

Linkage disequilibrium under soft sweeps

Cornelia Borck
joint work with Peter Pfaffelhuber

September 23, 2010

C.Borck, 2010. The Effect of Recurrent Mutation on the Linkage Disequilibrium under a Selective Sweep, submitted,
arXiv:1009.3150v1.

Linkage disequilibrium helps us to detect (hard) selective sweeps.
(Pfaffelhuber, Lehnert, Stephan)

Can linkage disequilibrium detect soft selective sweeps, too???

Ignoring mutations during the sweep, but **with recombination** see at the end of the sweep:

Sweep with recombination

Ind1	G-....T-....G-....C-....T-....C-....T-G
Ind2	G-....T-....G-....C-....T-....C-... -T-....G
Ind3	G-....T-....G-....C-....T-....C-....T-....G
Ind4	G-....T-....G-....C-....T-....C-....T-G
Ind5	G-....T-....G-....C-....T-....C-....T-G
Ind6	A-...-G-....G-....C-....T-....C-....T-G
Ind7	G-....T-....G-....C-....T-....C-....T-....G
Ind8	G-....T-....G-....C-....T-....C-....T-G
Ind9	G-....T-....G-....C-....T-....C-....A-A
Ind10	G-....T-....G-....C-....T-....C-... -T-....G



- two SNP-loci lying on the **same** side of the selected locus are **tightly linked**
- two SNP-loci lying on **different** sides are **unlinked**

If the mutation rate of the beneficial allele is relatively high (e.g. $\theta \sim 1$), it may also happen the following...

Soft sweep without recombination

Ind1	A-...-T-...-G-...-T-... -T-...-A -...-A-...-T
Ind2	G-...-T-...-C-...-A-...-T-...-T- ...-T-...-G
Ind3	G-...-T-...-C-...-A-...-A-...-A-...-T-...-G
Ind4	A-...-G-...-C-...-T-...-A-...-A-...-A-...-T-... -T
Ind5	A-...-G-...-G-...-A-...-T-...-T- ...-T-...-G
Ind6	G-...-T-...-G-...-A-... -T-...-A-...-A-...-G
Ind7	G-...-T-...-G-...-T-... -A-...-C-...-A-...-A
Ind8	A-...-G-...-G-...-T-... -T-...-A-...-A-...-G
Ind9	A-...-G-...-G-...-A-...-T-...-C-...-A-...-T
Ind10	G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1 A-...-T-...-G-...-T-... -T-...-A-...-A-...-T
Ind2 G-...-T-...-C-...-A-...-T-...-T- ...-T-...-G
Ind3 G-...-T-...-C-...-A-...-A-...-A-...-T-...-G
Ind4 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5 A-...-G-...-G-...-A-...-T-...-T- ...-T-...-G
Ind6 G-...-T-...-G-...-A-... -T-...-A-...-A-...-G
Ind7 G-...-T-...-G-...-T-... -A-...-C-...-A-...-A
Ind8 A-...-G-...-G-...-T-... -T-...-A-...-A-...-G
Ind9 A-...-G-...-G-...-A-...-T-...-C-...-A-...-T
Ind10 G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1 A-...-T-...-G-...-T-... -T-...-A -...-A-...-T
Ind2 G-...-T-...-C-...-A-...-T-...-T- ...-T-...-G
Ind3 G-...-T-...-C-...-A-...-A-...-A-...-T-...-G
Ind4 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5 A-...-G-...-G-...-A-...-T-...-T- ...-T-...-G
Ind6 G-...-T-...-G-...-A-... -T-...-A-...-A-...-G
Ind7 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind8 A-...-G-...-G-...-T-... -T-...-A-...-A-...-G
Ind9 A-...-G-...-G-...-A-...-T-...-C-...-A-...-T
Ind10 G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1 A-...-T-...-G-...-T-... -T-...-A -...-A-...-T
Ind2 G-...-T-...-C-...-A-...-T-...-T- ...-T-...-G
Ind3 G-...-T-...-C-...-A-...-A-...-A-...-T-...-G
Ind4 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5 A-...-G-...-G-...-A-...-T-...-T- ...-T-...-G
Ind6 G-...-T-...-G-...-A-... -T-...-A-...-A-...-G
Ind7 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind8 A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind9 A-...-G-...-G-...-A-...-T-...-C-...-A-...-T
Ind10 G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind2	G-...-T-...-C-...-A-...-T-...-T-...-T-...-G
Ind3	G-...-T-...-C-...-A-...-A-...-A-...-T-...-G
Ind4	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind6	G-...-T-...-G-...-A-... -T-...-A-...-A-...-G
Ind7	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind8	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind9	A-...-G-...-G-...-A-...-T-...-C-...-A-...-T
Ind10	G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind2	G-...-T-...-C-...-A-...-T-...-T- ...-T-...-G
Ind3	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind4	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind6	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind7	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind8	A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind9	A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind10	G-...-T-...-G-... -T-...-A-...-T-...-A-...-A

Soft sweep without recombination

Ind1 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind2 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind3 A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind4 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind5 A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind6 A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind7 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind8 A-...-G-...-G-...-C-...-T-...-A-...-A-...-G
Ind9 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T
Ind10 A-...-G-...-C-...-C-...-A-...-A-...-T-... -T

If more than one individual contributes to the fixation of the beneficial allele, the sweep is called soft (Hermisson and Pennings).

