

Evolution at two loci

April 12, 2010

Haplotypes and genotypes

- Consider two diploid, biallelic loci. The first locus has alleles A and a , and the second one alleles B and b .
- There are four **haplotypes** (AB , Ab , aB and ab) and 16 diploid genotypes (which reduce to 10 if we do not distinguish between, e.g., $AB|ab$ and $Ab|aB$).
- Assuming random mating, the haplotypes combine to genotypes following Hardy-Weinberg proportions.
- Thus, we can focus on tracking haplotype frequencies X_{AB} , X_{Ab} , X_{aB} and X_{ab} .

Recombination

- Even in the absence of selection, haplotype frequencies change due to recombination.
- Recombination occurs due to chromosome assortment and cross-over during meiosis.
- Let r be the **recombination rate**.
- $r = 0.5$ for loci on different chromosomes.
- $r \leq 0.5$ for loci on the same chromosome.

Modeling recombination

After one round of recombination, the frequency of haplotype AB changes to

$$\begin{aligned}X'_{AB} &= X_{AB}^2 + \frac{1}{2}2X_{AB}X_{Ab} + \frac{1}{2}2X_{AB}X_{aB} \\ &\quad + \frac{1}{2}2X_{AB}X_{ab}(1-r) + \frac{1}{2}2X_{Ab}X_{aB}r \\ &= X_{AB}(X_{AB} + X_{Ab} + X_{aB} + X_{ab}) - r(X_{AB}X_{ab} - X_{Ab}X_{aB}) \\ &= X_{AB} - r(X_{AB}X_{ab} - X_{Ab}X_{aB}) = X_{AB} - rD\end{aligned}$$

The quantity $D = X_{AB}X_{ab} - X_{Ab}X_{aB}$ is called **linkage disequilibrium**.

Linkage disequilibrium

D can be written in three alternative ways:

- $D = X_{AB}X_{ab} - X_{Ab}X_{aB}$ (“pure” vs. “mixed” haplotypes)
- $D = X_{AB} - p_A p_B$, the difference between the true frequency of a “pure” haplotype and the one expected without LD (here, p_A and p_B are allele frequencies).

Proof: Use the fact that $p_A = X_{AB} + X_{Ab}$, $p_B = X_{AB} + X_{aB}$.

- $D = \text{Cov}(A, B)$, the covariance between the alleles at the two loci (if alleles A and B are coded as 1 and alleles a and b as 0).

Proof: Follows directly from the definition of covariance, $\text{Cov}(x, y) = E(xy) - E(x)E(y)$.

Linkage equilibrium

The opposite of linkage disequilibrium (LD) is **linkage equilibrium** (LE). Under LE, the two loci are independent:

$$X_{AB} = p_A p_B$$

$$X_{Ab} = p_A q_B$$

$$X_{aB} = q_A p_B$$

$$X_{ab} = q_A q_B$$

where $q = 1 - p$.

Linkage versus linkage disequilibrium

The usual terminology may lead to confusion.

- Two loci are **linked**, when they are on the same chromosome (physical linkage). Otherwise, they are **unlinked**.
- Two loci are at **linkage disequilibrium**, when their allele frequencies are not independent.

Note that two linked loci can be linkage equilibrium, and two unlinked loci can be at linkage disequilibrium.

It is for this reason that Rice (2004) prefers the alternative term **gametic (dis)equilibrium**.

Back to evolution of haplotype frequencies

Repeating the previous calculation for all four haplotypes, we find that

$$X_{AB} = p_A p_B + D$$

$$X_{Ab} = p_A q_B - D$$

$$X_{aB} = q_A p_B - D$$

$$X_{ab} = q_A q_B + D$$

$$X'_{AB} = X_{AB} - rD$$

$$X'_{Ab} = X_{Ab} + rD$$

$$X'_{aB} = X_{aB} + rD$$

$$X'_{ab} = X_{ab} - rD$$

before and after recombination, respectively.

