Neutral Evolution at Two Loci

Alan R. Rogers

March 30, 2020

Linkage disquilibrium (LD)

	Locus	
Gamete	1	2
1	A	B
2	A	B
3	A	B
4	A	B
5	A	B
6	A	b
7	a	B
8	a	B
9	a	b
10	a	b

	A	a	
B	5	2	7
b	1	2	3
	6	4	10

▶ B is more common among A-gametes than a-gametes.

▶ A is more common among B-gametes than b-gametes.
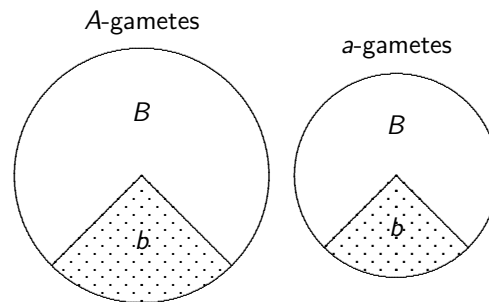
▶ This is LD.

Linkage equilibrium (LE)

	Locus	
Gamete	1	2
1	A	B
2	A	B
3	A	B
4	A	B
5	A	b
6	A	b
7	a	B
8	a	B
9	a	b

	A	a	
B	4	2	6
b	2	1	3
	6	3	9

▶ B is equally common among A-gametes and a-gametes.

▶ A is equally common among B-gametes and b-gametes.

▶ This is LE.

Linkage Equilibrium $\iff$ shaded fractions equal
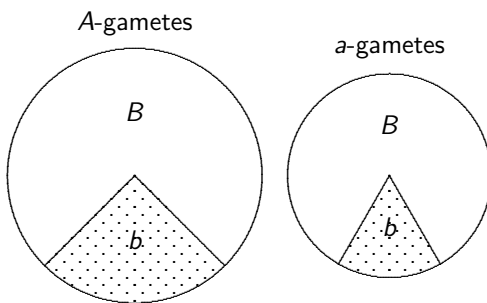

A-gametes

a-gametes

LE: Knowledge about one locus tells nothing about other.

Here, b is equally common among a-gametes and A-gametes.

(Also, a is equally common among b-gametes and B-gametes.)

LD $\iff$ shaded fractions unequal

A-gametes

a-gametes

LD: Knowledge about one locus helps predict the other.

Here, b is more common among A-gametes than among a-gametes.

(Also, a is more common among B-gametes than among b-gametes.)

People sometimes confuse LD with epistasis

|	
Epistasis	In a comparison of diploid genotypes, the *phenotypic effect* of one locus depends on another locus.
Linkage disequilibrium	In a comparison of haploid chromosomes, the *allele frequency* at one locus depends on another locus.

You can see LD in sequence data
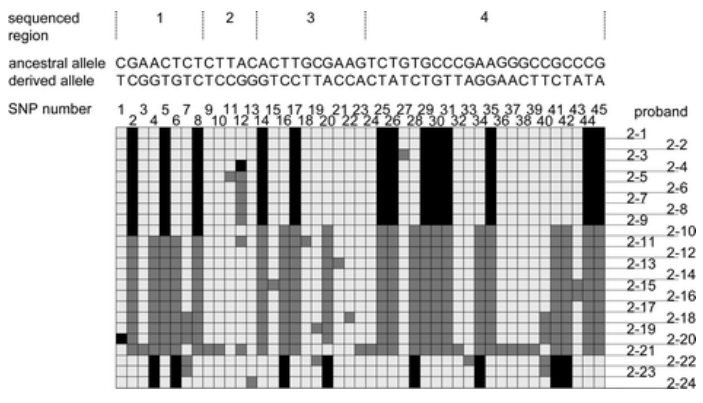

▶ Dots: identical to chimp sequence.

▶ Sites not independent.

▶ A at site 1343 predicts G at 1951

▶ This is linkage disequilibrium (LD).

