Introductory seminar on "Mathematical Population Genetics" winter term 2017/18

Prof. Joachim Hermisson
Ilse Höllinger

3 Recombination and Drift

3.1 Haldane's Mapping Function

The recombination fraction r between two loci on the same chromosome is the probability of an odd number of recombination events between these loci. Ignoring interference between adjacent recombination events (in biological reality a recombination events reduces the probability of another recombination events in its close proximity), r relates to the genetic map distance d via *Haldane's mapping function*. Thereby d is the average number of recombination events in a given interval.

1. Derive *Haldane's mapping function*

$$r = \frac{1}{2}(1 - \exp[-2d])$$

2. Why is it useful to have a map function which converts recombination distance r into mapping distance d and vice versa? What might be easier to measure in data?

3. When could it be appropriate to use the approximation $r = d$? What does that mean in the biological context?

Tips for the derivation (1):
We denote the expected proportions of $0, 1, \ldots, k$ recombination events between to loci $p_0, p_1, \ldots, p_k$ and ignoring interference between recombination events we can assume that they are Poisson distributed, with $P[k] = \frac{\lambda^k \exp[-\lambda]}{k!}$. Then r and d can be written as functions of p_i. Show that $d = \lambda$ and then you can use this result to derive the mapping function.

3.2 Recombination with additive fitness in discrete time

We consider a haploid population, where at two loci, $\mathcal{A}$ and $\mathcal{B}$ there segregate two alleles each, denotes by A_1 and A_2 and B_1 and B_2. Fitness is additive as shown in the following scheme:

	B_1	B_2
A_1	$a_1 + b_1$	$a_1 + b_2$
A_2	$a_2 + b_1$	$a_2 + b_2$

- Write down mean fitness $\overline{\omega}$ and show that it can be written as the sum of single-locus mean fitness.
- Show that the time derivative of mean fitness does not depend on r or D. Show that mean fitness is non-decreasing (hint: use Jensen's inequality).
- Show that all equilibria are in linkage equilibrium (LE)
- Show that linkage equilibrium $D = 0$ is not maintained, if the population is not at equilibrium, *e.g.* give an example.

3.3 Covariance(genotype, fitness) governs dynamics

We consider the discrete time dynamics of a diploid population with two loci $\mathcal{A}$ and $\mathcal{B}$. The dynamics for the four haplotypes $A_1B_1, A_1B_2, A_2B_1, A_2B_2$ (alternative notation ab, aB, Ab, AB), with frequencies $P_\bullet$ and marginal fitness $\omega_\bullet$ indexed by $\bullet = *_{11}, *_{12}, *_{21}, *_{22}$ read

$$P'_{ij} = \frac{\omega_{ij}}{\overline{\omega}} P_{ij} + \eta_{ij} r \underbrace{\frac{\omega_{A_1B_1A_2B_2}}{\overline{\omega}} D}_{\hat{D} \text{ for diploids}} \tag{1}$$

where $\eta_{ij} = 1$ if $i = j$, and otherwise -1.

We define a measure g_{ij} for the genotype, as the frequency of haplotype A_iB_j within a (diploid) genotype $(A_kB_l|A_mB_n)$. Thus g_{ij} can take only the values $0, 0.5, 1$. Show that we can write the change in haplotype frequency, $\Delta P_{ij} = P'_{ij} - P_{ij}$, in terms of the covariance between the measure of the genotype g_{ij} and fitness, *i.e.*

$$\Delta P_{ij} = \frac{1}{\overline{\omega}} \Big(\underbrace{\mathrm{cov}(\omega, g_{ij})}_{=P_{ij}(\omega_{ij} - \overline{\omega})} + \eta_{ij} r \hat{D} \Big) \tag{2}$$

******

Hint: Set $g_{ij} = g_{11}$ for the change in P_{11} and show $\mathrm{cov}(\omega, g_{11}) = P_{11}(\omega_{11} - \overline{\omega})$ for this measure of genotype. This then easily generalises to arbitrary i, j.

g_{11}: The measure for the genotype gives the proportion of how many times I find a haplotype in a genotype: never (0), once (0.5) or twice (1). Note that with this measure for the haplotype $A_1B_1 \Rightarrow g(A_1B_2|A_2B_1) = 0$.

Definition: If there are n possible realizations of pairs of random variables (X, Y), namely (x_i, y_i) for $i = 1, ..., n$, with possibly unequal probabilities p_i , then the covariance is

$$\mathrm{cov}(X, Y) = \sum_i^n p_i \big(x_i - \mathrm{E}(X) \big)\big(y_i - \mathrm{E}(Y) \big).$$

3.4 (Shorter Exercise) Homo- and heterozygosity without replacement

In the lecture we have defined homozygosity F_t and heterozygosity H_t as the probability that two randomly haploid alleles (single chromosomes with diploid individuals) are identical or different by descent, respectively. Sampling occurs with replacement, such that a chosen allele, can be sampled twice with rat $\frac{1}{2N}$.

Let us now consider the case, where sampling occurs without replacement in a population of size $2N$ with k alleles. Derive the formulas for F_t and H_t as expressions of the allele frequencies, as well as the recursion equations. What can be said about the long term evolution of these measures in comparison when sampling occurs with replacement?