

# **Incorporating Molecular Genetic Information in Genetic Improvement Programs for Livestock**

Based in part on: Dekkers et al. (2001), Dekkers and Hospital (2002), and Dekkers and Settar (2003)

Substantial advances have been made in the genetic improvement of agriculturally important animal and plant populations through artificial selection on quantitative traits. Most of this selection has been on observable phenotype, without knowledge of the genetic architecture of the selected characteristics, which is treated as a black box, with no knowledge of the number of genes that affect the trait, let alone of the effects of each gene or their locations in the genome.

Despite the obvious flaws of this model, the tremendous rates of genetic improvement that have been achieved attest to the utility of the quantitative genetic approach. Nevertheless, quantitative genetic selection has several limitations: phenotype is an imperfect predictor of an individual's breeding value; phenotype may not be observed on both genders or prior to the time when selection decisions must be made; and phenotype is not very effective in resolving negative associations between genes, e.g. those caused by linkage or epistasis. The ideal situation for quantitative genetic selection is that the trait has high heritability and that the phenotype can be observed on all individuals prior to reproductive age. This ideal is hardly ever achieved, which limits the effectiveness of quantitative genetic selection.

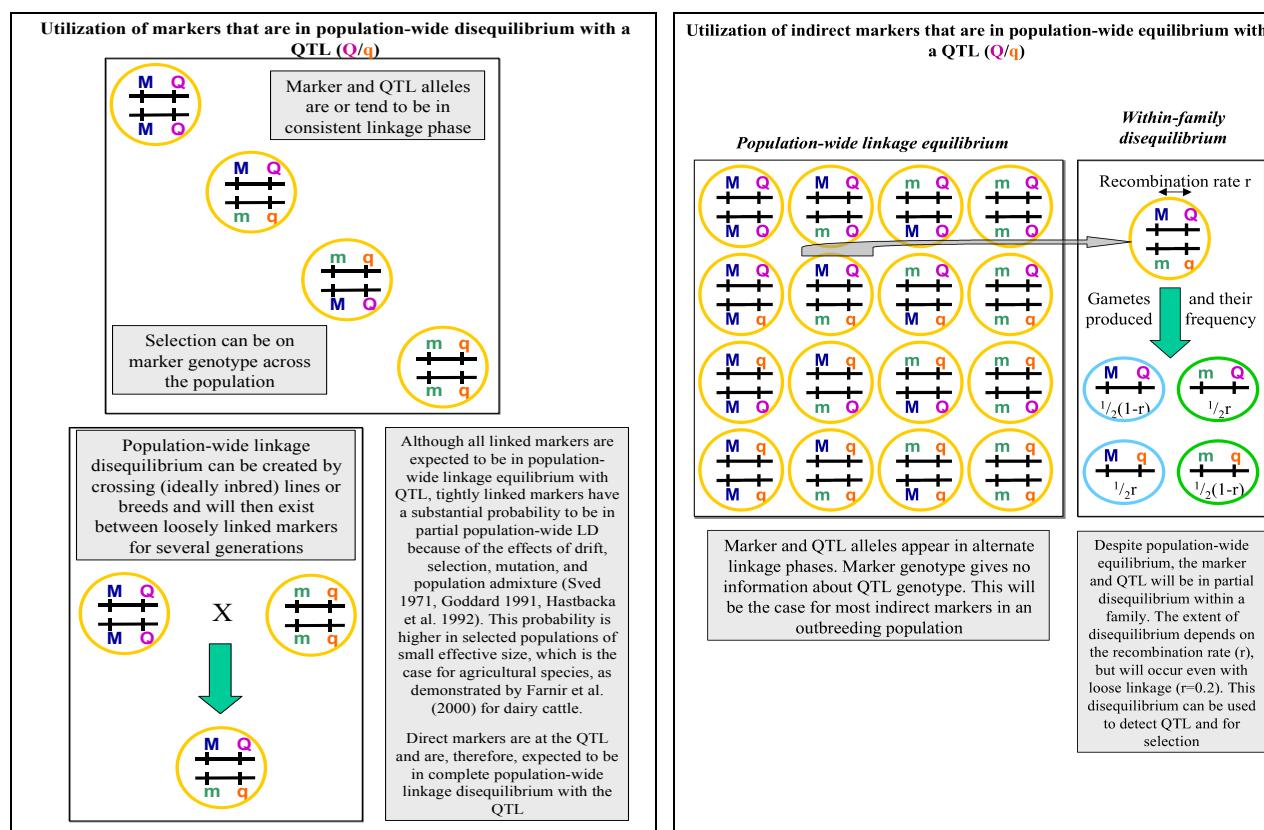
Andersson (2001) and Mauricio (2001) reviewed how molecular genetics can be used to discern the genetic nature of quantitative traits in animals and plants, respectively, by identifying genes or chromosomal regions that affect the trait — so-called quantitative trait loci or QTL. This has enabled identification and characterization of at least some of the genes that contribute to genetic variation in quantitative traits. Because DNA can be obtained at any age and on both genders, molecular genetics can alleviate some of the limitations of quantitative genetic selection, as will be discussed below. Thus, the genes and genetic markers that are being discovered can be used to enhance genetic improvement of breeding stock through marker-assisted selection. The purpose of this Chapter is to show how this information can be used to enhance genetic improvement. Emphasis will be on utilization of natural variation within a species, rather than on the introduction of new genetic variation through genetic modification, although some of the programs reviewed, such as introgression, also play an important role in the introduction of transgenes into breeding populations (see e.g. Gama et al. 1992).

<h3>Applications of Molecular Data</h3> <ul style="list-style-type: none"> <li>➢ Parental identification / verification</li> <li>➢ Traceability</li> <li>➢ Evaluation of Genetic diversity</li> <li>➢ Introgression of desirable genes Marker-Assisted Introgression (MAI)</li> <li>➢ Enhance selection within outbred populations Marker-Assisted Selection (MAS)</li> </ul>	<h3>Use of Molecular Data in Selection</h3> <pre> graph TD     A[Unknown genes Molec. genetics] --&gt; B[Identified or marked QTL]     C[Phenotypic data EBV] --&gt; B     D[Genotypic data] --&gt; B     E[Selection strategy] --&gt; B   </pre>						
<h3>Benefit from use of molecular data</h3> <ul style="list-style-type: none"> <li>➢ Higher <math>h^2</math> than phenotypic data</li> <li>➢ Expressed in both sexes</li> <li>➢ Expressed at early age (embryo stage)</li> <li>➢ Explains within-family variation       <ul style="list-style-type: none"> <li>➢ traces Mendelian sampling terms</li> </ul> </li> </ul> $g_x = \frac{1}{2}g_{sire} + \frac{1}{2}g_{dam} + RA_{sire} + RA_{dam}$ <p>Information providing records</p> <table border="1"> <tr> <td>Ancestral records</td> <td>Half-sib records</td> <td>Own phenotype Progeny phenotype Marker/genotypic data on X</td> </tr> <tr> <td>Full-sib records</td> <td></td> <td></td> </tr> </table>	Ancestral records	Half-sib records	Own phenotype Progeny phenotype Marker/genotypic data on X	Full-sib records			<h3>Factors Affecting Extra Response from MAS</h3> <ul style="list-style-type: none"> <li>➢ Effects of identified QTL (% of genetic variance)</li> <li>➢ Recombination rates between markers and QTL</li> <li>➢ Effectiveness of Phenotypic Selection       <ul style="list-style-type: none"> <li>➢ Heritability</li> <li>➢ Restrictions on phenotyping (measurement)           <ul style="list-style-type: none"> <li>➢ in one sex only (sex-limited traits)               <ul style="list-style-type: none"> <li>➢ EBV based on relatives for one sex</li> </ul> </li> <li>➢ late in life (after selection)               <ul style="list-style-type: none"> <li>➢ EBV based on relatives</li> </ul> </li> <li>➢ not on live animal (meat quality traits)               <ul style="list-style-type: none"> <li>➢ EBV from relatives, reduced intensity</li> </ul> </li> <li>➢ difficult to measure (disease traits)</li> </ul> </li> </ul> </li> </ul>
Ancestral records	Half-sib records	Own phenotype Progeny phenotype Marker/genotypic data on X					
Full-sib records							
<h3>Traits</h3> <ul style="list-style-type: none"> <li>• Single gene traits       <ul style="list-style-type: none"> <li>• Genetic defects/disorders</li> <li>• Appearance</li> </ul> </li> <li>• Quantitative traits</li> </ul>	<h3>Quantitative Traits</h3> <table border="1"> <thead> <tr> <th>Ease of QTL detection</th> <th>Potential gain from MAS/GAS</th> </tr> </thead> <tbody> <tr> <td> <ul style="list-style-type: none"> <li>• Routine recorded           <ul style="list-style-type: none"> <li>• both sexes</li> <li>• sex-limited</li> <li>• late in life</li> </ul> </li> <li>• Difficult to record           <ul style="list-style-type: none"> <li>• feed intake</li> <li>• product quality</li> </ul> </li> <li>• Unrecorded / low <math>h^2</math> <ul style="list-style-type: none"> <li>• disease resistance</li> </ul> </li> </ul> </td> <td> </td> </tr> </tbody> </table>	Ease of QTL detection	Potential gain from MAS/GAS	<ul style="list-style-type: none"> <li>• Routine recorded           <ul style="list-style-type: none"> <li>• both sexes</li> <li>• sex-limited</li> <li>• late in life</li> </ul> </li> <li>• Difficult to record           <ul style="list-style-type: none"> <li>• feed intake</li> <li>• product quality</li> </ul> </li> <li>• Unrecorded / low <math>h^2</math> <ul style="list-style-type: none"> <li>• disease resistance</li> </ul> </li> </ul>			
Ease of QTL detection	Potential gain from MAS/GAS						
<ul style="list-style-type: none"> <li>• Routine recorded           <ul style="list-style-type: none"> <li>• both sexes</li> <li>• sex-limited</li> <li>• late in life</li> </ul> </li> <li>• Difficult to record           <ul style="list-style-type: none"> <li>• feed intake</li> <li>• product quality</li> </ul> </li> <li>• Unrecorded / low <math>h^2</math> <ul style="list-style-type: none"> <li>• disease resistance</li> </ul> </li> </ul>							

