Lecture 10
Mass, Family, and Line Selection

Bruce Walsh lecture notes
Tucson Winter Institute
7 - 9 Jan 2013
Topics

• Breeder’s equation
  - Outcrossing population ($h^2$)
  - Clones ($H^2$)
  - Truncation selection, selection intensity
  - Permanent vs. transient response

• Family selection
  - Different types of family selection

• Selfing
  - Selection while selfing

• Line selection
  - SSD, Pedigree, Bulk, and DH schemes
  - Early vs. late testing
Selection

• Basic goal is to develop elite genotypes
  - With an outcrossed population, this is done by increasing the frequency of favorable alleles
    • Selection on additive (A), rather than interaction (D), effects
    • In a large population, continual improvement (response) expected over a number of generations
  - With inbred populations, this is done by generating a series of inbreds from a cross and picking the elite lines.
    • Selection on genotypic values, hence additive + interaction effects
    • Further progress depends upon generating new variation through additional crosses
Outcrossed Populations

• Improvement in an outcrossed (or open pollinated) population akin to what animal breeders do for improvement (called recurrent selection by plant breeders)
  - A within-generation change (the increase in trait mean among the selected individuals) is translated into a between-generation change.

• Individuals are chosen either on the basis of
  - their phenotypic value (mass selection)
  - the performance of their offspring (progeny testing) or relatives (such as sib or family selection)
  - Idea is to use such information to obtain estimates of the breeding values of individuals
Response to Selection

• Selection can change the distribution of phenotypes, and we typically measure this by changes in mean
  - This is a within-generation change, namely the selection differential \( S = \mu^* - \mu \)

• Selection can also change the distribution of breeding values
  - This is the response to selection, the change in the trait in the next generation (the between-generation change) \( R(t) = \mu(t+1) - \mu(t) \)
The Breeders’ Equation: Translating S into R

Recall the regression of offspring value on midparent value

\[ y_O = \mu_P + h^2 \left( \frac{P_f + P_m}{2} - \mu_P \right) \]

Averaging over the values of the selected midparents,

\[ E[ (P_f + P_m)/2 ] = \mu^* \]

Likewise, averaging over the regression gives

\[ E[ y_o - \mu ] = h^2 (\mu^* - \mu) = h^2 S \]

Since \( E[ y_o - \mu ] \) is the change in the offspring mean, it represents the response to selection, giving:

\[ R = h^2 S \]

The Breeders’ Equation (Jay Lush)
• Note that no matter how strong $S$, if $h^2$ is small, the response is small

• $S$ is a measure of selection, $R$ the actual response. One can get lots of selection but no response

• If offspring are asexual clones of their parents, the breeders’ equation becomes
  \[ R = H^2 S \]

• If males and females subjected to differing amounts of selection,
  \[ S = \frac{S_f + S_m}{2} \]
Pollen control

• Recall that $S = (S_f + S_m)/2$
• An issue that arises in plant breeding is pollen control --- is the pollen from plants that have also been selected?
• Not the case for traits (i.e., yield) scored after pollination. In this case, $S_m = 0$, so response only half that with pollen control
• Tradeoff: with an additional generation, a number of schemes can give pollen control, and hence twice the response
  - However, takes twice as many generations, so response per generation the same
Selection on clones

• Although we have framed response in an outcrossed population, we can also consider selecting the best individual clones from a large population of different clones (e.g., inbred lines)

• $R = H^2S$, now a function of the board sense heritability. Since $H^2 > h^2$, the single-generation response using clones exceeds that using outcrossed individuals

• However, the genetic variation in the next generation is significantly reduced, reducing response in subsequent generations
  - In contrast, expect an almost continual response for several generations in an outcrossed population.
The Selection Intensity, $i$

The selection intensity $i$ is the selection differential expressed in terms of phenotypic standard deviations

$$i = \frac{S}{\sqrt{V_P}} = \frac{S}{\sigma_p}$$

Consider two traits, one with $S = 10$, the other $S = 5$. Which trait is under stronger selection? Can’t tell, because $S$ is a function of the phenotypic variance of the trait.

In contrast, $i$ is scaled measure and hence allows for fair comparisons over different traits.
Truncation selection

- A common method of artificial selection is through selection --- all individuals whose trait value is above some threshold (T) are chosen.
- Equivalent to only choosing the uppermost fraction $p$ of the population.

