

# Why LD Helps Us Find Selective Sweeps

Alan R. Rogers

March 14, 2024

## LD helps us detect ongoing selective sweeps

This is curious, because we have known for over 50 years that selection at a single locus does not cause LD (Felsenstein 1965).

This lecture will explain why selective sweeps are *associated* with LD even though they don't cause it.

We begin with a graphical argument that shows why selection at a single locus doesn't cause LD.

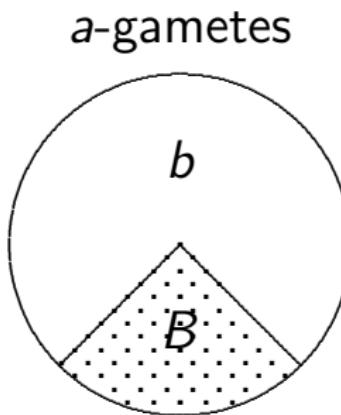
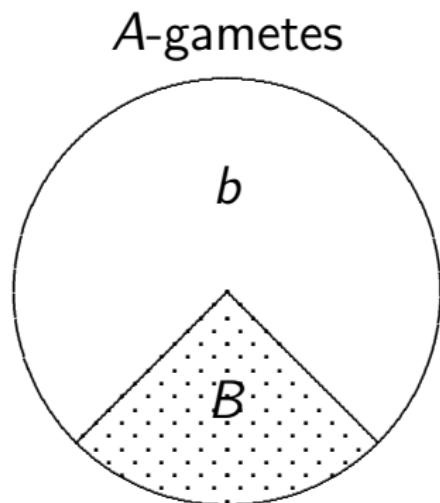
## LD helps us detect ongoing selective sweeps

This is curious, because we have known for over 50 years that selection at a single locus does not cause LD (Felsenstein 1965).

This lecture will explain why selective sweeps are *associated* with LD even though they don't cause it.

We begin with a graphical argument that shows why selection at a single locus doesn't cause LD.

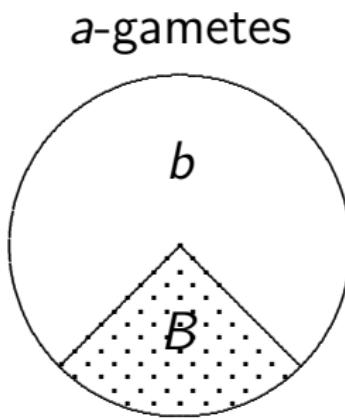
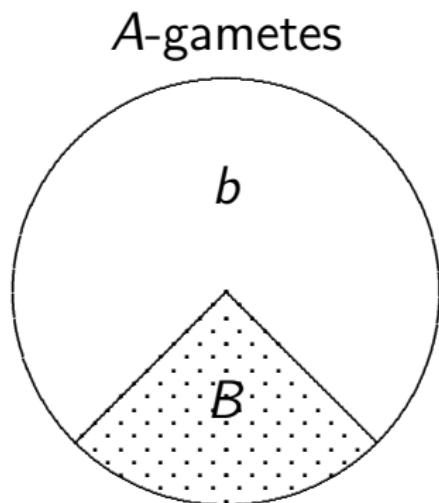
# Linkage equilibrium $\iff$ shaded fractions equal



LE: Neither locus predicts other

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

Suppose allele  $A$  is favored;  $B/b$  are neutral



Selection inflates left circle.

Shrinks right one.

Shaded wedges unchanged.

System remains in LE.

If selection doesn't *cause* LD, why does LD help us detect selection?

$D$ , the conventional measure of LD, doesn't help us here, because it *is* affected by selection—but not in a way that is helpful.

To explain this puzzle, I introduce another measure, which does not change in response to selection at a single locus.

# Conditional allele frequencies

Gamete type	$AB$	$Ab$	$aB$	$ab$
Frequency	$x_1$	$x_2$	$x_3$	$x_4$

Freq of  $B$  among  $A$ -gametes

$$p_{B|A} = \frac{x_1}{x_1 + x_2}$$

The size of one pie slice.

Freq of  $B$  among  $a$ -gametes

$$p_{B|a} = \frac{x_3}{x_3 + x_4}$$

The size of the other slice.

Neither is affected by selection on  $A/a$ .

# The Nei-Li measure of linkage disequilibrium

$$d = p_{B|A} - p_{B|a}$$

(Nei & Li, 1980)

Difference in frequency of  $B$  between  $A$ - and  $a$ -gametes.

Selection on  $A/a$  affects neither  $p_{B|A}$ ,  $p_{B|a}$ , nor  $d$ .

# The Nei-Li measure of linkage disequilibrium

$$d = p_{B|A} - p_{B|a}$$

(Nei & Li, 1980)

Difference in frequency of  $B$  between  $A$ - and  $a$ -gametes.

Selection on  $A/a$  affects neither  $p_{B|A}$ ,  $p_{B|a}$ , nor  $d$ .