(GARRIGAN ET AL 2004)
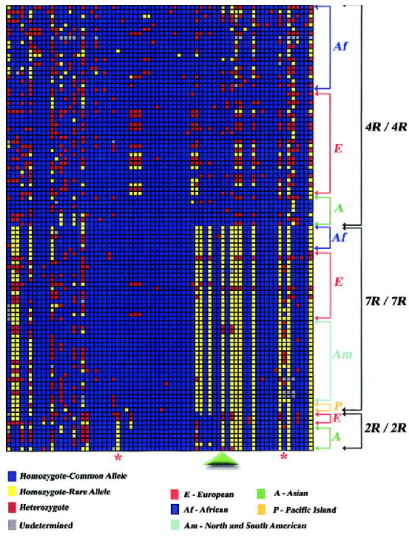
▶ Columns are SNPs

▶ Rows are diploid genotypes

▶ Blue: common homozygote

▶ Yellow: rare homozygote

▶ Red: heterozygote

▶ Note LD w/i 7R genotypes
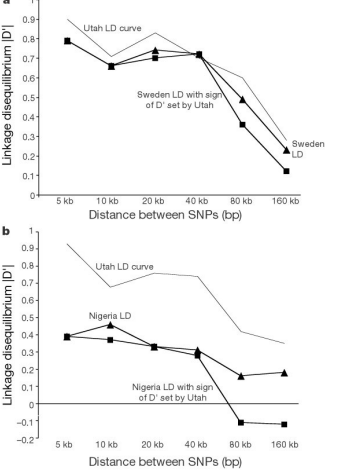

LD at the NF1 locus (Schmegner et al 2005)

DNA sequences from region of human lactase gene
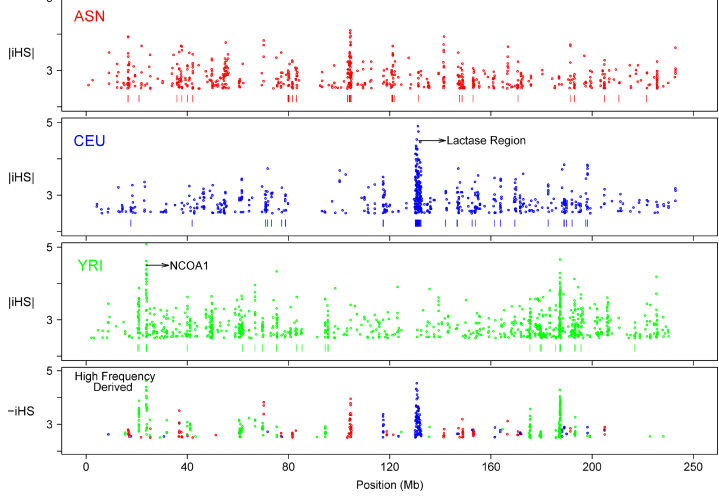

More LD in Europe than Africa

▶ LD declines with distance along chromosome

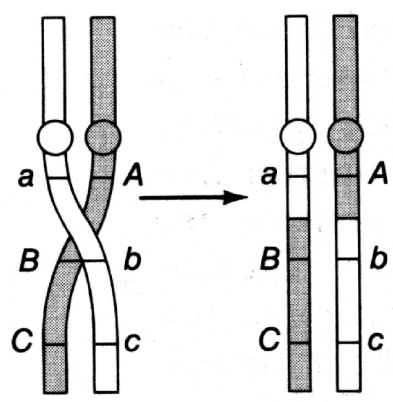▶ More LD in Europe than Africa

▶ Why?

