

STAT246 : Statistical Genetics

Terry Speed

Lecture 2 - January 19, 2006

Scribe: Soyeon Ahn

EXERCISES:

Ex1. Derive the Hardy-Weinberg Equilibrium (**HWE**) frequencies for >2 alleles.

Ex2. Suppose that your brother has genotype $\{u, v\}$ at an STR locus. What is the probability that you are also $\{u, v\}$ at that locus? Consider the case $u = v$ and $u \neq v$ separately. Assume “random mating” at this locus.

More about the HWE

In the **HWE** derivation we ignored the possibility of

- mutation (mutation rate for STRs)
- selection, i.e we implicitly assumed approx all genotypes were equally viable
- finite population size (ignoring ‘genetic drift’)
- population structure, e.g, ethnic/racial/geographic subdivision
- realities of human populations overlapping generations family structure (dependence), etc.

We explicitly ruled out “assortative mating” in relation to the locus under study.

HWE in practice: usually o.k , frequently not. In reality, we don’t know until we look.

Possible reasons for failure of **HWE** : See above, & genotyping errors (see later).

Recombination fractions

One locus(Figure 1(1)) can be generalizing to two loci(Figure 1(2)) in the notion of recombination and linkage. We need a story for the transmission from parents to gametes of variants of two loci (Generalization of the second law of Mendel).

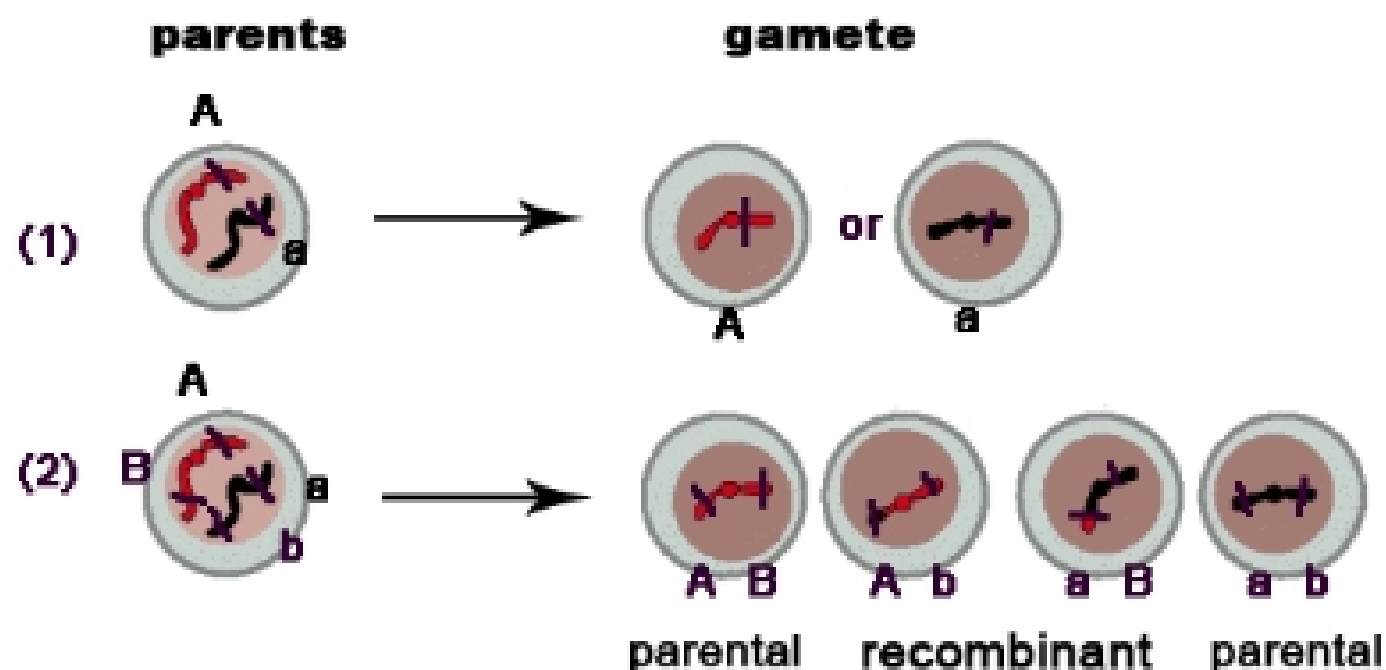


Figure 1: (1) One locus, gamete can be A or a with probability $\frac{1}{2}$ (2) two loci, the first and last gamete are parental and middle ones are recombinant at these two loci.