![Diagram showing individuals culled and allowed to reproduce](image-url)
Truncation selection

• The fraction $p$ saved can be translated into an expected selection intensity (assuming the trait is normally distributed),
  
  - allows a breeder (by setting $p$ in advance) to choose an expected value of $i$ before selection, and hence set the expected response

  
  $i = \frac{S}{\sigma} = \frac{\varphi(z_{[1-p]})}{p}$

  
  - $R$ code for $i$: `dnorm(qnorm(1-p))/p`

<table>
<thead>
<tr>
<th>$p$</th>
<th>0.5</th>
<th>0.2</th>
<th>0.1</th>
<th>0.05</th>
<th>0.01</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>0.798</td>
<td>1.400</td>
<td>1.755</td>
<td>2.063</td>
<td>2.665</td>
<td>2.892</td>
</tr>
</tbody>
</table>
Selection Intensity Versions of the Breeders’ Equation

\[ R = h^2 S = h^2 \frac{S}{\sigma_p} \sigma_p = i h^2 \sigma_p \]

Since \( h^2 \sigma_p = (\sigma^2_A / \sigma^2_p) \sigma_p = \sigma_A (\sigma_A / \sigma_p) = h \sigma_A \)

\[ R = i h \sigma_A \]

Since \( h = \) correlation between phenotypic and breeding values, \( h = r_{PA} \)

\[ R = i r_{PA} \sigma_A \]

Response = Intensity * Accuracy * spread in Va

When we select an individual solely on their phenotype, the accuracy (correlation) between BV and phenotype is \( h \)
Accuracy of selection

More generally, we can express the breeders equation as

\[ R = i r_{uA} \sigma_A \]

Here we select individuals based on the index \( u \) (for example, the mean of \( n \) of their sibs).

\( r_{uA} \) = the accuracy of using the measure \( u \) to predict an individual's breeding value = correlation between \( u \) and an individual's BV, \( A \)
Example 10.4. **Progeny testing**, using the mean of a parent’s offspring to predict the parent’s breeding value, is an alternative predictor of an individual’s breeding value. In this case, the correlation between the mean $x$ of $n$ offspring and the breeding value $A$ of the parent is

$$\rho(x, A) = \sqrt{\frac{n}{n + a}}, \quad \text{where} \quad a = \frac{4 - h^2}{h^2}$$

From Equation 10.11, the response to selection under progeny testing is

$$R = i\sigma_A \sqrt{\frac{n}{n + a}} = i\sigma_A \sqrt{\frac{h^2 n}{4 + h^2(n - 1)}}$$

Note that for very large $n$ that the accuracy approaches one. Progeny testing gives a larger response than simple selection on the phenotypes of the parents (**mass selection**) when

$$\sqrt{\frac{n}{4 + h^2(n - 1)}} > 1, \quad \text{or} \quad n > \frac{4 - h^2}{1 - h^2}$$

In particular, $n > 4, 5, \text{ and } 7, \text{ for } h^2 = 0.1, 0.25, \text{ and } 0.5$. Also note that the ratio of response for progeny testing ($R_{pt}$) to mass selection ($R_{ms}$) is just

$$\frac{R_{pt}}{R_{ms}} = \frac{1}{h} \sqrt{\frac{h^2 n}{4 + h^2(n - 1)}} = \sqrt{\frac{n}{4 + h^2(n - 1)}}$$

which approaches $1/h$ for large $n$. 
Improving accuracy

• Predicting either the breeding or genotypic value from a single individual often has low accuracy --- $h^2$ and/or $H^2$ (based on a single individuals) is small
  - Especially true for many plant traits with high $G \times E$
  - Need to replicate either clones or relatives (such as sibs) over regions and years to reduce the impact of $G \times E$
  - Likewise, information from a set of relatives can give much higher accuracy than the measurement of a single individual
Stratified mass selection

- In order to accommodate the high environmental variance with individual plant values, Gardner (1961) proposed the method of **stratified mass selection**
  - Population stratified into a number of different blocks (i.e., sections within a field)
  - The best fraction p within each block are chosen
  - Idea is that environmental values are more similar among individuals within each block, increasing trait heritability.
Family selection

• Low heritabilities of traits a major issue
  - High $G \times E$, esp. year-to-year
  - Single phenotypes very poor predictors of $A, G$
  - Hence, often grow out relatives in field trails
    (multiple plots over a range of regions and years
    -- better sampling of $G \times E$)

• Within- versus between-family selection
• Response under general family selection
• Lush’s family index
Different types of family-based selection

Uppermost fraction \( p \) chosen, \( m \) families each with \( N \) sibs

1. **Between-family selection**: Individuals are selected solely on the basis of their family means, \( \bar{z}_i \), so that all individuals from the same family have the same selective rank. Here, the best \( pm \) families are chosen.

2. **Strict within-family (WF) selection**: The best \( pn \) individuals from each family are chosen, so that individuals are ranked within each family.

3. **Selection on within-family deviations (FD)**: Individuals are ranked solely on the basis of their within-family deviation \( z_{ij} - \bar{z}_i \). The \( pN \) individuals with the largest deviations (regardless of family) are chosen.