(REICH ET AL 2001)

LD unevenly distributed within genome

Facts we need to understand

- LD decays w/ distance along chromosome.
- Populations differ.
- Unevenly distributed w/i genome

Cross-overs shuffle DNA

- occur during reproduction.
- shuffle parental chromosomes.
- sites far apart more likely to recombine
- result: "recombinant" chromosomes

Why loci are independent on recombinants

```
-------------------------------------------------
|                                               |
|.....A........b.....    A recombinant chromosome. |
|_____/_____/                           |
| from dad  from mom                            |
|                                               |
|                       Gamete from Dad carried A. |
|                       Gamete from Mom carried b. |
|                                               |
|Probability of this?   p_A p_b under random mating|
|_____|
```

Ingredients of a model

$$
\begin{aligned}
x_1 &= \text{frequency of } AB\text{-gametes among parents} \\
p_A &= \text{frequency of } A\text{-gametes among parents} \\
p_B &= \text{frequency of } B\text{-gametes among parents} \\
c &= \text{prob of recombination between the two loci}
\end{aligned}
$$

In any generation, there are two kinds of AB gamete:

1. non-recombinants: these were ABs in the last generation
   Frequency: $(1-c)x_1$

2. recombinants: formed from an A gamete and a B gamete, drawn at random. Frequency: cp_Ap_B

Next step: sum these contributions.

Model with random mating, no selection

$$
\begin{aligned}
x_1 &= \text{frequency of } AB\text{-gametes among parents} \\
p_A &= \text{frequency of } A\text{-gametes among parents} \\
p_B &= \text{frequency of } B\text{-gametes among parents} \\
c &= \text{prob of recombination between the two loci}
\end{aligned}
$$

Change in frequency of AB-gametes during one generation:

$$
\begin{aligned}
x_1' &= \overbrace{(1-c)x_1}^{\text{nonrecombinants}} + \overbrace{cp_Ap_B}^{\text{recombinants}} \\
&= x_1 - c(x_1 - p_Ap_B) \\
&= x_1 - cD
\end{aligned}
$$

Several equivalent definitions of D

The previous slide defined D, a measure of LD:

Gamete		Definition
AB	$D =$	$x_1 - p_Ap_B$
Ab	$-D =$	$x_2 - p_Ap_b$
aB	$-D =$	$x_3 - p_ap_B$
ab	$D =$	$x_4 - p_ap_b$

If the association between A and B is positive, then that between A and b must be negative. A more convenient formula:

$$
D = x_1x_4 - x_2x_3
$$

They all give the same answer.

Calculating D

	Locus	
Gamete	1	2
1	A	B
2	A	B
3	A	B
4	A	B
5	A	B
6	A	b
7	a	B
8	a	B
9	a	b
10	a	b

$$
\begin{array}{cccc}
AB & Ab & aB & ab \\
x_1 & x_2 & x_3 & x_4
\end{array}
$$

	A	a	
B	5	2	7
b	1	2	3
	6	4	10

$$
\begin{aligned}
D &= x_1 x_4 - x_2 x_3 \\
&= \frac{5}{10} \cdot \frac{2}{10} - \frac{1}{10} \cdot \frac{2}{10} \\
&= \frac{2}{25}
\end{aligned}
$$

All four gametes, still no selection

Gamete	Recurrence		
AB	x_1'	$=$	$x_1 - cD$
Ab	x_2'	$=$	$x_2 + cD$
aB	x_3'	$=$	$x_3 + cD$
ab	x_4'	$=$	$x_4 - cD$

How recombination affects D

After one generation,
$$
\begin{aligned}
D' &= x_1' x_4' - x_2' x_3' \\
&= (x_1 - cD)(x_4 - cD) - (x_2 + cD)(x_3 + cD) \\
&= (1 - c)D
\end{aligned}
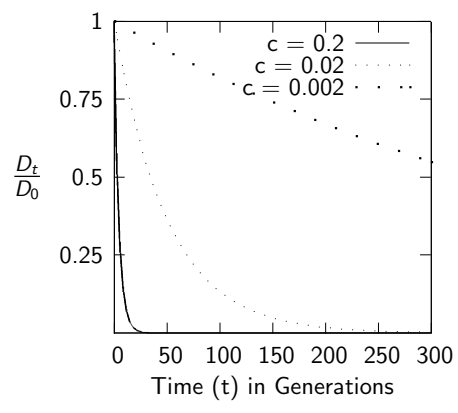$$