Figure 1 (2) Parental : $AB[\frac{1}{2}(1-r)]$, $ab[\frac{1}{2}(1-r)]$, recombinant : $Ab[\frac{1}{2}r]$, $aB[\frac{1}{2}r]$

r : recombination fraction between the two loci. (some people use θ , others use R), $r \leq \frac{1}{2}$

$r = \frac{1}{2}$ for loci on different chromosome (unlinked)

$r < \frac{1}{2}$ (linked).

Terminology

Haplotype : allelic assignment at 2 or more linked loci

ex) -A-B- -A-b- -a-B- -a-b-

The haplotype analogue of **HWE** for 2 loci is called (badly) **Linkage Equilibrium (LE)**.

The haplotype frequencies in the population are products of the associated allele frequencies, i.e, independence across loci of allele.

frequency f_{AB} of haplotype $AB = p_A \times q_B$, where p_A = allele frequency of A , q_B = allele frequency of B.

Similarly under **LE** , $f_{Ab} = p_A q_b$, etc.

Unlike the simple derivation of **HWE** where equilibrium is reached in 1 round of random mating (and no mutation, etc), **LE** is only reached asymptotically.

Let $f_{ij}(t)$ = frequency of haplotype $a_i b_j$ in the population at generation t . (See Figure 2), where alleles at A are a_1, a_2, \dots, a_i and the alleles at B are b_1, b_2, \dots, b_j with frequencies $\{p_i\} \{q_j\}$ respectively.

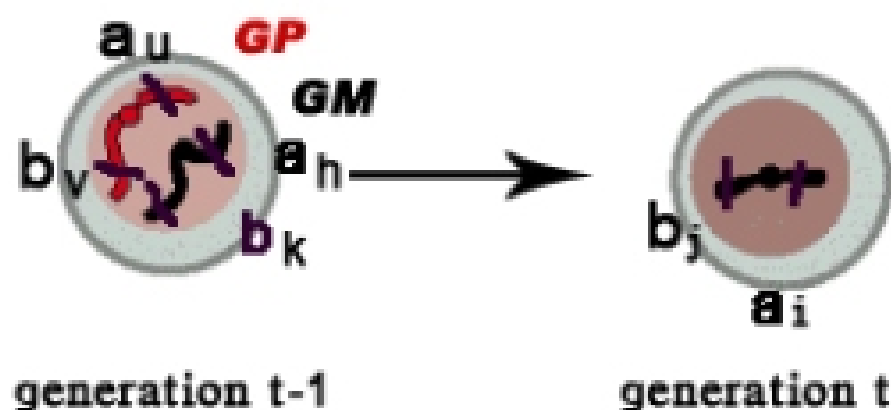


Figure 2: Under Linkage Equilibrium, GM : Grand maternal, GP : Grand paternal

Two cases R : recombination occurs , NR : no recombination occurs in producing the gamete.

In this 2 locus context, the random mating assumption is on pairs of allele of the loci. Let H denote the gamete's haplotype.

$$\begin{aligned}
& pr(H = a_i b_j \text{ at generation } t) \\
&= \sum_{GM, GP} pr(GM, GP, R, H = a_i b_j) + \sum_{GM, GP} pr(\dots, NR, \dots) \\
&= \sum_{GM, GP} f_{GM}(t-1) \times f_{GP}(t-1) \times \tau \times pr(H = a_i b_j | GM, GP, R) + \sum_{GM, GP} \text{similar with NR} \\
&\hspace{15em} (1) \\
&= \sum_{GM, GP} f_{GM}(t-1) \times f_{GP}(t-1) \times \tau \times pr(\dots) + \sum_{GM, GP} f_{GM}(t-1) \times f_{GP}(t-1) \times \frac{1}{2}(1-\tau) \times pr(\dots) \\
&\hspace{15em} (2) \\
&= \tau \times p_i q_j + (1-\tau) \times f_{ij}(t-1), i.e
\end{aligned}$$

$$f_{ij}(t) = (1-\tau)f_{ij}(t-1) + \tau \times p_i q_j$$

$$\begin{aligned}
f_{ij}(t) - p_i q_j &= (1-\tau)(f_{ij}(t) - p_i q_j) \\
&= (1-\tau)^t (f_{ij}(0) - p_i q_j)
\end{aligned} \hspace{10em} (3)$$

Equation (1) : $\tau = pr(rec. | GM, GP)$

Equation (2) :

first term $f_{GM}(t-1) : -i - k -$ and $f_{GP}(t-1) : -u - j -$ or

$f_{GM}(t-1) : -h - j -$ and $f_{GP}(t-1) : -i - v -$

second term $f_{GM}(t-1) : -i - j -$ and $f_{GP}(t-1) : -u - v -$ or

$f_{GM}(t-1) : -h - k -$ and $f_{GP}(t-1) : -i - j -$

Hence independence in the limit.