Molecular genetic analyses of quantitative traits lead to the identification two broadly different types of genetic loci that can be used to enhance genetic improvement programs: causal mutations and presumed non-functional genetic markers that are linked to QTL (indirect markers). Causal mutations for quantitative traits are hard to find, difficult to prove, and few examples are available (Andersson 2001). Non-functional or anonymous polymorphisms are abundant across the genome and their linkage with QTL can be established by evidence of empirical associations of marker genotypes with trait phenotype. Two approaches are used to identify indirect markers (Andersson 2001): directed searches using candidate gene approaches in unstructured populations (Rothschild and Soller 1997); and undirected genome-wide searches

in specialized populations, such as F2 crosses or half-sib family populations. Because candidate gene markers focus on polymorphisms within a gene that is postulated to affect the trait, they are often tightly linked to the QTL. A candidate gene marker can represent the functional polymorphism, although this is difficult to prove (Andersson 2001). Genome scans, on the other hand, only identify regions of chromosomes that affect the trait. The length of these regions is typically 10 to 20 cM, but the exact position and number of QTL within the region is unknown.

Whereas a causative polymorphisms give direct information about genotype for the QTL, use of indirect markers for QTL mapping and for selection is based on existence of linkage or gametic phase disequilibrium (LD) between the marker and the QTL. Marker-QTL LD can exist at the population level but always exists within families, even between loosely linked loci. Although two loci are expected to be in population-wide equilibrium in large random-mating populations, partial population-wide LD can exist by chance between tightly linked loci in breeding populations that are under selection. Population-wide LD can also be created by crossing lines or breeds. Although LD will then exist even between loosely linked loci, this LD will erode rapidly over generations. Indirect markers that are identified using the candidate gene marker approach are expected to be in substantial LD with the QTL in which they reside. Unless the functional polymorphism has been identified, however, linkage phase of a candidate gene marker with the functional variant can differ from one population to the next and must, therefore, be assessed in the population in which it will be used. Although more abundant and extensive, within-family LD is more difficult to use because linkage phases between the markers and QTL will not be the same in all families and must, therefore, be assessed on a within-family basis.



<h3>Three Types of Molecular Information</h3> <p>1) Genotype for functional gene ➤ polymorphism = causative mutation</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>BB</td><td>Bb</td><td>bb</td></tr> </table> <p>2) Genotype for a direct marker ➤ polymorphism is in population-wide linkage disequilibrium with causative mutation</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>Marker - QTL</td></tr> <tr><td>haplotypes present</td></tr> <tr><td>MB</td><td>mb</td><td>MB</td><td>mb</td></tr> <tr><td>++</td><td>++</td><td>XX</td><td>XX</td></tr> </table> <p>3) Genotype for linked genetic markers ➤ polymorphism in linkage equilibrium across population</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr><td>Marker - QTL</td></tr> <tr><td>haplotypes present</td></tr> <tr><td>MB</td><td>mb</td><td>Mb</td><td>mb</td></tr> <tr><td>++</td><td>++</td><td>++</td><td>++</td></tr> </table> <p><math>E(\mu_{MM}) = E(\mu_{Mm}) = E(\mu_{mm})</math></p> <p>Use within-family disequilibrium</p>	BB	Bb	bb	Marker - QTL	haplotypes present	MB	mb	MB	mb	++	++	XX	XX	Marker - QTL	haplotypes present	MB	mb	Mb	mb	++	++	++	++	<h3>Use of within-family linkage disequilibrium for QTL mapping and MAS</h3> <p>Sire: <math>\begin{array}{c} \leftarrow r \\ M-B \\ \hline m-b \end{array}</math></p> <p>Random dams: <math>\begin{array}{c} ? ? \\ ? ? \\ ? ? \end{array}</math></p> <p>M progeny: <math>\begin{array}{c} M-B \mu^{+/r}\alpha \\ ? ? \frac{1}{2}(1-r) \\ M-b \mu^{-/r}\alpha \\ ? ? \frac{1}{2}r \end{array}</math></p> <p>m progeny: <math>\begin{array}{c} m-b \mu^{-/r}\alpha \\ ? ? \frac{1}{2}(1-r) \\ m-B \mu^{+/r}\alpha \\ ? ? \frac{1}{2}r \end{array}</math></p> <p>Non-recombinants</p> <p>Recombinants</p> <p>Average <math>\mu^{+/r}(1-2r)\alpha</math></p> <p><math>\mu^{-/r}(1-2r)\alpha</math></p> <p>Contrast <math>\mu_M - \mu_m = (1-2r)\alpha</math></p>
BB	Bb	bb																						
Marker - QTL																								
haplotypes present																								
MB	mb	MB	mb																					
++	++	XX	XX																					
Marker - QTL																								
haplotypes present																								
MB	mb	Mb	mb																					
++	++	++	++																					
<p>Linkage Disequilibrium can persist many generations for tightly linked loci</p> <p>Measure of Disequilibrium <math>= D_{M,B} = freq(MB) - freq(M)*freq(B)</math></p> <p>Random mating in large population: <math>D_{M,B}(t+1) = (1-r) D_{M,B}(t) = (1-r)^t D_{M,B}(0)</math></p> <p><math>D_{M,B}(t)</math></p> <p>Generation</p>																								

The use of molecular genetics in selection programs rests on the ability to determine the genotype of individuals for causal mutations or indirect markers using DNA analysis. This information is then used to assess the genetic value of the individual, which can be captured in a molecular score that can be used for selection. This removes some of the limitations of quantitative genetic selection discussed above.

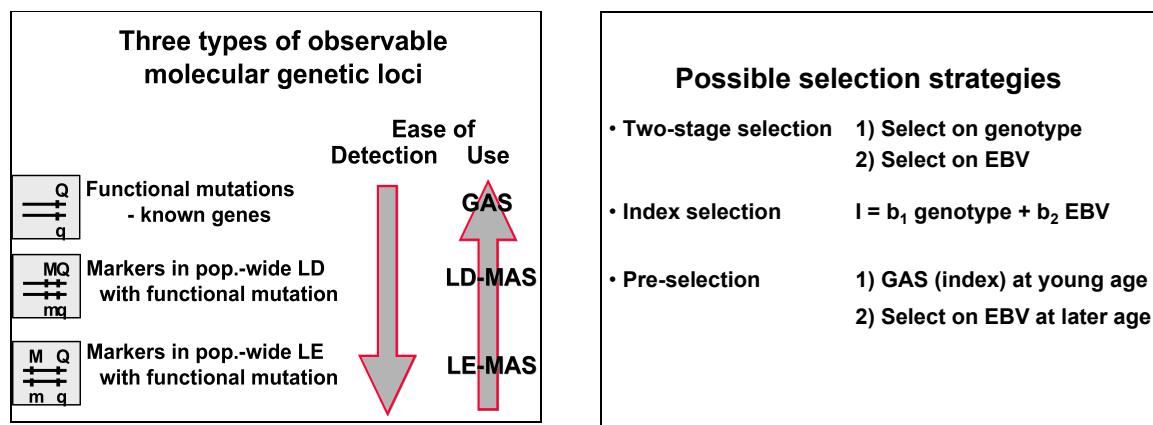
It is clear that the use of molecular data for genetic improvement would be most effective if the genetic architecture of a quantitative trait was completely transparent such that we knew the number, positions, and effects of all genes involved. In that case, the process of selection would be reduced to a simple ‘building block’ problem (genotype building) of selection and mating to create individuals with the right combination of alleles at each QTL. However, this situation is far from reality and may never be achieved; although advances in molecular genetics have been able to partially dissect the black box of quantitative traits, the information provided by molecular data is far from complete, for three main reasons. First, in most cases only a limited number of genes that affect the trait has been identified, albeit the ones with the largest effects. Nevertheless, a substantial part of the black box remains obscure and selection exclusively on genotype for identified QTL would not result in maximum response to selection. Instead, selection on molecular score must be combined with selection on phenotype, which reflects the collective action of all genes, including those that have not been identified. Second, with indirect

markers, selection is not directly on the QTL, but on the marker, via LD. As LD erodes in the course of the selection program due to recombination, efficiency of selection is reduced. Third, for both causal and indirect markers, the effects of the QTL must be estimated empirically on the basis of statistical associations between markers and phenotype. Estimation requirements are particularly high for markers that are not in population-wide LD and for which within-family LD must be used. In that case, marker-QTL linkage phase and effects must be estimated on a within-family basis. Thus, the use of molecular information does not remove the need for phenotypic information and, therefore, suffers to some degree from the same limits as quantitative genetic selection.

