Linkage, QTL, Mapping Populations and Marker-Assisted-Selection

Dave Douches, Allen Van Deynze and Dave Francis

PAA SolCAP Workshop
August 2009
Detecting QTLs using Marker-based analysis (MBA) requires genotyping of individuals.

**Steps toward QTL analysis using MBA:**

1- Segregating populations

2- Measure phenotypic variation
   - The more accurate measurement the more precise QTL identification.

3- Measure genotypic variation
   - Genotype individuals
   - Genetic map construction

4- Statistical analysis to find QTLs and QTL validation
Common Population Structures

- F2;
- F3;
- F4;
- Recombinant Inbred (RI);
- F1-derived, intermated recombinant inbred (IRI);
- Doubled haploids (DH);
- Backcross (BC1);
- Association study in the base germplasm;
- Near Isogenic Lines (NIL).
QTL mapping in an F2

…is an association study in a population where the confounding effect of pedigree has been removed.
Mapping Populations in Potato

- 4x crosses (between heterozygous parents)
  - Mixed segregation patterns
  - Simplex and duplex segregation patterns

- 2x crosses
  - Testcross, F2, tri-allelic ratios
  - Doubled monoploid cross
    - Segregation in one parent - testcross
Mapped disease resistance genes and QTL
Potato: Ultra-high density (UHD) genetic map

- 70,344 unigenes
- 150 SSRs (1000 more in ESTs)
- 10,000+ AFLP markers (from 381 primer combinations)
- ordered on 136 progeny of diploid RH x SH cross
- most dense genetic map in any organism

https://gabi.rzpd.de/projects/Pomamo/
Walter De Jong, Cornell University
The level of resolution needed depends on intended application

- Combining QTLs between lines within a segregating population.
- Moving chromosome segments within heterotic pools.
- Make germplasm wide inferences/claims about a particular chromosome segment (assuming IBD).
- Moving chromosome segments across heterotic pools of elite germplasm.
- Introgression of special trait (e.g. disease resistance) from exotic germplasm.
- QTL cloning.
Type of population depends on resolution needed

Comparison of resolution and research time for various approaches to dissect quantitative variation. The research times assume the target species has only two generations per year. NIL, near-isogenic line; RIL, recombinant inbred line

Buckler and Thornsberry (2002)
Linkage

- Association of two or more loci on a chromosome with limited recombination

Linkage Disequilibrium or Gametic Phase Disequilibrium

- Non random association of alleles at two or more loci not necessarily on the same chromosome
- Measures co-segregation of alleles in a population
- Mendel’s pea traits – showed complete linkage equilibrium and hence independent assortment
- Can arise from intermixture of populations with different gene frequencies
- Can also be produced or maintained by selection favoring one combination of alleles over the other – e.g. selection for yield in a breeding population

Falconer and McKay (1996)
Linkage disequilibrium
Linkage Disequilibrium Decay

$r = \text{Recombination Rate between two loci}$

$r = 0.5 = \text{Two loci are unlinked}$

$r = 0 = \text{Two loci are completely linked and do not independently assort}$

Falconer and Mackay (1996)
What are QTL?

- Genes are regions of DNA controlling a trait or phenotype

- Quantitative Trait Loci (QTL) are regions of DNA associated with quantitative traits
  - They can be defined using molecular markers that highlight specific DNA segments or genes

This can change how we manipulate these traits in breeding
A trait is quantitative/polygenic if its inheritance is controlled by two or more genes and their interaction with environment.

1- A trait having a continuous distribution

2- Controlled by the interaction of many genes

Cont…
A trait is quantitative/polygenic if its inheritance is controlled by two or more genes and the their interaction with environment.

