

Published on *Plant Breeding E-Learning in Africa* (<u>https://pbea.agron.iastate.edu</u>) <u>Home</u> > <u>Course Materials</u> > <u>Quantitative Genetics</u> > Linkage

Linkage



By William Beavis (ISU)

(CC) EXCEPT otherwise noted, this work is licensed under a <u>Creative Commons Attribution-NonCommercial 4.0 International License</u>.

Objectives

- Demonstrate understanding that linkage and linkage disequilibrium are properties of populations, not individuals.
- Distinguish gamete from linkage disequilibrium
- Demonstrate ability to estimate recombination and disequilibrium statistics

Introduction

Plant breeding populations, by definition, employ methods that force populations into states of disequilibrium. Plant breeders do not mate infinite (or even large) numbers of parents, thus drift has a major impact on population disequilibrium. They select the parents that will be used in matings, thus selection, linkage and pleiotropy affect the population structure. New lines from external breeding projects are often introduced to the breeding nurseries, thus migration affects the structure of plant breeding populations. After the passage of the plant variety protection act, plant breeders working in the commercial sector began to keep breeding records for purposes of protecting intellectual property. An unintended consequence has been the application of linear mixed models to produce predictors of performance, originally developed by animal breeders. These methods are predicated on use of coefficients of relationship among cultivars with known performance and progeny with unknown or limited information on performance.

Herein we introduce gametic and linkage disequilibrium as measures of deviation (disequilibrium) from Hardy-Weinberg Equilibrium. In other words the estimation of these population parameters are based on a reference population and the reference population must be defined or else the calculated values have no meaning.

Disequilibrium Definitions

The motivation is to 'map' genetic loci based on how they are most likely to be inherited relative to each other. If alleles at two loci are on the same chromosome in close proximity to each other, then they will be inherited together more often than not. It was recognized in the 1920's (Sax, 1923) that markers could have value for selecting phenotypes that are difficult to assay, but 60 years passed before the theory could be evaluated on a genome-wide scale. Linkage represents a mechanism that results in Disequilibrium among alleles at more than a single locus on the same chromosome. It is also possible that Disequilibrium among alleles at more than a single locus can result from mechanisms other than linkage, e.g., selection and drift. Unfortunately, the term "linkage disequilibrium" has been applied to all forms of multi-locus disequilibrium. Herein we try to use the term "linkage disequilibrium" only for cases where we know alleles are on the same chromosome and "gametic disequilibrium" for situations when we do not know whether the loci are on the same chromosome.

Disequilibrium Example

Consider parent 1 with genotype A1A1B1B1C1C1D1D1 and parent 2 with A2A2B2B2C2C2D2D2. Loci A, B and C are on a homologous chromosome and D on the separate chromosome (Fig. 1).



Fig. 1 A, B, C and D loci on two pairs of homologous chromosomes.

The genotype of the F1 generation resulting from the cross between parent 1 and 2 will be A1A2 B1B2 C1C2 D1D2. Loci A and D are located on different chromosomes and will segregate independently according to random segregation of chromosomes into gametes. For two different alleles at each locus four possible combinations can occur, each with a chance of 25%. A and C are unlinked on the same chromosome they are so far away from each other, that recombination occurs between them in 50% of the meioses. The frequencies of all gametes involving alleles at the A and C locus (A1C1, A1C2, A2C1, A2C2) is 0.25; just as it is for the alleles for the A and D loci and the B and D loci . Since locus A and C assort independently, the frequency of double homozygous dominant and double homozygous recessive genotypes (A1A1C1C1, A2A2C2C2) is 0.25x0.25 and frequency of double heterozygous genotypes (A1A2C1C2) is 0.5 x 0.5.

Loci A and B are linked because they are located in close proximity on the same chromosome resulting in recombination frequencies that are less than 0.5, e.g., 0.1. The difference between expectation for unlinked loci, and the estimated recombination frequency can be used to classify linkage, i.e., the likelihood of two loci being inherited together. To estimate recombination frequencies non parental gametes can be counted and divided by the total number of gametes.

