Introduction to different measures of linkage disequilibrium (LD) and their calculation

By

Dr. M. Awais Khan
University of Illinois, Urbana-Champaign

Calculation of linkage disequilibrium

To understand the calculation of linkage disequilibrium consider following example

Suppose there are two genes on Chromosome 5 of apple, each with two alleles

ACTGGTAT……GATCAACCAG
ACTCGTAT……GATCAACCAG
ACTCGTAT……GATCATCCAG

SNP1 **SNP2**

Showing only alleles for both SNPs

Alleles	SNP1	SNP2
Allele 1	G	A
Allele 2	C	T

Steps in LD calculation

For better understanding of LD calculation, it is divided into five steps

Step 1) Calculate allele frequencies

If p1 and p2 =frequency of the alleles at SNP1 and q1 and q2 =frequency of the alleles at SNP2 then in tabular form it could be written as follows

SNP1		SNP2	
Allele	Frequency	Allele	Frequency
G	p1	A	q1
C	p2	T	q2

Step 2) Calculate haplotype frequencies

From our example of two SNPs each with two alleles all possible haplotypes are

		SNP2	
	Allele	A	T
SNP1	**G**	**GA**	**GT**
	C	**CA**	**CT**

Suppose haplotype frequencies are as follows

Haplotype	**Frequency**	**Haplotype**	**Frequency**
GA	p11	GT	p12
CA	p21	CT	q22

Step 3) Linkage equilibrium

When haplotype frequencies are equal to the product of their corresponding allele frequencies, it means the loci are in linkage equilibrium

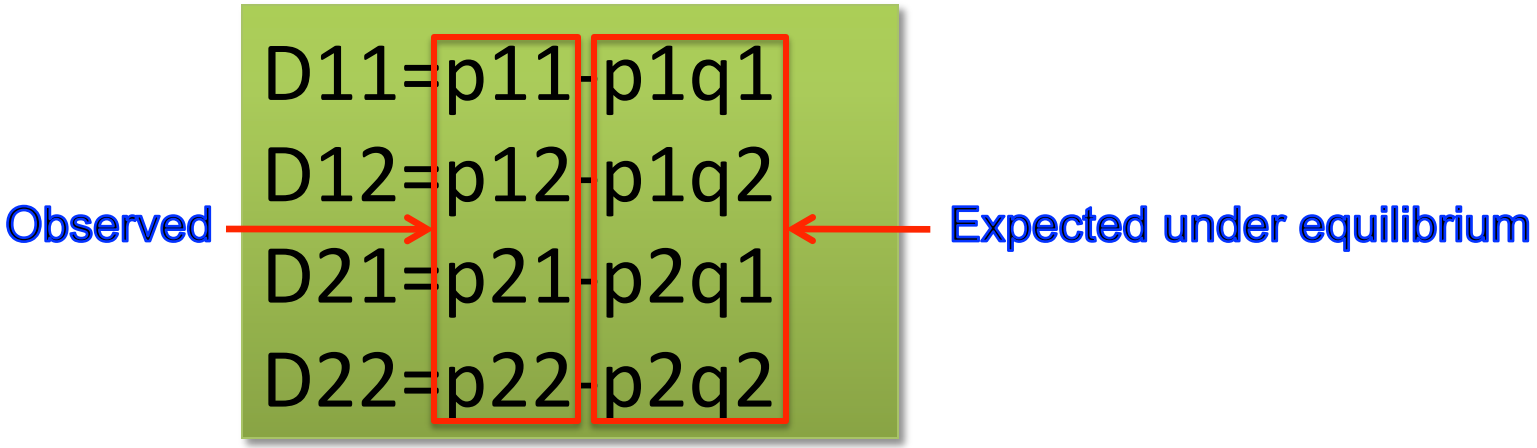
Haplotype frequency		Product of allelic frequency
p11	=	p1q1
p12	=	p1q2
p21	=	p2q1
p22	=	p2q2

Step 4) Linkage disequilibrium

We can deduce linkage disequilibrium for each haplotype as the deviation of observed haplotype frequency from its corresponding allelic frequencies expected under equilibrium

		SNP2		
		1	2	
SNP1	1	p1q1+D	p1q2-D	p1
	2	p2q1-D	p2q2+D	p2
		q1	q2	1

After solving above for D, we get as follows:



Step 5) Calculation of Linkage disequilibrium measure D

Commonly used measure of linkage disequilibrium, D equals to p11 p22 p12 p21 and we can prove it by solving the four equations from previous slide

a) $p_{11}p_{22} = (p_1 q_1 + D)(p_2 q_2 + D)$

$= p_1q_1p_2q_2 + p_1q_1D + p_2q_2D + D^2$

b) $p_{12}p_{21} = (p_1 q_2 - D)(p_2 q_1 - D)$

$= p_1q_1p_2q_2 - p_2q_1D - p_1q_2D + D^2$

c) Subtracting $(p_1q_1p_2q_2 + p_1q_1D + p_2q_2D + D^2)$ – $(p_1q_1p_2q_2 - p_2q_1D - p_1q_2D + D^2)$