Despite the limitations outlined above, molecular genetic information can be used to enhance several breeding strategies through what is broadly referred to as Marker-Assisted Selection (MAS). All strategies for MAS are based on the use a molecular score, although the composition of this score differs from application to application. In addition to those described below, the application of molecular data in genetic programs includes their use for parentage verification or identification (for example, when mixed semen is used in artificial insemination) and in genetic conservation programs to identify unique genetic resources and quantify genetic diversity.

The type of genetic information that is available, and its association with the functional mutation (population-wide LD or within-family LD), has important consequences for the use of molecular information in selection programs. On this basis, the following three types of selection programs using molecular information can be distinguished:

- Gene-assisted selection (GAS) – selection based on the functional mutation for the QTL
- Marker-assisted selection based on population-wide LD (LD-MAS) – selection based on markers or marker haplotypes that are in population-wide disequilibrium with the QTL
- Marker-assisted selection based on within-family LD (LE-MAS) – selection based on markers or marker haplotypes that are in population-wide equilibrium with the QTL but in LD with the QTL on a within-family basis.



For each of the three types of selection (GAS, LD-MAS, and LE-MAS), there are two basic strategies for combining the molecular information with phenotypic information in a selection strategy:

- 1) Two-stage selection, in which selection is on the molecular score in the first stage and on phenotype or a (polygenic) EBV in the second stage
- 2) Index selection, in which selection is on an index of molecular score and phenotypic information.

In addition, molecular information could be used primarily for pre-selection of young animals for further testing.

Methods to derive indexes that combine molecular and phenotypic information will be presented in the next section, followed by methods to predict responses to selection with presence of QTL of large effects. We will then compare alternative selection strategies for utilization of QTL information between and within breeds, and finish with an economic analysis of MAS and opportunities for the redesign of breeding programs to more fully capture the benefits of MAS.

## 12.1 Including QTL Information in Estimated Breeding Values

When distinguishing QTL that have been mapped from other background genes that affect the trait, which will be referred to as polygenes, the genetic value  $g_i$  of an individual  $i$ , can be partitioned into the sum of genetic values at the QTL,  $g_{Qi}$ , and the sum of genetic values at polygenes,  $g_{pi}$ :

$$g_i = g_{Qi} + g_{pi}$$

Molecular genetic information provides information that can be used to estimate  $g_{Qi}$ , whereas an individual's phenotype provides information on the collective effect of all genes. Unless all QTL that affect the trait have been identified, selection on QTL must be combined with selection on phenotypic information, to ensure simultaneous improvement of both  $g_{Qi}$  and  $g_{pi}$ . Lande and Thompson (1990) suggested that QTL and phenotypic information should be combined in an index of the following form:

$$I_i = b_Q \hat{g}_{Qi} + b_P P_i$$

where  $\hat{g}_{Qi}$  is the molecular score for individual  $i$ , i.e. the individual's estimated breeding value for the QTL,  $P_i$  is the individual's phenotype, and  $b_Q$  and  $b_P$  are index weights. The molecular score,  $\hat{g}_{Qi}$ , can be computed as the sum over QTL or markers of estimates of effects on phenotype based on the individual's QTL or marker genotypes. An example is in Table 12.1.

Lande and Thompson (1990) showed that index weights could be derived by standard selection index theory, given the proportion of genetic variance explained by the QTL or markers ( $q = \sigma_Q^2 / h^2 \sigma_P^2$ ), and the (total) heritability of the trait ( $h^2 = \sigma_g^2 / \sigma_P^2$ ):

$$\begin{bmatrix} b_Q \\ b_P \end{bmatrix} = \mathbf{P}^{-1} \mathbf{G} = \begin{bmatrix} \sigma_Q^2 & \sigma_Q^2 \\ \sigma_Q^2 & \sigma_P^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_Q^2 \\ \sigma_g^2 \end{bmatrix} = \begin{bmatrix} \frac{1-h^2}{1-qh^2} \\ h^2 \frac{(1-q)}{1-qh^2} \end{bmatrix}$$

Thus, the relative weight on the molecular score relative to phenotype is:  $\frac{b_Q}{b_P} = \frac{\sqrt{h^2} - 1}{1 - q}$

Table 12.1. Example of the calculation of molecular score and index of phenotype and molecular score with 3 additive QTL with allele substitution effects (allele A vs. B) of +10, +5, and -10 for QTL 1, 2, and 3, respectively. The QTL jointly explain 50% of the genetic variance for a trait with heritability 0.5. Resulting index weights on molecular score and phenotype are  $2/3$  and  $1/3$ , respectively (after J. Holland, 1998).

Animal	QTL 1		QTL 2		QTL 3		Molecular score	Phenotype	Index value
	Genotype	Value	Genotype	Value	Genotype	Value			
1	AA	10	AA	5	AA	-10	5	35	15.0
2	AA	10	AA	5	BB	10	25	-10	13.3
3	AB	0	BB	-5	AB	0	-5	-15	-8.3
4	AB	0	BB	-5	AA	-10	-15	15	-5.0
5	BB	-10	AA	5	AB	0	-5	25	5.0

Index of QTL and own phenotype (Lande & Thompson, 1990 Genetics 124:743)	Index of QTL and own phenotype (Lande & Thompson, 1990 Genetics 124:743)
<p>Index: <math>I = b_Q g_Q + b_P P</math></p> $\begin{bmatrix} b_Q \\ b_P \end{bmatrix} = P^{-1} G = \begin{bmatrix} \sigma_Q^2 & \sigma_Q \sigma_P \\ \sigma_Q \sigma_P & \sigma_P^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_Q^2 \\ \sigma_g^2 \end{bmatrix}$ <p>Accuracy <math>r_{g,I} = (b'G)^{-1}/\sigma_g</math></p>	<p>Index of QTL and own phenotype (Lande &amp; Thompson, 1990 Genetics 124:743)</p> <p><math>g = g_Q + g_{pol}</math>      <math>g_Q = \text{QTL/marker BV}</math>      <math>g_{pol} = \text{Polygenic BV}</math></p> <p><math>\sigma_g^2 = \text{Total genetic var.} = \sigma_Q^2 + \sigma_{pol}^2</math>  <math>q = \text{fraction of genetic variance due to QTL/marker} = \sigma_Q^2/\sigma_g^2</math>  <math>h^2 = \text{total heritability} = \sigma_g^2/\sigma_p^2</math></p> <p>Index: <math>I = b_Q g_Q + b_P P</math></p> <p>Selection index theory:</p> <p><math>\rightarrow b_Q = (1-h^2)/(1-qh^2) \quad b_P = h^2(1-q)/(1-qh^2)</math>  <math>b_Q/b_P = (1/h^2 - 1)/(1-q)</math>  <math>\text{Efficiency} = r_{g,I}/r_{g,P} = [(q/h^2) + (1-q)^2/(1-h^2q)]^{1/2}</math></p> <p>Can be expanded to multiple QTL and multiple phenotypic records using standard selection index theory</p>

Example relative weights are in Table 12.2, which shows that the index gives more weight to the molecular score as heritability decreases and as the proportion of variance explained by the QTL increases.

Table 12.1 also gives index values for the example animals. This illustrates that different selection decisions would be made based on molecular score alone, based on phenotype alone, and based on the index.

The Lande and Thompson (1990) formulation of the index is easily extended to situations where indexes of phenotypes of relatives are used. Indexes can also be extended to multiple-trait situations.

Table 12.2. Index weight on molecular score relative to phenotype ( $b_Q/b_P$ ) for different heritabilities and proportions of genetic variance explained by the QTL (after J. Holland, 1998).

Heritability ( $h^2$ )	Proportion of genetic variance explained by QTL ( $q$ )				
	0.10	0.25	0.50	0.75	1.00
0.10	10	12	18	36	Total weight
0.25	3.33	4	6	12	Total weight
0.50	1.11	1.33	2	4	Total weight
0.75	0.37	0.44	0.67	1.33	Total weight
1.00	0	0	0	0	Either

It is useful to note that the above index can be reparameterized into an equivalent index of molecular score and phenotype adjusted for the molecular score as follows:

$$I'_i = b'_Q \hat{g}_{Qi} + b'_P P'_i$$

Where  $P'_i = P_i - \hat{g}_{Qi}$ . Using selection index theory and defining polygenic heritability as the

heritability of phenotype adjusted for molecular score: 
$$h_{pol}^2 = \frac{\sigma_g^2 - \sigma_Q^2}{\sigma_p^2 - \sigma_Q^2} = \frac{h^2(1-q)}{1-qh^2}$$

weights for this index can then be derived to be independent of  $r$  and equal to:  $b'_Q = 1$  and

$$b'_P = h_{pol}^2 : \quad \begin{bmatrix} b'_Q \\ b'_P \end{bmatrix} = \mathbf{P}^{-1} \mathbf{G} = \begin{bmatrix} \sigma_Q^2 & 0 \\ 0 & \sigma_p^2 - \sigma_Q^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_Q^2 \\ \sigma_g^2 - \sigma_Q^2 \end{bmatrix} = \begin{bmatrix} 1 \\ h_{pol}^2 \end{bmatrix}$$

