Oat adaptation to Mediterranean climate: a critical analysis for association studies

Spanish Research Council (CSIC) Spain

Nicolas Rispail, Gracia Montilla-Bascón, Javier Sánchez-Martín, Fernando Flores Amero Emeran, Josefina C Sillero, Tim Langdon, Catherine Howarth, Diego Rubiales, Elena Prats.
WHY IS NEEDED TO IMPROVE ADAPTATION OF OATS IN MEDITERRANEAN AREA?
In the last 20 years...

Northern Europe

- 8500 Ha decrease per year

Mediterranean Basin

- 7500 Ha increase per year

Increasing interest in Mediterranean area for oat cultivation

FAO, 2015
Oats in Mediterranean Basin vs Northern Europe

Average data of the last 20 years (FAO, 2015)

Agroclimatic conditions in the North may favour high yields ……

**BUT**

…… there is plenty of room to increase yield under southern conditions
Oats in Mediterranean Basin

Reasons for the low yield

- yield instability (biotic and abiotic stresses)
- limited adaptation of many of the varieties used bred in northern countries as spring varieties and used in the Mediterranean area as a winter crop
- Low breeding effort (low yearly gain)
DIFFICULTIES FOR BREEDING COMPLEX TRAITS SUCH AS YIELD RELATED COMPONENTS AND STRESS RESISTANCE

\[ \text{VALIDATION OF AGRONOMIC COMPONENTS AT FIELD IT IS CRUCIAL} \]

\[ \text{GENOTYPE } \times \text{ ENVIRONMENT INTERACTIONS } (G \times E) \]
Selection & Validation at field: G x E interactions

- 174 oat accessions
- 4 Mediterranean countries
- 6 locations
- 2 years experiments
- 12 environments

- Identification of mega-environments
- Selection of locations with good discrimination power and repeatability for each/all traits
- Selection of good and stable cultivars

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

✓ Identification of mega-environments

Spain and Tunisia are in different ME that Egypt and Palestine

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

✓ Selection of locations: i.e. Yield

PC1 = 44%, PC2 = 22%, Sum = 66%
\((G+GE)/(E+G+GE) = 0.46\)

In this example:

- Salamanca has not good discrimination power
- Cordoba has good discrimination power, is representative of average environment and is repeatable
- Escacena is similar to Cordoba but some years are not so repeatable

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

✓ Selection of cultivars: i.e. Rust

Resistant and stable genotypes

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

- Selection of locations of cultivars: i.e. Rust

Susceptible genotypes

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

✓ Selection of locations of cultivars: i.e. Rust

Inverse behaviour may indicate a different rust isolate & genotypes with major resistance genes

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
Selection & Validation at field: G x E interactions

✓ Selection of locations of cultivars: i.e. Rust

Similar behaviour may indicate diverse resistance responses usually poligenic controlled = durable resistance

This is supported by the controlled conditions experiments

Sánchez-Martín et al., 2014, Field Crop Research
Sánchez-Martín et al., Submitted
BIPlOTS ANALYSIS ARE VERY USEFUL BUT COMPLEX

↓

MARKERS FOR AGRONOMIC TRAITS DERIVED FROM THESE WOULD BE AN ADVANTAGE
GWAS for development of markers associated with agronomic traits

Collection with high genetic diversity

- Phenotyping
- Genotyping

Statistical models for marker-trait correlation (GLM, MLM)

- Linkage disequilibrium
- Population structure and relatedness

Identification marker-trait associations

Montilla Bascón et al., 2015, Frontiers Plant Sci
GWAS for agronomic traits

- Yield,
- Biomass,
- H Index,
- Flowering date
- Rust resistance
Development of physiological and molecular markers

GWAS Scheme

Phenotyping

Genotyping

- **Genotyping**: 15000 DArT and 31 SSR
- **Data curation**: 1872 polymorphic markers

Montilla Bascón et al., 2015, Frontiers Plant Sci
GWAS for agronomic traits

GWAS Scheme

Phenotyping
Genotyping
Linkage disequilibrium

Most of the marker pairs (507042) showed a significant LD
277920 marker pairs showed a $r^2 < 0.1$ indicating a good genome coverage of the markers

Montilla Bascón et al., 2015, Frontiers Plant Sci
Development of physiological and molecular markers

GWAS Scheme

- Phenotyping
- Genotyping
- Linkage disequilibrium
  - Population structure and relatedness

✔ Population Structure:

- Varieties
- Red Oats
- White oats adapted to low altitude
- White oats adapted to high altitude

Montilla Bascón et al., 2013, Plant Mol Biol Rep
Montilla Bascón et al., 2015, Frontiers Plant Sci
Development of physiological and molecular markers

GWAS Scheme

- Phenotyping
- Genotyping
- Linkage disequilibrium

Statistical models for marker-trait correlation (GLM, MLM)

Population structure and relatedness

Models:
- GLM,
- (GLM+PCA)
- (GLM+Q matrix)
- (GLM+ PCA + Q matrix)
- (MLM)

Statistical Tests:
- Marker-\(p\)
- \(q\)-value (False Discovery Rate FDR)
- \(p\)-value (Bonferroni)

To remove additional false positives family-related.

