Identify and use of candidate genes and other molecular markers linked to QTL controlling milling quality and resistance to sheath blight disease in rice

Henry Nguyen, University of Missouri-Columbia
Sally Leong, USDA/ARS, Wisconsin

**General Objectives**

- Genome-wide discovery of new genetic markers
- Provide marker information and genotyping assays to other mapping groups

**Specific Contributions**

Identification and use of candidate genes and other molecular markers linked to QTL controlling milling quality (Nguyen lab) and resistance to sheath blight disease (Leong lab)
Strategic Work Plan Year 1
(Nguyen Lab)

Identify SFP markers for milling quality using rice gene-chip

• Genotypes: Cypress and RT0034 (MY1 population)
• Array hybridization using rice gene-chip
• 2 genotypes x 5 replications = 10 slides
• Data analysis using SAM and others
• Marker validation

Strategic Work Plan Year 1
(Leong Lab)

Identify SFP markers for sheath blight disease resistance using rice gene-chip

• Genotypes: Rosemont and Pecos (SB1 population)
• Array hybridization using rice gene-chip
• 2 genotypes x 5 replications = 10 slides
• Data analysis using SAM and others
• Marker validation
Development of New Markers

(1) Focus on SFPs in QTL regions where markers are needed.

(2) Target SFPs associated with candidate genes based on transcript profilings (e.g. SAGE or microarray analysis), leading to the development of SNP markers.

Rationale

• The SNP mapping GeneChip is unavailable for rice.
• High-density rice Affymetrix GeneChip is an alternative for genome-wide discovery of new markers.
Single Feature Polymorphisms (SFPs)

- A SFP is defined if a feature has a significant hybridization signal change between two hybridized genotypes.
- Expression arrays have been used to detect and score allelic variation in yeast (genome size: 12MB) and Arabidopsis (genome size: 120MB) via direct hybridization of labeled total genomic DNA.
- The discovered SFPs have been successfully used for mapping and cloning a QTL related to high temperature growth in yeast and a mutation in Arabidopsis.
Applications and Prediction

- Arabidopsis: Large Scale Identification of SFP in Complex Genomes - 3806 SFPs (~4%) at a 5% false discovery rate were identified between Col and Ler genotypes.
- Yeast: Genetic Diversity - 11,115 SFPs (3.9%) existing in one or more of 14 different yeast strains.

Affymetrix Rice GeneChip

| _class   | _organism                      | probe_sets | unique_probe_set | gene_probe_set |
|----------+--------------------------------+------------+------------------+----------------|
| main     | O. sativa                      | 1679       | 1140             | 70             |
| main     | O. sativa (japonica cultivar-group) | 54168   | 43673            | 1282           |
| rRNA     | O. sativa                      | 3          | 1                | 0              |
| main     | O. sativa (indica cultivar-group) | 1347     | 1005             | 55             |
| control  | O. sativa (japonica cultivar-group) | 10       | 7                | 0              |
| reporter |                                 | 65         | 43               | 0              |
| control  | M. grisea                      | 10         | 5                | 0              |
|          | (7 rows)                       |            |                  |                |
Advantages

1. Large-scale identification of SFPs or deletions in complex genomes.
2. Reliability is high, which relies on independent replicates.
3. Eliminate the need to amplify specific loci by PCR.
4. The physical location of each SFP is known, compared to anonymous markers.
5. Cost-effective for the discovery of new markers ($0.20 per unvalidated SFP; ~$50 per validated marker)
6. Genotyping is very attractive for QTL mapping if the cost of GeneChips and hybridization decreases.

Limitations

1. The SFP markers can not tell us the exact nucleotide variation present within a gene.
2. The cost of GeneChips and hybridization prohibits genotyping the whole mapping population, but may be feasible using Bulk Segregant Analysis (BSA).
3. For complex genomes, the number of replicates has to be increased to reduce false positives.
4. The nature of target used may limit the detection of SFPs.
Data Analysis

• “Fold change” of intensity provides no statistical measurements.

• Conventional t tests can provide the probability that a difference in hybridization intensity occurred by chance. But here we are testing 1,008,018 features, a significant level of alpha=0.01 would identify 10,008 SFPs by chance.

• SAM (Significance Analysis of Microarray) algorithm overcomes the above drawbacks, and is well used for microarray data analysis to find differentially expressed genes.

Principles of SAM

• Assign each feature a score based on its t-statistics.

\[
d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

• Features with scores greater than a threshold are deemed potentially significant.

• A fudge factor S\(0\) is added to pooled variance so that d score is independent of expression level.

• Permutation testing is used to estimate the percentage of genes identified.
**SAM Results with Normalized Dataset**

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Observed</th>
<th>Permut</th>
<th>Difference</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>26603</td>
<td>5085.15</td>
<td>21517.85</td>
<td>0.191</td>
</tr>
<tr>
<td>0.15</td>
<td>15739</td>
<td>1797.9</td>
<td>13941.1</td>
<td>0.114</td>
</tr>
<tr>
<td>0.25</td>
<td>5950</td>
<td>477.3</td>
<td>5472.7</td>
<td>0.080</td>
</tr>
<tr>
<td>0.3</td>
<td>3709</td>
<td>260.05</td>
<td>3448.95</td>
<td>0.070</td>
</tr>
<tr>
<td>0.35</td>
<td>2005</td>
<td>129.85</td>
<td>1875.15</td>
<td>0.065</td>
</tr>
<tr>
<td>0.4</td>
<td>1152</td>
<td>68.55</td>
<td>1083.45</td>
<td>0.0595</td>
</tr>
<tr>
<td>0.5</td>
<td>436</td>
<td>23.85</td>
<td>412.15</td>
<td>0.0547</td>
</tr>
</tbody>
</table>

**Application of SFPs**

- Option 1: Direct mapping of unique SFPs by searching the gene position in pseudo-chromosomes, or BAC clones on the rice genetic/physical maps.
- Option 2: Design primers, direct sequencing of PCR products from the regions containing SFPs, then convert the polymorphisms into SNP or other PCR-based markers.
Validation of New SNP Markers

- Design SNP primers or probes corresponding to the SNP detection platform.
- Test the parents and some progeny of mapping populations.
- Provide primers (or probes) and genotyping assay information to QTL mapping groups.

Backup Strategies for Genetic Marker Discovery

- SNP markers through genome scanning (Feltus et al. 2004)
- Diversity Arrays Technology (DArT) (Wenzl et al. 2004)
- EcoTilling (Comai et al. 2004)
- cDNA-AFLP
Diversity Arrays Technology (DArT)

- Detect and type DNA variation at several hundred genomic loci in parallel without relying on sequence information.
- Genotype call rate of 98% and a scoring reproducibility of 99.8%
- It can be effectively applied to genetic mapping and diversity analyses for a complex genome.

Wenzl et al., 2004, PNAS, 101:9915

Array containing individual fragments of a library that was built from a representation prepared from a mixture of genomic DNA sampled of a group of genotypes.
EcoTilling

Comai et al., 2004, Plant J. 37:778

RiceCAP

Flowchart of cDNA-AFLP approach

RiceCAP