Evolution of linkage disequilibrium

Using the above definitions results, we get

$$\begin{aligned}D' &= X'_{AB}X'_{ab} - X'_{Ab}X'_{aB} \\&= (X_{AB} - rD)(X_{ab} - rD) - (X_{Ab} + rD)(X_{aB} + rD) \\&= X_{AB}X_{ab} - X_{Ab}X_{aB} - rD(X_{AB} + X_{Ab} + X_{aB} + X_{ab}) \\&= D - rD \\&= D(1 - r)\end{aligned}$$

Thus, without selection, LD decreases exponentially to 0 (but unlike Hardy-Weinberg equilibrium, it is not reached immediately!).

Recombination and selection

With selection, the notation starts becoming somewhat complicated. Let genotype fitnesses be denoted by, e.g., $W_{AB|AB}$. As in the one-locus case, we can define the **marginal fitness** of a haplotype as, e.g.,

$$W_{AB}^* = X_{AB} W_{AB|AB} + X_{Ab} W_{AB|Ab} + X_{aB} W_{AB|aB} + X_{ab} W_{AB|ab}.$$

The **mean fitness** is then the weighted average of the marginal fitnesses:

$$\bar{W} = X_{AB} W_{AB}^* + X_{Ab} W_{Ab}^* + X_{aB} W_{aB}^* + X_{ab} W_{ab}^*.$$

Haplotype frequency change

After one round of selection and recombination, we get

$$X'_{AB} = \frac{1}{\bar{W}} \left[X_{AB} W_{AB}^* - rX_{AB}X_{ab} W_{AB|ab} + rX_{aB}X_{Ab} W_{aB|Ab} \right]$$

$$= \frac{1}{\bar{W}} (X_{AB} W_{AB}^* - rDW_H)$$

$$X'_{Ab} = \frac{1}{\bar{W}} (X_{Ab} W_{Ab}^* + rDW_H)$$

$$X'_{aB} = \frac{1}{\bar{W}} (X_{aB} W_{aB}^* + rDW_H)$$

$$X'_{ab} = \frac{1}{\bar{W}} (X_{ab} W_{ab}^* - rDW_H)$$

where $W_H = W_{AB|ab} = W_{aB|Ab}$ is the fitness of the **double heterozygote** (assuming absence of parental effects).

Haplotype frequency change

After one round of selection and recombination, we get

$$X'_{AB} = \frac{1}{\bar{W}} \left[X_{AB} W_{AB}^* - rX_{AB}X_{ab} W_{AB|ab} + rX_{aB}X_{Ab} W_{aB|Ab} \right]$$

$$= \frac{1}{\bar{W}} (X_{AB} W_{AB}^* - rDW_H)$$

$$X'_{Ab} = \frac{1}{\bar{W}} (X_{Ab} W_{Ab}^* + rDW_H)$$

$$X'_{aB} = \frac{1}{\bar{W}} (X_{aB} W_{aB}^* + rDW_H)$$

$$X'_{ab} = \frac{1}{\bar{W}} (X_{ab} W_{ab}^* - rDW_H)$$

where $W_H = W_{AB|ab} = W_{aB|Ab}$ is the fitness of the **double heterozygote** (assuming absence of parental effects).

Generalization to multiple loci

In a general multilocus model, the haplotype dynamics are given by

$$X'_k = \frac{\sum_{i,j} X_i X_j W_{ij} R(i, j \rightarrow k)}{\bar{W}}$$

where $R(i, j \rightarrow k)$ is the probability that genotype $i|j$ produces gametes with haplotype k .

When is linkage disequilibrium important?

- LD creates genetic correlations between traits.
- Often, this is important for overall adaptation (leading to **coadapted gene complexes**).
- Many models of non-allopatric speciation try to explain how LD can evolve between
 - A male display trait and a female preference.
 - A trait conferring local adaptation and a trait affecting mate choice.
 - A trait conferring adaptation to a habitat and preference for this habitat.

Ignoring or approximating LD

- If selection is much weaker than recombination, LD can often be ignored (**weak-selection approximation**). Then, the system can be described by following allele frequencies rather than haplotype frequencies. The validity of the approximation can be checked by simulations.
- If one is specifically interested in LD, but selection is weak, one can calculate a **quasi-linkage-equilibrium** (QLE, i.e. a small amount of LD that changes much slower than allele frequencies).