# Calculating $D$ and $d$

Gamete	Locus	
	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

$$D = \frac{5}{10} \cdot \frac{2}{10} - \frac{1}{10} \cdot \frac{2}{10} = \frac{2}{25}$$

$$d = p_{B|A} - p_{B|a} = \frac{5}{6} - \frac{1}{2} = \frac{1}{3}$$

$D$  depends on heterozygosity at locus  $A$ .

$$\begin{aligned}d &= \frac{x_1}{x_1 + x_2} - \frac{x_3}{x_3 + x_4} \\&= \frac{D}{p_A(1 - p_A)}\end{aligned}$$

$$D = dp_A(1 - p_A)$$

Selection at locus  $A$  affects  $p_A(1 - p_A)$  and therefore  $D$ .

$D$  depends on heterozygosity at locus  $A$ .

$$\begin{aligned}d &= \frac{x_1}{x_1 + x_2} - \frac{x_3}{x_3 + x_4} \\&= \frac{D}{p_A(1 - p_A)}\end{aligned}$$

$$D = dp_A(1 - p_A)$$

Selection at locus  $A$  affects  $p_A(1 - p_A)$  and therefore  $D$ .

# Mutation generated LD

A tiny population before and after mutation:

--a\_\_\_\_\_B\_\_

--a\_\_\_\_\_B\_\_

--a\_\_\_\_\_B\_\_

--Mutation-->

--A\_\_\_\_\_B\_\_

--a\_\_\_\_\_b\_\_

--a\_\_\_\_\_b\_\_

--a\_\_\_\_\_b\_\_

--a\_\_\_\_\_b\_\_

Mutation occurs on a  $B$  gamete. After mutation,  $\text{Pr}[B|A] = 1$  but  $\text{Pr}[B|a] = 1/3$ , so mutation has created LD.

## LD of a new mutant

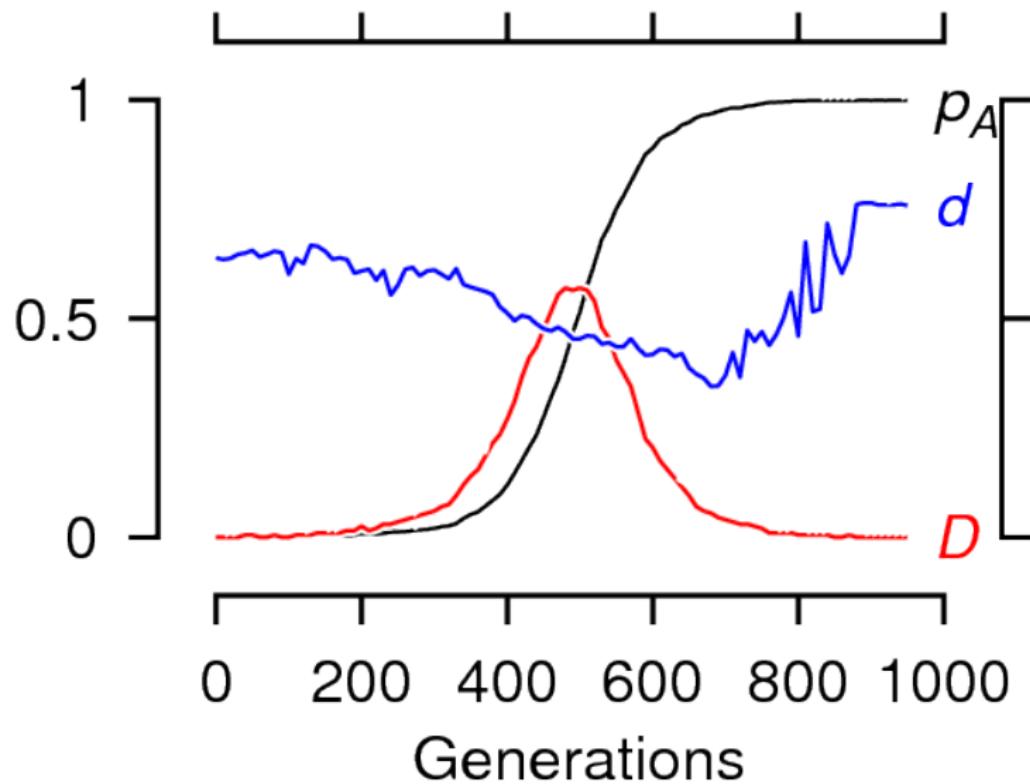
When allele  $A$  first arises by mutation, it will be rare, so  $p_A(1 - p_A) \approx 0$  and  $D \approx 0$ .

But  $d$  may be large: it is either  $1 - p_B$  or  $-p_B$ , depending on whether the mutant gamete is  $AB$  or  $Ab$ .

