

Day 4a Incorporating Markers in Breeding Value Estimation

Objectives

Illustrate concepts of incorporating marker data in BLUP breeding value estimation

- Using Linkage Disequilibrium information
- Using Linkage / co-segregation information.
- Using combined Linkage and Linkage Disequilibrium information

With examples in lab session

Marker-based genetic evaluation methods follow logically from the methods described for QTL detection using LD and linkage (in complex pedigrees)

1. **Marker-Assisted BLUP using linkage / co-segregation information**
2. **MA-BLUP using Linkage Disequilibrium markers**
3. **MA-BLUP using combined Linkage disequilibrium and linkage information**

1

1. MA-BLUP using linkage/co-segregation

Follow co-segregation of markers and phenotype (QTL) within families

1 By fitting QTL effect means (within family)

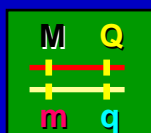
- used for QTL analysis in "simple" designs (e.g. HS)

2 By modeling covariances between relatives for QTL through IBD matrices

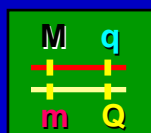
(Fernando & Grossman, 1989)

• Used in extended pedigrees

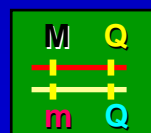
Sire 1



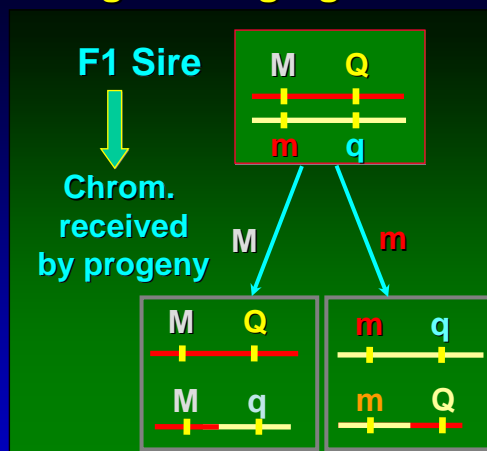
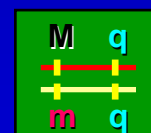
Sire 2



Sire 3



Sire 4

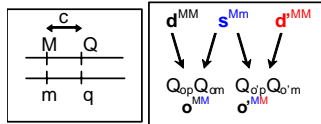


2

Marker-assisted BLUP using linkage / co-segregation information

(Fernando and Grossman, GSE, 1989)

Marker and QTL in LE across the population – only use within-family linkage (cosegregation)



Q_{ip} = QTL allele received by i from sire

Q_{im} = QTL allele received by i from dam

Genetic model:

$$g_i = u_i + v_i^p + v_i^m$$

u_i = residual genetic or polygenic effects not in LD with marker
 v_i^p = effect of QTL allele received by i from its sire (paternal) ($=\alpha_{ip}$)
 v_i^m = effect of QTL allele received by i from its dam (maternal) ($=\alpha_{im}$)

Additive Animal Model: $y_i = \text{fixed effects} + u_i + v_i^p + v_i^m + e_i$

3

Matrix formulation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \mathbf{e}$$

Total additive var. =

$$\sigma_a^2 = \sigma_R^2 + 2\sigma_v^2$$

\mathbf{y} = vector of phenotypic records (y_i)
 $\boldsymbol{\beta}$ = vector of fixed effect solutions
 \mathbf{X} = incidence matrix relating records to fixed effects (mean, hys, age)
 \mathbf{u} = vector of polygenic breeding values (u_i) – assumed distributed $\sim N(0, \mathbf{A}\sigma_R^2)$
 with \mathbf{A} = add.relationship matrix and σ_R^2 = polygenic variance = $\sigma_g^2 - 2\sigma_v^2$
 \mathbf{Z} = incidence matrix relating records to polygenic breeding values in \mathbf{u}
 \mathbf{v} = vector of QTL allele breeding values (v_i^p, v_i^m) – assumed distrib. $\sim N(0, \mathbf{G}_v\sigma_v^2)$
 with \mathbf{G}_v = IBD matrix derived from marker data using linkage / co-segregation
 and $\sigma_v^2 = 1/2$ of the variance contributed by QTL = $1/2(2pq\alpha^2)$
 \mathbf{W} = incidence matrix relating records to breeding values of QTL alleles in \mathbf{v}
 \mathbf{e} = vector of residuals (e_i) – assumed distributed $\sim N(0, \mathbf{I}\sigma_e^2)$

Henderson's MME:

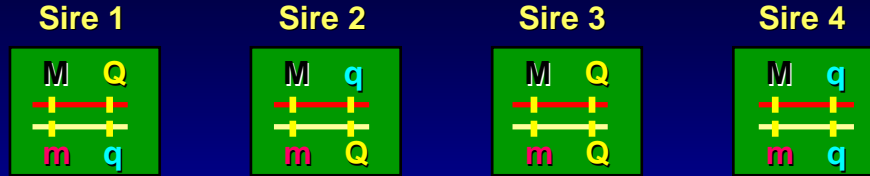
$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \lambda_v\mathbf{G}_v^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{C}^{\mathbf{X}\mathbf{X}} & \mathbf{C}^{\mathbf{X}\mathbf{Z}} & \mathbf{C}^{\mathbf{X}\mathbf{W}} \\ \mathbf{C}^{\mathbf{Z}\mathbf{X}} & \mathbf{C}^{\mathbf{Z}\mathbf{Z}} & \mathbf{C}^{\mathbf{Z}\mathbf{W}} \\ \mathbf{C}^{\mathbf{W}\mathbf{X}} & \mathbf{C}^{\mathbf{W}\mathbf{Z}} & \mathbf{C}^{\mathbf{W}\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

