QTL Mapping, MAS, and Genomic Selection

Dr. Ben Hayes
Department of Primary Industries
Victoria, Australia

A short-course organized by
Animal Breeding & Genetics
Department of Animal Science
Iowa State University
June 4-8, 2007

With financial support from
Pioneer Hi-bred Int.

USE AND ACKNOWLEDGEMENT OF SHORT COURSE MATERIALS
Materials provided in these notes are copyright of Dr. Ben Hayes and the Animal Breeding and Genetics group at Iowa State University but are available for use with proper acknowledgement of the author and the short course. Materials that include references to third parties should properly acknowledge the original source.
Linkage Disequilibrium to Genomic Selection
Course overview

- Day 1
  - Linkage disequilibrium in animal and plant genomes
- Day 2
  - QTL mapping with LD
- Day 3
  - Marker assisted selection using LD
- Day 4
  - Genomic selection
- Day 5
  - Genomic selection continued
Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
Marker Assisted Selection using LD

- Marker assisted selection (MAS) can be based on DNA markers
  - in linkage equilibrium with a QTL (LE-MAS)
  - in linkage disequilibrium with a QTL (LD-MAS)
  - actual mutation causing QTL effect (Gene-MAS).
- All three types of MAS are currently used in the livestock industries (Dekkers 2004).
Table 1. Examples of gene tests used in commercial breeding for different species (D = dairy cattle, B = beef cattle, C = poultry, P = pigs, S = sheep) by trait category and type of marker.

<table>
<thead>
<tr>
<th>Trait category</th>
<th>Direct marker</th>
<th>Linkage disequilibrium marker</th>
<th>Linkage equilibrium marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congenital defects</td>
<td>BLAD (D^a)</td>
<td>Citrulinaemia (D,B^b)</td>
<td>DUMPS (D^c)</td>
</tr>
<tr>
<td>Maple syrup urine</td>
<td>Maple syrup urine (D,B^e)</td>
<td>Mannosidosis (D,B^f)</td>
<td>RYR (P^g)</td>
</tr>
<tr>
<td>Appearance</td>
<td>CKIT (P^h)</td>
<td>MC1R/MSHR (P,B^i,D^j)</td>
<td>MGF (B^k)</td>
</tr>
<tr>
<td>Milk quality</td>
<td>-Casein (D^m)</td>
<td>-Casein (D^n)</td>
<td>FMO3 (D^o)</td>
</tr>
<tr>
<td>Meat quality</td>
<td>RYR (P^q)</td>
<td>RN/PRKAG3 (P^r)</td>
<td>A-FABP/FABP4 (P^s)</td>
</tr>
<tr>
<td>&gt;15 PICmarq (P^u)</td>
<td>CAST (P^v)</td>
<td>THYR (B^w)</td>
<td>Leptin (B^x)</td>
</tr>
<tr>
<td>Feed intake</td>
<td>MC4R (P^y)</td>
<td>B blood group (C^z)</td>
<td>K58 (P^aa)</td>
</tr>
<tr>
<td>Disease</td>
<td>Prp (S^ab)</td>
<td>F18 (P^ac)</td>
<td>Booroola (S^ad)</td>
</tr>
<tr>
<td>Reproduction</td>
<td>Booroola (S^ae)</td>
<td>Booroola (S^af)</td>
<td>Inverdale (S^ag)</td>
</tr>
<tr>
<td></td>
<td>Hanna (S^ah)</td>
<td>PRLR (P^ai)</td>
<td>RBP4 (P^aj)</td>
</tr>
<tr>
<td>Growth and composition</td>
<td>MC4R (P^ak)</td>
<td>CAST (P^al)</td>
<td>QTL (P^am)</td>
</tr>
<tr>
<td></td>
<td>IGF-2 (P^an)</td>
<td>IGF-2 (P^ao)</td>
<td>QTL (B^ap)</td>
</tr>
<tr>
<td></td>
<td>Myostatin (B^aq)</td>
<td>Myostatin (B^ar)</td>
<td>QTL (B^as)</td>
</tr>
<tr>
<td></td>
<td>Callipyge (S^at)</td>
<td>Callipyge (S^au)</td>
<td>Carwell (S^av)</td>
</tr>
<tr>
<td>Milk yield and composition</td>
<td>DGAT (D^aw)</td>
<td>PRL (D^ax)</td>
<td>QTL (D^ay)</td>
</tr>
<tr>
<td></td>
<td>GRH (D^az)</td>
<td>-Casein (D^ba)</td>
<td></td>
</tr>
</tbody>
</table>
Marker Assisted Selection using LD

- LE-MAS is most difficult to implement.
  - marker-QTL phase within each family must be established before an increase in selection response can be realised.