3- Influenced by the environment

Examples

- Yield
- Lycopene content in tomato
- Early Blight (EB) resistance
Trait Distribution in the Progeny

If we could observe directly the QTL we could see the 3 underlying trait distributions.
Modeling a Marker Locus and a Linked QTL

Single Marker Analysis modeling, assume:
- One QTL locus A,
- One Marker locus M,

Measure of association between A and M
- Self pollinated crops: compare AA to aa (no dominance effect)
- Cross Pollinated crop: compare AA_T to aA_T (dominance and additive effects are confounded)
### Assigned values of the genotypes at A and M

<table>
<thead>
<tr>
<th>Genotyping</th>
<th>Phenotyping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Genotyped</td>
<td>Cloned Progeny</td>
</tr>
<tr>
<td>AA</td>
<td>AA =&gt; a</td>
</tr>
<tr>
<td>Aa</td>
<td>Aa =&gt; d</td>
</tr>
<tr>
<td>aa</td>
<td>aa =&gt; -a</td>
</tr>
</tbody>
</table>

Let $\delta = d$, $d/2$, and 0 for cloned, selfed, and tescrossed progenies, respectively.
Need to collect good data: Good phenotypes are important.

- **Mature Green**
- **Breaker**
- **Pink**
- **Light Red**
- **Red Ripe**

**Ripening**

0 dpi

1 dpi

2 dpi

3 dpi

**Infection**
What information do we need?

- Information pertinent to the association between marker genotypes and phenotypic values.
  - Total number of individuals genotyped, $N$ (and number in each class, $n_{MM}$, $n_{Mm}$, and $n_{mm}$)
  - Values of $a$ and $\delta$.
  - Values of residual genetic variance and error variance.
Detection of a QTL

- Detection of linkage between A and M, will occur only if both are segregating in the population and if they are physically linked;
- We cannot extrapolate results from one mapping population to the rest of our germplasm without considering disequilibrium.
Do we have disequilibrium between a linked marker and a QTL?

- We cannot really assess very well the extent of gametic phase disequilibrium between loci in the population without very extensive mapping studies.

- With SNPs we can look at marker-marker associations for a much smaller cost: we just need to genotype our germplasm.
The marker density needed depends on the population.
Potato SNP map

- Potentially 7600 markers
  - Candidate genes and randomly distributed
- High SNP polymorphism rate
- Potentially over 600 markers per chromosome
- We need over 3000 random markers to saturate the 4x genome
  - Bradshaw et al 2008
- We need less than 1000 markers to saturate the 2x genome
Sample (population) size

- $\sigma^2 = \text{error } \sigma^2 + \text{residual genetic } \sigma^2$
- “error $\sigma^2$” and “residual genetic $\sigma^2$” are based on overall mean for each entry.
- Increasing the number of field replications will reduce error $\sigma^2$ but not the between-line genetic $\sigma^2$.
- To further reduce the denominator in the t-test. We need to add more genetic entries.

- Typical split plot:
  - Sub-plot error: *plot-level error $\sigma^2$*
  - Plot error: *genetic $\sigma^2 + (\text{plot-level error } \sigma^2)/(\text{rep number})$*
Figure 1 Comparisons of the statistical powers obtained from Monte Carlo simulations (open circles and triangles) with those predicted from the theory (dotted and solid lines). The plots with open circles and solid line (upper plots) represent the situation of $h^2=0.10$, while the plots with open triangles and dotted line (lower plots) represent the situation of $h^2=0.05$. 
**Figure 2** Changes of statistical power as QTL position changes from one end to the other end of a marker interval of 40 cM in length. The sample size is $n = 100$ and the size of the QTL is $h^2 = 0.05$. The lowest power occurs when the QTL is in the middle of the marker interval.
Most popular QTL analysis methods for MBA

- Single-Marker Analysis (SMA)
- Interval Mapping (IM)
  - Simple Interval Mapping (SIM)
    - MapMaker/QTL
    - QGENE
    - R/QTL
    - MapQTL5
  - Composite Interval Mapping (CIM)
    - QTL Cartographer
    - R/QTL
    - MapQTL5
    - QTL Network 2.0
- Multiple Interval Mapping (MIM)
  - WinQTL cartographer
SMA is being used less nowadays due to its disadvantages and availability of more advanced statistical methods.