Thus, the resulting index is:  $I'_i = \hat{g}_{Qi} + h_{pol}^2 P'_i$

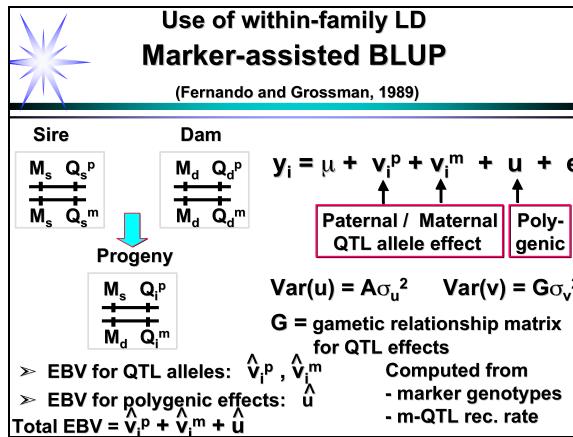
One important advantage of index  $I'$  over index  $I$  is that its index weights remain constant over generations, whereas weights for index  $I$  must be updated each generation as QTL frequencies, and therefore the proportion of genetic variance explained by the QTL, change. This index also allows easy extension to indexes based on BLUP EBV. To see this, note that the second term in this index,  $h_{pol}^2 P'_i$ , represents the individual's estimated breeding value for polygenes,  $\hat{g}_{pi}$ , based on own phenotype adjusted for the QTL. This index can be expanded to BLUP EBV from a model that includes QTL or markers as a fixed or random effect (see Fernando and Grossman, 1989, for methodology to include marked QTL as random effects in a BLUP animal model). Such models result in estimates of molecular scores,  $\hat{g}_{Qi}$ , and EBV for polygenic effects,  $\hat{g}_{pol,i}$ , with accuracy  $r_{pol}$ . Index weights for combining these two estimates, realizing that the variance of polygenic EBV is equal to  $r_{pol}^2 \sigma_{pol}^2$ , where  $\sigma_{pol}^2 = h_{pol}^2 (\sigma_p^2 - \sigma_Q^2)$  is the polygenic variance, can be derived as:

$$\begin{bmatrix} b'_Q \\ b'_P \end{bmatrix} = \mathbf{P}^{-1} \mathbf{G} = \begin{bmatrix} \sigma_Q^2 & 0 \\ 0 & r_{pol}^2 \sigma_{pol}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_Q^2 \\ r_{pol}^2 \sigma_{pol}^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus the index is:

$$I'_i = \hat{g}_{Q_i} + \hat{g}_{pol,i}$$

<p><b>Index of QTL and own phenotype</b> Alternative (but equivalent) formulation</p> <p><math>P^*</math> = phenotype adjusted for QTL/marker</p> <p><math>P^* = P - g_Q</math></p> <p><math>\sigma_{P^*}^2 = \sigma_P^2 - \sigma_Q^2</math></p> <p><math>h_{pol}^2 = \text{polygenic heritability} = \sigma_{pol}^2/\sigma_P^2</math></p> <p>Index: <math>I = b_Q g_Q + b_{P^*} P^*</math></p> <p><b>Selection index theory:</b></p> <p><math>\rightarrow b_Q = 1 \quad b_{P^*} = h_{pol}^2</math></p> <p><math>I = g_Q + h_{pol}^2 P^*</math></p> <p>↑ overall EBV    ↑ QTL EBV    ↑ Polygenic EBV</p>	<p><b>Index of QTL and Phenotypic information</b> Generalization to BLUP EBV</p> <p><math>\hat{g}_Q = \text{EBV based on (multiple) markers/QTL}</math> <math>= \sum \hat{g}_{Qi}</math> for multiple markers/QTL</p> <p><math>\hat{g}_{pol} = \text{BLUP for polygenic BV}</math></p> <p>Estimates can be obtained from BLUP-QTL animal models (Fernando &amp; Grossman, 1989 Genet. Sel. Evol. 21:467)</p> <p><math>I = \hat{g}_Q + \hat{g}_{pol}</math></p>
--	--



If the phenotypic EBV is from a regular animal model and not from a model that includes the marker or QTL as separate effects, derivation of the index can only be approximated by correcting the EBV for effects of the QTL. This could be done by regressing the regular EBV,  $\hat{g}_i$ , on the molecular score (or QTL genotype(s)) using:

$$\hat{g}_i = \beta \hat{g}_{Q_i} + e_i$$

Residuals from this model then provide approximate estimates of polygenic EBV, i.e.  $\hat{g}_{pol,i} \approx \hat{e}_i$ , which can be used in the index described above. Note that, although  $\hat{g}_{Q_i}$  may represent an unbiased estimate of the QTL effects, the estimate of the regression coefficient  $\beta$  will be less than 1. The reason is that when estimating EBV  $\hat{g}_i$ , all effects, including the QTL effects are regressed back toward zero. In theory, the extent of regression can be approximated by the square of the accuracy of the EBV, i.e.  $r^2$ . This can be most readily seen for EBV based on own phenotype alone, in which case:

$$\hat{g}_i = h^2 P_i = h^2 \hat{g}_{Q_i} + h^2 (P_i - \hat{g}_{Q_i})$$

Thus in this case the regression factor is  $\beta = h^2 = r^2$  since  $r = h$  for selection on own phenotype. This relationship  $\beta = r^2$  is, however, only an approximation when phenotypic information from relatives contributes to the EBV because, the extent of regression of phenotypic information from relatives is not equal to  $r^2$ . This is most easiest seen from table 4.1 in Chapter 4, when comparing index coefficients  $\mathbf{b}$  to the square of the accuracy  $r_{HI}$ .

In addition, if animals with EBV with different accuracy are included in the analysis, a single regression coefficient will not suffice. This could be accommodated by a weighted least squares analysis or by first de-regressing EBV. These problems are, however, all circumvented when the marker information is included directly in the genetic evaluation model, which is the preferred method.

It is useful to note that selection based on own phenotype (without molecular information) can also be written as selection on an index of breeding values for the QTL and polygenes by noting that selection on  $P_i$  is equivalent to selection on  $h_p^2 P_i$ , which can be written as

$$\hat{g}_i = h_{pol}^2 P_i = h_{pol}^2 g_{Q_i} + h_{pol}^2 P'_i = h_{pol}^2 g_{Q_i} + \hat{g}_{pol,i}.$$

Thus, with phenotypic selection, the emphasis on the molecular score relative to the EBV for polygenes is equal to the polygenic heritability,  $h_{pol}^2$ , instead of 1 as in MAS. Similarly, for more complex EBV based on phenotypic records, as shown earlier, the EBV can be approximated by:

$$\hat{g}_i \approx r^2 g_{Q_i} + \hat{g}_{pol,i}$$

and the implicit weight on the molecular score is approximately equal to the square of accuracy,  $r^2$ .

## 12.2 Predicting Response to Selection with QTL Information

Consider the previously derived selection index of molecular score and own phenotype, when the molecular score explains a fraction  $q$  of the additive genetic variance:

$$I_i = b_Q \hat{g}_{Q_i} + b_P P_i$$

The accuracy of this index and response to selection can be derived by standard selection index theory (Chapter 4) as:

$$\begin{aligned} r_{g,I} &= \sqrt{\frac{\mathbf{b}' \mathbf{G}}{\sigma_g^2}} = \sqrt{\left[ \frac{1-h^2}{1-qh^2} \quad h^2 \frac{(1-q)}{1-qh^2} \right] \left[ q \right]} \\ &= \sqrt{\frac{q - 2qh^2 + h^2}{1 - qh^2}} = \sqrt{q + h^2 \frac{(1-q)^2}{1 - qh^2}} \end{aligned}$$

Similarly for the alternate index parameterization:

$$I'_i = b_Q' \hat{g}_{Q_i} + b_P' P'_i$$

$$\text{and } r_{g,I'} = \sqrt{\frac{\mathbf{b}' \mathbf{G}}{\sigma_g^2}} = \sqrt{\left[ 1 \quad h_{pol}^2 \right] \left[ \frac{q}{1-q} \right]} = \sqrt{q + h_{pol}^2(1-q)}$$

Using  $h_{pol}^2 = \frac{h^2(1-q)}{1-qh^2}$  it can easily be shown that  $r_{g,I'} = r_{g,I}$ , i.e. the two indexes are equivalent

Assuming equal selection in males and females, with selection intensity  $i$ , response to selection can be predicted as:

$$R_{MAS} = i r_{g,I} \sigma_g$$

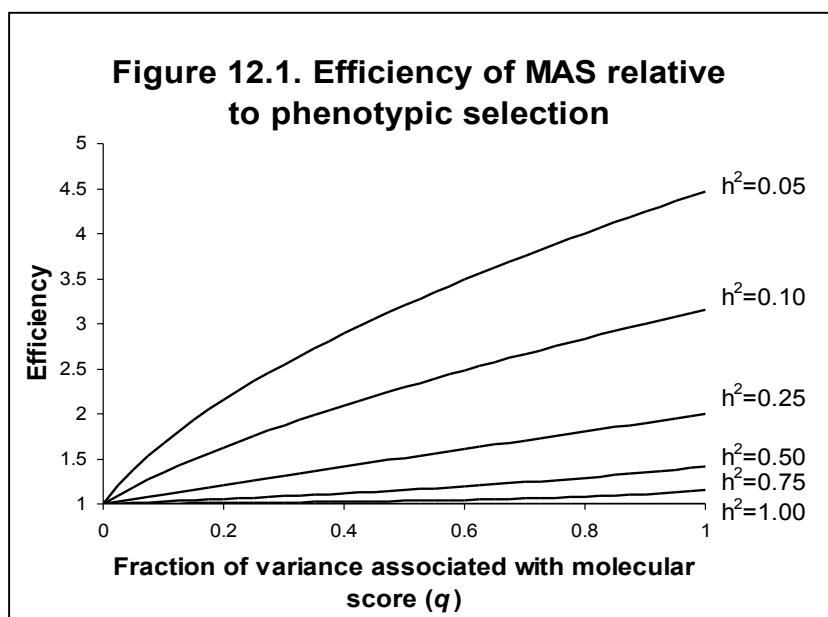
Response to phenotypic selection without QTL information is:  $R_P = i r_{g,P} \sigma_g$   
With  $r_{g,P} = h$ , the efficiency of selection using marker information, defined as response to MAS relative to response without marker information, is given by:

$$E = \frac{R_{MAS}}{R_P} = \frac{r_{g,I}}{r_{g,P}} = \sqrt{\frac{q}{h^2} + \frac{(1-q)^2}{1-qh^2}}$$