- LD-MAS now very attractive due to very large numbers of single nucleotide polymorphism (SNP) markers suitable for LD mapping now available.

- Gene-MAS requires enormous amount of work and resources!!
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.

- In many cases, the selection candidates will have no phenotypic information of their own, e.g., young dairy bulls which are progeny test candidates.
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
LD-MAS with single markers

- Estimate effects of marker or markers in reference population

\[ y = 1_n \mu + X g + Zu \]
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
LD-MAS with single markers

- Predict breeding values using marker information:

\[
\hat{MEBV} = \hat{u} + X \hat{g}
\]
**LD-MAS with single markers**

- **Example**

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Phenotype</th>
<th>SNP allele 1</th>
<th>SNP allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3.53</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3.54</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3.83</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.87</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.91</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2.34</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2.65</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3.76</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>6</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>7</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
LD-MAS with single markers

• The data was simulated as a SNP effect of 1 for 2 allele plus effect of sire 1 of 3 and sire 5 of -3 + random effect
LD-MAS with single markers

- **Example**

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Phenotype</th>
<th>SNP allele 1</th>
<th>SNP allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3.53</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3.54</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3.83</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.87</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.91</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2.34</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2.65</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3.76</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>6</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>7</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
LD-MAS with single markers

- **Build:**

\[
\begin{bmatrix}
\begin{array}{c}
\mu \\
\hline
\overline{g} \\
\hline
\overline{u}
\end{array}
\end{bmatrix} = 
\begin{bmatrix}
1_n'1_n & 1_n'X & 1_n'Z \\
X'1_n & X'X & X'Z \\
Z'1_n & Z'X & Z'Z + A^{-1} \lambda
\end{bmatrix}^{-1}
\begin{bmatrix}
1_n'y \\
X'y \\
Z'y
\end{bmatrix}
\]
LD-MAS with single markers

- Example
- $1_n$ and $X$

<table>
<thead>
<tr>
<th>record</th>
<th>1n</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Phenotype</th>
<th>SNP allele 1</th>
<th>SNP allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3.53</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3.54</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3.83</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.87</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.91</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2.34</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2.65</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3.76</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>6</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>7</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
### LD-MAS with single markers

- Example
- Z

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Phenotype</th>
<th>SNP allele 1</th>
<th>SNP allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3.53</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3.54</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3.83</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.87</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.91</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2.34</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2.65</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3.76</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>3.69</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>6</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>7</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>8</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**animal**

<table>
<thead>
<tr>
<th>record</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
LD-MAS with single markers

- Example
- $A$
- $\lambda = 1/2$
LD-MAS with single markers

- Example
- Solve equations.

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{g} \\
\hat{u}
\end{bmatrix} = 
\begin{bmatrix}
1_n'1_n & 1_n'X & 1_n'Z \\
X'1_n & X'X & X'Z \\
Z'1_n & Z'X & Z'Z + A^{-1}\lambda
\end{bmatrix}^{-1} 
\begin{bmatrix}
1_n'y \\
X'y \\
Z'y
\end{bmatrix}
\]

\[
\begin{array}{c|c}
\hat{\mu} & 2.69 \\
\hat{g} & 0.87 \\
\hat{u} & \begin{array}{c|c|c|c|c|c}
1 & 0.56 \\
2 & -0.01 \\
3 & 0.19 \\
4 & 0.3 \\
5 & -1.1 \\
6 & -0.23 \\
7 & -0.03 \\
8 & 0.14 \\
9 & 0.09 \\
10 & 0.09 \\
11 & 0.28 \\
12 & 0.43 \\
13 & -0.67 \\
14 & -0.56 \\
15 & -0.48
\end{array}
\end{array}
\]
Marker Assisted Selection using LD

- LD-MAS as a two step procedure.
  - Step 1. Effects of a marker or set of markers are estimated in a reference population.
  - Step 2. The breeding values of a group of selection candidates are calculated using the marker information.
LD-MAS with single markers

- Predict breeding values using marker information:

\[ \text{MEBV} = u + Xg \]
LD-MAS with single markers

- Predict breeding values using marker information:

\[ \hat{MEBV} = \hat{u} + X \hat{g} \]

