**Breeding Values and Markers**

Jeffrey Endelman

Agron/Hort 812, Spring 2020

Recall that we are pursuing the decomposition of genotypic value into additive and non-additive values:

|  |  |
| --- | --- |
|  | [1] |

Previously, we considered one QTL with an arbitrary number of alleles. Now, we will restrict treatment to bi-allelic QTL but allow for multiple loci.

**Bi-allelism, Single Locus**

Denote the two alleles as *B*/*b*, with the frequency of *B* denoted by *p* and *q* = 1-*p*. The solution for the additive effects satisfies (cf Eq. 2 from previous lecture)

|  |  |
| --- | --- |
|  | [2] |

Defining the **allele substitution effect** as , the additive effects can be written as

and the additive value of the three possible genotypes is

|  |  |
| --- | --- |
|  | [3] |

Letting *x* denote the dosage (0,1,2) of the *B* allele, the additive value *a*is proportional to the allele substitution effect:

|  |  |
| --- | --- |
|  | [4] |

The proportionality constant is the "centered genotype" *w*, which is the allele dosage relative to the population mean: .

**Note:** In classical quantitative genetics, the additive effects are parameters of the model, not random variables. The allele dosage (and thus *w*) is a random variable due to the random properties of segregation.

The additive genetic variance is

Substituting Eq. 4 into Eq. 1 shows that the allele substitution effect is the regression coefficient when regressing the genotypic value on the centered genotype:

|  |  |
| --- | --- |
|  | [5] |

**Multiple Bi-Allelic QTL**

The extension of Eq. 1 and 4 to a trait with Q QTL is

|  |  |
| --- | --- |
| + *r* | [6] |

Eq. 6 expresses genotypic value as a regression on the centered genotypes at Q QTL, with denoting the allele substitution effect at locus *k.* The bold symbols after the second equality are vectors of length Q. From regression theory, the solution is

|  |  |
| --- | --- |
|  | [7] |

Where is a symmetric Q x Q matrix and **c** is a Q x 1 vector. As shown in Eq. 7, the off-diagonal elements of can be written in terms of the correlation coefficient between the QTL genotypes, which highlights the influence of linkage disequilibrium (LD) between the QTL on the solution.

The additive genetic variance is

|  |  |
| --- | --- |
|  | [8] |

Note that Eq. 8 is not a sum of locus-specific variances, i.e., it is not

|  |  |
| --- | --- |
|  | [9] |

Eq. 9 is only true when the **QTL are in linkage equilibrium**. In this case, is a diagonal matrix, and the allele substitution effect at each locus is the same as the single locus result (Eq. 5):

|  |  |
| --- | --- |
|  | [10] |

**Marker Effects**

When the QTL genotypes are unknown, the genotypic values can be regressed on the centered genotypes, denoted by the vector **z**, with the regression coefficients (i.e., **additive marker effects**) denoted by **b**. For a strictly additive trait (zero residual in Eq. 6), the regression equation is

|  |  |
| --- | --- |
|  | [11] |

The residual in Eq. 11 represents additive genetic variance that is not captured by the markers due to incomplete marker-QTL LD. The solution for the marker effects is

|  |  |
| --- | --- |
|  | [12] |

For *m* markers, is a *m* x *m* matrix that depends on the LD between markers, while is a *m* x *Q* matrix representing marker-QTL LD. If there is at least one marker in perfect LD with every QTL (e.g., when whole-genome sequence data are available), then the marker effects at the causal loci equal the true allele substitution effects, and all other markers have zero effect (for a proof, see de los Campos et al. 2015).

Multiplying Eq. 12 by the marker genotypes produces **genomic breeding values**:

|  |  |
| --- | --- |
|  | [13] |

where is the best linear predictor of allele dosage at the QTL. The GBV variance, which offers a lower bound on the true additive variance, is

|  |  |
| --- | --- |
|  | [14] |

Note that, even when there is no LD between QTL, there is still LD between the genome-wide markers. As a result, is not diagonal and, unlike the true additive variance, the GBV variance cannot be written as a sum over marker loci (except when there is perfect LD between markers and QTL, as discussed above).

For a single marker-QTL pair, Eq. 14 reduces to

|  |  |
| --- | --- |
|  | [15] |

Which shows that the proportion of the additive variance captured by the marker equals the squared correlation between the marker and QTL genotypes (i.e., marker-QTL LD).

**References (not required reading)**

de los Campos G, Sørensen D, Gianola D. 2015. Genomic Heritability: What is it?. PloS Genetics 11(5):e1005048.