Lecture 1.

This course, ‘breeding with molecular markers’, will examine the role of ‘marker assisted selection’ or ‘genome assisted selection’ in improving plant varieties. In the last two years the draft genome sequence of soybean, corn, cucumber, potato, and tomato have been completed and made available. More genomes will follow. For example current activities in China seek to use ‘Next Generation’ sequencing to sequence 100 tomato varieties. The question we will try to address in this class is ‘how can we translate genome sequence data into new plant varieties?’.

The development of more productive plant varieties, combined with improved methods of production offers a solution to meet the caloric and nutritional needs of a growing world population. Increasing the efficiency of agricultural food production while reducing the environmental ‘footprint’ is of enormous importance. Each year humans re-create the food supply that feeds approximately 7 billion people. FAO estimates that by the close of the seasons ending in 2008 world stocks of cereals were 420 million Tons, the lowest since 1983 (www.fao.org/docrep/010/ah877e/ah877e04.htm). Wheat stores have declined 11 percent, to the lowest level since the UN began keeping records in 1980 (FAO), while the US Department of Agriculture (USDA) has reported that world wheat stocks have fallen to 47-year lows. It is estimated that these stocks represent 12 weeks worth of global consumption. These statistics are not adjusted for increases in population, and they serve to remind us that the world food supply can be precarious.

As the human population grows, land and water resources available for food production shrink. At the same time old plant diseases re-emerge as epidemics; new plant diseases emerge with increased adoption of no-till practices and high-tunnel production, and diets change as economies develop. The challenges of re-creating the world’s food supply on an annual basis are therefore profound.

We need to make more progress to meet the needs of a growing population. Technology that increases the ‘efficiency’ of crop improvement must be put into practice. Marker-assisted-selection (MAS) is promoted as a way to increase the efficiency of the breeding process. In this class we will examine different measures of ‘efficiency’. In order to make valid comparisons, we need to review some basic concepts from classical breeding and link those to marker-assisted breeding.

**Gain under selection as a measure of progress.**

Progress in plant breeding can be measured objectively. There is a theoretical basis for measuring this progress as ‘gain under selection’ or ‘response to selection’ as well as practical consequences in evaluating the effectiveness of breeding programs. In addition, measures of selection, whether theoretical or empirical provide a framework to assess the efficacy of different selection strategies. Therefore an understanding of some concepts from conventional plant breeding provide an important foundation. In order to effectively use these concepts to monitor progress, a breeding program must: (1) Clearly articulated goals for key traits; (2) collect and archive objective data; (3) monitor progress over time.
Gain under selection

\[ \Delta G = h^2(\sigma_p)K \]  
(Eq. 1)

Where \( h^2 \) is the ‘narrow sense heritability’ (note this symbol reads as ‘h superscript 2’ rather than h-squared and is a measure of the additive genetic variation relative to total phenotypic variation; in contrast the symbol \( H \) denotes broad sense heritability, a measure of total genetic variation relative to total phenotypic variation; \( K \) is the selection differential (Mean of unselected population – Mean of selected population in standard deviation units); \( \sigma_p \) is the phenotypic variation of the population

Narrow Sense Heritability \( \left( h^2 = \frac{V_{aG}}{V_p} \text{ or } \frac{\sigma_{aG}^2}{(\sigma_p)^2} \right) \)  
Eq (2)

Broad Sense Heritability \( \left( H = \frac{V_G}{V_p} \text{ or } \frac{\sigma_G^2}{(\sigma_p)^2} \right) \)  
Eq (3)

**Figure 1.** Distribution of Horsfall-Barrett Scores for disease severity in a segregating population rated for bacterial spot disease.

<table>
<thead>
<tr>
<th>H-B Scale</th>
<th>Unselected</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>7.41</td>
<td>3.71</td>
</tr>
<tr>
<td>STDEV</td>
<td>1.86</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Figure 1 presents the distribution of disease ratings for severity of bacterial spot infection in a tomato population. Ratings are expressed using the Horsfall-Barrett scale, where low values indicate lower infection. Black bars represent the distribution of the total population and white bars represent the distribution of the ‘resistant’ selections. The mean of the selected population is ~2 standard deviation units lower than the mean of the unselected population.
Table 1. Selection differential for a normally distributed population.