4. **Family-index selection**: Individuals are ranked using an index weighting within- and between-family components,

\[
I = b_1 (z_{ij} - \bar{z}_i) + b_2 \bar{z}_i \\
= b_1 z_{ij} + (b_2 - b_1) \bar{z}_i
\]
Within- vs. Between-family selection

The selection response for a particular family-based scheme depends on how the additive-genetic (breeding value) and total (phenotypic)variances are partitioned within and between families. When the number of sibs per family \( n \) is large, these variances are partitioned as

<table>
<thead>
<tr>
<th></th>
<th>Within-family</th>
<th>Between-family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breeding values</td>
<td>( (1 - r) \sigma_A^2 )</td>
<td>( r \sigma_A^2 )</td>
</tr>
<tr>
<td>Phenotypic values</td>
<td>( (1 - t) \sigma_z^2 )</td>
<td>( t \sigma_z^2 )</td>
</tr>
</tbody>
</table>

where \( t \) and \( r \) are the phenotypic and additive-genetic correlations between sibs (\( r = 1/4 \) for half-sibs, 1/2 for full-sibs). When the number of measured sibs within each family is small, \( t_n = t + (1 - t)/n \) replaces \( t \) and \( r_n \) (similarly defined) replaces \( r \).

Between-family response

\[
R_b = \frac{r_n \sigma_A^2}{t_n \sigma_z^2} S = \sigma_A \left( \frac{\sigma_A}{\sigma_z} \right) \left( \frac{r_n}{\sqrt{t_n}} \right) \left( \frac{S}{\sqrt{t_n \sigma_z}} \right) = \sigma_A h \frac{r_n}{\sqrt{t_n} \sigma_z}
\]

Within-family response

\[
R_{FD} = \frac{(1 - r_n) \sigma_A^2}{(1 - t_n) \sigma_z^2} S = \sigma_A h \frac{1 - r_n}{\sqrt{1 - t_n} \sigma_z}
\]
Within, Between or Individual selection?

Which scheme is best departs on the trait heritability and the intraclass correlation $t$ among sibs.

In the absence of epistasis,

$$t = r h^2 + c^2, \quad \text{where} \quad \sigma^2_{Zc} = \begin{cases} \sigma^2_{Ec(HS)} & \text{half-sibs} \\ \frac{1}{4} \sigma^2_D + \sigma^2_{Ec(FS)} & \text{full-sibs} \end{cases}$$

where $c^2$ is the residual between-family variance (upon removal of any additive variance).

Between-family response > individual when

$$c^2 < \begin{cases} \frac{1}{16} (1 - 4 h^2) & \text{half-sibs (for family, sib selection)} \\ \frac{1}{4} (1 - h^2) & \text{half-sibs (for parental and S1 seed selection)} \\ \frac{1}{4} (1 - 2h^2) & \text{full-sibs} \end{cases}$$

Low heritability, small common-family variance
Within-family response > individual when

\[ c^2 > 1 - (1 - r)^2 - rh^2 = \begin{cases} \frac{7}{16} - \frac{h^2}{4} & \text{half-sibs} \\ \frac{3}{4} - \frac{h^2}{2} & \text{full-sibs} \end{cases} \]

Requires low heritability, and a large c.

\( c^2 \text{Var}(z) \) = between family common variance thus accounts for much of the total trait variance
General response

- The basic idea is that a parent P generates a set of relatives $x_1$ through $x_n$ whose phenotypes are measured (selection unit or group)
- Based on the performance of the selection group, relatives $R_i$ of the best parents is chosen to represent $P_i$ in forming the next generation (recombination unit or group)
  - $R$ can be
    - The parent itself
    - Progeny (measured or unmeasured) of the parent, could be half- or full-sibs to the selection group, could also be the seed from selfing the parent
    - Other relatives of P
Key: The covariance $\sigma(x_i, y)$ between a member in the selection group and the offspring is critical to predicting selection response, closely related to $\sigma(x_i, A R_i)$, the covariance between selection unit and BV of $R_i$. Expected response is the average breeding values of the $R_i$. 

Crossed to make offspring

Chooses parents

Response in next generation
Response under general family selection

• Recall the accuracy version of the breeder’s equation, \( R = i r_{uA} \sigma_A \)

• In our context,
  - \( i \) is the selection intensity between selection units
  - \( u \) is the value of the selection unit
  - \( A \) the breeding value in the corresponding member \( R \) of the recombination unit, \( \sigma^2_A \) the additive variance in the recombination unit

• The correlation \( r_{uA} \) is obtained from standard resemblance between relatives calculations (full details in WL Chapter 19)
A) Family Selection
\( \mathcal{R} \) is a measured sib

\[ x_i = \mathcal{R} \]

\[ x_1 \ldots x_i = \mathcal{R} \quad \ldots \quad x_n \]