An equivalent equation can be derived using the alternate index  $I'$ :

$$E = \frac{R_{MAS}}{R_P} = \frac{r_{g,I'}}{r_{g,P}} = \frac{1}{h} \sqrt{q + h_{pol}^2(1-q)}$$

Figure 12.1 shows the impact of heritability and proportion of variance explained by the molecular score on efficiency of MAS. This Figure shows that MAS will be most beneficial for traits with low heritability and when the molecular score explains a large proportion of the genetic variance.



Similar procedures, using selection index theory, can be used to derive accuracy and efficiency of MAS for more complex EBV that use information from relatives and/or multiple traits (see Lande and Thompson, 1989). Efficiency of such indexes is approximately equal to those illustrated in Figure 12.1, but with  $h^2$  replaced by accuracy squared,  $r^2$ . This shows that, in general, for a given proportion of variance explained by QTL, MAS will be most efficient for cases in which regular selection is relatively ineffective. This includes traits with low heritability, sex-limited traits, traits that are observed late in life (after selection), and traits that require sacrificing the animal to observe phenotype (e.g. carcass quality traits).

## 12.3 Reparameterized selection index formulation of MAS

Methods proposed by Lande and Thompson (1990) are not in a form that is suitable for use in standard selection index procedures, such as the program SelAction (Rutten et al. 2002), which require specification of phenotypic traits with heritabilities, standard deviations and phenotypic and genetic correlations as input parameters. The Lande and Thompson formulation also does not easily allow assessment of the effect of having marker information on relatives in addition to, or instead of on the individual itself, which is important for a cost-benefit analysis of genotyping specific individuals or groups of individuals. The purpose of what follows is to present a reparameterization of the Lande and Thompson index that does accommodate these issues. The main concept is to model marker-based EBV as a separate correlated genetic trait with a heritability equal to one. The justification for using this approach is provided in the following.

### 12.3.1 Modeling marker-based EBV as a genetic trait with heritability one.

When individuals are genotyped for markers that are in LD with QTL across the population (LD markers), marker effects can be estimated across families from an analysis of phenotype and marker genotype data obtained from the population. Here, it will be assumed that estimates are obtained from fitting markers or haplotypes as random rather than fixed effects, i.e. they represent BLUP EBV, similar to what is obtained for polygenic EBV. Such a model was described by Meuwissen et al. (2001) for genomic selection, with estimates derived from phenotypic data and high-density SNP genotypes in a single generation and used for several subsequent generations. When based on multiple regions of the genome, or on all regions of the genome, as with genomic selection, estimates of the breeding value of an individual based on the effect of its marker genotypes on performance can be computed as the sum of EBV across alleles or haplotypes for each genomic region  $j$  as:

$$\hat{Q} = \sum_j (\hat{g}_j^{\text{pat}} + \hat{g}_j^{\text{mat}})$$

where  $\hat{g}_j^{\text{pat}}$  and  $\hat{g}_j^{\text{mat}}$  are the BLUP estimates of the effects of the paternal and maternal marker or QTL allele for interval  $j$ , or of the paternal and maternal haplotypes. When same estimates of allele or haplotype effects are used for several generations, the marker-based EBV of a progeny can be written as the average of the marker-based EBV of its parents plus the sum of deviations for alleles or haplotypes that are transmitted to the progeny:

$$\hat{Q}_{\text{progeny}} = \frac{1}{2}\hat{Q}_{\text{sire}} + \frac{1}{2}\hat{Q}_{\text{dam}} + \left( \sum_j \hat{g}_{ij}^{\text{pat}} - \frac{1}{2}\hat{Q}_{\text{sire}} \right) + \left( \sum_j \hat{g}_{ij}^{\text{mat}} - \frac{1}{2}\hat{Q}_{\text{dam}} \right).$$

Note that this is equivalent to the Mendelian genetic model that is assumed for polygenic breeding values, with the latter two terms representing the Mendelian sampling terms. When based on multiple QTL regions and markers, these Mendelian sampling terms will approximately follow a normal distribution, which is what is assumed for polygenic traits. Further, because an individual's marker-based EBV is fixed conditional on marker genotypes and previously derived estimates of marker effects, it has no non-genetic random residual term. Thus, it can be observed without error based on the marker genotypes of the individual. Note that this does assume that (if needed) parental origin of alleles or haplotypes can be determined without error and that estimates of marker or haplotype effects remain consistent across several generations. Thus, although marker-based EBV represent estimates, they can be viewed and modeled as a genetic

trait that is inherited in a polygenic manner and that can be observed on individuals without error (i.e. no environmental effect).

Modeling marker-based EBV as a genetic trait with heritability one provides a convenient basis for including marker information as a correlated trait in selection index calculations. It also allows marker information from relatives to be incorporated in a natural way. With heritability equal to one, correlations between marker-based EBV on relatives will be equal to the additive genetic relationship between relatives. Additive genetic relationships quantify the correlation between additive effects of relatives for single and across multiple loci (Falconer and McKay, 1998) and can, therefore, also be used to model correlations of marker-based EBV between relatives, and to predict marker-based EBV of an individual based on the marker-based EBV of relatives.

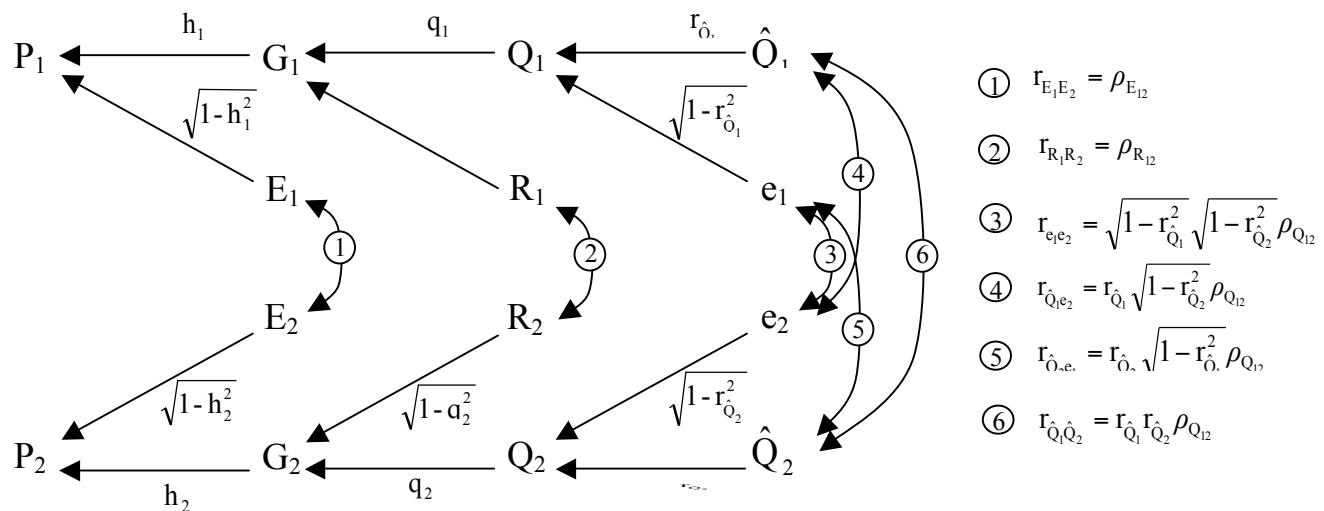
Incorporating marker-based EBV as a correlated genetic trait in selection index calculations does, however, require knowledge of correlations with phenotypic information. In what follows, these correlations will first be derived for an individual trait and then extended to multiple traits. The general theory that will be developed herein applies to cases in which a selected group of markers is used for selection, e.g., based on prior QTL studies. Use of genomic selection will, however, result in several simplifying assumptions, because of the random associations of markers with QTL across the genome, and this will be explored as a special case.

### 12.3.2 Single trait MAS index formulation

With availability of LD markers, the total additive genetic value of trait 1 ( $G_1$ ) can be partitioned into genetic effects that are correlated with markers through LD ( $Q_1$ ) and residual genetic effects ( $R_1$ ) that are independent of markers. Note that, in addition to QTL that are not in LD with markers,  $R$  also includes effects resulting from incomplete LD of QTL with markers. This partitioning results in the following model for phenotypes,

$$P_1 = G_1 + E_1 = Q_1 + R_1 + E_1$$

where  $E_1$  represents random environmental effects.



