RECOMBINATION AND HARDY-WEINBERG READING: Nielsen & Slatkin, pp. 107-116

- Aim: To understand how **linkage** (and genetic **recombination**) between different genetic loci affects frequencies of multilocus genotypes (genome structure), single locus genotypes, and their constituent alleles.

MULTIPLE LOCI, LINKAGE, AND RECOMBINATION

- Expression of most characters depends on more than 1 locus

- Saw that, assuming H-W conditions, diploid genotype frequencies can be found from allele frequencies after 1 generation of random mating.

- **Questions**: Given H-W conditions

  (1) Can multilocus **gamete** frequencies be computed using single locus allele frequencies?

  (2) If so, will this be the case after a single round of random mating?

- Consider the simplest diploid case: 2 loci ("A" = flower color and "B" = flower shape), 2 alleles each (*A, a* and *B, b*)

  – 4 possible gametes: *AB*, *Ab*, *aB*, *ab*

  - gamete frequencies P_{AB} , P_{Ab} , P_{aB}, P_{ab} (Note: $P_{AB} + P_{Ab} + P_{aB} + P_{ab} = 1$)

  – 10 possible genotypes: *AB*/*AB, AB*/*Ab,* etc.

  - ***Practice* Exercises**:
    1) List the other 8 possible genotypes.
    2) Show that there are 16 genotypes if maternally and paternally inherited gametes can be distinguished.

  – Can the situation can be simplified?

  - Treat each gamete as a different allele (with names "AB","Ab", etc.)
    – then, if parents mate randomly, offspring genotype frequencies will be Freq(*Ab*/*Ab*)= P_{Ab}^2 , Freq(*AB*/*Ab*)=2 P_{AB} P_{Ab} , etc.

  - Lesson: with random mating, need to keep track of just 4 (or 3 independent) *gamete* frequencies to follow the 2-locus genotype frequencies

  – **Question**: Can we describe two-locus gamete frequencies with just 2 allele frequencies?
    p_A , p_a = allele frequencies at locus A; p_B , p_b = allele frequencies at locus B.
    - Note: $p_A = P_{AB} + P_{Ab}$, $p_B = P_{AB} + P_{aB}$, etc.

Answer: In general, no unless the population is in a state of **linkage equilibrium** in which case:

$P_{AB} = p_A p_B$, $P_{Ab} = p_A p_b$, $P_{aB} = p_a p_B$, $P_{ab} = p_a p_b$.

i.e., frequency of each gamete = product of frequencies of constituent alleles.

– **Question** : Are random mating populations in linkage equilibrium?

Answer Not necessarily. Consider, e.g., a random mating population with $P_{AB} = \frac{1}{2}$, $P_{Ab} = 0$, $P_{aB} = 0$, $P_{ab} = \frac{1}{2}$.

Then $p_A = p_a = p_B = p_b = \frac{1}{2}$, but $P_{AB} = \frac{1}{2} \neq p_A p_B = \frac{1}{4}$, $P_{Ab} = 0 \neq p_A p_b = \frac{1}{4}$, etc.

– Under H-W conditions populations will approach linkage equilibrium.

– Consider how this occurs...

• First, need to measure a population's degree of **linkage disequilibrium** (note, terminology is problematic - but well entrenched in the literature).

– measured by a magic number called D (the "coefficient of disequilibrium") which is defined as follows:

$$D = P_{AB} - p_A p_B = P_{ab} - p_a p_b = p_A p_b - P_{Ab} = p_a p_B - P_{aB}$$
$$= P_{AB} P_{ab} - P_{Ab} P_{aB}$$

(Yes, these definitions are all mathematically interchangeable: see Nielsen & Slatkin, p. 110.)

– If $D = 0$, then $P_{AB} = p_A p_B$, $P_{Ab} = p_A p_b$, etc. (i.e., linkage equilibrium)

– In example above, $D = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) - 0 \cdot 0 = \frac{1}{4}$;

– D can be negative or positive;

• Upper and low limits of D depend on allele frequencies at both loci:
Lower: larger of $-p_A p_B$ and $-p_a p_b$
Upper: smaller of $p_A p_b$ and $p_a p_B$

• E.g., if $p_A = 1/4$, $p_B = 1/2$, then $-1/8 \leq D \leq 1/8$.