<table>
<thead>
<tr>
<th>( \hat{u} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
</tr>
<tr>
<td>0.43</td>
</tr>
<tr>
<td>-0.67</td>
</tr>
<tr>
<td>-0.56</td>
</tr>
<tr>
<td>-0.48</td>
</tr>
</tbody>
</table>
**LD-MAS with single markers**

- Predict breeding values using marker information:

\[
\hat{\text{MEBV}} = \hat{u} + X\hat{g}
\]

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{u})</td>
<td>(u)</td>
<td>(X)</td>
</tr>
<tr>
<td>0.28</td>
<td>1</td>
<td>0.87</td>
</tr>
<tr>
<td>0.43</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>-0.67</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>-0.56</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>-0.48</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
LD-MAS with single markers

- Predict breeding values using marker information:

\[
\hat{\text{MEBV}} = \hat{\mathbf{u}} + \mathbf{X}\hat{\mathbf{g}}
\]

<table>
<thead>
<tr>
<th>(\hat{\mathbf{u}})</th>
<th>(\mathbf{X})</th>
<th>(\hat{\mathbf{g}})</th>
<th>MEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>1</td>
<td>0.87</td>
<td>1.14</td>
</tr>
<tr>
<td>0.43</td>
<td>1</td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td>-0.67</td>
<td>+</td>
<td>0</td>
<td>=</td>
</tr>
<tr>
<td>-0.56</td>
<td>1</td>
<td></td>
<td>-0.67</td>
</tr>
<tr>
<td>-0.48</td>
<td>2</td>
<td></td>
<td>1.26</td>
</tr>
</tbody>
</table>
LD-MAS with single markers

- The data was simulated as a SNP effect of 1 for 2 allele plus effect of sire 1 of 3 and sire 5 of -3 + random effect

<table>
<thead>
<tr>
<th>( ^\wedge )</th>
<th>u</th>
<th>X</th>
<th>( ^\wedge )</th>
<th>g</th>
<th>MEBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>1</td>
<td>0.87</td>
<td></td>
<td></td>
<td>1.14</td>
</tr>
<tr>
<td>0.43</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td>-0.67</td>
<td>+</td>
<td>0</td>
<td></td>
<td>=</td>
<td>-0.67</td>
</tr>
<tr>
<td>-0.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>-0.48</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1.26</td>
</tr>
</tbody>
</table>
LD-MAS with single markers

- Corr(MEBV,TBV) = 0.93

<table>
<thead>
<tr>
<th>$\hat{u}$</th>
<th>$X$</th>
<th>$\hat{g}$</th>
<th>MEBV</th>
<th>TBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>1</td>
<td>0.87</td>
<td>1.14</td>
<td>1.75</td>
</tr>
<tr>
<td>0.43</td>
<td>1</td>
<td></td>
<td>1.3</td>
<td>1.75</td>
</tr>
<tr>
<td>-0.67</td>
<td>+</td>
<td>0</td>
<td>-0.67</td>
<td>-0.75</td>
</tr>
<tr>
<td>-0.56</td>
<td>1</td>
<td></td>
<td>0.3</td>
<td>0.25</td>
</tr>
<tr>
<td>-0.48</td>
<td>2</td>
<td></td>
<td>1.26</td>
<td>1.25</td>
</tr>
</tbody>
</table>
LD-MAS with single markers

- Corr(MEBV, TBV) = 0.93
- Corr(EBV, TBV) = ?

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{\mathbf{u}}
\end{bmatrix} = \begin{bmatrix}
1_n'1_n & 1_n'\mathbf{Z} \\
\mathbf{Z}'1_n & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda
\end{bmatrix}^{-1} \begin{bmatrix}
1_n'\mathbf{y} \\
\mathbf{Z}'\mathbf{y}
\end{bmatrix}
\]
LD-MAS with single markers

- $\text{Corr}(\text{MEBV}, \text{TBV}) = 0.93$
- $\text{Corr}(\text{EBV}, \text{TBV}) = 0.88$

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{\mathbf{u}}
\end{bmatrix} = \begin{bmatrix}
1_n'1_n & 1_n'\mathbf{Z} \\
\mathbf{Z}'1_n & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda
\end{bmatrix}^{-1} \begin{bmatrix}
1_n'y \\
\mathbf{Z}'y
\end{bmatrix}
\]
Marker Assisted Selection using LD

