Linkage Equilibrium and Disequilibrium

David H. Alexander

August 18, 2010

Two-locus dynamics: linkage equilibrium (LE) and disequlibrium (LD)

Consider two linked loci 1 and 2. Let the population frequencies satisfy

$$p_A = \Pr(\text{allele } A \text{ at } 1),$$
$$p_B = \Pr(\text{allele } B \text{ at } 2).$$

Linkage equilibrium (LE) between 1 and 2 says that the alleles an individual has at the two loci are independent, i.e. the haplotype AB has population frequency

$$p_{AB} \doteq \Pr(AB) = p_A p_B.$$

Linkage disequilibrium (LD) is non-independence between the markers.

Causes of Linkage Disequilibrium (LD)

- Linkage equilibrium holds under the following assumptions: a) infinite population size, b) no inbreeding, c) random mating, d) no mutation, e) no migration, and f) no selection.
- In contrast to HWE, linkage equilibrium is reached over many generations.
- Disequilibrium is usually attributed to: a) population bottlenecks and founder effects, b) the recent admixture of two populations, c) selection involving the interaction between alleles at different loci, and d) favorable mutations and hitchhiking.

Measures of LD

All LD measures assume that loci are biallelic.

- D
- D'
- r^2

These LD summary statistics assume *phased* genotype data (*haplotypes*), which are hard to come by—requires computational *phasing*.

Measures of LD: D

$$D = D_{AB} \doteq p_{AB} - p_A p_B$$

- A raw measure of LD
- $D = 0$ indicates linkage equilibrium
- Problem: range of D varies depending on the allele frequencies of A and B, which makes it hard to interpret nonzero values of D

Measures of LD: D'

$$D' = \begin{cases} \dfrac{D}{\min\{p_A p_b, p_a p_B\}}, & D \geqslant 0; \\ \dfrac{D}{\min\{p_A p_B, p_a p_b\}} & D < 0 \end{cases}$$

- A standardized D statistic, $|D'| \leqslant 1$
- $D' = 0$ indicates LE, $|D'| = 1$ indicates *perfect* LD
- Under *perfect* LD, only three of the four possible haplotypes occur.