Gametic Disequilibrium

Disequibrium can be created by self pollination, crossing relatives within a breeding population, mutation, drift, selection and migration. For example, consider alleles at loci A and D. Let's assume that each contribute to phenotypic variability in flower initiation in an additive manner. Let's assume selection for earlier flowering (conferred by the A1 and D2 alleles). The impact will be a negative covariance between the alleles at loci A and D , which redues the genetic variances and creates disequilibrium between those loci. Even though A1 and D2 alleles are physically independent, they become correlated by selection which results in $D_{A1,D2} > 0$. This also is reffered to as the Bulmer effect.

Although individual loci achieve HWE after one generation of random mating, genotype frequencies at two or more loci do not achieve equilibrium jointly after one generation of random mating.

To illustrate this point, consider two populations, one consisting of entirely AABB genotypes and the other consisting entirely of aabb genotypes. Assume they are mixed equally and allowed to randomly mate. The first generation would consist of the three genotypes AABB, AaBb, and aabb in the proportions $\frac{1}{4}$: $\frac{1}{2}$: $\frac{1}{4}$. However, for two loci with two alleles, nine genotypes are possible.

(For n alleles at each locus and k loci, there are $(rac{n(n+1)}{2})^k$ possible genotypes).

Continued random mating would produce the missing genotypes, but they would not appear at the equilibrium frequencies immediately.

Disequilibrium Table

Consider the following table:

Alleles	А	а	В	b
Allele Frequencies	р _А	1-p _A	р _В	1-p _B
Gametic Types	AB	Ab	aB	ab
Frequencies at Equilibrium	Рав	р _А (1-р _В)	(1-p _A)p _{AB}	(1-p _A)(1-p _B)
Actual Frequeicies	R	S	Т	U
Difference from Equilibrium	D _{AB}	-D _{AB}	-D _{AB}	+D _{AB}

A coupling heterozygote would be AB/ab and occurs with frequency 2RU and the repulsion heterozygote would be Ab/aB occurring with frequency 2ST. If the frequency of these two genotypes is equal, the population is in equilibrium and:

D = RU - ST = 0

It can be shown that after t generations of random mating, the disequilibrium is given by:

$$D_t = D_0 \left(1 - c\right)^t$$

where, D_0 is the disequilibrium in generation 0 and c is the recombination frequency between the loci, where c = $\frac{1}{2}$ for independently segregating loci.

Dissipation of Disequilibrium

The dissipation of disequilibrium relative to generation 0 is given in the figure below:



Fig. 2 Dissipation of disequilibrium.

Deviations from independence at multiple loci are often referred to as linkage disequilibrium, even if linkage is not the cause. Unless two loci are known to reside on the same chromosome the term Gametic Disequilibrium is a less ambiguous term to describe disequilibrium among loci.

Estimation and Testing

Disequilibrium at the A and B loci is a comparison of gametic frequency, p_{AB} , with the product of allele frequencies, $p_A p_B$; i.e.,

Invalid Equation

Note the similarities to \hat{D}_{A} . Thus,

Invalid Equation

Chi-Square Statistic

Again a chi-square statistic for the hypothesis of no disequilibrium can be based on

$$Z^{2} = \chi^{2}_{AB} = \frac{2nD^{2}_{AB}}{p_{A}(1-p_{A})p_{B}(1-p_{B})}$$

Gamete	AB	AB	AB	AB
Observed	n _{AB}	n _{Ab}	n _{aB}	n _{AB}
Expected	$2n\hat{p}_A\hat{p}_B$	$2n\hat{p}_A\hat{p}_b$	$2n\hat{p}_a\hat{p}_B$	$2n\hat{p}_a\hat{p}_b$

Acknowledgements

This module was developed as part of the Bill & Melinda Gates Foundation Contract No. 24576 for Plant Breeding E-Learning in Africa.

Quantitative Genetics Linkage Author: William Beavis (ISU)

Multimedia Developers: Gretchen Anderson, Todd Hartnell, and Andy Rohrback (ISU)

How to cite this module: Beavis, W. 2016. Linkage. *In* Quantitative Genetics, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from <u>https://pbea.agron.iastate.edu</u>.

Source URL: https://pbea.agron.iastate.edu/course-materials/quantitative-genetics/linkage?cover=1