A path diagram of this model is in **Figure 1**.

Note that when markers used for marker-assisted genetic evaluation are randomly located across the genome, as would be the case for genomic selection, effects included in Q and R represent a random partitioning of QTL effects into effects that are associated with markers through LD (= Q) and effects that are independent of marker genotypes (= R).

Let  $h_1^2$  denote total heritability for trait 1, and  $q_1^2$  the proportion of genetic variance contributed by  $Q_1$ . The proportion  $q_1^2$  depends on the genetic variance contributed by QTL that are in LD with markers and the extent of LD between markers and QTL. For an individual QTL linked to a single marker,  $q_1^2$  is equal to the product of LD between the marker and the QTL, as measured by  $r^2$  (Hill and Robertson, 1968), and the proportion of total genetic variance that is contributed by the QTL. When the QTL is in LD with multiple markers,  $q_1^2$  will depend on the maximum  $r^2$  of the QTL with any of its surrounding markers and on the structure of LD around the QTL.

Marker-based EBV,  $\hat{Q}_1$ , are estimates of genetic effects  $Q_1$ . Using properties of BLUP EBV (Henderson, 1984), the relationship between marker-based EBV and genetic effects Q can be modeled as:  $Q_1 = \hat{Q}_1 + e_1$ , where  $e_1$  represents the prediction error for the marker-based EBV. Let  $r_{\hat{Q}_1}$  denote the accuracy of  $\hat{Q}_1$  as a predictor  $Q_1$ , i.e. the correlation between  $Q_1$  and  $\hat{Q}_1$ . Then path coefficients associating  $Q_1$ ,  $\hat{Q}_1$  and  $e_1$  can be derived and are presented in Figure 1. The correlation of  $\hat{Q}_1$  with  $G_1$  then is equal to  $r_{MG_1} = q_1 r_{\hat{Q}_1}$ . This correlation represents the accuracy of the marker-based EBV as a predictor of the total genetic value  $G_1$ , and represent the accuracies of marker-based EBV that were obtained by Meuwissen et al. (2001) for genomic selection.

The use of BLUP to estimate marker-based EBV results in a zero correlation between  $\hat{Q}_1$  and  $e_1$ , as reflected in Figure 1, i.e. BLUP EBV and their prediction errors are uncorrelated (Henderson, 1984). Then, using the path coefficient diagram in Figure 1, the following correlations that are required for inclusion of marker-based EBV as a trait in selection index calculations can be determined:

The genetic correlation between the trait and marker-based EBV is:  $r_{G_1\hat{Q}_1} = q_1 r_{\hat{Q}_1} = r_{MG_1}$

The corresponding phenotypic correlation is:  $r_{P_1\hat{Q}_1} = h_1 q_1 r_{\hat{Q}_1} = h_1 r_{MG_1}$ .

Note that in correlations depend only on the accuracy of the marker-based EBV,  $r_{MG_1}$ , which depends on the extent and pattern of LD between markers and QTL and the amount and accuracy of data available to estimate marker effects, which are all under the control of the breeder. Deterministic methods to predict  $r_{MG_1}$  for a given marker density, LD structure, and amount of phenotypic information have not yet been developed, but they can be derived by stochastic simulation, as in Meuwissen et al. (2001). For the purposes of the work presented herein, the effect of different levels of  $r_{MG_1}$  on responses to selection will be evaluated.

### 12.3.3 Extension to multiple traits

Here, the theory developed above for a single phenotypic trait will be extended to two phenotypic traits, accommodating correlations between the traits. Similar procedures can be used to model any data on any pair of traits.

Let  $\rho_{G_{12}}$ ,  $\rho_{R_{12}}$ , and  $\rho_{Q_{12}}$  be the genetic correlations traits 1 and 2 for the genetic components G, R, and Q. The partitioning of genetic effects into Q and R results in  $Q_1$  to be uncorrelated to  $R_1$  and  $Q_2$  to  $R_2$ . Genetic correlations between  $Q_1$  and  $Q_2$  and between  $R_1$  and  $R_2$  are expected to be equal to the genetic correlation between  $G_1$  and  $G_2$  ( $E(\rho_{R_{12}}) = E(\rho_{Q_{12}}) = E(\rho_{G_{12}})$ ) if the same markers are used for MA-genetic evaluation for both traits, and markers have not been pre-selected based on associations with phenotype, as would be the case for genomic selection. This holds because genetic effects associated with markers will then be comprised of a random proportion of genetic effects that contribute to each trait. The correlation between environmental components contributing to traits 1 and 2 is denoted by  $\rho_{E_{12}}$ .

Using the path coefficient diagram in Figure 1, the following phenotypic and genetic correlations between the phenotypic and the marker-based traits that are necessary for derivation of selection indices can be determined:

$$r_{G_i \hat{Q}_j} = q_i r_{\hat{Q}_i} r_{\hat{Q}_i \hat{Q}_j} + q_i \sqrt{1 - r_{\hat{Q}_i}^2} r_{\hat{Q}_i e_i} = q_i r_{\hat{Q}_i} r_{\hat{Q}_i \hat{Q}_j} r_{\hat{Q}_j} \rho_{Q_{12}} + q_i \sqrt{1 - r_{\hat{Q}_i}^2} r_{\hat{Q}_j} \sqrt{1 - r_{\hat{Q}_j}^2} \rho_{Q_{12}} = q_i r_{\hat{Q}_j} \rho_{Q_{12}}$$

$$r_{P_i \hat{Q}_j} = h_i r_{G_i \hat{Q}_j} = h_i q_i r_{\hat{Q}_j} \rho_{Q_{12}}$$

With random allocation of markers, the proportion of genetic variance that is associated with markers,  $q_i^2$  is expected to be equal for both traits. Thus:  $E(q_1^2) = E(q_2^2) = q^2$ . However, the accuracy of marker-based EBV as a predictor of  $Q_i$ , and therefore also the accuracy of marker-based EBV as a predictor of  $G_i$  ( $r_{MG_i} = q_i r_{\hat{Q}_i}$ ) can differ between traits because it not only depends on the amount of phenotypic data but also on the accuracy, i.e. heritability, of the phenotypic data that is available to estimate marker effects. Nevertheless, if it is assumed that  $q_1 = q_2$  and  $r_{\hat{Q}_1} = r_{\hat{Q}_2}$ , and thus  $r_{MG_1} = r_{MG_2} = r_{MG}$ , and  $\rho_{Q_{12}} = \rho_{G_{12}}$ , and then correlations simplify to:

$$r_{G_i \hat{Q}_j} = r_{MG} \rho_{G_{12}}$$

$$r_{P_i \hat{Q}_j} = h_i r_{MG} \rho_{G_{12}}$$

These parameters and their simplifications are summarized in Table 1.

**Table 1.** Genetic parameters<sup>1</sup> for four traits considered for derivation of selection criteria: phenotype for trait (P<sub>1</sub>) and trait 2 (P<sub>2</sub>), and marker-based EBV for trait 1 ( $\hat{Q}_1$ ) and trait 2 ( $\hat{Q}_2$ ) marker-based EBV.

	P <sub>1</sub>	P <sub>2</sub>	$\hat{Q}_1$	$\hat{Q}_2$
P <sub>1</sub>	$h_1^2$	$\rho_{P_{12}}$	$h_1 q_1 r_{\hat{Q}_1} =^2 h_1 r_{MG_1}$	$h_1 q_1 r_{\hat{Q}_2} \rho_{Q_{12}} = h_1 r_{MG_2} \rho_{Q_{12}}$
P <sub>2</sub>	$\rho_{G_{12}}$	$h_2^2$	$h_2 q_2 r_{\hat{Q}_1} \rho_{Q_{12}} = h_2 r_{MG_1} \rho_{Q_{12}}$	$h_2 q_2 r_{\hat{Q}_2}$
$\hat{Q}_1$	$q_1 r_{\hat{Q}_1} = r_{MG_1}$	$q_2 r_{\hat{Q}_1} \rho_{Q_{12}} = r_{MG_1} \rho_{Q_{12}}$	1	$r_{\hat{Q}_1} r_{\hat{Q}_2} \rho_{Q_{12}}$
$\hat{Q}_2$	$q_1 r_{\hat{Q}_2} \rho_{Q_{12}} = r_{MG_2} \rho_{Q_{12}}$	$q_2 r_{\hat{Q}_2} = r_{MG_2}$	$r_{\hat{Q}_1} r_{\hat{Q}_2} \rho_{Q_{12}}$	1

<sup>1</sup>  $h_i^2$  = heritability of phenotype for trait i

$q_i^2$  = proportion of genetic variance associated with markers for trait i

$r_{\hat{Q}_i}$  = accuracy of  $\hat{Q}_i$  as a predictor of marker-associated genetic effects, Q<sub>i</sub>.