- LD-MAS with a single marker
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
How many QTL to use in LD-MAS

- Advantage of MAS over non-MAS approximately proportional to proportion of total genetic variance explained by QTL
- Estimates of number of QTL per trait between 100 and 200
- Do we need to track all these with markers?
How many QTL to use in LD-MAS

Percentage of QTL (ranked in order of size)

Percentage of genetic variance accounted for

Pigs

Dairy
How many QTL to use in LD-MAS

- If we use 10-20 QTL per trait in our LD-MAS program, we will exploit ~50% of the genetic variance.
- Assumes we have perfect knowledge of the QTL alleles.
- The proportion of genetic variance captured at each QTL in LD-MAS depends on the extent of linkage disequilibrium between the marker and the QTL.
How many QTL to use in LD-MAS

• Use multiple regression to estimate vector of SNP effects with multiple markers

\[ y = \mathbf{1}_n \mu + X_1 g_1 + X_2 g_2 + e \]
How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{\mu}_1 \\
\hat{\mu}_2 \\
\hat{\mu}_u
\end{bmatrix} = 
\begin{bmatrix}
1_n'1_n & 1_n'X_1 & 1_n'X_2 & 1_n'Z \\
X_1'1_n & X_1'X_1 & X_1'X_2 & X_1'Z \\
X_2'1_n & X_2'X_1 & X_2'X_2 & X_2'Z \\
Z'1_n & Z'X_1 & Z'X_2 & Z'Z + A^{-1}\lambda
\end{bmatrix}^{-1}
\begin{bmatrix}
1_n'y \\
X_1'y \\
X_2'y \\
Z'y
\end{bmatrix}
\]
How many QTL to use in LD-MAS

• Use multiple regression to estimate vector of SNP effects with multiple markers
• Accounts for the fact that some SNPs may be picking up the same QTL
LD-MAS with single markers

- Predict breeding values using marker information:

\[ \hat{\text{MEBV}} = u + X_1 \hat{g}_1 + X_2 \hat{g}_2 + \ldots. \]
How many QTL to use in LD-MAS

- Use multiple regression to estimate vector of SNP effects with multiple markers (random?)

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{g}_1 \\
\hat{g}_2 \\
\hat{u}
\end{bmatrix} =
\begin{bmatrix}
1_n'1_n & 1_n'X_1 & 1_n'X_2 & 1_n'Z \\
X_1'1_n & X_1'X_1 + I\lambda_1 & X_1'X_2 & X_1'Z \\
X_2'1_n & X_2'X_1 & X_2'X_2 + I\lambda_2 & X_2'Z \\
Z'1_n & Z'X_1 & Z'X_2 & Z'Z + A^{-1}\lambda
\end{bmatrix}^{-1}
\begin{bmatrix}
1_n'y \\
X_1'y \\
X_2'y \\
Z'y
\end{bmatrix}
\]

- Use variance component estimation to get SNP effects
Marker Assisted Selection using LD

- LD-MAS with a single marker
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term
Accounting for bias in QTL effects

- Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term.
- This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.
Accounting for bias in QTL effects

• Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term.

• This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.

• If the QTL effect is over-estimated, the advantage of MAS can be eroded substantially (eg LD-MAS with a single marker).
Accounting for bias in QTL effects

• Strong tendency to overestimate QTL effects in a genome scan, as these effects can exceed significance thresholds if the estimate is larger than the actual effect due to a large positive error term

• This over-estimation is more pronounced in genome scans of low power, positive error term must be large to overcome the significance threshold.

• If the QTL effect is over-estimated, the advantage of MAS can be eroded substantially (eg LD-MAS with a single marker)

• Must regress QTL effects prior to use in MAS
Accounting for bias in QTL effects

Accuracy of predicting MEBV

- Pedigree only data set 2
- Pedigree and Markers data set 2
Accounting for bias in QTL effects

• Options for estimating unbiased estimates of QTL effect
  – Best method is to estimate QTL effects in a population which is completely independent of the sample used in the original genome scan where the QTL were first detected.
  – This will also validate that the markers are not an artefact of the statistical model used in the genome scan or some unaccounted for population stratification.
  – But maybe too expensive
  – Use prior knowledge of distribution of QTL effects to regress effects
  – Cross validation
Accounting for bias in QTL effects

- Use prior knowledge of distribution of QTL effects to regress effects
- Then for a given size of experiment and estimated size of effect, we can calculate the true effect
Accounting for bias in QTL effects