At the end of a soft sweep with two ancestors see:

Soft sweep without recombination

Ind1	A....G....C....C....A....A....T.... -T
Ind2	A....G....C....C....A....A....T.... -T
Ind3	A....G....G....C....T....A....A....G
Ind4	A....G....C....C....A....A....T.... -T
Ind5	A....G....G....C....T....A....A....G
Ind6	A....G....G....C....T....A....A....G
Ind7	A....G....C....C....A....A....T.... -T
Ind8	A....G....G....C....T....A....A....G
Ind9	A....G....C....C....A....A....T.... -T
Ind10	A....G....C....C....A....A....T.... -T



-SNP-loci are tightly linked
(for loci lying on the same side AND lying on different sides)

The **linkage disequilibrium** measures the non-independence of two loci with alleles ℓ and L at the first locus and r and R at the second locus. We measure the linkage disequilibrium by

$$\hat{\sigma}^2 := \frac{\mathbb{E}[\hat{D}^2]}{\mathbb{E}[\hat{q}_L(1 - \hat{q}_L)\hat{q}_R(1 - \hat{q}_R)]}.$$

a quantity introduced by Ohta and Kimura in 1969,
denoting by

- \hat{q}_L - fraction of the sample carrying the allele L
- \hat{q}_R - analogously
- \hat{q}_{LR} - fraction of the sample carrying both alleles L and R
- $\hat{D} = \hat{q}_{LR} - \hat{q}_L\hat{q}_R$

Observe

$\mathbb{E}[q_L(1 - q_L)q_R(1 - q_R)]$ = the probability to draw an unlinked heterozygous pair w.r.t. the L-locus and a unlinked heterozygous pair w.r.t. the R-locus.

$\mathbb{E}[D^2]$ - exist similar interpretation with linked and unlinked heterozygous pairs.

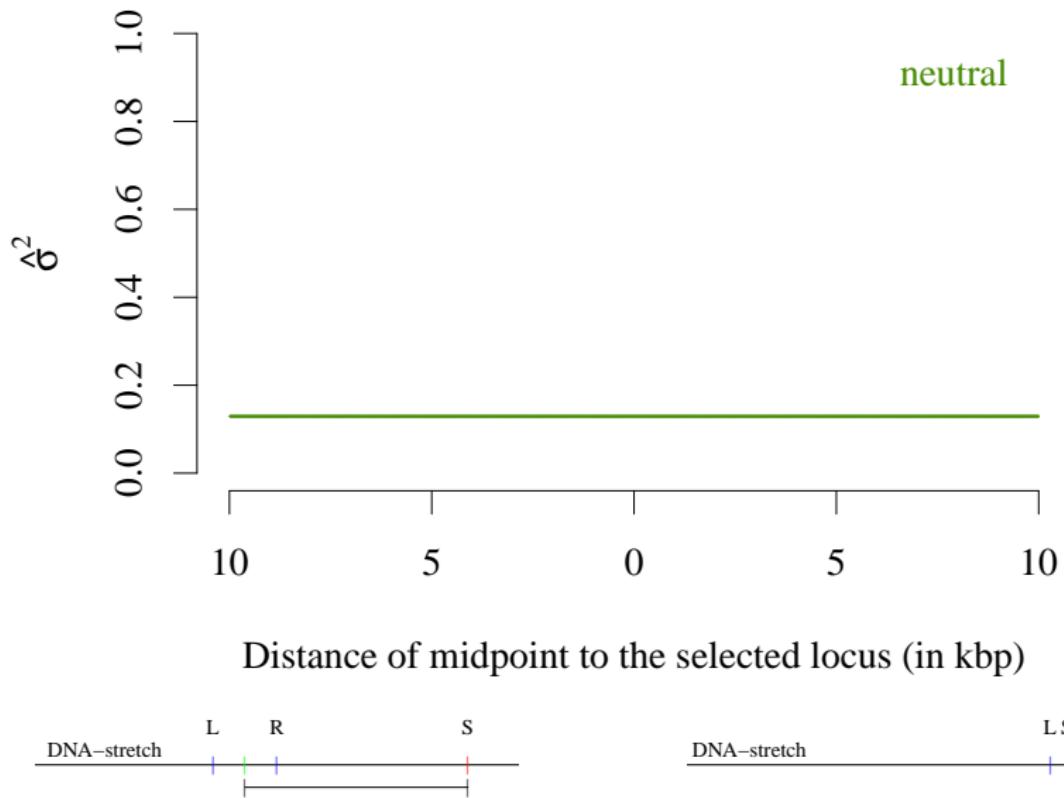
Approximate the genealogy of the sample by a star-like genealogy, i.e.