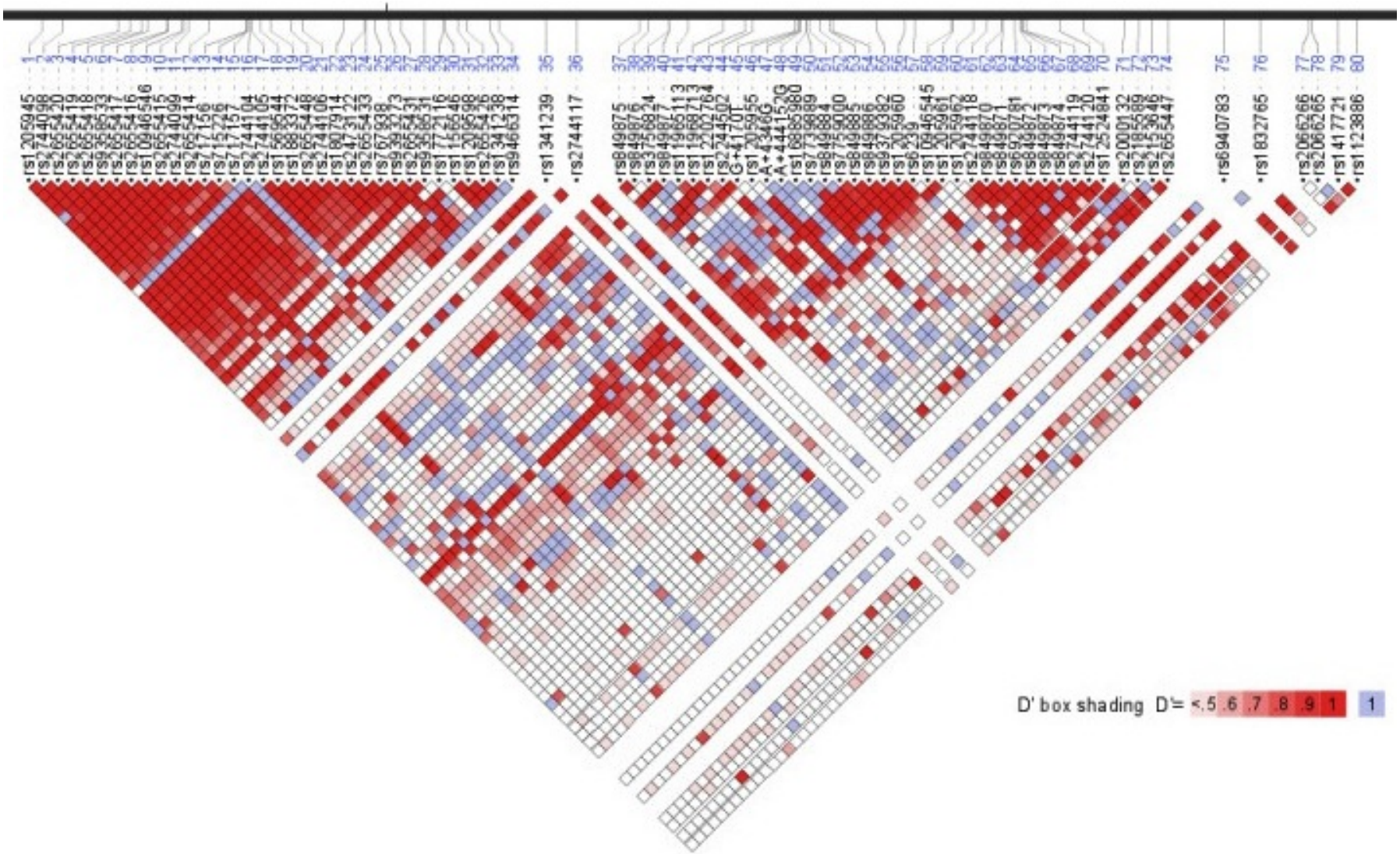
Measures of LD: r^2

$$r^2 = \frac{D^2}{p_A(1-p_A)p_B(1-p_B)}$$

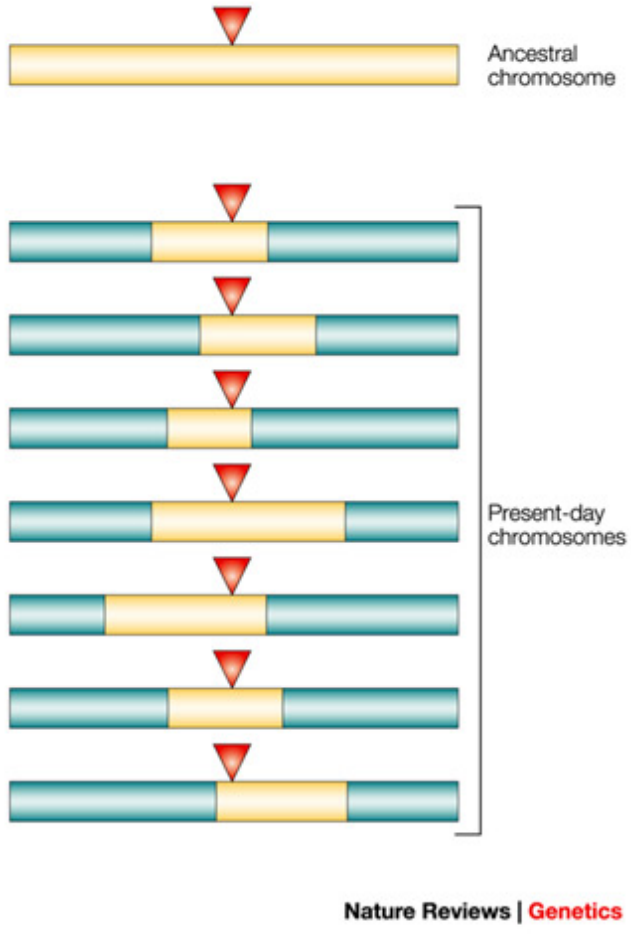
- Standardized, $0 \leqslant r^2 \leqslant 1$
- Same as Pearson correlation coefficient between allele indicator variables

Example: LD plot

Summary plot of pairwise LD statistic (D', r^2)



Decay of Linkage Disequilibrium via Recombination



A new mutation is in perfect LD with all loci on its chromosome.