D declines each generation by a factor of $1 - c$.
After t generations,
$$
D_t = D_0 (1 - c)^t
$$

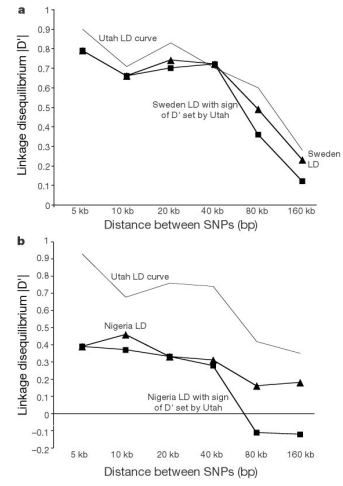## D declines gradually toward zero

$c = 0.2$
- Loci far apart
- Loose linkage
- LD declines rapidly

$c = 0.02$
- Loci closer
- slower decline

$c = 0.002$
- Loci closer still
- even slower decline

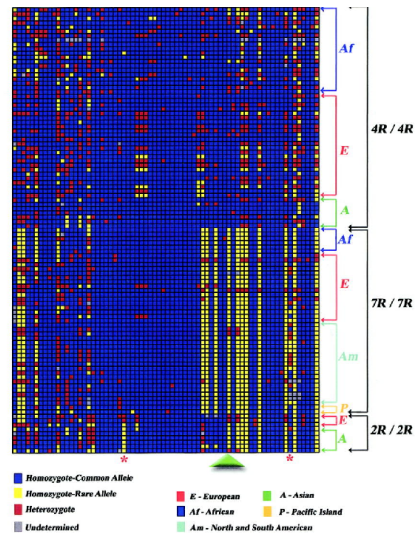Is this theory enough to explain the data?

More LD in Europe than Africa

- c increases w/ distance along chromosome.
- Therefore LD should decline.
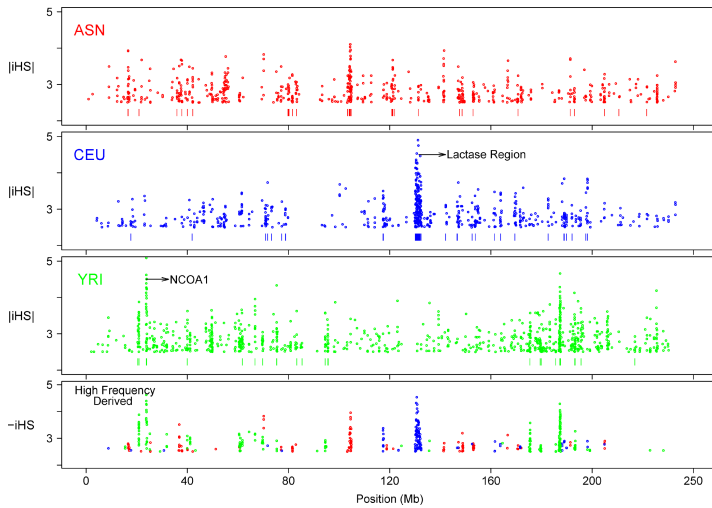- But why more LD in Europe?

(REICH ET AL 2001)

Slide 25:

(GARRIGAN et al 2004)

- Short DNA sequence.
- tight linkage: c is small
- LD decays very slowly
- But why is it not zero?

Slide 26:

- Also a short sequence
- But why is there any LD?

Slide 27:

Why is LD unevenly distributed?

Slide 28:

Summary

- Our theory explains why D declines w/ distance along chromosome.
- If loci are far apart on chromosome, c is high and D declines rapidly.
- It tells us nothing about the forces that generate LD.
- Nothing about population differences.
- Nothing about variation across the genome.