- Use prior knowledge of distribution of QTL effects to regress effects
- Then for a given size of experiment and estimated size of effect, we can calculate the true effect
- See Weller et al. 2005 for distributions of QTL effects across traits
Accounting for bias in QTL effects

• Cross validation
  – split data set in two
  – regress solutions from data set two on data set one to get $b_{x_1x_2}$
  – then the regression of the true effects of the SNPs on the solutions from the full data set is
    • $b_{u,xt} = \frac{2b_{x_1x_2}}{1+b_{x_1x_2}}$
Accounting for bias in QTL effects

- Cross validation

\[ y = 0.2821x + 0.1594 \]

\[ R^2 = 0.1097 \]
Accounting for bias in QTL effects

• Cross validation
  – split data set in two
  – regress solutions from data set two on data set one to get $b_{x1x2}$
  – then the regression of the true effects of the SNPs on the solutions from the full data set is
    • $b_{u,xt} = \frac{2b_{x1x2}}{1+b_{x1x2}}$
    • = 0.44
Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
LD-MAS with haplotypes

- Model:
  \[^\wedge\wedge MEBV = u + Xg\]

- \(g\) is a vector of haplotype effects, eg.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>-0.12</td>
</tr>
<tr>
<td>3</td>
<td>-0.11</td>
</tr>
<tr>
<td>4</td>
<td>0.21</td>
</tr>
</tbody>
</table>
LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes
    - Number of haplotype effects to estimate
    - Number of phenotypic records
    - Accuracy of inferring haplotypes
LD-MAS with haplotypes

• Accuracy of LD-MAS with haplotypes
  – Depends on
    • Proportion of QTL variance explained by haplotypes
    • Number of haplotype effects to estimate
    • Number of phenotypic records
    • Accuracy of inferring haplotypes
LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
  - Proportion of QTL variance explained by haplotypes

\[ r^2 (h, q) = \frac{n \sum_{i=1}^{n} \frac{D_i^2}{p_i}}{q_1 q_2} \]
LD-MAS with haplotypes

• Example:
• 10,000 SNPs genotyped in 379 Angus animals
• Select a SNP from the 10,000 at random to be a “QTL”
  – determine
    • Nearest marker
    • 2, 4 or 6 marker haplotypes
      – Haplotypes estimated using PHASE program (Stephens et al. 2001)
      – This takes into account LD structure in the cattle populations
• Calculate the proportion of QTL variance explained by the marker haplotypes.
Results

![Graph showing the proportion of QTL variance explained by marker haplotypes against the number of markers in the haplotype. The graph illustrates a positive correlation, with the proportion increasing as the number of markers increases. Notable points include:

- At 0 markers, the proportion is 0.
- At 1 marker, the proportion is approximately 0.1.
- At 2 markers, the proportion is slightly above 0.2.
- At 3 markers, the proportion is around 0.3.
- At 4 markers, the proportion is near 0.4.
- At 5 markers, the proportion jumps to approximately 0.5.
- At 6 markers, the proportion reaches just below 0.6.

The trend suggests that the contribution of marker haplotypes to QTL variance increases with the number of markers included.]
Results

Number of markers in haplotype

Proportion of QTL variance explained by marker haplotypes

1 Q
1 q
Results

The graph shows the proportion of QTL variance explained by haplotypes as a function of the number of markers in the haplotype.

- The x-axis represents the number of markers in the haplotype.
- The y-axis represents the proportion of QTL variance explained by the haplotype.

Two haplotypes are highlighted:

- **1 Q**: This haplotype has a proportion of QTL variance explained by markers ranging from 0.1 to 0.6.
- **1 q**: This haplotype has a proportion of QTL variance explained by markers ranging from 0.5 to 0.6.

The graph indicates a clear trend where the proportion of QTL variance increases with the number of markers in the haplotype.
LD-MAS with haplotypes

- Example:

<table>
<thead>
<tr>
<th></th>
<th>Proportion of QTL variance explained</th>
<th>Maximum number of haplotypes</th>
<th>Observed number of haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest marker</td>
<td>0.10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Best marker</td>
<td>0.20</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2 Marker haplotypes</td>
<td>0.15</td>
<td>4</td>
<td>3.4</td>
</tr>
<tr>
<td>4 Marker haplotypes</td>
<td>0.28</td>
<td>16</td>
<td>9.4</td>
</tr>
<tr>
<td>6 Marker haplotypes</td>
<td>0.55</td>
<td>64</td>
<td>20.8</td>
</tr>
</tbody>
</table>
LD-MAS with haplotypes