B) Sib Selection
\( \mathcal{R} \) is a unmeasured sib

\[ x_1 \ldots x_n \]

\[ \mathcal{R} \]

\[ y \]

C) Parental Selection
(progeny testing)

\[ P = \mathcal{R} \]

\[ x_1 \ldots x_n \]

\[ y \]

D) \( S_1 \) Seed Selection
\( \mathcal{R} \) is the selfed progeny (\( S_1 \)) of \( P \)

\[ x_1 \ldots x_n \]

\[ \mathcal{R} \]

\[ y \]
Between-Family Selection

**Family Selection**
- Half-sib Family Selection
- Full-sib Family Selection

**Sib Selection / Remnant Seed**
- Half-sib Sib Selection
- Full-sib Sib Selection

**Parental Selection / Progeny testing**
- Parent $P$
- $S_1$ Seed of $P$

**$S_1$ Seed Selection**
- Half-sib $S_1$ Seed Selection
- Full-sib $S_1$ Seed Selection

Within-Family Selection

**Family Deviation (FD) Selection**
- Half-sib Family Deviation Selection
- Full-sib Family Deviation Selection

**Strict Within-Family (WF) Selection**
- Half-sib Strict Within-Family Selection
- Full-sib Strict Within-Family Selection

<table>
<thead>
<tr>
<th>Recombination Unit $R$</th>
<th>Selection Unit $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured sib</td>
<td>$\overline{z}_{HS}$</td>
</tr>
<tr>
<td>$\overline{z}_{FS}$</td>
<td></td>
</tr>
<tr>
<td>Unmeasured sib</td>
<td>$\overline{z}_{HS}$</td>
</tr>
<tr>
<td>$\overline{z}_{FS}$</td>
<td></td>
</tr>
<tr>
<td>Parent $P$</td>
<td>$\overline{z}_{HS}$</td>
</tr>
<tr>
<td>$S_1$ Seed of $P$</td>
<td>$\overline{z}_{HS}$</td>
</tr>
<tr>
<td>$\overline{z}_{FS}$</td>
<td></td>
</tr>
<tr>
<td>Measured Sib</td>
<td>$z_{ij} - \overline{z}_{HS}$</td>
</tr>
<tr>
<td>$z_{ij} - \overline{z}_{FS}$</td>
<td></td>
</tr>
<tr>
<td>Measured Sib</td>
<td>$z_{ij} - \mu_{HS}$</td>
</tr>
<tr>
<td>$z_{ij} - \mu_{FS}$</td>
<td></td>
</tr>
</tbody>
</table>
Between-family Selection:

**Family selection** ($R_i$ is a measured sib from family $i$)

\[
\sigma(\bar{z}_i, y| R_1) = r_n (\sigma_A^2/2) = \begin{cases} 
(1 + 3/n) (\sigma_A^2/8) & \text{half-sibs} \\
(1 + 1/n) (\sigma_A^2/4) & \text{full-sibs}
\end{cases}
\]

**Sib Selection/ Remnant Seed** ($R_1$ is an unmeasured sib from family $i$)

\[
\sigma(\bar{z}_i, y| R_1) = r (\sigma_A^2/2) = \begin{cases} 
\sigma_A^2/8 & \text{half-sibs} \\
\sigma_A^2/4 & \text{full-sibs}
\end{cases}
\]

**Parental Selection/ Progeny testing** ($R_1$ is the parent of the sibs)

\[
\sigma(\bar{z}_i, y| R_1) = \sigma_A^2/4
\]

**S_1 seed design** ($R_1$ is a selfed progeny of the parent of the sibs)

\[
\sigma(\bar{z}_i, y| R_1) = \sigma_A^2/4
\]
Within-family Selection:

Selection on Family Deviations (FD)

\[ \sigma(z_{ij} - \bar{z}_i, y | R_1) = (1 - r_n) (\sigma^2_A/2) \]

\[ = \begin{cases} 
(1 - 1/n) (3/8) \sigma^2_A & \text{half-sibs} \\
(1 - 1/n) (\sigma^2_A/4) & \text{full-sibs} 
\end{cases} \]

Strict Within-family Selection (FW)

\[ \sigma(z_{ij} - \mu_i, y | R_1) = (1 - r) (\sigma^2_A/2) \]

\[ = \begin{cases} 
(3/8) \sigma^2_A & \text{half-sibs} \\
\sigma^2_A/4 & \text{full-sibs} 
\end{cases} \]
Variance in the selection unit

Half-sib between-family variance

$$\sigma^2(\bar{z}_{HS}) = \frac{\sigma_A^2}{4} + \frac{(3/4)\sigma_A^2 + \sigma_D^2 + \sigma_{E_s}^2}{n} + \sigma_{E_e(HS)}^2$$