Successive recombinations break up the haplotypes containing the mutation.

Decay of LD: the math

Consider the raw measure of linkage disequilibrium,

$$D_{AB} = p_{AB} - p_A p_B.$$

Under the equilibrium conditions, at generation n we have

$$\begin{aligned} D_{AB}^n &= (1-\theta) D_{AB}^{n-1} \\ &= (1-\theta)^n D_{AB}^0 \end{aligned}$$

so that LD decays geometrically fast, and bigger $\theta \implies$ faster convergence to LE. But $\theta \leqslant 0.5$, so:

Recombination *slowly* breaks up blocks of linkage disequilibrium.

LD is the new normal

- It used to be that markers were so far apart that LE was the norm, and was often assumed. (And there were tests for deviation from LE that could be employed to make sure this was a safe assumption.)
- Now, with very dense marker sets and full sequence data (small interlocus θ), LD has become the new norm. Need to take care that multimarker analyses do not make independence assumptions.

Tests for LD

MENDEL offers permutation test procedures to test the null hypothesis of linkage equilibrium. The test statistic is the Pearson χ^2 statistic for independence.

Permutation procedure for haplotypes

To simulate haplotype under the null hypothesis, we can randomly permute data within the columns, shuffling each column separately from the others:

	Marker 1	Marker 2
Haplotype 1	*A*	*b*
Haplotype 2	*a*	*B*
⋮		
Haplotype 2n	*A*	*b*

Shuffle ⇝

	Marker 1	Marker 2
Haplotype 1	*a*	*b*
Haplotype 2	*A*	*b*
⋮		
Haplotype 2n	*A*	*B*

Test for independence of alleles at different markers.

Permutation procedure for genotypes

Same permutation procedure can be applied even when we don't have *phase* information!

	Marker 1	Marker 2
Multi-genotype 1	aa	Bb
Multi-genotype 2	Aa	bb
⋮		
Multi-genotype n	aa	BB

Shuffle ⇝

	Marker 1	Marker 2
Multi-genotype 1	Aa	BB
Multi-genotype 2	aa	Bb
⋮		
Multi-genotype n	aa	bb