• Widest limits on D occur when $p_A = p_B = 1/2$: $-1/4 \leq D \leq 1/4$

- ***Practice* Exercise**: What are the <u>gamete</u> frequencies when D is at its extremes in these cases?

- Will now show that, under H-W conditions, $D \to 0$.

- Recursion for D: $D = P_{AB} - p_A p_B$; $D' = P'_{AB} - p'_A p'_B$

  – Under H-W, allele frequencies don't change so $p'_A = p_A$, $p'_B = p_B$
  – Just need to know how P_{AB} changes.

- <u>DIVERSION</u>: Salient features of genetic <u>recombination</u>:

  - Consider individual with genotype Ab/aB
  - let r = the rate of recombination between locus A and locus B
  - What gametes are produced? What ratios?
    Ans. All 4: $(1-r)/2\ Ab : (1-r)/2\ aB : r/2\ AB : r/2\ ab$
  - if loci are very close on a chromosome, then $r \approx 0$ ("tightly linked")
  - if loci are far apart or on different chromosomes, $r = \frac{1}{2}$ ("loosely linked"/"unlinked")

- <u>Back to our story</u>, describing changes in P_{AB} (remember the goal is to find D')

- Consider the frequencies of parents that can produce AB gametes and the fraction of their gametic output which actually consists of AB gametes: **Handout I.3 Two-locus gamete production**

- Observation: Gamete frequencies affected by recombination only in <u>double heterozygotes</u>

- Adding up the 3rd column and simplifying:

$$
\begin{aligned}
P'_{AB} &= P_{AB}^2 + P_{AB}P_{Ab} + P_{AB}P_{aB} + (1-r)P_{AB}P_{ab} + rP_{Ab}P_{aB} \\
&= P_{AB}\left(P_{AB} + P_{Ab} + P_{aB} + P_{ab}\right) - r\left(P_{AB}P_{ab} - P_{Ab}P_{aB}\right) \\
&= P_{AB} - rD
\end{aligned}
$$

- Finally

  $D' = P'_{AB} - p'_A p'_B = \left(P_{AB} - rD\right) - p_A p_B = D - rD$

  $D' = (1-r)D$

  – Important: Derivation assumes parental population itself was formed by random mating.
  – Similar reasoning shows that $P'_{Ab} = P_{Ab} + rD$, $P'_{aB} = P_{aB} + rD$, and $P'_{ab} = P_{ab} - rD$.

- <u>Summary</u> :

(1) allele frequencies p_A, p_B don't change
(2) gamete frequencies can increase or decrease
(3) Linkage disequilibrium D decreases by a factor $(1-r)$ each generation.

• Implications:

(1) Genetic equilibrium is not reached in 1 generation (contra H-W equilibrium for single locus)—even if loci are on different chromosomes ($r = 0.5$) !!

(2) As long as $r > 0$, $D \to 0$.
– D does not oscillate toward zero
– Rate of approach depends on r:
  • If $r = 0.5$, D will have only 3% of its original value after 5 generations

  • If $r = 0.05$, D will still have 77% of its original value after 5 generations

• What is the MEANING of D?
– measures *statistical* rather than *physical* association between alleles at different loci
  • $D = 0 \Rightarrow$ "no statistical association between loci"
    – i.e., if sampled gamete has A allele, chance it carries B allele is p_B.
  • $D = 1/4 \Rightarrow$ gamete with A (a) will also carry B (b)
– D can be viewed as the covariance between alleles at A and B loci

• Why doesn't recombination instantly randomize things like segregation did?
– **Punch line**: Approach to linkage equilibrium limited by the number of *double heterozygotes* .

• Why allele frequencies and D are more useful than gamete frequencies
[NOTE: descriptions are mathematically equivalent.]

(1) If population is in linkage equilibrium, its genetic composition is easier to describe using allele frequencies and D vs. gamete frequencies.

(2) Easier to comprehend evolution in terms of changes in p_A, p_B, and D versus gamete frequencies:

$$P'_{AB} = \left(P_{AB} + P_{Ab}\right)\left(P_{AB} + P_{aB}\right) + (1-r)\left(P_{AB}P_{ab} - P_{Ab}P_{aB}\right), \text{etc.}$$

vs.

$$p'_A = p_A, \quad p'_B = p_B, \quad D' = (1-r)D$$

• Biological Implications of D:

(1) If $D \neq 0$, events affecting one locus will incidentally affect the other locus.

(2) $D \neq 0$ may reveal a population's history.