Evolution of LD

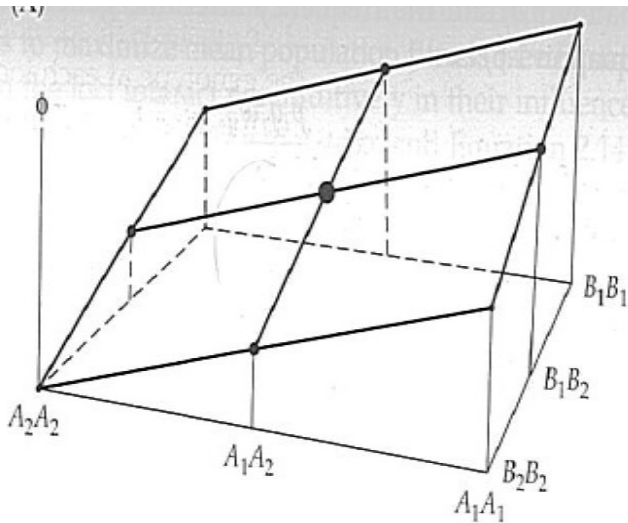
What happens to linkage disequilibrium under selection?

- If fitnesses are **additive** (i.e. $W_{ij|kl} = \alpha_{ik} + \beta_{jl}$), LD = 0 at equilibrium.
- Only in this case is mean fitness guaranteed to be maximized.
- With non-additive fitnesses (i.e. **epistasis**), LD is maintained, and mean fitness is not maximized.

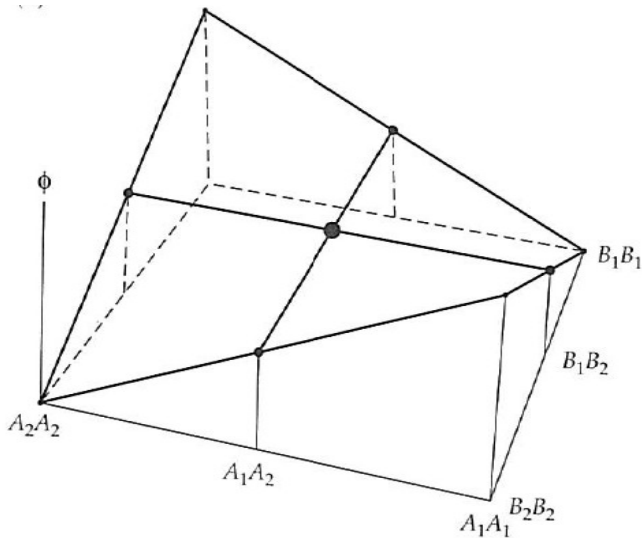
Epistasis

- “**Functional epistasis** occurs whenever the effect of a genetic substitution (on one or multiple loci) depends on the **genetic background** (i.e., the state of other loci in the genotype)” (Hansen and Wagner 2001). It arises as a “result of physical interactions among biomolecules within gene regulatory networks and biochemical pathways ” (Moore and Williams 2005) and is a property of the **genotype-phenotype map**.
- “**Statistical epistasis** is defined as deviation from additivity in a mathematical model summarizing the relationship between multilocus genotypes and phenotypic variation in a population” (MW 2005).