To verify the significance of the associations

Montilla Bascón et al., 2015, Frontiers Plant Sci
Controversy about specific approaches:

- Different set of data:
  - environments (results of each location and year): 12
  - locations (results of the different year in the same location): 5
  - megaenvironments: grouping similar environments: 2

- Different possible estimates: Means, BLUPs (best linear unbiased predictors)

- Different models: GLM, GLM corrected, MLM

- Different statistical tests: Lod 3, FDR, Bonferroni, etc.
GLM (corrected with Q or PCA) vs MLM models
GWAS for agronomic traits

**MLM models does not always fit better than GLM models**

Significant markers < 1 false positive according to FDR

<table>
<thead>
<tr>
<th>Marker</th>
<th>marker_(p)</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>oPt-794463</td>
<td>1,16E-10</td>
<td>1,89E-07</td>
</tr>
<tr>
<td>Merge2</td>
<td>2,90E-06</td>
<td>0,0023664</td>
</tr>
<tr>
<td>oPt-3224</td>
<td>4,64E-04</td>
<td>0,252416</td>
</tr>
<tr>
<td>oPt-0657</td>
<td>7,03E-04</td>
<td>0,286824</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>p</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>oPt-794463</td>
<td>7,31E-09</td>
<td>1,19E-05</td>
</tr>
<tr>
<td>Merge2</td>
<td>2,98E-05</td>
<td>0,0243168</td>
</tr>
<tr>
<td>oPt-0657</td>
<td>3,84E-04</td>
<td>0,208896</td>
</tr>
<tr>
<td>oPt-5986</td>
<td>5,23E-04</td>
<td>0,213384</td>
</tr>
<tr>
<td>oPt-3224</td>
<td>9,95E-04</td>
<td>0,324768</td>
</tr>
</tbody>
</table>
GWAS for agronomic traits

MLM models does not always fit better than GLM models

Good fit, with significant markers

Model do not fit
Red line under diagonal=no significant markers
GWAS for agronomic traits

What is a good fit?

Slopes closer to 1 as possible and no higher than $1 \pm 0.20$ ($\lambda<1.2$)

A good fit is essential for next steps...
GWAS for agronomic traits

How the data of each environment (year + site) affect the markers raised for a locality or Mega-environment?

Cordoba (20)
Co9 (5)
Co10 (16)

Phenotypic distribution

environments with a better distribution of phenotypes have more weight in average environments
BLUPs vs Means

BLUPs do not always fit better than means although many times they can improve data.

When you have a strong marker this usually is significant independently of the data used.
GWAS for agronomic traits

Statistical Tests:

✓ Marker-p $\rightarrow$ LOD 3??

✓ q-value (False Discovery Rate FDR) $\rightarrow$ 0.25??

✓ p-value (Bonferroni) $\rightarrow$ 0.05 / nº markers

I think this is a completely arbitrary criteria, but if a marker is significant according to other criteria in several environments and it is under LOD 3 in others this may support the strength of the marker.....

✓ Marker-p $\rightarrow$ LOD 3?? (p=0.001)

✓ p-value (Bonferroni) $\rightarrow$ too strict???
GWAS for agronomic traits

Take into account the distribution of $p$

$q$-value should be determined for each model

$\sqrt q$-value (FDR test)

$\text{cut\_off} = 0.18$

$\text{cut\_off} = 0.33$
GWAS for agronomic traits

**Significant markers: YIELD**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Cordoba</th>
<th>Escacena</th>
<th>Salamanca</th>
<th>MegaEnvironment</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>oPt-10734</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>0,06</td>
</tr>
<tr>
<td>oPt-12924</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>0,073</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>0,054</td>
</tr>
<tr>
<td>437</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>0,056</td>
</tr>
<tr>
<td>oPt-9844</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>0,057</td>
</tr>
<tr>
<td>oPt-5217</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>0,06</td>
</tr>
<tr>
<td>oPt-8261</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>0,058</td>
</tr>
<tr>
<td>oPt-9348</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>0,04</td>
</tr>
<tr>
<td>Merge53</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>0,041</td>
</tr>
<tr>
<td>oPt-0246</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,054</td>
</tr>
<tr>
<td>oPt-17855</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,047</td>
</tr>
<tr>
<td>oPt-0894</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,052</td>
</tr>
<tr>
<td>Merge126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,075</td>
</tr>
<tr>
<td>oPt-0802</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,064</td>
</tr>
<tr>
<td>oPt-3038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,065</td>
</tr>
<tr>
<td>oPt-4136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,057</td>
</tr>
<tr>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,036</td>
</tr>
<tr>
<td>Merge85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,047</td>
</tr>
<tr>
<td>oPt-7795</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,053</td>
</tr>
<tr>
<td>oPt-13101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,075</td>
</tr>
<tr>
<td>Merge13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,061</td>
</tr>
<tr>
<td>Merge2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,1385</td>
</tr>
<tr>
<td>oPt-5986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,096</td>
</tr>
<tr>
<td>oPt-3224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,059</td>
</tr>
<tr>
<td>oPt-0657</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,069</td>
</tr>
<tr>
<td>oPt-794463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,293</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>0,064</td>
</tr>
</tbody>
</table>

Explaining in total aprox 65% of the variance
### GWAS for agronomic traits

#### Significant markers: RUST RESISTANCE

<table>
<thead>
<tr>
<th>Marker</th>
<th>Cordoba</th>
<th>Escacena</th>
<th>Salamanca</th>
<th>ME1</th>
<th>ME2</th>
</tr>
</thead>
<tbody>
<tr>
<td>oPt-9546</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>AM30-178</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>oPt-11494</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>oPt-388855</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>oPt-10121</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>oPt-388933</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>oPt-794313</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Merge136</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>oPt-14210</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>oPt-3145</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
Thanks a lot for your attention