$r_{MG_i}$  = accuracy of  $\hat{Q}_i$  as a predictor of the total genetic value, G<sub>i</sub>

$\rho_{G_{12}}$  = genetic correlation between traits 1 and 2

$\rho_{P_{12}}$  = phenotypic correlation between traits 1 and 2

$\rho_{Q_{12}}$  = correlation between Q<sub>1</sub> and Q<sub>2</sub>

$\rho_{R_{12}}$  = correlation between residual genetic effects for traits 1 (R<sub>1</sub>) and 2 (R<sub>2</sub>)

<sup>2</sup> Results after the equality signs assume  $q_1=q_2$  and use  $q_i r_{\hat{Q}_i} = r_{MG_i}$

$$\rho_{G_{12}} = q_1 q_2 \rho_{Q_{12}} + \sqrt{1 - q_1^2} \sqrt{1 - q_2^2} \rho_{R_{12}}$$

$$\rho_{P_{12}} = h_1 h_2 \left( q_1 q_2 \rho_{Q_{12}} + \sqrt{1 - q_1^2} \sqrt{1 - q_2^2} \rho_{R_{12}} \right)$$

The final correlation that is needed is the correlation between the marker-based EBV for the two traits. Based on the assumption that phenotypic data that contribute to  $\hat{Q}_1$  and  $\hat{Q}_2$  are independent, which will strictly not be valid if traits 1 and 2 are measured on the same animals, but will approximately be true if sufficient data are available, these correlations can be derived to be equal to:

$$r_{\hat{Q}_1 \hat{Q}_2} = \frac{\text{Cov}(\hat{Q}_1, \hat{Q}_2)}{\sqrt{\text{Var}(\hat{Q}_1) \text{Var}(\hat{Q}_2)}} = \frac{r_{\hat{Q}_1}^2 r_{\hat{Q}_2}^2 \rho_{Q_{12}}}{r_{\hat{Q}_1} r_{\hat{Q}_2}} = r_{\hat{Q}_1} r_{\hat{Q}_2} \rho_{Q_{12}} = r_{\hat{Q}_1} r_{\hat{Q}_2} \rho_{G_{12}}$$

The latter equality assuming  $\rho_{Q_{12}} = \rho_{G_{12}}$ , as for genomic selection. This correlation cannot be further simplified to depend only on  $r_{MG}$  and  $\rho_{G_{12}}$ .

It should also be noted that, although the use of BLUP to estimate marker-based EBV results in a zero correlation between the marker-based EBV for a trait,  $\hat{Q}_i$ , and its prediction error,  $e_i$ , when marker-based EBV are obtained from single-trait procedures, which is what is assumed here, prediction errors of an individual's marker-based EBV for trait 1 (2) will be correlated to prediction errors of its marker-based EBV for trait 2 (1). In addition, prediction errors for marker-based EBV for trait 1 (2) will also be correlated with the marker-based EBV for trait 2 (1).

$$\begin{aligned} r_{\hat{Q}_1 e_2} &= \frac{\text{Cov}(\hat{Q}_1, Q_2 - \hat{Q}_2)}{\sqrt{\text{Var}(\hat{Q}_1)\text{Var}(e_2)}} = \frac{\text{Cov}(\hat{Q}_1, Q_2) - \text{Cov}(\hat{Q}_1, \hat{Q}_2)}{\sqrt{\text{Var}(\hat{Q}_1)\text{Var}(e_2)}} = \frac{r_{Q_1}^2 \rho_{Q_{12}} - r_{Q_1}^2 r_{Q_2}^2 \rho_{Q_{12}}}{r_{\hat{Q}_1} \sqrt{1 - r_{\hat{Q}_2}^2} \rho_{Q_{12}}} = r_{\hat{Q}_1} \sqrt{1 - r_{\hat{Q}_2}^2} \rho_{Q_{12}} \\ r_{\hat{Q}_2 e_1} &= r_{\hat{Q}_2} \sqrt{1 - r_{\hat{Q}_1}^2} \rho_{Q_{12}} \\ r_{e_1 e_2} &= \sqrt{1 - r_{\hat{Q}_1}^2} \sqrt{1 - r_{\hat{Q}_2}^2} \rho_{Q_{12}} \end{aligned}$$

Note that, using path diagram theory (Lynch and Walsh, 1998) and the path diagram in Figure 1, these correlations result in the correct correlation between  $Q_1$  and  $Q_2$ :

$$r_{Q_1 Q_2} = r_{\hat{Q}_1} r_{\hat{Q}_1 \hat{Q}_2} r_{\hat{Q}_2} + r_{\hat{Q}_1} r_{\hat{Q}_1 e_2} \sqrt{1 - r_{\hat{Q}_2}^2} + \sqrt{1 - r_{\hat{Q}_1}^2} r_{\hat{Q}_2 e_1} r_{\hat{Q}_1} + \sqrt{1 - r_{\hat{Q}_1}^2} r_{e_1 e_2} \sqrt{1 - r_{\hat{Q}_2}^2},$$

which, when substituting the previous equations for correlations among EBV and prediction errors, simplifies to  $\rho_{Q_{12}}$ .

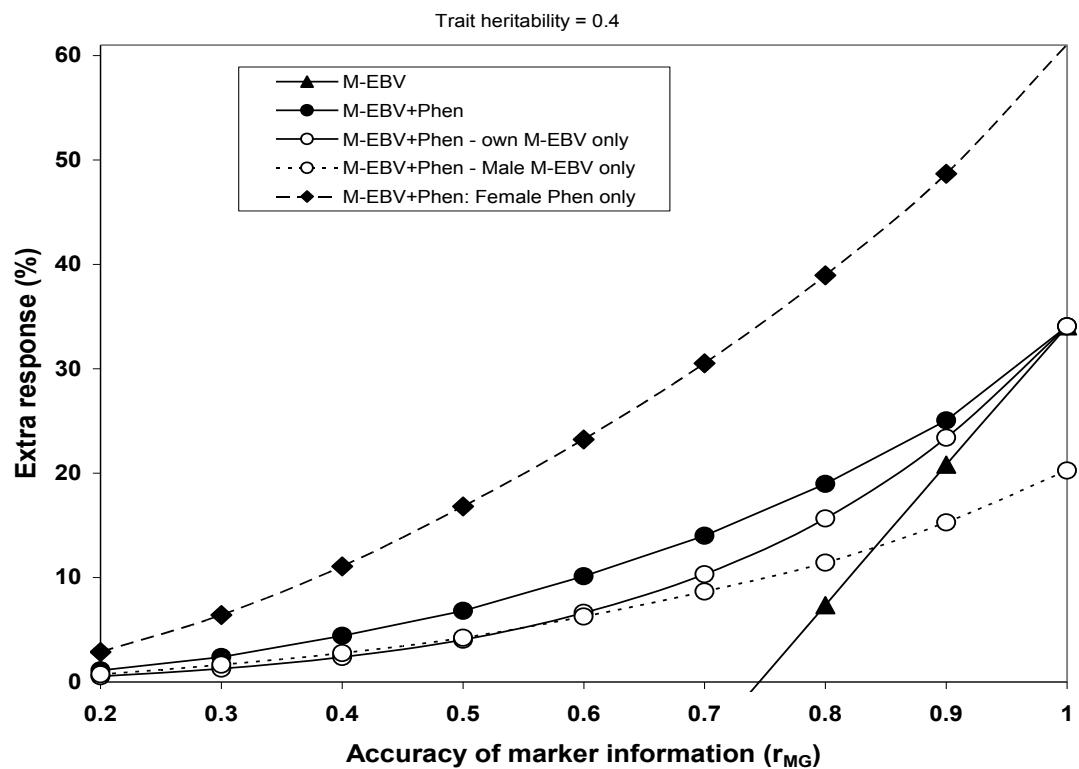
#### 12.3.4 Incorporating the Bulmer effect and predicting rates of inbreeding