Full-sib between-family variance

$$\sigma^2(\bar{z}_{FS}) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{4} + \frac{(1/2)\sigma_A^2 + (3/4)\sigma_D^2 + \sigma_{E_s}^2}{n} + \sigma_{E_e(FS)}^2$$

Half-sib with nested full-sibs between-family variance

($n_f$ females per male, $n_s$ offspring/female, $n = n fn_s$)

$$\sigma^2(\bar{z}_{HS(FS)}) = \frac{\sigma_A^2}{4} \left( 1 + \frac{1}{n_f} + \frac{2}{n} \right) + \frac{\sigma_D^2}{4n_f} \left( 1 + \frac{3}{n_s} \right) + \frac{\sigma_{E_s}^2}{n} + \frac{\sigma_{E_e(FS)}^2}{n_f} + \sigma_{E_e(HS)}^2$$

Half-sib within-family variance

$$\sigma^2(z_{ij} - \bar{z}_i \mid HS) = \left( 1 - \frac{1}{n} \right) \left( \frac{3}{4} \sigma_A^2 + \sigma_D^2 + \sigma_{E_s}^2 \right)$$

Full-sib within-family variance

$$\sigma^2(z_{ij} - \bar{z}_i \mid FS) = \left( 1 - \frac{1}{n} \right) \left( \frac{1}{2} \sigma_A^2 + \frac{3}{4} \sigma_D^2 + \sigma_{E_s}^2 \right)$$
Table 19.5. Variance-component expressions of the expected response with between-family selection schemes using outbred sibs. The number \( n \) of measured sibs is assumed sufficiently large that terms of order \( 1/n \) can be ignored (i.e., \( r_n \approx r \) and \( t_n \approx t \)). We also assume no epistasis and the simple structure \( E = E_c + E_s \) for environmental values.

<table>
<thead>
<tr>
<th>Family, Sib</th>
<th>Half-sibs</th>
<th>Full-sibs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \frac{(\sigma_A^2/8)(\bar{i}<em>{xm} + \bar{i}</em>{xf})}{\sqrt{\sigma_A^2/4 + \sigma_{Ec}(HS)}} )</td>
<td>( \frac{(\sigma_A^2/4)(\bar{i}<em>{xm} + \bar{i}</em>{xf})}{\sqrt{\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{Ec}(FS)}} )</td>
</tr>
<tr>
<td>Parental, S(_1) seed</td>
<td>( \frac{(\sigma_A^2/4)(\bar{i}<em>{xm} + \bar{i}</em>{xf})}{\sqrt{\sigma_A^2/4 + \sigma_{Ec}(HS)}} )</td>
<td>( \frac{(\sigma_A^2/4)(\bar{i}<em>{xm} + \bar{i}</em>{xf})}{\sqrt{\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{Ec}(FS)}} )</td>
</tr>
</tbody>
</table>

\[
R_{WF} = \begin{cases} 
\frac{(3/8)\sigma_A^2}{\sqrt{(3/4)\sigma_A^2 + \sigma_D^2 + \sigma_{Es}^2}} (\bar{i}_{xm} + \bar{i}_{xf}) & \text{half-sibs} \\
\frac{(1/4)\sigma_A^2}{\sqrt{\sigma_A^2/2 + (3/4)\sigma_D^2 + \sigma_{Es}^2}} (\bar{i}_{xm} + \bar{i}_{xf}) & \text{full-sibs}
\end{cases}
\]
Specific schemes: ear-to-row

- A common scheme in corn breeding is to plant the seeds from an ear as rows
  - Each row is thus a half-sib family (this is the selection unit)
  - Some seed from ear saved (these form the recombination unit)
  - Suppose total $N$ seeds per ear grown as $n_p$ rows of $n_s$ sibs in $n_e$ environments

$$R_{ER} = \sqrt{\frac{\sigma^2_A}{2} + \frac{\sigma^2_{F \times E}}{n_e} + \frac{\sigma^2_{E_p}}{n_e n_p} + \frac{\sigma^2_{E_{w(p)}}}{N}}$$

Family x E interaction
Modified ear-to-row

• Lonnquist (1964) proposed ear-to-row
• Combines ear-to-row (between family) with selection within row (within-family selection)
• Plant the seeds from the ear into two sets of rows. One set is several rows over multiple environments. Select best ears based on this performance
• Grow out residual seed from these best ears in a single row, then select best from each family within each row
Best individuals within the families selected from field trials are used for recombination.

