Association genetics in barley

Ildikó Karsai, Hungarian Academy of Sciences, Martonvásár, Hungary

Frank Ordon, Dragan Perovic, Julius Kühn Institute, Quedlinburg, Germany

Günther Schweizer, Bianca Büttner, Bayerische Landesanstalt für Landwirtschaft, Freising, Germany

Jutta Förster, Saaten Union Biotech, Germany

José Luis Molina, IRTA Lleida, Spain

Marian Moralejo, Universitat de Lleida, Spain

Francisco Ciudad, ITACyL Valladolid, Spain

Ana Casas, Samia Yahiaoui (INRA Algerie), Alfonso Cuesta-Marcos (Oregon State Univ, Corvallis), Cristina Casao (MPIZ, Kölln), Cristina Silvar (Univ A Coruña), Pilar Gracia, José Manuel Lasa, E Igartua
Association in the Spanish Barley Core Collection (SBCC)
Choice of germplasm. Emphasis on application for breeding

http://www.eead.csic.es/EEAD/barley/
WHAT IS A CORE COLLECTION?

Frankel (1984) introduced the concept and defined it as 'a limited set of accessions which represents, with a minimum of repetitiveness, the genetic diversity of a crop species and its wild relatives'.

Core collections are increasingly used as diversity panels to mine for useful traits and genes, using genome wide association. They have the advantage of being a permanent resource, available to the scientific community, therefore fostering widespread use.

Excellent potential for high throughput phenotyping

Plant material: 185 genotypes, including 159 landrace-derived inbred lines (148 6-row, 11 2-row) from the Spanish Barley Core Collection, and 26 cultivars (13 6-row, 13 2-row)

Markers: Barley OPA1 (1536 SNPs), Infinium iSelect (9000 SNPs), functional markers for flowering time genes, approx. 4000 polymorphic markers with map position. Marker selection: <10% missing, MAF>5%

Phenotypes: 50 traits, agronomic (recorded at 10 field trials in Spain), morphological, and disease resistance (field evaluations in Spain and controlled inoculations in Germany, JKI and LfL)

Association: TASSEL, model K (for Kinship). K matrix from TASSEL, done with 500 best quality SNPs, equally distributed.
SBCC: Diversity

Lasa et al. 2002, Hereditas; Silvar et al. 2010, Plant Breeding

OPTICHINA, Barcelona, September 2012
SBCC: Population structure

- Previously unknown underlying populations revealed by cluster analysis (STRUCTURE)
- Spanish 6-row landraces are constituted by two main populations, close to other Mediterranean cultivars, but different from cultivars from more Northern countries. Spanish 2-row landraces are not much different from cultivars from other origins.

Yahiaoui et al. 2008, Theor Appl Genet
Association mapping issues

• Genome wide association mapping (GWAS) procedures must be fine tuned to each experiment: “One size does NOT fit all”

• Choice of genotypic panel: phenotypic diversity for target traits must be sufficient

• Population structure induces high rates of false associations (false positives). Avoid disconnected genotypic groups known a priori

• Density of markers must be enough to provide a good genome cover. Even then, density can be highly variable

• Ascertainment bias. Selection of SNPs in relevant groups of germplasm

• SNP selection: polymorphisms targeted may not be functional, or even representative of the loci

• Not for every trait (more later)
**Association mapping issues: density**

![Graph: Average LD (r²)](image)

**Significant LD up to 3 cM**
Association mapping issues: landscape of genomic diversity. Ascertainment bias in SNPs?

Igartua et al. 2012, Genet Res Crop Evol
Level of interchromosomal LD: number of loci in linkage disequilibrium (LD, $r^2$) with loci from other chromosomes. Ideally, the profile should be flat to avoid distortion in association patterns.
Methods to control population structure

- **Nothing.** OK if structure is weak

- **Cluster analysis** (STRUCTURE) = Q. Usually regarded as too liberal.

- **Genomic control.** Not widely used.

- **Principal component analysis** (Eigenstrat). Widely used.

- **Mixed models** (Yu et al. 2005, Nature Genet)
  - Q
  - K. Widely used
  - Q+K. Widely used

Methods to control for multiple testing

- **Nothing.** Not recommended, but still used in some cases

- **Estimated proportion of significant tests** (Pasam et al, 2012, BMC plant Biology). Liberal, but acceptable if used with strict control of population structure

- **Multiple regression**, with sequential introduction of significant markers. Seldom used.

- Threshold set by *random variables* analyzed with mixed models (Comadran et al, 2011, Theor Appl Genet)

- **False Discovery Rate, FDR.** Very popular, implemented in standard statistical packages (SAS, Genstat, R?)

- **Bonferroni** correction, full (too strict) or for an estimated effective number of independent tests (Varshney et al 2012, Field Crops Res)
Comparison of models to control for population structure for row type. 224 spring barleys, worldwide. Expected versus observed P values for the association tests. Pasam et al. 2012. BMC Plant Biology
Comparison of 2 models to control for population structure for 25 traits of the SBCC. Expected versus observed P values for the association tests.
Population structure. Performance of models

Linear correlation coefficients between P of association tests and several genetic features

<table>
<thead>
<tr>
<th></th>
<th>Naïve</th>
<th>Q model</th>
<th>K model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long range LD</td>
<td>0.29</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Short range LD</td>
<td>0.23</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>Diversity</td>
<td>-0.19</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Omitted from the original presentation: Genome-wide graphs, –log10Probability vs cM for:

- Powdery mildew tolerance, average
- Powdery mildew tolerance, 7 isolates
- Leaf rust tolerance
- Scald
- Grain yield under low productivity
- Heading time in autumn sowings
- Row type
- Vernalization effect
Candidate genes for adaptation associated with geographical features and function

VrnH1

VrnH1 allele distribution over type of winter

<table>
<thead>
<tr>
<th>Type of Winter</th>
<th>VrnH1 allele</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VrnH1-6</strong></td>
<td>8 74 10 0</td>
</tr>
<tr>
<td><strong>VrnH1-4</strong></td>
<td>1 19 29 2</td>
</tr>
</tbody>
</table>

Casao et al. 2011, Molec Breeding
<table>
<thead>
<tr>
<th>Population</th>
<th>Type</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBCC097 x Plaisant</td>
<td>RIL</td>
<td>Powdery mildew, flowering</td>
</tr>
<tr>
<td>SBCC145 x Beatrix</td>
<td>DH</td>
<td>Scald, powdery mildew, flowering, height</td>
</tr>
<tr>
<td>SBCC154 x Beatrix</td>
<td>DH</td>
<td>Scald, flowering, spike morphology</td>
</tr>
<tr>
<td>SBCC042 x Cierzo</td>
<td>BC2 RIL</td>
<td>Yield</td>
</tr>
<tr>
<td>SBCC073 x Cierzo</td>
<td>BC2 RIL</td>
<td>Yield</td>
</tr>
<tr>
<td>SBCC016 x Esterel</td>
<td>RIL</td>
<td>Vernalization</td>
</tr>
<tr>
<td>SBCC157 x Hispanic</td>
<td>BC3</td>
<td>Vernalization, cold tolerance</td>
</tr>
<tr>
<td>SBCC058 x Plaisant</td>
<td>BC3</td>
<td>Vernalization, cold tolerance</td>
</tr>
</tbody>
</table>
Candidate genes for adaptation associated with geographical features and function

VrnH3

Casas et al. 2011, Theor Appl Genet

OPTICHINA, Barcelona, September 2012
Candidate genes for adaptation associated with geographical features and function

### VrnH3

<table>
<thead>
<tr>
<th></th>
<th>&lt;40° N</th>
<th>&gt;40° N</th>
</tr>
</thead>
<tbody>
<tr>
<td>VrnH3-TC</td>
<td>67</td>
<td>18</td>
</tr>
<tr>
<td>VrnH3-AG</td>
<td>13</td>
<td>60</td>
</tr>
</tbody>
</table>

**Latitude**

Adapted from Casas et al. 2011, Theor Appl Genet

135-139: polymorphism at the promoter
AG-TC: polymorphism at the first intron

---

OPTICHINA, Barcelona, September 2012
Validation, QTL search, fine mapping: biparental populations

VrnH3. Three alleles confronted in two populations with one common parent. Effect of the polymorphisms at the promoter and first intron apparently confirmed with markers developed within the gene.

Ponce-Molina et al., 2012, Crop Sci
Validation, QTL search, fine mapping: biparental populations

SBCC097 x Plaisant

SBCC145 x Beatrix

Powdery mildew. Three alleles found in two populations with two different accessions as donors, apparently new for cultivated barley. One, on 6HL extreme, confirms association results

Silva et al. 2010, Molec Breeding; Silva et al., 2011, Theor Appl Genet
Candidate genes for adaptation associated with geographical features and function

PpdH2, winter barleys

<table>
<thead>
<tr>
<th>Latitude</th>
<th>PpdH2</th>
<th>ppdH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;44° N</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>44°-50° N</td>
<td>11</td>
<td>42</td>
</tr>
<tr>
<td>&gt;50° N</td>
<td>6</td>
<td>46</td>
</tr>
</tbody>
</table>

Days to heading, field

Vernalization Treatment (days)
GWAS in plants feasible, but not for every trait. Very difficult for traits directly related with history of selection, or correlated with population structure.

The choice of germplasm is critical. Core collections are a good starting point, but further elaboration may be needed to achieve an optimum panel.

Association with geo-referenced accessions offers further layers of information to extract biological meaning.

Validation in biparental populations provides further proof for trait discovery.

Ascertainment bias can be an issue for SNP platforms. GBS will be unbiased.