**Advantage:**
- Simplest approach for detecting QTLs
- Quick and less memory intensive

**Disadvantages:**
- May not detect loosely linked QTLs
- QTL effects are often underestimated
- Does not determine the distance between marker and QTL
- Does not determine the direction of QTL from marker
SIM was one the most commonly used methodology in QTL mapping

**Advantage:**

The likelihood map represents the position of the QTL at various points of the genome

The probable position of the QTL is given by support intervals

Requires less progenies than the SMA

**Disadvantages:**

The number of QTL cannot be resolved

The locations of the QTL are sometimes not well resolved, the exact positions of the QTL cannot be determined

The statistical power is still relatively low
CIM has at least four advantages over SMA and SIM

**Advantages:**
- By focusing on one genome region, the multidimensional search is reduced to one-dimensional search
- The resolution of QTL locations obtained is much higher than SMA and IM
- There are more variables in the model, making the model more efficient
- Markers can be used as boundary conditions to narrow down the most likely QTL position
Comparison of SMA, SIM and CIM for EB resistance in tomato
Trait-based analysis (Selective Genotyping)

*Solanum lycopersicum* X *S. habrochaites*

(EB Susceptible) (EB Resistant)

F₁ X *S. lycopersicum*

BC₁

~850 BC₁ Plants

EB Resistant

EB Susceptible
QTL mapping step by step: A practical example
A raw file is used for the genetic map construction

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data type ri self</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>172 294 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4LC34L10a</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>-</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>*TG301</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>-</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>*TOS16116</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>-</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>*LC34L10b</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>*LEC71F17</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>*LEC72D1</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>10</td>
<td>*TG125</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>11</td>
<td>*CLE1616</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>12</td>
<td>*TOF3A14</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>13</td>
<td>*TG70</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>14</td>
<td>*LC66G13b</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>15</td>
<td>*TOF20P4</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>16</td>
<td>*LEG9N2</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>17</td>
<td>*LEC802</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>18</td>
<td>*CT132PSU</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>19</td>
<td>*TG273</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>20</td>
<td>*LES9N20</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>21</td>
<td>*TG59</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>H</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>22</td>
<td>*CT191</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>H</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>23</td>
<td>*TOE6F10</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>H</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>24</td>
<td>*LG61B21a</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>25</td>
<td>*LEG3714</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>26</td>
<td>*TG559a</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>
QTL mapping step by step: A practical example
A genetic map is constructed
QTL mapping step by step: A practical example

