

Selection at Two Loci

Alan R. Rogers

March 5, 2020

Model with random mating, no selection

x_1 = frequency of AB -gametes

p_A = frequency of A -gametes

p_B = frequency of B -gametes

c = probability of recombination

Change in frequency of AB -gametes during one generation:

$$x'_1 = x_1 - cD$$

All four gametes, still no selection

Gamete	<i>Recurrence</i>		
<i>AB</i>	x'_1	=	$x_1 - cD$
<i>Ab</i>	x'_2	=	$x_2 + cD$
<i>aB</i>	x'_3	=	$x_3 + cD$
<i>ab</i>	x'_4	=	$x_4 - cD$

Selection affecting gametes

Gamete		Recurrence	
AB	x'_1	$=$	$w_1(x_1 - cD)/\bar{w}$
Ab	x'_2	$=$	$w_2(x_2 + cD)/\bar{w}$
aB	x'_3	$=$	$w_3(x_3 + cD)/\bar{w}$
ab	x'_4	$=$	$w_4(x_4 - cD)/\bar{w}$

where $\bar{w} = \sum x_i w_i$ is mean fitness.

What if selection acts on adults?

The effect of recombination

What gametes are produced by the following genotypes?

Genotype	Heterozygous loci	Gametes produced			
		AB	Ab	aB	ab
AB/AB	0	1			
AB/Ab	1	$1/2$	$1/2$		
AB/ab	2	$\frac{1-c}{2}$	$\frac{c}{2}$	$\frac{c}{2}$	$\frac{1-c}{2}$
Ab/aB	2	$\frac{c}{2}$	$\frac{1-c}{2}$	$\frac{1-c}{2}$	$\frac{c}{2}$

Only double heterozygotes make recombinant gametes. If these genotypes have low fitness, few recombinants appear.

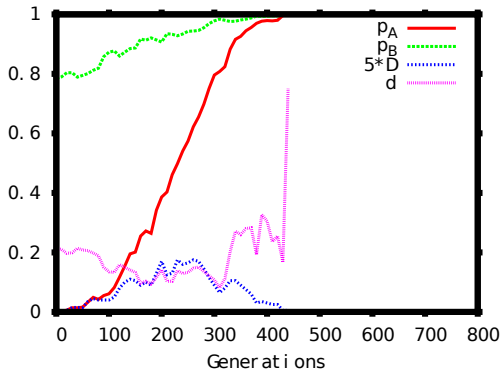
Selection affecting diploid adults

Gamete		<i>Recurrence</i>	
<i>AB</i>	x'_1	=	$\bar{w}_1(x_1 - cw_h D)/\bar{w}$
<i>Ab</i>	x'_2	=	$\bar{w}_2(x_2 + cw_h D)/\bar{w}$
<i>aB</i>	x'_3	=	$\bar{w}_3(x_3 + cw_h D)/\bar{w}$
<i>ab</i>	x'_4	=	$\bar{w}_4(x_4 - cw_h D)/\bar{w}$

- ▶ Fitnesses become \bar{w}_i : weighted mean over genotypes in which gamete i appears.
- ▶ Recombination limited by the fitness (w_h) of double heterozygotes: only these contribute recombinant gametes.
- ▶ Useful as a recipe for calculation.

A sweeps; B hitch-hikes

Parameters: $s = 0.02$, $c = 0.001$, $N = 5000$



Selective sweep of allele A .

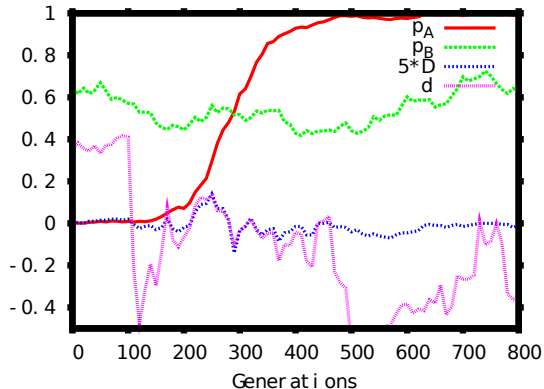
Allele B hitch-hikes to fixation.

D high when p_A has high heterozygosity.

d high throughout

Linked allele may fail to increase

Parameters: $s = 0.02$, $c = 0.001$, $N = 5000$

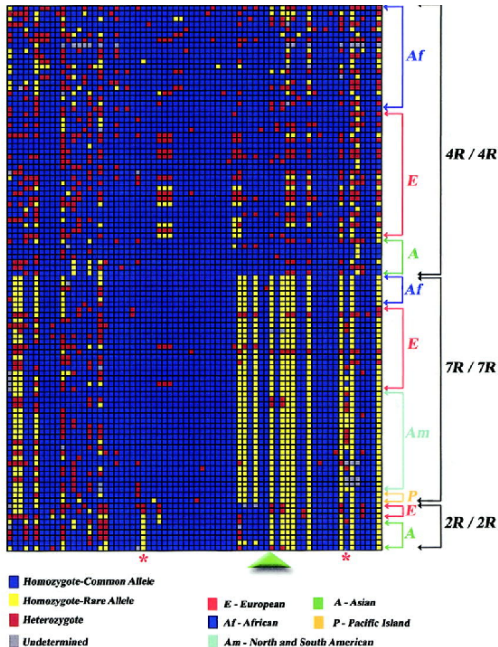


Allele A sweeps to fixation.