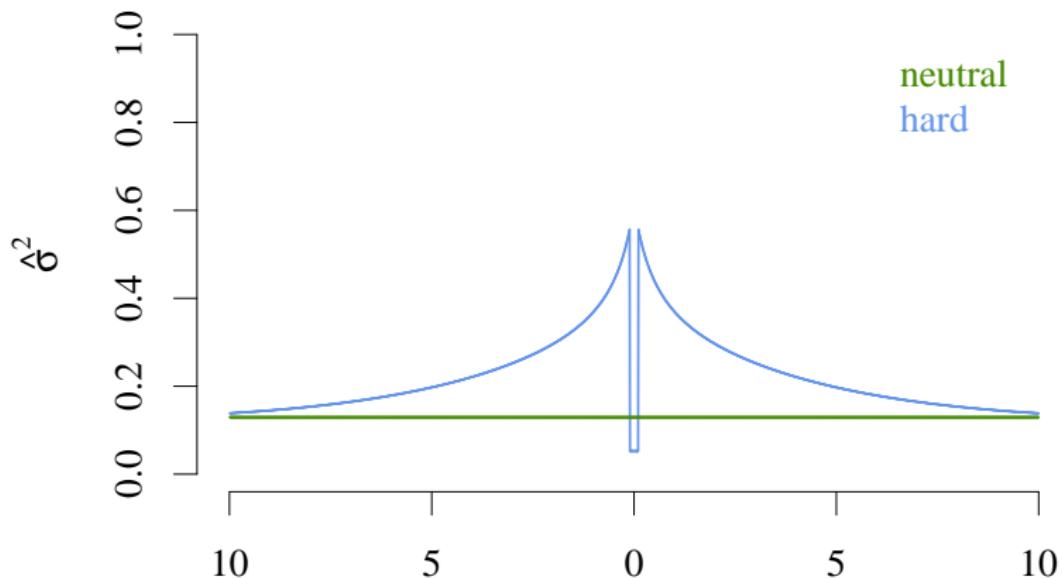
- recombination events affect only single lines → independent recombination events
- all coalescence events occur at the beginning of the sweep: each founder of the sweep is represented by one star

Follow linked and unlinked heterozygous pairs at the end of the sweep till the beginning of the sweep.

Obtain in the star-like approximation $\hat{\sigma}^2$ at fixation as a function depending on

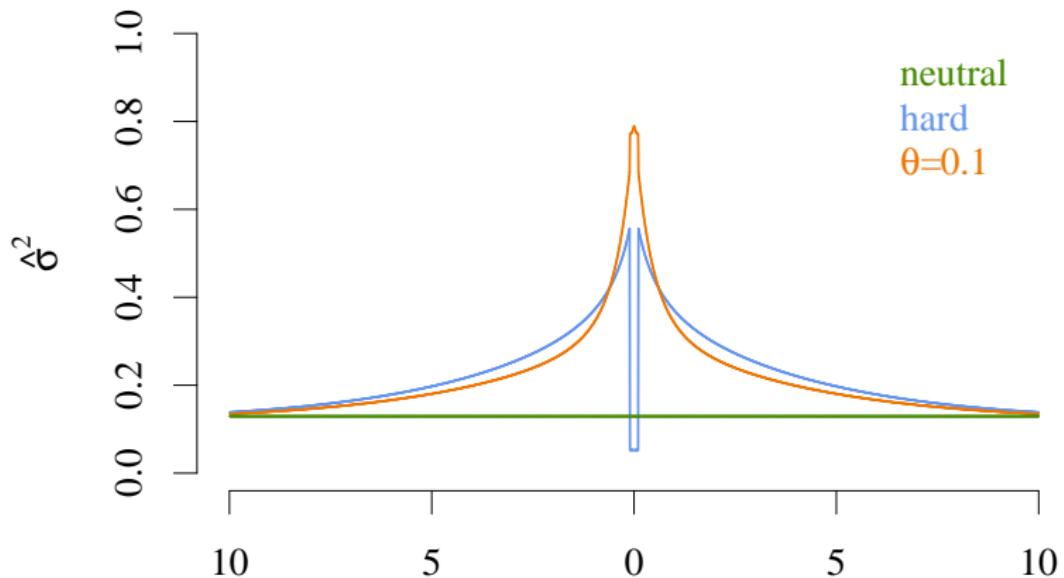
- recombination rate $\rho = 4Nr$ ($=0.025$) per bp
- distance between two neutral loci d ($=200\text{bp}$)
- distance to selected locus
- selection strength $\alpha = 2Ns$ ($= 1000$)
- recurrent mutation rate to the beneficial allele $\theta = 4N\mu$ ($0, 0.1$ respectively)
- sample size n ($= 20$)
- neutral mutation rate $\theta_{neutral} = 4N\mu_{neutral}$ ($=0.005$) per bp





Distance of midpoint to the selected locus (in kbp)





Distance of midpoint to the selected locus (in kbp)



Linkage disequilibrium helps us to distinguish between soft and hard sweeps and neutral evolution!

Thank you for your attention!