Phenotypic data is collected and incorporated into the file

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>#</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>*LC34L10a</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>-</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>*TG301</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>H</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>*TOS16l16</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>-</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>*LC34L10b</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>*LC34L1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>352</td>
<td>Frt-We04</td>
<td>8.12</td>
<td>31.75</td>
<td>29.08</td>
<td>8.84</td>
<td>17.93</td>
<td>34.48</td>
<td>8.37</td>
<td>13.04</td>
<td>18.38</td>
<td>9.25</td>
<td>5.91</td>
<td>27.54</td>
<td>10.40</td>
<td>16.47</td>
</tr>
<tr>
<td>354</td>
<td>Frt-SS04</td>
<td>7.1</td>
<td>6.35</td>
<td>6.2</td>
<td>8</td>
<td>8.2</td>
<td>6.65</td>
<td>6.95</td>
<td>6.43</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>7.7</td>
<td>6.3</td>
<td>6.55</td>
</tr>
<tr>
<td>356</td>
<td>Frt-We05R2</td>
<td>8.03</td>
<td>14.35</td>
<td>22.96</td>
<td>11.55</td>
<td>8.85</td>
<td>26.47</td>
<td>8.31</td>
<td>29.46</td>
<td>26.2</td>
<td>10.93</td>
<td>25.46</td>
<td>43.14</td>
<td>15.06</td>
<td>20.53</td>
</tr>
<tr>
<td>357</td>
<td>Frt-We05AveR1R2</td>
<td>7.51</td>
<td>14.63</td>
<td>22.13</td>
<td>12.39</td>
<td>7.68</td>
<td>33.75</td>
<td>8.07</td>
<td>31.05</td>
<td>27.51</td>
<td>10.5</td>
<td>25.34</td>
<td>36.96</td>
<td>14.9</td>
<td>18.73</td>
</tr>
<tr>
<td>358</td>
<td>Frt-pH05R1</td>
<td>3.59</td>
<td>4.16</td>
<td>3.9</td>
<td>3.97</td>
<td>3.97</td>
<td>4.34</td>
<td>4.14</td>
<td>4.38</td>
<td>4.28</td>
<td>3.98</td>
<td>4.1</td>
<td>5.17</td>
<td>4.9</td>
<td>4.24</td>
</tr>
<tr>
<td>359</td>
<td>Frt-pH05R2</td>
<td>3.99</td>
<td>3.97</td>
<td>3.98</td>
<td>3.78</td>
<td>3.8</td>
<td>4.54</td>
<td>4.27</td>
<td>4.5</td>
<td>4.09</td>
<td>3.8</td>
<td>3.88</td>
<td>4.28</td>
<td>4.03</td>
<td>4.2</td>
</tr>
<tr>
<td>360</td>
<td>Frt-pH05AveR1R2</td>
<td>3.79</td>
<td>4.07</td>
<td>3.94</td>
<td>3.88</td>
<td>3.89</td>
<td>4.44</td>
<td>4.21</td>
<td>4.44</td>
<td>4.19</td>
<td>3.89</td>
<td>3.99</td>
<td>4.73</td>
<td>4.47</td>
<td>4.22</td>
</tr>
<tr>
<td>361</td>
<td>Frt-SS05R1</td>
<td>7.9</td>
<td>7.9</td>
<td>7.6</td>
<td>6.1</td>
<td>7.1</td>
<td>6.4</td>
<td>8</td>
<td>6.7</td>
<td>7</td>
<td>8.4</td>
<td>8.1</td>
<td>7</td>
<td>7.4</td>
<td>7.1</td>
</tr>
<tr>
<td>362</td>
<td>Frt-SS05R2</td>
<td>8.2</td>
<td>7.8</td>
<td>7.2</td>
<td>6.5</td>
<td>7.2</td>
<td>6.3</td>
<td>7.3</td>
<td>6.8</td>
<td>6.7</td>
<td>7.8</td>
<td>8.4</td>
<td>6.3</td>
<td>6.8</td>
<td>7.2</td>
</tr>
<tr>
<td>363</td>
<td>Frt-SS05AveR1R2</td>
<td>8.05</td>
<td>7.85</td>
<td>7.4</td>
<td>6.3</td>
<td>7.15</td>
<td>6.35</td>
<td>7.65</td>
<td>6.75</td>
<td>6.85</td>
<td>8.1</td>
<td>8.25</td>
<td>6.65</td>
<td>7.1</td>
<td>7.15</td>
</tr>
</tbody>
</table>
QTL mapping step by step: A practical example
A map file is then created for QTL analysis

*Chromosomes: 12
*chrom1 31 -1 0 0 -505.285
  0 1 2 3 4 5 6 7 8 9
  10 11 12 13 14 15 16 17 18 19
  20 21 22 23 24 25 26 27 28 29 30
  0.0004 0.1367 0.0614 0.0580 0.0103 0.0004 0.0004 0.0424 0.0818 0.0184
  0.1190 0.0141 0.0069 0.0070 0.0544 0.0553 0.0146 0.0885 0.0639 0.0032
  0.0004 0.0004 0.0818 0.0421 0.0439 0.1213 0.0244 0.0270 0.0555 0.0110
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
*chrom2 32 -1 0 0 -394.876
  31 32 33 34 35 36 37 38 39 40
  41 42 43 44 45 46 47 48 49 50
  51 52 53 54 55 56 57 58 59 60
  61 62
  0.0004 0.0004 0.0004 0.0004 0.0031 0.0004 0.0004 0.0004 0.0004 0.0032
  0.0064 0.0732 0.0623 0.0033 0.1090 0.0674 0.0004 0.0727 0.0310 0.0788
  0.0469 0.0004 0.0004 0.0385 0.0128 0.0274 0.0556 0.0278 0.1470 0.0438
  0.0004
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
*chrom3 22 -1 0 0 -418.717
  63 64 65 66 67 68 69 70 71 72
  73 74 75 76 77 78 79 280 80 81
  82 281
Key points of QTL mapping

- QTL’s mapped in a bi-parental cross may not be appropriate for MAS in all populations
  - Marker allele and trait may not be linked in all populations.
  - Genetic background effects may be population specific.
  - Original association may be spurious.
  - QTL detection is dependent on magnitude of the difference between alleles and the variance within marker classes.