- Accuracy of estimating QTL allele effects from haplotypes:

\[
r(q, \hat{h}) = \sqrt{r(h, q) \sum_{i=1}^{n} \frac{p_i^2}{p_i + \lambda / T}}
\]

\[
\lambda = \frac{\sigma_e^2}{\sigma_h^2}
\]
LD-MAS with haplotypes

- Accuracy of estimating QTL allele effects from haplotypes:
LD-MAS with haplotypes

- Accuracy of LD-MAS with haplotypes
  - Depends on
    - Proportion of QTL variance explained by haplotypes
    - Number of haplotype effects to estimate
    - Number of phenotypic records
    - Accuracy of inferring haplotypes??
Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
  - LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
LD-MAS with the IBD approach

- MEBVs:

\[ \text{MEBV} = u + v \]
LD-MAS with the IBD approach

- MEBVs:

\[ \text{MEBV} = \hat{u} + \hat{v} \]

\[
\begin{bmatrix}
1_n'1_n & 1_n'Z & 1_n'W \\
Z'1_n & Z'Z + A^{-1}\lambda_1 & Z'W \\
W'1_n & W'Z & W'W + G^{-1}
\end{bmatrix}

\begin{bmatrix}
\mu \\
u \\
g
\end{bmatrix}

= \begin{bmatrix}
1_n'y \\
Z'y \\
W'y
\end{bmatrix}

- Where W is a matrix allocating records to QTL allele effects
LD-MAS with the IBD approach

• Has the potential to be most accurate method for LD-MAS because can capture linkage information as well
  – Particularly with sub-optimal markers densities
Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
Gene Assisted Selection

- Greatest increases in response (not limited by LD)
- Simplest, cheapest to implement in breeding program
  - No need to establish phase within families
  - Cost of discovery very high
  - Number of examples now (Dekkers 2004)
  - May become apparent that mode of inheritance is not additive
  - Eg. IGF2 mutation in pigs is imprinted (only expressed if mutated allele from father)
Marker Assisted Selection using LD

- LD-MAS with single markers
- How many QTL to use in LD-MAS?
- Bias in QTL effects
- LD-MAS with marker haplotypes
- LD-MAS with the IBD approach
- Gene assisted selection
- Optimising the breeding scheme with marker information
Optimising the breeding scheme with MAS

- Which traits
- Age at selection?
## Optimising the breeding scheme with MAS

- **Expected response from MAS**
  - Traits measured on both sexes before selection
  - Traits measured on one sex before selection
  - Traits measured after selection
  - Traits measured on relatives

<table>
<thead>
<tr>
<th>Traits measured before selection</th>
<th>Traits measured on one sex before selection</th>
<th>Traits measured after selection</th>
<th>Traits measured on relatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>Feed intake</td>
<td>Pigs born alive</td>
<td>Carcass quality</td>
</tr>
<tr>
<td>Fatness</td>
<td>Milk production</td>
<td>Fertility</td>
<td>Disease resistance (fish)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disease resistance (cattle)</td>
</tr>
</tbody>
</table>
Optimising the breeding scheme with MAS

• Which traits
• Age at selection
  - \( G = i \sigma_g / L \)
  - where \( G \) = genetic gain
  - \( i \) is the intensity of selection
  - \( r \) is the accuracy of selection
  - \( \sigma_g \) is the genetic standard deviation and
  - \( L \) is the generation length
Optimising the breeding scheme with MAS

• Age at selection
  – We have already discussed improving $r$
  – What about $L$?
• Accuracy of traditional EBVs increase as animal ages and it and its relatives acquire phenotypic data.
• But animals can be typed for markers at any age
• Gain in accuracy from markers greatest at young age.
• So if selection optimised, marker data should lead to a decrease in generation length
• Eg. in dairy cattle selected for milk production, MAS leads to greater gains if selection of yearling bulls and cows is practiced than if a traditional progeny testing system is adhered to
• Reproductive technologies?
Take home points

- Markers in LD with QTL relatively easy to use in breeding programs
- Using haplotypes may improve accuracy?
- IBD approach allows linkage information to be used as well
- **Response:** Traits measured on both sexes before selection $<<$ traits measured on one sex before selection $<<$ traits measured after selection $<<$ traits measured on relatives
- Optimal use of marker information with selection at younger ages