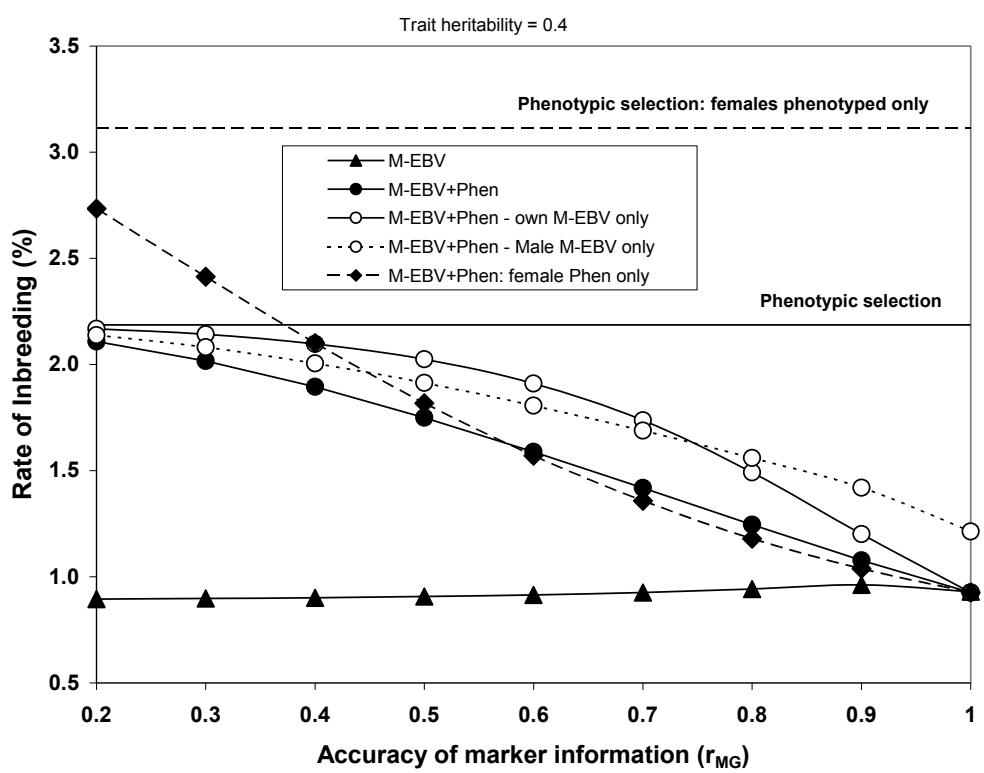
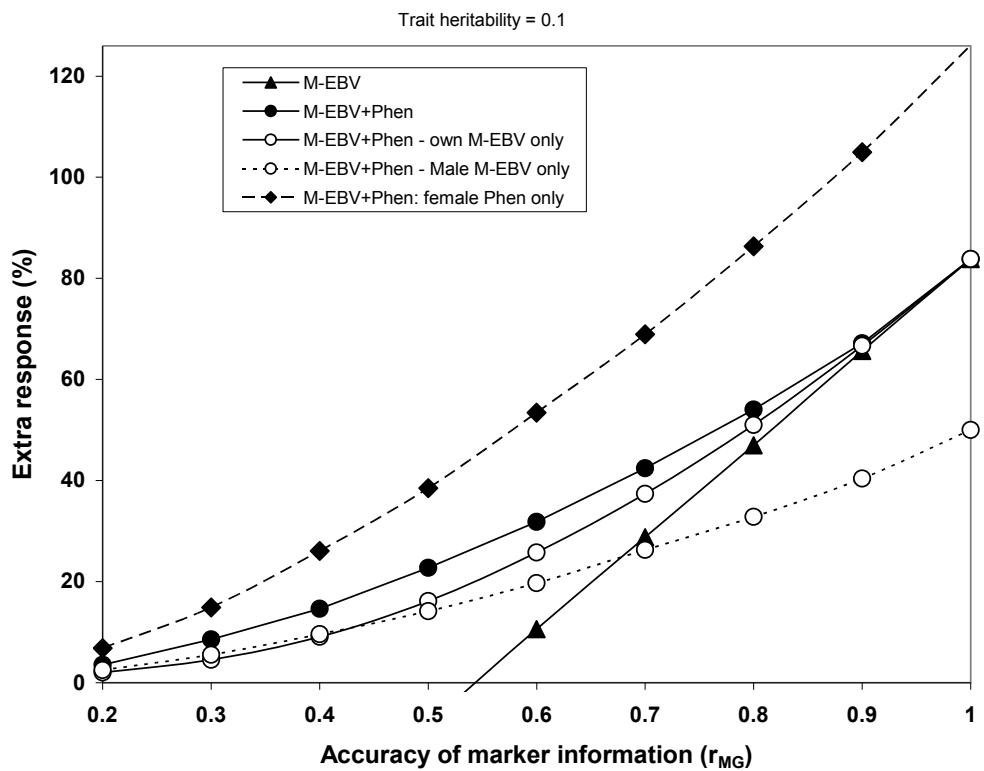
Reparameterizing the model by specifying marker-based EBV as correlated traits with heritability equal to one, in principle allow methods that have been developed for predicting response to selection for polygenic traits to be applied. This includes pseudo-BLUP selection index methods for deterministic modeling of selection on Animal Model EBV, incorporation of the effects of selection on variance-covariance structures through the Bulmer effect (Wray and Hill (1989) and Villanueva et al. (1993)), effects of co-selection of relatives on selection intensities (Meuwissen, ), etc. In addition, methods developed for prediction of rates of inbreeding based on long-term contribution theory can be used (Woolliams and Bijma (2000)). Such methods have been implemented in the selection index software package SelAction (Rutten et al.) and this software will be used to demonstrate use of the model and potential benefits of MAS. When applying these methods and this software, it must be realized that use of a trait with heritability equal to one may push the validity of the developed methods and that all predictions are based on the infinitesimal model and, therefore, do not account for changes in gene frequencies, which will be more important with MAS. In addition, methods assume that the genetic basis of traits remains constant.

#### 12.3.4 Choice of Parameters

To illustrate the use of the developed methodology to predict the potential benefit of MAS, markers used were assumed to be randomly allocated across the genome, reflecting genomic selection, thus  $\rho_{R_{12}} = \rho_{Q_{12}} = \rho_{G_{12}}$ . This same assumption also causes the expected proportion of genetic variance that is associated with markers to be equal for all traits, thus  $q_1 = q_2$ , which leads

to  $q_i r_{\hat{Q}_i} = q_j r_{\hat{Q}_j} = r_{MG_i}$ , which is the correlation of the marker-based EBV with the total genetic value. Under these assumptions and using  $r_{MG_i}$  as an input parameter, phenotypic and genetic correlations between phenotypes and marker-based EBV depend only on  $r_{MG_i}$ , and not on its partition into  $q_i$  and  $r_{\hat{Q}_i}$ . This makes the results more general and applicable to different combinations of  $q_i$  and  $r_{\hat{Q}_i}$  for a given level of accuracy of marker-based EBV ( $r_{MG_i}$ ). Resulting equations are in Table 1. The correlation between marker-based EBV ( $r_{\hat{Q}_p \hat{Q}_c} = r_{\hat{Q}_p} r_{\hat{Q}_c} \tilde{n}_{Q_{pc}}$ ) does depend on  $r_{\hat{Q}_i}$ . To evaluate the impact of marker information, various levels of  $r_{MG_i}$ , ranging from 0.2 to 0.9, were evaluated. In all cases,  $r_{MG_i}$  was equal for both traits.





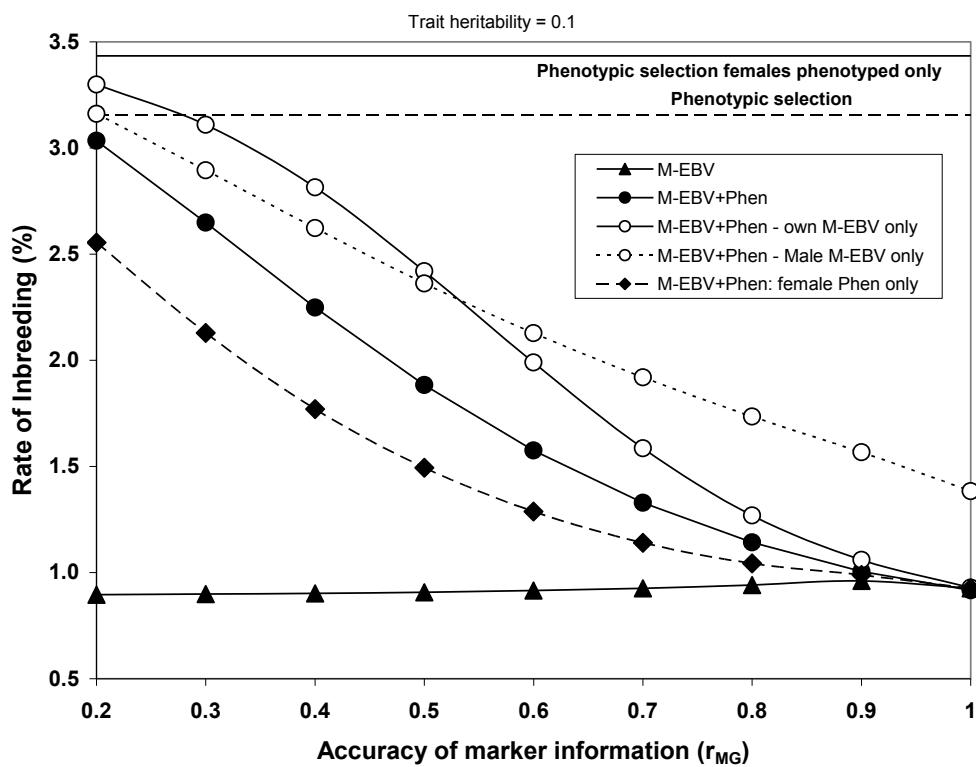


Table 2. Genetic parameters for selection on a breeding goal of two traits ( $P_1$  and  $P_2$ ) with and without marker information and resulting responses to selection in individual traits and the breeding goal ( $\Delta H$ ) and rates of inbreeding ( $\Delta F$ ). Marker-based EBV (Error! Objects cannot be created from editing field codes. and Error! Objects cannot be created from editing field codes.) have accuracies of 0.8, based on markers explaining 62.4% of the genetic variance.

Correlations <sup>1</sup>	$P_1$	$P_2$	$\hat{Q}_1$	$\hat{Q}_2$		
$P_1$	--	-0.5	0.438	-0.131		
$P_2$	-0.3	--	-0.076	0.253		
$\hat{Q}_1$	0.8	-0.24	--	-0.243		
$\hat{Q}_2$	-0.24	0.8	-0.243	--		
Heritability	0.3	0.1	1	1		
Phenotypic SD	1	1	0.8	0.8		
Economic value	1	1	0	0		
Response to selection					$\Delta H$	$\Delta F(%)$
Phenotype only	0.408	0.041	0.394	0.052	0.448	2.36
Markers only	0.418	0.068	0.655	0.167	0.486	0.94
Combined	0.469	0.074	0.582	0.148	0.543	1.29

<sup>1</sup> Phenotypic correlation above the diagonal; genetic correlations below the diagonal