

Genomic & Phenotypic Distances

bb

11/8/2021

Assume a trait is encoded by L biallelic loci in a haploid population. Normalize the effects of alleles so major alleles correspond to effect size zero. At locus ℓ the additive effect on trait value of the minor allele is a_ℓ . Consider the haploid genomes of sampled individuals. For the i th individual this can be represented by a binary string (I_1^i, \dots, I_L^i) , where $I_\ell^i = 0$ if it carries the major allele at the ℓ th locus and 1 if it carries the minor allele. Then the breeding value of the i th individual is given by $g_i = \sum_\ell a_\ell I_\ell^i$. The distance between the breeding values of the i th and j th individuals is naturally defined by $d_{bv}(g_i, g_j) = |g_i - g_j|$. This expands to

$$d_{bv}(g_i, g_j) = \left| \sum_\ell a_\ell (I_\ell^i - I_\ell^j) \right|.$$

Similarly, the genetic distance between sampled haploid genomes is commonly computed as the number of allelic differences per locus:

$$\begin{aligned} \pi_{ij} &= \frac{1}{L} \sum_\ell (1 - I_\ell^i) I_\ell^j + I_\ell^i (1 - I_\ell^j) \\ &= \frac{1}{L} \sum_\ell I_\ell^j - I_\ell^j I_\ell^i + I_\ell^i - I_\ell^i I_\ell^j = \frac{1}{L} \sum_\ell I_\ell^j - 2I_\ell^i I_\ell^j + I_\ell^i. \end{aligned}$$

Hmmm, dang. Was hoping some interesting relation would pop out. Not seeing anything though...