Field trials to select best families. Each family replicated over several environments.
Response

Total $R = R$ from ear-to-row + $R$ from within-row

$$R_{ER(m)} = R_{ER} + R_{ER(w)}$$

$$= \bar{v} \frac{\sigma_A^2/8}{\sigma(z_{HS})} + \bar{v} \frac{(3/8)\sigma_A^2}{\sigma_w(HS)}$$

$$R_{ER(m)} = \sqrt{\frac{\bar{v} \sigma_A^2/8}{\sigma_{GF} + \frac{\sigma_F \times E}{n_e} + \frac{\sigma_{EP}^2}{n_e n_p} + \frac{\sigma_{E_{w(p)}^2}}{N}}} + \frac{\bar{v} (3/8)\sigma_A^2}{\sqrt{\sigma_{Gw(HS)}^2 + \sigma_{E_w(p)}^2}}$$
Lush’s family index

Finally, Lush suggested that an index weighting both within- and between-family values is optimal

\[ I = b_1 (z_{ij} - \bar{z}_i) + b_2 \bar{z}_i \]

The optimal weights are given by

\[ b_1 = \frac{1 - r}{1 - t}, \quad b_2 = \frac{1 + (n - 1) r}{1 + (n - 1) t} \]

Ratio of response/individual selection

\[ \frac{R_{LI}}{R_m} = \sqrt{1 + \frac{(r - t)^2(n - 1)}{(1 - t)[1 + (n - 1)t]}} > 1 \]
Half-sibs

More emphasis on within-family deviations

More emphasis on between-family deviations

b2 / b1

t

n = 2
n = 4
n = 10
n = 25
n = 50
More emphasis on within-family deviations

More emphasis on between-family deviations

Full-sibs

- $n = 2$
- $n = 4$
- $n = 10$
- $n = 25$
- $n = 50$
Relative response of Index/Individual selection

Half-Sibs

Phenotypic correlation $t$ between sibs

Response relative to Mass Selection

- $n = 2$
- $n = 5$
- $n = 20$
Relative response of Index/Individual selection

![Graph showing relative response of Index/Individual selection with different values of n (n = 2, n = 5, n = 20) and phenotypic correlation t between sibs.](image)
Permanent Versus Transient Response

Considering epistasis and shared environmental values, the single-generation response follows from the midparent-offspring regression

\[ R = h^2 S + \frac{S}{\sigma_z^2} \left( \frac{\sigma_A^2}{2} + \frac{\sigma_{AA}^2}{4} + \cdots + \sigma(E_{sire}, E_o) + \sigma(E_{dam}, E_o) \right) \]

Breeder’s Equation

Response from epistasis

Response from shared environmental effects

Permanent component of response

Transient component of response --- contributes to short-term response. Decays away to zero over the long-term
Permanent Versus Transient Response

The reason for the focus on $h^2$S is that this component is permanent in a random-mating population, while the other components are transient, initially contributing to response, but this contribution decays away under random mating.

Why? Under HW, changes in allele frequencies are permanent (don’t decay under random-mating), while LD (epistasis) does, and environmental values also become randomized.
Response with Epistasis

The response after one generation of selection from an unselected base population with $A \times A$ epistasis is

$$R = S \left( h^2 + \frac{\sigma_{AA}^2}{2 \sigma_z^2} \right)$$

The contribution to response from this single generation after $\tau$ generations of no selection is

$$R(1 + \tau) = S \left( h^2 + (1 - c)^\tau \frac{\sigma_{AA}^2}{2 \sigma_z^2} \right)$$

c is the average (pairwise) recombination between loci involved in $A \times A$
Response with Epistasis

\[ R(1 + \tau) = S \left( h^2 + (1 - c)\tau \frac{\sigma^2_{AA}}{2\sigma^2_z} \right) \]

Response from additive effects \((h^2 \ S)\) is due to changes in allele frequencies and hence is permanent. Contribution from \(A \times A\) due to linkage disequilibrium

Contribution to response from epistasis decays to zero as linkage disequilibrium decays to zero
Why unselected base population? If history of previous selection, linkage disequilibrium may be present and the mean can change as the disequilibrium decays.

More generally, for $t$ generation of selection followed by $\tau$ generations of no selection (but recombination)

$$R(t + \tau) = th^2 S + (1 - c)^\tau R_{AA}(t)$$

$R_{AA}$ has a limiting value given by

$$\tilde{R}_{AA} = \lim_{t \to \infty} R_{AA}(t) = \frac{1}{c} \left( S \frac{\sigma^2_{AA}}{2 \sigma^2_z} \right)$$

Time to equilibrium a function of $c$

$$t_{1/2} = \frac{-\ln(2)}{\ln(1 - c)}$$
What about response with higher-order epistasis?

\[
\frac{S \sigma^2(A^i)}{\sigma_z^2} = \frac{1}{c} \left( S \frac{\sigma^2_{AA}}{2 \sigma_z^2} \right)
\]