$p_{11}p_{22} - p_{12}p_{21} = D(p_1 q_1 + p_2 q_1 + p_2 q_2 + p_1 q_2)$

$= D \times (1) = D$

Estimate of D in case of Linkage Equilibrium

If allele frequencies of p1 and q1 are both 0.5 and equilibrium occurs (only Ab and aB exist in the population)

$P11 = p1q1 = 0.5 \times 0.5 = 0.25$

$P22 = p2q2 = 0.5 \times 0.5 = 0.25$

$P12 = p1q2 = 0.5 \times 0.5 = 0.25$

$P21 = p2q1 = 0.5 \times 0.5 = 0.25$

$D = (P11)(P22) - (P12)(P21)$

$D = (0.25)(0.25) - (0.25)(0.25) = 0$

Estimate of D in case of Linkage Disequilibrium

If allele frequency of p1 and q1 are both 0.5 and there is complete non-random association (only AB and ab exist in the population) with equal allele frequencies at all loci

$$P11 = p1q1 + D = 0.25 + D = 0.5$$

$$P22 = p2q2 + D = 0.25 + D = 0.5$$

$$P12 = p1q2 - D = 0.25 - D = 0$$

$$P21 = p2q1 - D = 0.25 - D = 0$$

$$D = (P11)(P22) - (P12)(P21)$$

$$D = (0.5)(0.5) - (0)(0) = 0.25$$

Standardization of D

Sometimes, depending on allele frequency of two loci, the value of D can be negative, but actual gametic frequencies cannot be negative

To overcome this issue, standardization methods have been proposed

Standardization of D

In a common standardization method, a relative measure of disequilibrium (D) compared to its maximum is used:

$D' = D / Dmax$

When D is positive

$Dmax = \min [(p1q2) \text{ or } (p2q1)]$

When D is negative

$Dmax = \min [(p1q1) \text{ or } (p2q2)]$

This standardization makes D-values range between 0 and 1

Correlation coefficient as a measure of LD

Another commonly used measure to calculate LD between loci is Pearson coefficient of correlation (r)

$r = D/ (p1p2q1q2)^{1/2}$

However, squared coefficient of correlation (r^2) is often used to remove the arbitrary sign introduced

Testing significance of LD

To test if LD is statistically significant we can do a χ^2 test

$\chi^2=\Sigma(obs-exp)^2/ exp$

expected is random associations between alleles

However, r can be conveniently used for chi-test, as

$\chi^2 = r^2 N$

where N is the number of chromosomes in the sample

Example

Let's assume that we have genotypic data for the two SNPs with two alleles each (same example used to deduce the equations for different LD measures)

Genotypic data

GA = 474 GT = 611 CA = 142 CT = 773

Total = 2000

Calculation of haplotype and allele frequencies

Haplotype Frequencies

GA = 474 / 2000 = .2370
GT = 611 / 2000 = .3055
CA = 142 / 2000 = .0710
CT = 773 / 2000 = .3865

Allele frequencies

G = 0.542
C = 0.457
A = 0.308
T = 0.692

Input values in the equation for D to calculate linkage disequilibrium

D = (P11 P22) - (P12 P21)

D = (0.2370 x 0.3865) - (0.3055 x 0.0710) = 0.0699

To estimate Dmax input allelic frequencies and value for D in the following equation

Dmax = min [(p1q2) or (p2q1)]

Dmax = (0.5425 x 0.692) = 0.375

or = (0.4575 x 0.308) = 0.141

Now calculate D' input value of D and Dmax calculated in previous step in the following equation

$D' = D / Dmax$

$D' = 0.0699 / 0.141 = 0.496 = 0.5$

To calculate coefficient of correlation (r), input value of D and allele frequencies calculated in previous steps in the following equation

$r = D / (p1p2q1q2)^{½}$

$r = 0.0699 /(0.5425 \times 0.4575 \times 0.308 \times 0.692)^{1/2}$

$r = 0.0699 / 0.23 = 0.304$

$r^2 = (0.304)^2 = 0.092$

To check the significance of LD between loci use following equation

$\chi^2 = r^2 N$

$\chi^2 = 0.092 \times 2000 = 184.8$ (1 df)

At 184.8 and df of 1, P-value is 0.0001

So, we can conclude based on our calculations that there is a significant LD between loci and it is 50% of the theoretical maximum

Also note that two SNPs are in complete LD (not separated by recombination) when D ' = 1 or r^2=1

Reference:

Lewontin R.C. 1988. On Measures of Gametic Disequilibrium. Genetics, 120(3): 849-852.

Devlin B., Risch N. 1995. A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. Genomics 29 (2): 311-322.