with $\lambda = \frac{\sigma_e^2}{\sigma_R^2}$ $\lambda_v = \frac{\sigma_e^2}{\sigma_v^2}$ $\begin{bmatrix} \mathbf{C}^{\mathbf{X}\mathbf{X}} & \mathbf{C}^{\mathbf{X}\mathbf{Z}} & \mathbf{C}^{\mathbf{X}\mathbf{W}} \\ \mathbf{C}^{\mathbf{Z}\mathbf{X}} & \mathbf{C}^{\mathbf{Z}\mathbf{Z}} & \mathbf{C}^{\mathbf{Z}\mathbf{W}} \\ \mathbf{C}^{\mathbf{W}\mathbf{X}} & \mathbf{C}^{\mathbf{W}\mathbf{Z}} & \mathbf{C}^{\mathbf{W}\mathbf{W}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \lambda_v\mathbf{G}_v^{-1} \end{bmatrix}$

Total EBV = $\hat{g}_i = \hat{u}_i + \hat{v}_i^p + \hat{v}_i^m = [\mathbf{Z}_i \quad \mathbf{W}_i] \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{v}} \end{bmatrix}$ where \mathbf{Z}_i and \mathbf{W}_i are row i of \mathbf{Z} and \mathbf{W}

PEV of $\hat{g}_i = \text{PEV}(\hat{g}_i) = [\mathbf{Z}_i \quad \mathbf{W}_i] \begin{bmatrix} \mathbf{C}^{\mathbf{Z}\mathbf{Z}} & \mathbf{C}^{\mathbf{Z}\mathbf{W}} \\ \mathbf{C}^{\mathbf{W}\mathbf{Z}} & \mathbf{C}^{\mathbf{W}\mathbf{W}} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_i' \\ \mathbf{W}_i' \end{bmatrix} \sigma_e^2 \Rightarrow \text{Accuracy} = r = \sqrt{1 - \frac{\text{PEV}(\hat{g}_i)}{\sigma_g^2}}$

MA-BLUP using linkage has seen limited application Linkage phase inconsistent between sires



QTL effects estimated based on family info only

Requires family information

- marker genotypes
- phenotypes
- Complex logistics
- 'Complex' analysis
- Limited accuracy

➔ Need LD markers

$$y_i = \mu + v_i^p + v_i^m + u + e$$

Pat / Mat Poly-genic

$\text{Var}(u) = A\sigma_a^2$ $\text{Var}(v) = G_v\sigma_v^2$

Fernando & Grossman, 1989

5

MA-BLUP using LD-markers

Several options exist:

1) Fit IBD matrix derived using LD

Methodology similar as Fernando and Grossman, except G_v is now based on LD-IBD (Meuwissen & Goddard GSE 2001)

2) Fit marker genotype as fixed effect

$$y_i = \text{fixed effects} + \text{marker genotype} + u_i + e_i$$

Several possibilities exist for the way in which marker genotype is fitted:

- as a covariate: bX_i where X_i = number of "1" alleles carried by i (minus 1) = -1 or 0 or 1 and b = allele substitution effect at the marker
 - Fits only additive effect
 - With >2 alleles, or haplotypes – fit $\sum_{h=1}^m b_h X_{ih}$ with h = alternate allele/haplotype
- as a class variable – allowing for a different mean for each genotype
 - Allows for dominance and can be extended to fitting haplotypes

3) Fit marker genotype effect as a random effect

$$y_i = \text{fixed effects} + bX_i + u_i + e_i \quad \text{with } b \sim N(0, \sigma_b^2)$$

The difference between fitting random effects using IBD (1) versus (3) is that (1) models covariances between haplotypes even if they are not IBS, based on haplotype similarity

Method 3 essentially assumes IBD=0 if two haplotypes are not IBS

6

MA-BLUP using combined LD-LA info

Several options again exist:

- 1) **Fit random QTL effects using IBD matrix derived using combined LD-LA**
Method of Fernando and Grossman but matrix \mathbf{G}_v now is the LD-LA based IBD matrix

- 2) **Fit fixed marker genotype effect to capture LD and random effect to capture linkage / co-segregation**

$$y_i = \text{fixed effects} + bX_i + v_i^p + v_i^m + u_i + e_i \quad V(\mathbf{v}) = \mathbf{G}_v\sigma_v^2$$

with \mathbf{G}_v = linkage-based IBD matrix

- 3) **Same as 2) but now fit marker effect as random**

$$y_i = \text{fixed effects} + bX_i + v_i^p + v_i^m + u_i + e_i \quad \text{with } b \sim N(0, \sigma_b^2)$$

- 4) **Mixture models that model founder QTL alleles and follow them through the pedigree** (Rohan Fernando)

The difference between fitting random effects using IBD (1) versus (3) again is that method 3 essentially assumes IBD=0 if two haplotypes are not IBS

7

Day 4a

Incorporating Markers in Breeding Value Estimation

Objectives

Illustrate concepts of incorporating marker data in BLUP breeding value estimation

- Using Linkage Disequilibrium information
- Using Linkage / co-segregation information.
- Using combined Linkage and Linkage Disequilibrium information

With examples in lab session

Marker-based genetic evaluation methods follow logically from the methods described for QTL detection using LD and linkage (in complex pedigrees)

1. **Marker-Assisted BLUP using linkage / co-segregation information**
2. **MA-BLUP using Linkage Disequilibrium markers**
3. **MA-BLUP using combined Linkage disequilibrium and linkage information**

8