Fixed incremental difference that decays when selection stops

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AAA</th>
<th>AAAAA</th>
<th>AAAAAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \sigma^2(A^i)/\sigma_z^2$</td>
<td>0.500</td>
<td>0.250</td>
<td>0.125</td>
<td>0.063</td>
</tr>
<tr>
<td>$R(1)$</td>
<td>1.000</td>
<td>0.333</td>
<td>0.143</td>
<td>0.067</td>
</tr>
<tr>
<td>Limit</td>
<td>0.063</td>
<td>0.067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% $R(1)/$limit</td>
<td>50.0</td>
<td>75.0</td>
<td>87.5</td>
<td>93.8</td>
</tr>
</tbody>
</table>
Response in autotetraploids

- Autotraploids pass along two alleles at each locus to their offspring.
- Hence, dominance variance is passed along.
- However, as with $A \times A$, this depends upon favorable combinations of alleles, and these are randomized over time by transmission, so D component of response is transient.
Autotetraploids

P-O covariance

\[ \sigma(z_p, z_o) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{6}, \]

Single-generation response

\[ R = S \left( h^2 + \frac{\sigma_D^2}{3\sigma_z^2} \right) \]

Response to \( t \) generations of selection with constant selection differential \( S \)

\[ R(t) = th^2 S + R_D(t) \]

\[ R_D(t) = S \frac{3}{2} \left[ 1 - \left( \frac{1}{3} \right)^t \right] \frac{\sigma_D^2}{3\sigma_z^2} \]

Response remaining after \( t \) generations of selection followed by \( \tau \) generations of random mating

\[ th^2 S + \left( \frac{1}{3} \right)^\tau R_D(t) \]

Contribution from dominance quickly decays to zero
General responses

• As we have seen with both individual and family selection, the response can be thought of as a regression of some phenotypic measurement (such as the individual itself or its corresponding selection unit value x) on either the offspring value (y) or the breeding value $R_A$ of the individual in the recombination group.

• The regression slope for predicting y from x is $\frac{\sigma(x,y)}{\sigma^2(x)}$ and $\frac{\sigma(x,R_A)}{\sigma^2(x)}$ for predicting the BV $R_A$ from x.

• With transient components of response, these covariances now also become functions of time --- e.g. the covariance between x in one generation and y several generations later.
Ancestral Regressions

When regressions on relatives are linear, we can think of the response as the sum over all previous contributions.

For example, consider the response after 3 gens:

\[
R(3) = 8 \beta_{3,0} S_0 + 4 \beta_{3,1} S_1 + 2 \beta_{3,2} S_2
\]

- 8 great-grand parents
- \( S_0 \) is the selection differential
- \( \beta_{3,0} \) is the regression coefficient for an offspring at time 3 on a great-grandparent
- From time 0
- 4 grandparents
- Selection diff \( S_1 \)
- \( \beta_{3,1} \) is the regression of relative in generation 3 on their gen 1 relatives
- 2 parents
- 3 on their gen 1 relatives
Ancestral Regressions

More generally,

\[ \beta_{T,t} = \text{cov}(z_T, z_t) \]

The general expression \( \text{cov}(z_T, z_t) \), where we keep track of the actual generation, as oppose to \( \text{cov}(z, z_{T-t}) \) -- how many generations Separate the relatives, allows us to handle inbreeding, where the (say) P-O regression slope changes over generations of inbreeding.

Unless \( 2^t \beta_{\tau+t,\tau} \) remains constant as \( t \) increases, the contribution to cumulative response from selection on adults in generation \( \tau \) changes over time. For example, when loci are strictly additive (no dominance or epistasis), \( \sigma_G(\tau + t, \tau) = 2^{-t} \sigma_A^2(\tau) \) and thus \( 2^t \beta_{\tau+t,\tau} = h_r^2 \), the standard result from the breeders’ equation. However, unless \( 2^t \sigma_G(\tau + t, \tau) \) remains constant, any response contributed decays. Hence any term of \( \sigma_G(\tau + t, \tau) \) that decreases by more than 1/2 each generation contributes only to the transient response.
Selfing

- Finally, let’s consider selection when an F₁ is formed and then selfed for several generations to generate inbred lines.
- How best to advance lines to full inbreeding while selecting.
  - Should advancement occur first, and the we can chose among the result inbred lines.
  - Or should some selection (early testing) be occurring while the lines are being inbred.
- First, we will consider the response with simultaneous selfing and selection.
  - Inbreeding removes within-line variation, enhances between-line variation.
If dominance is present, the selection response under selfing has both a transient and a permanent component. When selection is relaxed, the mean potentially changes as the transient component decays. The expected total change in the mean after $n$ generations, the first $T$ of which were under selection (generations 0 to $T - 1$), is given by

$$R(\ n\ |\ T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(n, t)}{\sigma_G(t, t) + \sigma_e^2} = \sum_{t=0}^{T-1} \bar{n}_t \frac{\sigma_G(n, t)}{\sqrt{\sigma_G(t, t) + \sigma_e^2}}$$