Little change in linked allele. Why?

LD lost early on this run, so D near 0.

Loss of LD shows as big drop in d .



- ▶ LD at D4 dopamine receptor
- ▶ Rows are diploid genotypes
- ▶ Blue: common homozygote
- ▶ Yellow: rare homozygote
- ▶ Red: heterozygote
- ▶ Note LD w/i 7R genotypes

DNA sequences from region of human lactase gene

```
cgcttcaggcattcctatctaaacagaccaacgtaAgggtacaatgcctaaccagacgtttcaactct
20 .....
21 .....
22 .....
23 .....
24 .....
25 .....
26 .....
27 .....t.....
28 .....t.....
29 .....C.....
37 .....G..a.gt....t.....gac.c.tgtct.
38 ...ccgga...gat..at..gg..c....tc.gGaaa.g..ccttt...tg.....c...t.t...
39 ...ccgga...gat..at..gg..c....tc.gGaaa.g..ccttt...tg.....c...t.t...
40 ..tcc...agtag.t.cat..g....t..ttccgG..a.gt....t.....gac.c.tgtct.
41 ..tcc...agtag.t.cat..g....t..gttccgG..a.gt....t.....gac.c.tgtct.
42 ..tcc...agtag.t.cat..g....t..gttccgG..a.gt....t.....gac.c.tgtct.
43 ..tcc...agtag.t.cat..g....t.g.tc.gG..a.gt....t.....gac.c.tgtct.
44 ..tcc...agtag.t.cat..g....t..ttc.gG..acgt....t.....gac.c.tgtct.
45 ..tcc...agtag.t.cat..g....t.gttc.gG..a.gt....t.....gac.c.tgtct.
46 ...ccgga...gat..at..gg..c....tc.gGaaa.g..ccttt...tg.....cg.gt.t..c
47 ..tcc...agtag.t.cat..g....t..gttccgG..a.gt....t.....gac.c.tgtct.
48 ..tcc...agtag.t.cat..g....t..gttccgG..a.gt....t.....gac.c.tgtct.
49 ..tcc...agtag.t.cat..g....t..gttccgG..a.gt....t.....gac.c.tgtct.
50 tatccgga...g.tc.atcgg.tc.g.tg.tc.gG..a.g.g...tg...ggg...cg.gt.t..c
51 ta.ccgga...g.t..atcgg.tc.g.tg.tc.gG..a.g.g...tg...ggg...cg.gt.t..c
52 ta.ccgga...g.t..atc.g.tc.g.tg.tc.gG..a.g.g...tg...ggg...cg.gt.t..c
53 ta.ccgga...g.t..atcgg.tc.g.tg.tc.gG..a.g.g...tg...ggg...cg.gt.t..c
```

Why LD Helps Us Find Selective Sweeps

Alan R. Rogers

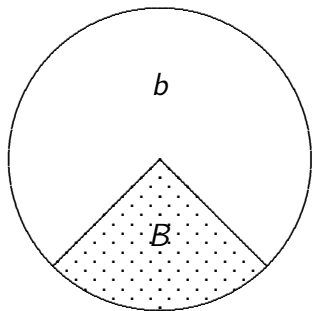
March 5, 2020

LD helps us detect ongoing selective sweeps

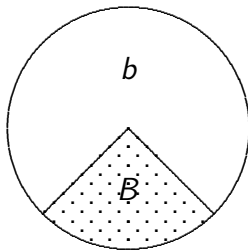
This is curious, because we have known for over 50 years that selection at a single locus does not cause LD (Felsenstein 1965).

Linkage equilibrium \iff shaded fractions equal

A-gametes



a-gametes

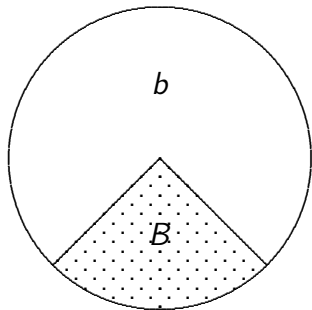


LE: Neither locus predicts other

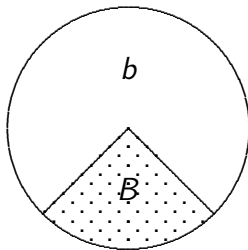
Here, *B* is equally common among *a*-gametes and *A*-gametes.

Suppose allele A is favored; B/b are neutral

A -gametes



a -gametes



Selection inflates left circle.

Shrinks right one.