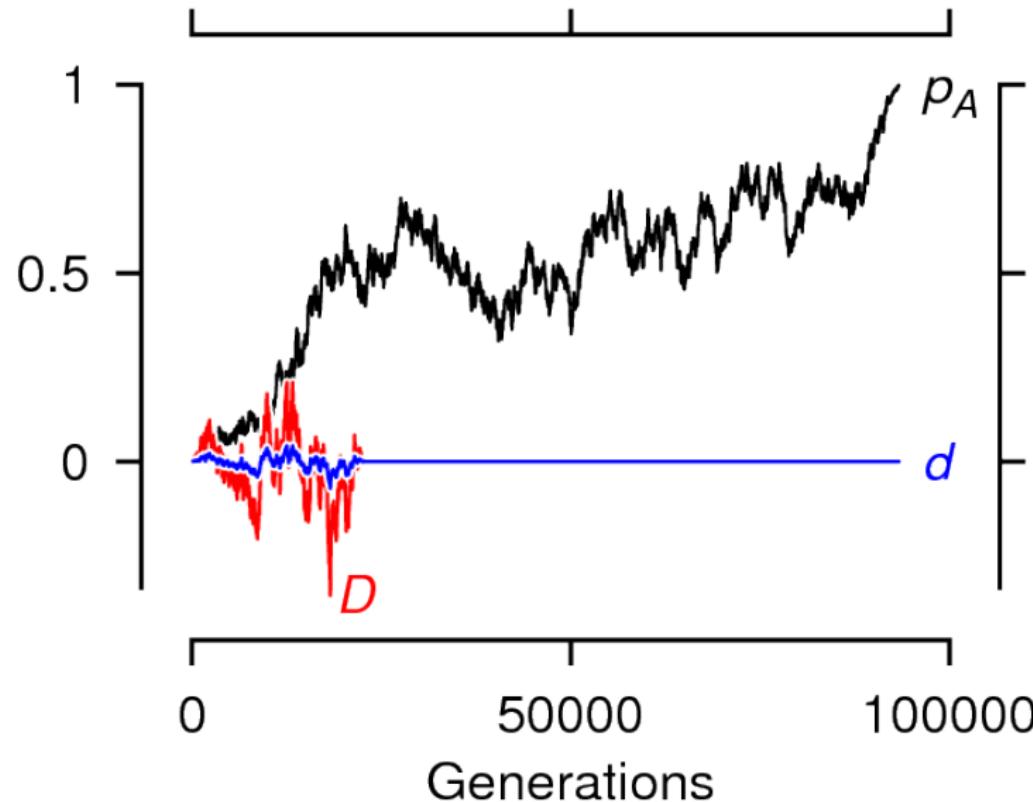
The value of  $d$  stays large if selection outruns recombination and drift.

*This is why LD helps us detect selection.*

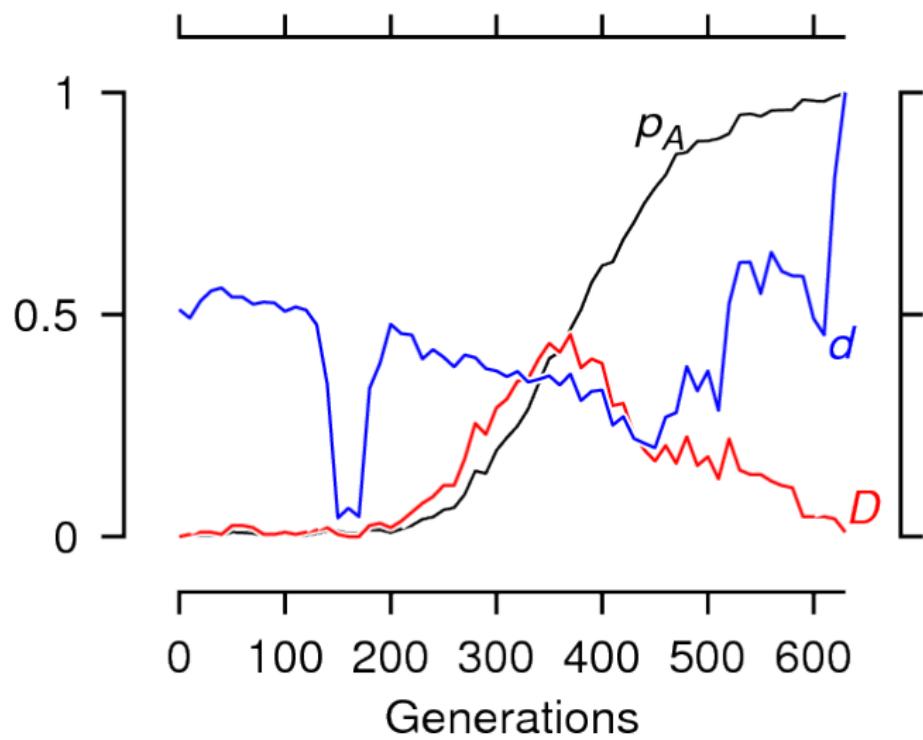
Selective sweep of  $A$ , with  $s = 0.02$ ,  $c = 0.001$ ,  
and  $2N = 50,000$



Neutral drift to fixation,  $c = 0.001$ , and  
 $2N = 50,000$



Selective sweep in small population ( $s = 0.02$ ,  
 $c = 0.001$ ,  $2N = 5000$ )



# Summary

- ▶ LD decays gradually in response to recombination.
- ▶ This is more obvious for  $d$  than for  $D$ , because  $d$  is insensitive to  $p_A$ .
- ▶ Advantageous alleles increase rapidly: there is little time for LD to decay.
- ▶ Neutral alleles increase slowly; plenty of time for LD to decay.