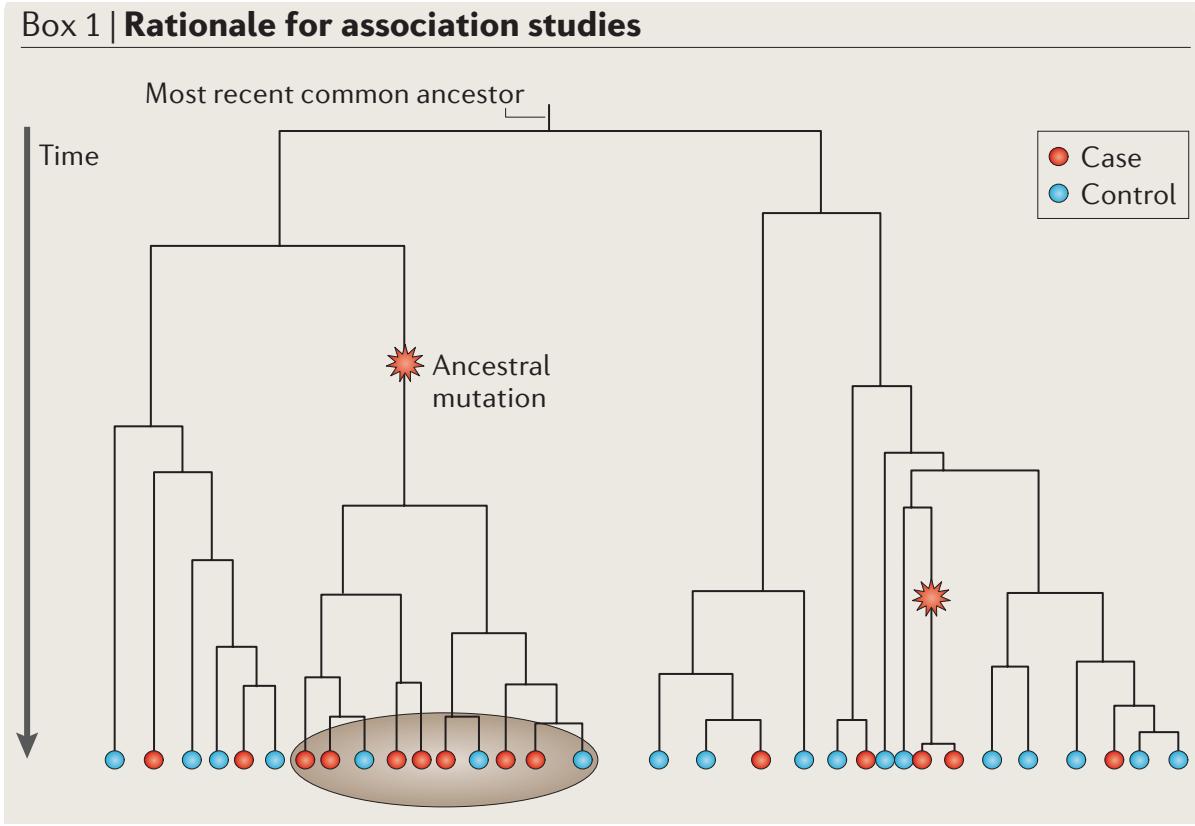
Now testing for independence of genotypes, not alleles.

The uses of LD

- Disease association mapping, i.e. *linkage disequilibrium mapping*
- Selection mapping—hunting for the signature of natural selection

Case-control association (CDCV hypothesis)

Case-control association testing is sometimes referred to as a method of "linkage disequilibrium mapping"—hunting for genotyped markers that are in LD with putative causal variant.



Mapping sites of selection

A favorable new mutation will start at frequency $1/2N$ and its population frequency will rise towards fixation...

...and it will bring its surrounding haplotype with it. So-called "hitchhiking effect."

Population geneticists recently have used patterns of LD to identify sites of natural selection. Tests of neutrality based on "extended haplotype homozygosity."

MENDEL Example

MENDEL Option 11 “Genetic Equilibrium” offers tests for Hardy-Weinberg and linkage equilibrium.

Try “Lecture 15c” example (uses Control11c.in), which tests for pairwise LD in a haplotype dataset

MENDEL Results

DATA TREATED AS HAPLOTYPES

PAIRWISE TESTS AND STATISTICS FOR LINKAGE EQUILIBRIUM

FIRST LOCUS	SECOND LOCUS	ESTIMATED P-VALUE		RANGE	POPULATION SIZE	ADJUSTED CHI-SQ	DPRIME
Marker1	Marker2	0.0000000			194	0.80014	0.94360
Marker1	Marker3	0.0000000			194	0.40252	0.87799
Marker1	Marker4	0.0000000			194	0.55622	0.94085
Marker1	Marker5	1.0000000			194	0.00028	-0.19167
Marker1	Marker6	0.0000000			194	0.29705	0.87067
Marker1	Marker7	0.6110000	+/-	0.0097505	194	0.00691	-1.00000
Marker1	Marker8	0.0000000			194	0.37464	0.87643

Recommended reading

M. Slatkin.
Linkage disequilibrium—understanding the evolutionary past and mapping the medical future.
Nature Reviews Genetics, 9(6):477–485, 2008.

B. Devlin and N. Risch.
A comparison of linkage disequilibrium measures for fine-scale mapping.
Genomics, 29(2):311–322, 1995.

JC Barrett, B. Fry, J. Maller, and MJ Daly.
Haploview: analysis and visualization of LD and haplotype maps.
Bioinformatics, 21(2):263, 2005.