The permanent response to $T$ generations of selection, $\tilde{R}(T)$, is given by

$$\tilde{R}(T) = R(\infty\ |\ T) = \sum_{t=0}^{T-1} S_t \frac{\sigma_G(\infty, t)}{\sigma_G(t, t) + \sigma_e^2} = \sum_{t=0}^{T-1} \bar{n}_t \frac{\sigma_G(\infty, t)}{\sqrt{\sigma_G(t, t) + \sigma_e^2}}$$

Details for computing these genetic covariances are given in Walsh & Lynch Chapter 20 (online)

Starting selection on the $F_2$'s and denoting this as generation 0, the response after $n \geq T$ generations, the first $T$ of which were under selection is

$$R(\ n\ |\ T) = \sum_{t=0}^{T-1} \bar{n}_t \frac{(2 - 2^{-t})(\sigma_A^2 + 2^{-n}\sigma_D^2)}{\sqrt{(2 - 2^{-t})(\sigma_A^2 + 2^{-t}\sigma_D^2) + \sigma_e^2}}$$
Selection of the best pure lines

• A very common setting in plant breeding is when two (or more) inbreds are crossed and the resulting $F_1$ continually selfed to form a series of inbred lines.

• This is different from selecting elite lines among a set of already inbred lines, as the breeder also has to advance the lines to fully inbreds, which often takes time, in addition to trying to select the best ones.
  
  - Most accurate measure of performance (given $G \times E$) are multiregional trails wherein a subset of the advancing lines are measured over a series of regions and years.
Advancing to full inbreds

• How best to combine advancing a line to being fully inbred while still selecting (testing) them.
  - Tradeoff between less accurate testing a large number of lines (but more variation kept) vs. more accurate testing of a smaller number of lines (representing less variation)

• Methods
  - Single seed descent (SSD)
  - Doubled haploids (DH)
  - Bulk Selection
  - Pedigree selection
SSD, DH selection

- Under **SSD** (*single-seed descent*), single seeds are used to advance a series of lines to full inbreeding, then selection (choosing among them) occurs, usually through multiregional trails.
- Single seeds are used to reduce any effects of selection during inbreed.
- Under **DH** (*double haploids*), inbred lines are formed in one generation.
  - Less chance for decay of any LD relative to SSD, but effect likely to be small.
Bulk Selection

- Seeds from natural selfers are grown and harvested in bulk over multiple generations.
  - One problem is the natural selection during the advancing of generations does select on yield (leaving more descendants) but also on other traits.
  - Often tall plants are naturally selected during the bulk over higher yielding short plants.
Atlas wins, but Vaughn best from an agricultural standpoint (higher yield -- 107% of Atlas, earlier heading date, better disease resistance)
Pedigree selection

• Not to be confused with the pedigree-based selection using BLUP (which very formally uses information for all relatives in selection decisions).

• Under pedigree selection (aka pedigree breeding), as individuals become more inbred, selection decisions shift from individual plants towards family-based performance.

• High heritability traits selected early (individual selection), lower heritability traits (e.g., those with high G x E) selected later (family selection allowing for replication over different G's).
Pedigree selection

Individual selection

Family-based selection

Individual selection

Family-based selection

Individual (single plant) selection

F_2

F_3

F_4

F_5

F_6

F_7

F_8

F_9

Three Generations of yield trails

Selection of best lines based on trials

Three Generations of yield trails

Selection of best lines based on trials
Early generation testing

• Much debate on the effectiveness of early generation testing
• Effectiveness depends on a high correlation between phenotypes in the tested generations and the final genotypes of their selfed descendants
  - Concerns: genotypes selected early can be different from their descendant full inbred offspring
  - Basic idea: OK for high heritability traits
  - Not so good for low-heritability traits
Effectiveness of different methods

**Soybeans** (*Glycine max*)
- No consistent F₆ differences between PS, BS
- No consistent F₆/F₇ differences between EGYT (F₄, F₅), PS, SSD
- No consistent F₈ differences between EGYT, PS, SSD
- SSD, BS showed no consistent differences
- BS, PS equal in 4 of 6 crosses
- SSD, PS equivalent
  - * BS significantly greater than PS equal in 2 of 6 crosses
  - * EGYT (F₄) superior to BS, PS

**Wheat** (*Triticum aestivum*)
- *SSD, SPS equivalent; both superior to BS*
- BS, SSD methods give identical results

**Barley** (*Hordeum vulgare*)
- BS, PS equally effective
- DH, SSD, PS equally effective

EGYT = early generation yield traits, BS = Bulk Selection, DH = Doubled haploids, SSD = single seed descent, PS = Pedigree selection