- Adding more genotypes is more efficient than replicating the same genotypes (provided that a minimum number of environments are sampled)

- $h^2$ and size of effect important

- Results are trait specific
SolCAP Mapping Populations

- **Germplasm panel**
  - Over 300 genotypes
  - Over 20 programs represented
    - Important advanced breeding lines
    - Current and Historical varieties
    - Parents of mapping populations
    - Important genetic stocks
  - Phenotyping at NY, WI and OR of core traits
    - Evaluation of specific gravity, glucose and sucrose, chip color, skin type, shape, vine maturity, tuber number, vitamin C, internal defects, etc.
    - Accessible through the SolCAP and Solanaceae Genome Network (SGN) websites.
- Association mapping
- Develop linkage hypotheses
- Phenotyping of other traits
SolCAP Mapping Populations

- **4x population**
  - Rio Grande X Premier Russet
  - 200 progeny
  - 7600 SNPs
  - 3 environments
    - North Carolina, Minnesota, Idaho
  - Identify QTL for core traits (CHO, Vit C)

- **Other populations from the community**
  - Core trait validation
  - Other traits of regional importance
SolCAP populations

- **2x population**
  - DM1-3-516 44 x 84SD22
  - 120 progeny
  - Testcross segregation
  - Map SNPs
  - Anchor scaffolds to the genetic map
    - Link mapping to PGSC
  - Map core traits

- **Other populations from the community**
Approach of MAS in Tomato

• Pathway approach for candidate gene identification and introduction to metabolic pathway databases.

• Identification of polymorphisms in data-based sequences

• Approach of potato breeding for the future
Example: QTL for color uniformity in elite crosses

<table>
<thead>
<tr>
<th>Chr 1</th>
<th>Chr 2</th>
<th>Chr 3</th>
<th>Chr 4</th>
<th>Chr 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist cM</td>
<td>Marker Name</td>
<td>Dist cM</td>
<td>Marker Name</td>
<td>Dist cM</td>
</tr>
<tr>
<td>17.1</td>
<td>CT233</td>
<td>13.9</td>
<td>TG67</td>
<td>13.4</td>
</tr>
<tr>
<td>2.1</td>
<td>LEOH36</td>
<td>10.7</td>
<td>CT205</td>
<td>12.4</td>
</tr>
<tr>
<td>20.0</td>
<td>LEOH36</td>
<td>10.9</td>
<td>TG165</td>
<td>18.5</td>
</tr>
<tr>
<td>2.0</td>
<td>LEOH36</td>
<td>10.0</td>
<td>LEOH23</td>
<td>18.5</td>
</tr>
<tr>
<td>24.7</td>
<td>TG125</td>
<td>18.4</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>7.2</td>
<td>CT62</td>
<td>18.2</td>
<td>CT244</td>
<td>18.5</td>
</tr>
<tr>
<td>12.8</td>
<td>TG273</td>
<td>6.1</td>
<td>TG608</td>
<td>18.5</td>
</tr>
<tr>
<td>9.3</td>
<td>TG59</td>
<td>5.2</td>
<td>CT205</td>
<td>18.5</td>
</tr>
<tr>
<td>9.9</td>
<td>CT191</td>
<td>10.0</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>10.7</td>
<td>TG14</td>
<td>6.1</td>
<td>CT205</td>
<td>18.5</td>
</tr>
<tr>
<td>5.8</td>
<td>TG465</td>
<td>10.1</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>1.1</td>
<td>TG260</td>
<td>10.1</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>15.1</td>
<td>LEOH7</td>
<td>10.1</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>10.1</td>
<td>TG255</td>
<td>18.2</td>
<td>TG14</td>
<td>18.5</td>
</tr>
<tr>
<td>18.2</td>
<td>TG580</td>
<td>18.2</td>
<td>TG14</td>
<td>18.5</td>
</tr>
</tbody>
</table>

