

Shannon Quinn
Feb 18, Scribe 11
Linkage Disequilibrium

Allele Frequencies

- Two allele loci: L_A and L_B
- $L_A \Rightarrow A$ or a
- $L_B \Rightarrow B$ or b

Allele Probabilities

- $P(A) + P(a) = 1$
- $P(B) + P(b) = 1$

Haplotype Frequencies

	B	b
A	$P(AB)$	$P(Ab)$
a	$P(aB)$	$P(ab)$

- Important to note from this matrix
 - 1st row: $P(AB) + P(Ab) = P(A)$
 - 2nd row: $P(aB) + P(ab) = P(a)$
 - 1st column: $P(AB) + P(aB) = P(B)$
 - 2nd column: $P(Ab) + P(ab) = P(b)$
 - $P(A) + P(a) = P(B) + P(b) = 1$

Under linkage equilibrium: $P(AB) = P(A) * P(B)$

- Since we assume the alleles are independent

Disequilibrium coefficient: D_{AB}

- Measures the degree to which linkage equilibrium does not hold

$$D_{AB} = P(AB) - P(A) * P(B)$$

- If D_{AB} is large, the observed haplotype frequency $P(AB)$ is very different from the expected frequency $P(A) * P(B)$, and linkage equilibrium does not hold
- We can write haplotype frequencies as a combination of D_{AB} and allele frequencies

$$P(AB) = P(A) * P(B) + D_{AB}$$

$$P(Ab) = P(A) * P(b) - D_{AB}$$

$$P(aB) = P(a) * P(B) - D_{AB}$$

$$P(ab) = P(a) * P(b) + D_{AB}$$

Range of D_{AB}

- If linkage equilibrium (LE) holds, $D_{AB} = 0$

- $\max(-P(A) * P(B), -P(a) * P(b)) \leq D_{AB} \leq \min(P(A) * P(b), P(a) * P(B))$
- range of D_{AB} depends on allele frequencies
- sign of D_{AB} is arbitrary

Examples

$$P(A) = 0.3, P(a) = 0.7$$

$$P(B) = 0.3, P(b) = 0.7$$

Theoretical bounds of D_{AB}

- $[-0.09 \leq D_{AB} \leq 0.21]$

$$P(A) = 0.2, P(a) = 0.8$$

$$P(B) = 0.1, P(b) = 0.9$$

Theoretical bounds of D_{AB}

- $[-0.02 \leq D_{AB} \leq 0.08]$

Normalizing D_{AB} : D'_{AB}

- scaled version of D_{AB}
- has a range of: $[-1 \leq D'_{AB} \leq 1]$

If $D_{AB} < 0$

$$D'_{AB} = D_{AB} / |\text{lower bound of } D_{AB}|$$

If $D_{AB} > 0$

$$D'_{AB} = D_{AB} / |\text{upper bound of } D_{AB}|$$

$D'_{AB} = \pm 1$ indicates no evidence of recombination, or very strong linkage disequilibrium (LD)

- this implies $D_{AB} = |\text{lower bound}|$
- also implies $P(AB) = 0$, as one of the four haplotypes did not occur

Can also normalize D_{AB} with r^2 , or Δ^2

$$r^2 = D_{AB}^2 / [P(A) * P(B) * P(a) * P(b)]$$

- r^2 does not have inherent bias that D'_{AB} does
- not dependent on haplotype frequencies
- has range of: $[0 \leq r^2 \leq 1]$
- $r^2 = 0$ implies LE ($D_{AB} = 0$)
- $r^2 = 1$ implies strong LD (no recombination; two haplotypes did not occur at all)

When does LE hold?

- Distance between markers/loci
 - Longer distances tend to imply more recombination
- Recombination rate
 - If this is slow enough, loci separated by large distances may still be under LD
- Random mating
- Genetic drift

LE should be reached eventually

r = recombination rate (note: not related to r^2)

- After N generations, recombination can be modeled with the equation:
$$D_{AB}^{(t+N)} = (1-r)^N * D_{AB}^t$$
- Depending on r , D_{AB} will decrease
 - Decay will be determined by r
- High recombination rate (r is close to 1) implies $1-r$ will be small, and LE will be quickly reached between loci A and B