

Mapping of QTL:

- The location of QTL can be found through candidate genes or association by linkage to molecular markers
- Candidate gene approach: assume a gene could have an effect, sequence it, and then check for possible associations
- Linkage by marker approach: if a QTL is linked to some marker, then an association can be found between allelic segregations of both
- Interval mapping approach: possible to identify QTL if present between a pair of markers that are a small distance apart. Get two highly inbred lines that differ for a character, then either carry out analysis for offspring of the F_2 generation, or in the backcross of the hybrid to one of the parents. The parental lines will produce the same genotype, while the hybrid cross will produce 8 different types of gametes with frequencies dependent on recombination frequency.
- Model for interval mapping: assuming additivity, then the phenotypic value of individual i is $P_i = M + \alpha g_i + e$, where α is the average effect of allelic substitution, M is the general mean, $e \sim \text{Norm}(0, \sigma^2)$ and g_i is the number of copies of allele a . The estimation of the effect size + position of the QTL is obtained by ML: $L(M, \alpha, \sigma^2) = \prod_i [G_i(0)L_i(0) + G_i(1)L_i(1)]$ where $L_i(x)$ is normally deviated with mean $= P_i - (M + \alpha x)$
- Decimal logarithm of the ratio between likelihood function above and that which assumes there is no QTL gives rise to LOD (log odds value, values higher than 2 indicate a QTL)
- GWAS: a regression of the form $y = 1\mu + Mg + e$. Basic idea: SNPs close to a given QTL will be in strong LD with it.
- LD is reduced by recombination. Therefore, the greater the distance between two given positions, the lower the expected value of disequilibrium (so only SNPs close to a QTL will have a strong LD $r^2 > 0.2$).
- Most often, GWAS is conducted on individual SNPs, but it can be conducted on groups of SNPs which may be theoretically more powerful for QTL detection
- Missing heritability problem: total contribution of SNPs to heritability is vastly smaller than the heritability of estimated character trait
- Reason 1 for missing heritability: GWAS cannot detect nucleotide variants because the number of QTLs found increases as sample size increases
- Reason 2 for missing heritability: gap between GWAS heritabilities and familial ones can occur because familial heritabilities are inflated by non-additive genetic components and environment
- For instance, for reason 2, an unknown fraction of genetic variation can be epistatic, though evidence has suggested that the magnitude of epistatic variance is always much smaller than additive variance.
- Detecting the footprint of selection: polymorphism or divergence
- Intraspecific polymorphism: we know that expected heterozygosity at a locus is $4N_e\mu$ and this can either be estimated by nucleotide diversity θ_π (the average number of nucleotide differences between two given sequences) or the Watterson estimator θ_w which is obtained by counting the number of polymorphic sites corrected for sample sizes.
- under a neutral model we expect $\theta_\pi = \theta_w$. Selective sweeps lead to a reduction in θ_π proportionally greater than that of θ_w . Balancing selection leads to a larger increase for θ_π and this can be induced by overdominance or frequency-dependent selection.
- Can detect footprint of natural selection by interspecific divergence (number of nucleotide divergences between two species) along with intraspecific polymorphism. Under neutrality, the relationship between the number of non-synonymous to synonymous polymorphic

changes within a species should be equal to the relationship between the number of non-synonymous/synonymous changes fixed between species (since the rate of change is a function of the neutral mutation rate).