<table>
<thead>
<tr>
<th>K</th>
<th>Percentage of population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>1.16</td>
<td>30%</td>
</tr>
<tr>
<td>1.40</td>
<td>20%</td>
</tr>
<tr>
<td>1.76</td>
<td>10%</td>
</tr>
<tr>
<td>2.06</td>
<td>5%</td>
</tr>
<tr>
<td>2.42</td>
<td>2%</td>
</tr>
<tr>
<td>2.64</td>
<td>1%</td>
</tr>
</tbody>
</table>

The relationship between heritability and the proportion of variance explained by a Marker.

Heritability is an key concept in plant breeding. It is a quantitative description of the proportion of variance explained by genetics. A trait or phenotype can be “(1) determined by genotype, or (2) transmitted from parent to offspring” (Nyquist, 1991). Therefore, heritability is a measure of the relative importance of heredity to the expression of a character. Heritability in the broad sense is the ratio of genotypic variance to phenotypic variance (Eq. 3, above); while heritability in the narrow sense is the ratio of additive genetic variance to phenotypic variance (Eq. 2, above). Heritability is the regression of breeding value on phenotypic value.

During the process of ‘Mendelizing a QTL’, we seek to use linkage of a qualitative character (the marker) to explain the genetics of a quantitative character (the trait). In genetic mapping and marker-assisted selection applications it is common to express the linkage or association of a DNA-based genetic marker and a trait or phenotype as the proportion of variance explained by the marker. This description is the same concept as heritability, and the methods used to estimate the variance explained by a marker are the same.

Estimating Heritability (parent-offspring regression)

Parent-offspring regression can be used to estimate the genetic components of variance by using covariances among relatives. The approach requires a mating design. For inbred crops, this usually involves regressing the trait values of F2 individuals with trait values for an F3 family derived from those F2 individuals (F2:3).

A brief review of regression analysis (for a straight line) follows:

\[ Y' = mX + b \]

where Y and X are variables and Y’ is an estimate of Y based on a value of X.
The regression coefficient $m$ is an expression of how much (on average) $Y$ may be expected to change per unit change of $X$.

$$m = \frac{\text{Sum}(X_i - X_{\text{ave}})(Y_i - Y_{\text{ave}})}{\text{Sum}(X_i - X_{\text{ave}})^2} = \text{change in } X = \text{the slope of the line} = \text{change in } Y$$

The statistical correlation coefficient “$r$” measures how closely two sets of data are associated. It is without units and as the limits of $-1.0$ to $+1.0$. The regression coefficient ($m$) and the correlation coefficient, $r$, always have the same sign. The correlation coefficient of $Y$ on $X$ is defined as the linear change of $Y$ in standard deviations, for each increase of one standard deviation in $X$. “$b$” is the $Y$ intercept.

$\text{Cov}(X,Y) = \frac{\text{Sum}(X_i - X_{\text{ave}})(Y_i - Y_{\text{ave}})}{n - 1}$

$$r = \frac{\text{Cov}(X,Y)}{\text{Sdev}_x \ast \text{Sdev}_y}$$

For PO regression

$$h^2 = r = \frac{\text{Cov}(P,O)}{(\text{Var } P \ast \text{Var } O)^{1/2}}$$

Note that $r$ and $m$ are related by: $m = r(S\text{dev}_x/\text{Sdev}_y)$. When the variances of $X$ and $Y$ are equal, $m = r$ and either can be used to estimate heritability. If the variances are unequal, standardized variables (sample has a mean of zero and a standard deviation of 1) can be used to insure that $m = r$.

**Figure 2. Examples of Parent-Offspring regression for Lycopene Content.**
Figure 2. Shows PO regression for a population phenotyped for lycopene content. Values a family of F3 plants grown from seed from a single F2 plant were regressed onto the values from the F2 individual. When data for the entire population is used, $r^2 = 0.6682$, and $r = 0.817$. Lycopene content is a high heritability trait in this population. Most plant breeders would not carry an entire F2 population to the next generation. Rather, the best plants would be selected and carried forward. The bottom graph in Figure 2 illustrates the P-O regression for selected individuals. Note that $r^2 = 0.3392$ and thus the estimate of $r$ decreases to 0.582.

This illustrates an important point: selection reduces genetic variation and thus reduces heritability. This is the classic dilemma of breeding programs.

Reference:
