Development of a High-Density SNP Genotyping Panel as a Community Resource for Genetic Analysis in Oat

Shiaoman Chao
USDA-ARS Fargo, ND
CORE Project (Collaborative Oat Research Enterprise)

- Project director: Eric Jackson.
- Large-scale SNP discovery based on Roche 454 reads, high throughput genotyping.
- Phenotyping CORE breeding lines, and association mapping.
- Implementing SNPs in marker assisted breeding.
- Database development
Resources for Oat SNP Discovery

• cDNAs – derived from roots, shoots, immature and mature embryos of 20 diverse hexaploid cultivars.
• DArTs – complexity reduced genomic fragments derived from 25 diverse hexaploid cultivars.
• Complexity reduced genomic fragments derived from two tetraploid accessions.
SNP Discovery Pipelines

- **Single template approach** (Gerry Lazo) - raw reads from Ogle, Hurdal, Assiniboia, and TAM O-301 were assembled to form four reference templates, which were used to align with reads from other accessions individually.

- **Composite template approach** (Nick Tinker) - raw reads from each accession were assembled, composite templates were formed to align with condensed reads from each accession individually.

- Filtering steps, read depth $\geq$ 5 to call SNP candidates.
OPA Development and SNP Genotyping

- Illumina GoldenGate genotyping assay.
- Sent ~11,000 *in silico* SNPs for assay design.
- Developed 4 pilot oat OPAs.
- Oat OPA1, 2 and 4 each had 1536 SNPs.
- Oat OPA3 had 3072 SNPs, but assay failed.
- Evaluated a total of 4,608 SNPs in three OPAs,
  - 3,930 cDNAs derived
  - 578 DArTs derived
  - 100 tetraploid derived
Samples Used for Evaluation

Total 576 samples evaluated,

• 109 germplasm originated from wide geographic regions in the world.

• 6 mapping populations (partial).

• A set of monosomic lines.
Genotyping Performance

- Dosage and cluster compression
- Heterozygote controls
- F2 or backcross mapping populations, early generation breeding lines
# SNP Discovery Methods

## Pilot Oat OPA1 and 2 Performance - based on 109 germplasm

<table>
<thead>
<tr>
<th>Discovery method</th>
<th>No. Tested</th>
<th>Good SNPs</th>
<th>Conversion rate (%)</th>
<th>No. SNP mapped</th>
<th>%</th>
<th>Polymorphic SNPs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA-STA</td>
<td>2270</td>
<td>991</td>
<td>44</td>
<td>738</td>
<td>33</td>
<td>878</td>
<td>89</td>
</tr>
<tr>
<td>cDNA-CTA</td>
<td>336</td>
<td>144</td>
<td>43</td>
<td>98</td>
<td>29</td>
<td>133</td>
<td>92</td>
</tr>
<tr>
<td>DArT-STA</td>
<td>300</td>
<td>121</td>
<td>40</td>
<td>86</td>
<td>29</td>
<td>108</td>
<td>89</td>
</tr>
<tr>
<td>DArT-Sanger’s</td>
<td>66</td>
<td>48</td>
<td>73</td>
<td>36</td>
<td>55</td>
<td>43</td>
<td>90</td>
</tr>
<tr>
<td>Genomic Tetraploid</td>
<td>100</td>
<td>76</td>
<td>76</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3072</strong></td>
<td><strong>1380</strong></td>
<td><strong>45</strong></td>
<td><strong>965</strong></td>
<td>31</td>
<td><strong>1169</strong></td>
<td>85</td>
</tr>
</tbody>
</table>
Pilot Oat OPA3 (3072-plex) Design Problem

- Pilot Oat OPA2
  - 400 genomic SNPs
  - 1136 cDNA SNPs
  - (1:3)

- Pilot Oat OPA3
  - 1476 genomic SNPs
  - 1596 cDNA SNPs
  - (1:1)

- 300 DArT
- 100 4X

- 1176 DArT
- 300 4X
Pilot Oat OPA4 (1536-plex) Design and Performance

• Filtered SNPs derived from highly abundant sequences by *in silico* prediction,
  o Searched 3 oligos designed by Illumina against all sequence reads
  o Searched ‘predicated’ oligos against all sequence reads
• Selected 212 DArT SNPs and 1324 cDNA SNPs (1:6).
• Scored 1348 good SNPs, 201 (95%) DArT SNPs and 1147 (87%) cDNA SNPs, 88% conversion rate (2X improvement).
Development of A Working Oat OPA

- Genotyping performance, cluster separation based on normalized theta value >0.3.
- Minor allele frequency >=2%.
- SNPs with less than 10% missing data among 109 germplasm, removing SNPs detecting null.
- Chromosome distribution.
Selection of Working Oat OPA SNPs

<table>
<thead>
<tr>
<th></th>
<th>POOPA1</th>
<th>POOPA2</th>
<th>POOPA4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA</td>
<td>446</td>
<td>311</td>
<td>812</td>
<td>1569</td>
</tr>
<tr>
<td>DArT</td>
<td>31</td>
<td>94</td>
<td>169</td>
<td>294</td>
</tr>
<tr>
<td>Tetraploid</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>477</td>
<td>410</td>
<td>981</td>
<td>1868</td>
</tr>
</tbody>
</table>
Minor Allele Frequency Distribution
Working SNP Candidates

Proportion of SNPs

<table>
<thead>
<tr>
<th>Allele Frequency</th>
<th>POOPA1</th>
<th>POOPA2</th>
<th>POOPA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.02-0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>&gt;0.05-0.10</td>
<td>0.10</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>&gt;0.10-0.20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>&gt;0.20-0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>&gt;0.30-0.40</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>&gt;0.40-0.50</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
</tbody>
</table>
SNP-based Consensus Maps of Cultivated Oat

21 linkage groups anchored to chromosomes
1000 SNPs mapped
Acknowledgements

Rebekah Oliver, USDA-ARS, Aberdeen, ID
Gerry Lazo, USDA-ARS, Albany, CA
Nick Tinker, AAFC, Ontario, Canada
Rick Jellen, BYU, Provo, UT
Jeff Maughan, BYU, Provo, UT
Rachel Redman, ex-BYU, Provo, UT
Carol Lange, TAMU, College Station, TX
Dawn Feltus, USDA-ARS, Fargo, ND
Mary Osenga, USDA-ARS, Fargo, ND

Funding Sources: General Mills, USDA NIFA, USDA-ARS CRIS, NAMA

Special thanks to Eric Jackson for making this work possible.