No (functional) epistasis

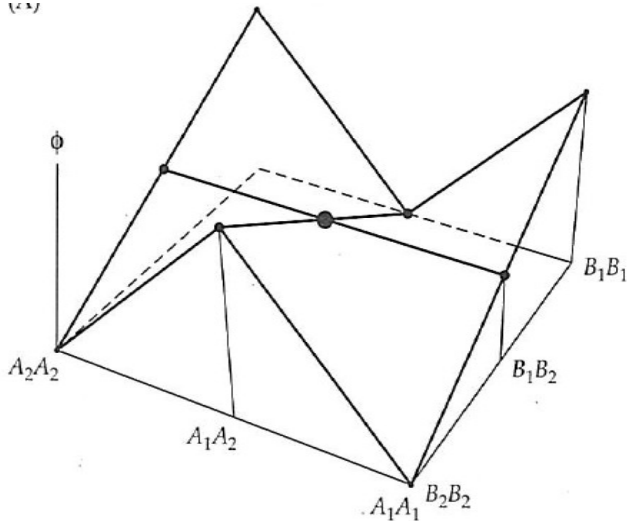


Additive \times additive epistasis

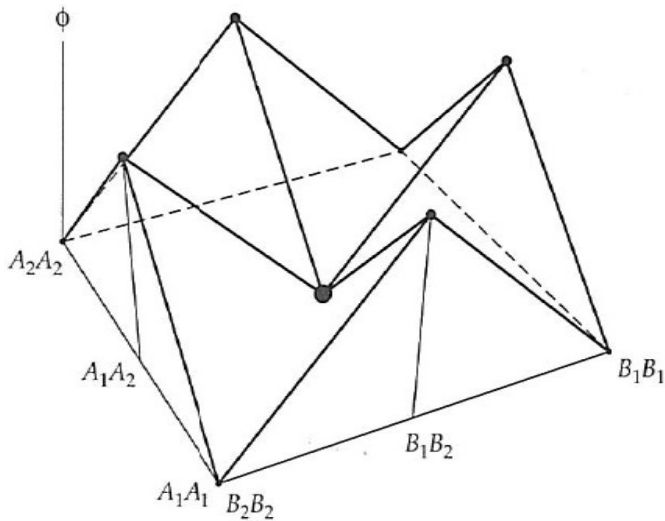


Additive \times dominance epistasis

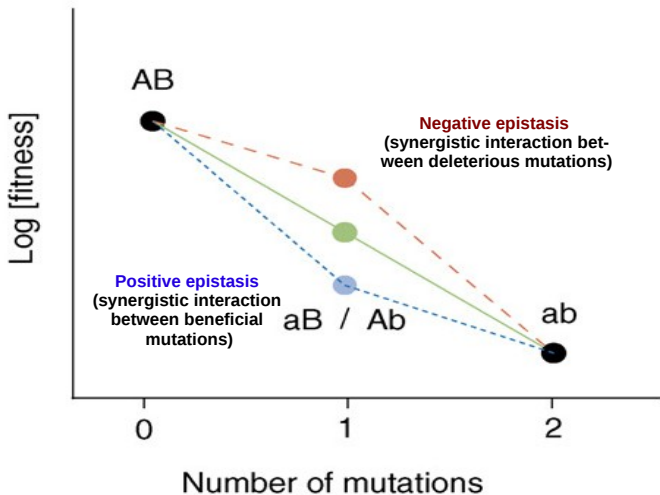
(c)



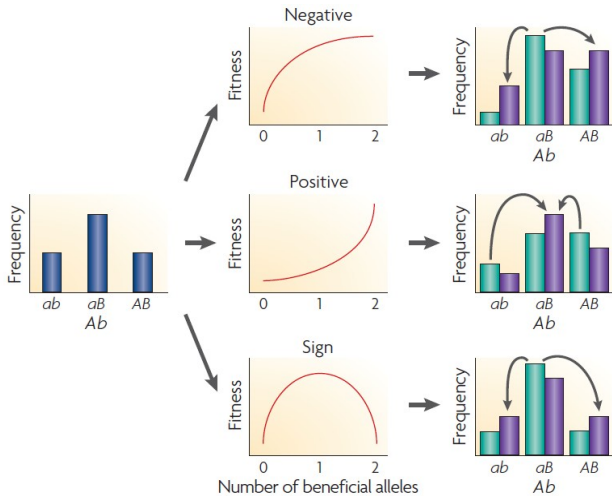
Dominance \times dominance epistasis



Positive and negative epistasis



Epistasis, LD and recombination



Epistasis, LD and recombination

Different forms of epistasis have different effects:

- **Positive epistasis creates positive LD**
(overrepresentation of extreme haplotypes).
Recombination decreases genetic variance.
- **Negative epistasis creates negative LD**
(overrepresentation of intermediate haplotypes).
Recombination increases genetic variance.

One theory for the evolution of sexual reproduction states that recombination increases the rate of evolution if there is negative LD (due to negative epistasis). However, there is little evidence that epistasis is predominantly negative.