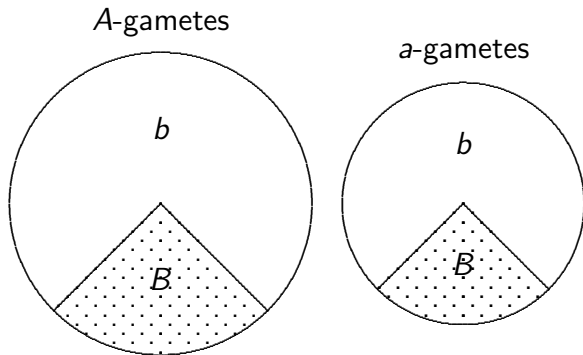
Shaded wedges unchanged.

System remains in LE.

Notation for gamete frequencies

Gamete type	AB	Ab	aB	ab
Frequency	x_1	x_2	x_3	x_4

Conditional allele frequencies



Freq of B among A -gametes

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

Freq of B among a -gametes

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

Selection on A/a affects neither.

The Nei-Li measure of linkage disequilibrium, d

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

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

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

(Nei & Li, 1980)

Linkage disequilibrium: 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 .

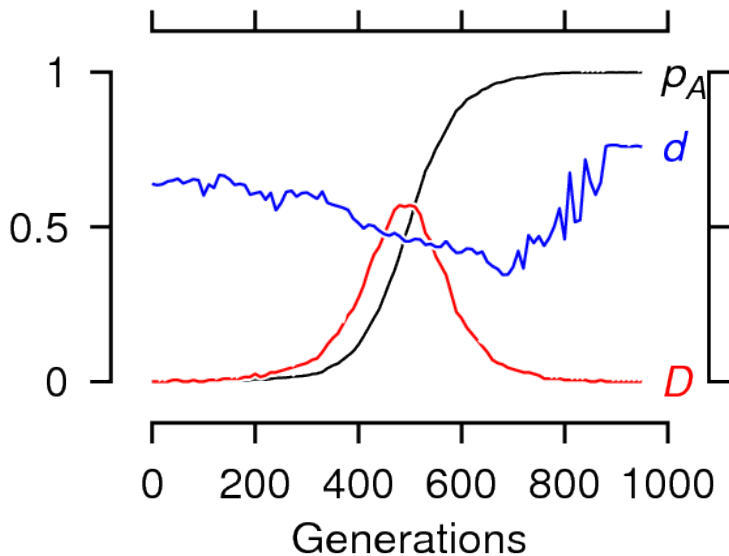
LD of a new mutant

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

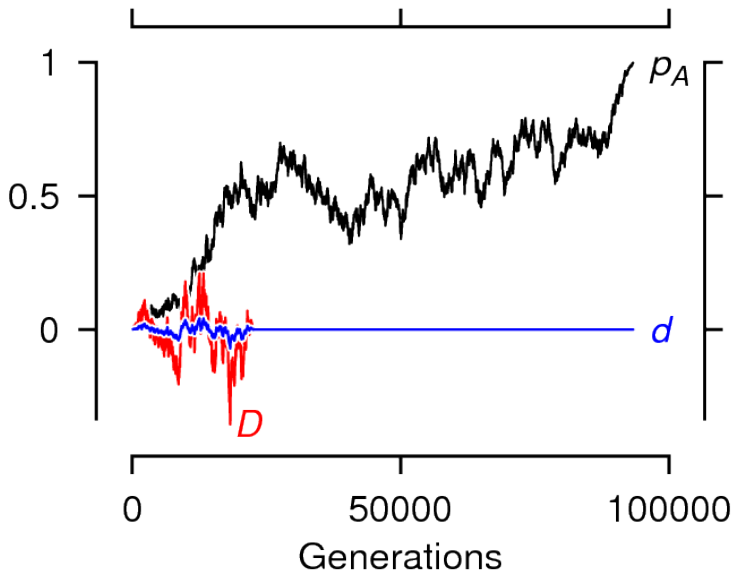
But d may be large: it is either $1 - p_B$ or $-p_B$, depending on whether the mutant gamete is AB or Ab .

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

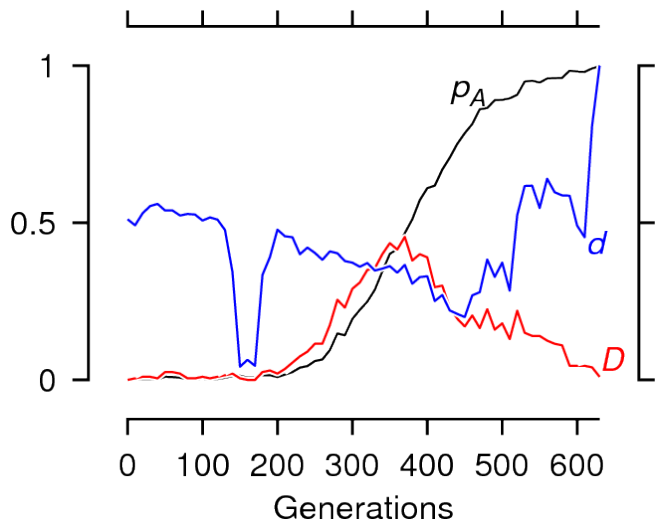
Selective sweep of A , with $s = 0.02$, $c = 0.001$, and $2N = 50,000$



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



Selective sweep in small population ($s = 0.02$, $c = 0.001$, $2N = 5,000$)



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.