- *dn/ds* test: can detect selection in phylogenetic branches that lead to ancestral forms that may have been extinct for millions of years. Compares the rates of substitutions for non-synonymous and synonymous positions using divergence between species. If the ratio is greater than 1, this implies the acceleration of fixation due to diversifying selection. If contra, then we have decreased fixation rate for non-syn mutations due to purifying selection.
- LD methods: drift and selection generate LD that decays by recombination as a function of distance between them (slower at lower recombination rates).
- But if selection fixes a beneficial allele rapidly, then this drags along linked neutral alleles which generates LD greater than expected using a neutral model. Decay with distance will also be smaller, generating a lower homozygosity in the general region of the selected locus (haplotype homozygosity)
- If selection is sufficiently strong, then the above occurs faster than recombination or mutation can act to break up the haplotype, and we observe a high haplotype homozygosity (extended haplotype homozygosity EHH)
- Specifics of EHH: look for segments of the genome with high density of SNPs that allow the identification of haplotypes. Then consider a broad region on both sides with lower density SNPs to evaluate the decay of LD with respect to distance. If the surrounding regions are under selection, then EHH decays more slowly with respect to distance versus a neutral allele.
- Relative EHH: scale each EHH to a combined EHH value considering all cores analyzed
- A method to reduce effects of variation in recombination rates along with demographic processes: compare the decay of EHH for ancestral alleles versus derived alleles of the core. EHH decay is added over extended region making up the integrated HH and is then calculated for the ancestral and derived alleles yielding the integrated haplotype statistic $iHS = \ln(iHH_A/iHH_D)$. Equal decay rates imply neutrality
- Detecting diversifying and convergent selection: consider finding diversifying selection among many populations. Cannot use LD as before, must use Q_{ST} (or F_{ST} for neutral markers). Convergent selection can be found when $Q_{ST} < F_{ST}$, but this does not inform the loci that are involved.
- Finding the loci requires a large number of neutral loci in LD with QTL. Selection will produce a drag of neutral variants that causes a high value of differentiation in allele frequencies between populations versus a neutral model
- Markers with high differentiation are grouped into large regions due to linkage forming 'islands of differentiation'
- Genomic selection: exploit LD between nearby loci from drift or selection to predict the additive value of individuals
- First conceptual step of genomic selection: estimate the effects of genetic markers in a reference population (difficult step, requiring both genotype and phenotype data, and a high number of markers)
- Then estimate the additive values of individuals
- Meuwissen proposed two Bayesian approaches. In A the a priori distribution of values of b , the vector of estimated effects, follows a t distribution, whereas in Bayes B a proportion of the markers has no effect
- GBLUP: BLUP with a genomic relationship matrix, accuracy of this method depends on N_e , the genetic length of the genome and the number of markers

Problems

- (1) The following table presents the 16 polymorphic nucleotides in five sequences of 500bp (other nucleotides are monomorphic). Calculate nucleotide diversity θ_π . What can be deduced from the comparison between the number of polymorphic sites and the nucleotide diversity?
See ch11.r for the answer to this question.
- (2) The polymorphism and divergence between synonymous nucleotide positions (they produce no change in amino acid) and replacement positions (non-synonymous: they produce a change) have been studied for the alcohol dehydrogenase (ADh) gene in samples from four *Drosophila* species. In these samples, 71 fixed and 106 polymorphic synonymous positions were found, as well as 23 fixed and 13 polymorphic replacement positions. Do these data indicate any kind of selection?
See ch11.r for the answer to this question.

Self Assessment

- (1) The procedure of QTL mapping by candidate genes requires partial sequencing of the relevant gene.
True, mapping QTL either entails the identification of candidate genes (sequencing part or the whole gene) or through association by linkage to other markers (again requiring sequencing).
- (2) Interval mapping of QTLs using genetic markers can only be carried out by crossing highly inbred lines.
False. Though the most suitable design for interval mapping is through the cross of two highly inbred lines, this is not necessary.
- (3) For the detection of a QTL by GWAS it is necessary that the magnitude of LD between markers and the QTL is greater than 0.2.
True, this is stated on page 254.
- (4) The complete sequencing of genomes allows the detection of QTLs of smaller effect and frequency than those found by SNP chips.
True, this is stated in the penultimate paragraph on page 256.
- (5) Epistatic interactions are a possible explanation of the observation that the total contribution to heritability of the QTLs found by GWAs is often lower than the heritability of the trait estimated by resemblance among relatives.
True, this is covered on page 257. Non-additive genetic components may be partly responsible for gaps between GWAS heritability estimates and broad-sense heritability.
- (6) If the value of the nucleotide diversity θ_π for a set of sequences of the same species is greater than the ratio S/a , where S is the number of polymorphic nucleotides in the sequences, this suggests the selective sweep of a beneficial allele.
False, it implies balancing selection as produced by overdominance or frequency-dependent selection.
- (7) A ratio $dn/ds > 1$ implies diversifying selection between species.
True, this is directly mentioned on page 259.
- (8) Genomic selection allows the generational interval to be shortened.
This is true, estimates can be made at an early age on the additive value of individuals.
- (9) Genomic selection can also provide a method for mapping QTLs.
True, it is not necessary to know the genomic position of markers for genomic selection

but if these are known this provides a method of mapping loci that affect the character.

- (10) The accuracy of GBLUP increases with the increase in the variance of genomic relationships of individuals around their genealogical value.
False, the accuracy of GBLUP increases with the decrease in the variance of genomic relationships of individuals around their genealogical value.