
Larger population sizes and higher heritabilities increase genomic selection accuracy

Hayes et al., 2009; Lorenzana and Bernardo, 2009; Lorenz et al., 2011; Asoro et al., 2011
Genomic Selection

B) TRAIT: per se

The trait is relevant not only for its heritability, but for the trait *per se*.

150 DH from Steptoe x Morex with 233 Markers (Hayes et al., 1993)
Genomic Selection

C) POPULATION STRUCTURE
AND D) RELATIONSHIP TO INDIVIDUALS

5 wheat populations with N= 92, 176, 148, 90, 176
for stem rust with DArTs

Crossa et al., 2013
B) POPULATION STRUCTURE AND D) RELATIONSHIP TO INDIVIDUALS

Older training populations have smaller or no prediction accuracy difference.

Related training and testing populations have larger prediction accuracy.

Mixed training populations have lower accuracy unless mixed populations are larger.

Asoro et al., 2011
E) NUMBER OF MARKERS

223 RIL maize BM-TC1 test-crossed 1339 SSR or RFLP

374 winter wheat F5 lines with 5000 DArTs

421-446 Oat lines, 1005 DArT markers

Lorenzana and Bernardo, 2009; Heffner et al., 2011; Asoro et al., 2011
Genomic Selection

PREDICTION METHOD

Prediction method (BayesC, RR-BLUP) did not affect accuracy as much as trait, marker density, and training population size, depth (increasing population size by including older lines increased selection accuracy) and age (older populations have lower accuracy for some traits).

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Marker density</th>
<th>Training population size</th>
<th>Training population depth</th>
<th>Training population age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait†</td>
<td>4</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Method‡</td>
<td>1</td>
<td>0.22</td>
<td>0.02</td>
<td>0.21</td>
<td>0.06</td>
</tr>
<tr>
<td>Design§</td>
<td>2</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Trait × method</td>
<td>4</td>
<td>0.03</td>
<td>0.14</td>
<td>0.56</td>
<td>0.31</td>
</tr>
<tr>
<td>Trait × design</td>
<td>8</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.0001</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Method × design</td>
<td>2</td>
<td>0.26</td>
<td>0.11</td>
<td>0.64</td>
<td>0.52</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

421-446 Oat lines, 1005 DArT markers

Asoro et al., 2011