QTL | Trait | Origin |
---|------|-------|
2 | L, YSD | S. lyc. |
4 | YSD | S. lyc. |
6 | L, Hue | og<sup>c</sup> |
7 | L, Hue | S. hab. |
11 | L, Hue | S. lyc. |
Tomato 'zeaxanthin epoxidase' 

<table>
<thead>
<tr>
<th>Locus details</th>
<th>Download GMOD XML</th>
<th>Note to Editors</th>
<th>Annotation guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[New] [Edit] [Delete]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus name</td>
<td>zeaxanthin epoxidase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbol</td>
<td>zep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene activity</td>
<td>zeaxanthin epoxidase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Enzyme that catalyzes the conversion of zeaxanthin to violaxanthin. Mutants are wilty due to abscisic acid deficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arm</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Locus synonyms 3: hp3 LeZEP ZE [Add/Remove]

Tomato-EXPEN 2000

Locus editors: Andrew Thompson; Navot Galnaaz
Comments on SGN databases:

• There is a wealth of information organized and available.
• Information is not always complete/up to date.
• Display is not always optimal, and several steps may be needed to go from pathway > gene > potential marker.
• Sequence data has error associated with it. eSNPs are not the same as validated markers.
• SolCAP will be asking for feedback on how best to improve the SGN database and access via the Breeders Portal.
Developing “Breeder Friendly” Tools

- Current SGN interfaces are aimed at the molecular biologist, with searches designed to facilitate molecular discovery
  - Need a portal that is trait and germplasm centered
- Ability to search by traits relevant to (and defined by) breeders
- When a marker is identified, a protocol for use in breeding will be provided.
- Search option that only yields polymorphic marker, phenotype, or QTL results from elite germplasm
- Ability to search for known parents or offspring of any given genotype
- Ability to generate a list of markers that are polymorphic between any two parents
- Detailed tutorial/definitions of terms and traits utilized within the database
Ovate
Databases

• Missing data in SGN
  • Limited ability to generate tables, PCR conditions sometimes incomplete, Enzyme sometimes missing, SNP not described.

• Missing data in Tomatomap.net
  • SNP and sequence context requires BMC genomics supplemental table, ASPE primers, GoldenGate primers.
    • 2007. BMC Genomics 8:465
      www.biomedcentral.com/content/pdf/1471-2164-8-465.pdf
Genotyping the core collections will impact breeding.

• Potential translational approaches:
  1) introgression from other populations (domesticated or wild)
  2) selection for coupling phase recombinants to establish linkage blocks of favorable alleles (e.g. disease resistance loci)
  3) population development designed to maximize variation w/in market classes
  4) association approaches
  5) whole genome approaches

• Other translational strategies will emerge under other CAPs or through innovation in public research.
Additional Thoughts from Tomato:

• Marker resources are currently not sufficient for QTL discovery in bi-parental or Association Mapping populations; in the future they should be.

• The best time to use genetic markers: early generation selection.

• Restructuring of breeding program to integrate markers may include:
  • 1) Increasing genotypic replication (population size) at the expense of replication (consider augmented designs).
  • 2) Collecting objective data.
Marker Assisted Selection in Potato

- Limited number of markers available
  - PVY, RB, GN, VW useful
  - Invertase, sucrose synthase and LB QTL not validated in our germplasm
  - More markers and QTL to validate from Li, et al. 2008 and Bradshaw et al. 2008

- Some are PCR-based
  - Need for marker conversion
  - Use of high-throughput DNA isolation methods

- Apply at early generation selection phase
  - Seedling stage
  - After year 1 or 2 selection

- Need markers linked to core traits (e.g. glucose and starch)
Small Grants Program

Address regional, individual program and emerging needs within the Solanaceae community through a small grants program.

Six key areas we intend to allocate resources to are:

- Genotyping mapping populations in the core facilities requested by the greater breeding community
- Marker conversion – developing SNP markers linked to QTL into easily assayed (e.g. CAPs or dCAPs) markers that end-users can readily apply in their own research programs
- QTL validation and MAS
- Population development to address emerging needs
- Extension or education special projects
- New directions not envisioned at the time of proposal submission.
References

- Van Deynze et al., 2007. BMC Genomics 8:465
  www.biomedcentral.com/content/pdf/1471-2164-8-465.pdf