Partial selective sweeps and genetic signatures of ongoing adaptation

Luca Ferretti, Alex Klassmann,
Sebastian Ramos-Onsins, Guillaume Achaz

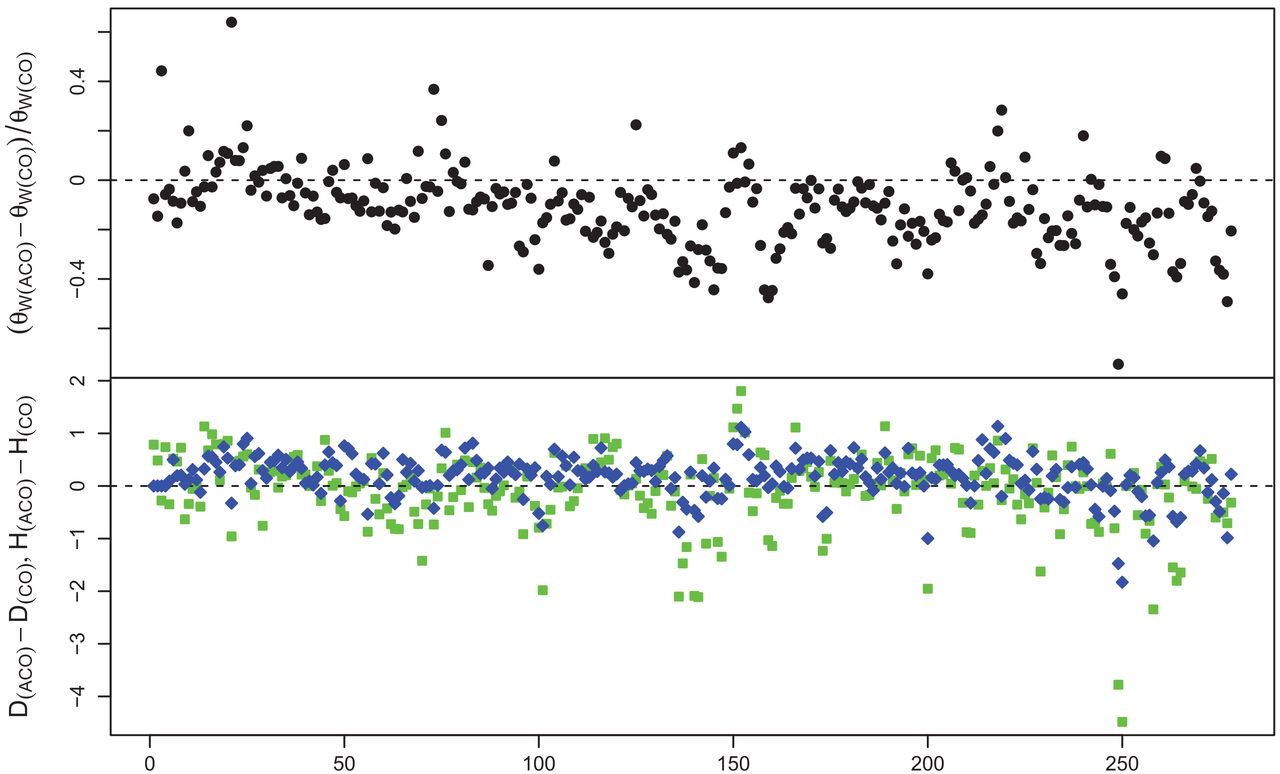
The Pirbright Institute
UPMC and Collège de France

The Pirbright Institute receives strategic funding from BBSRC.

www.pirbright.ac.uk

Signals of recent selection from sequences

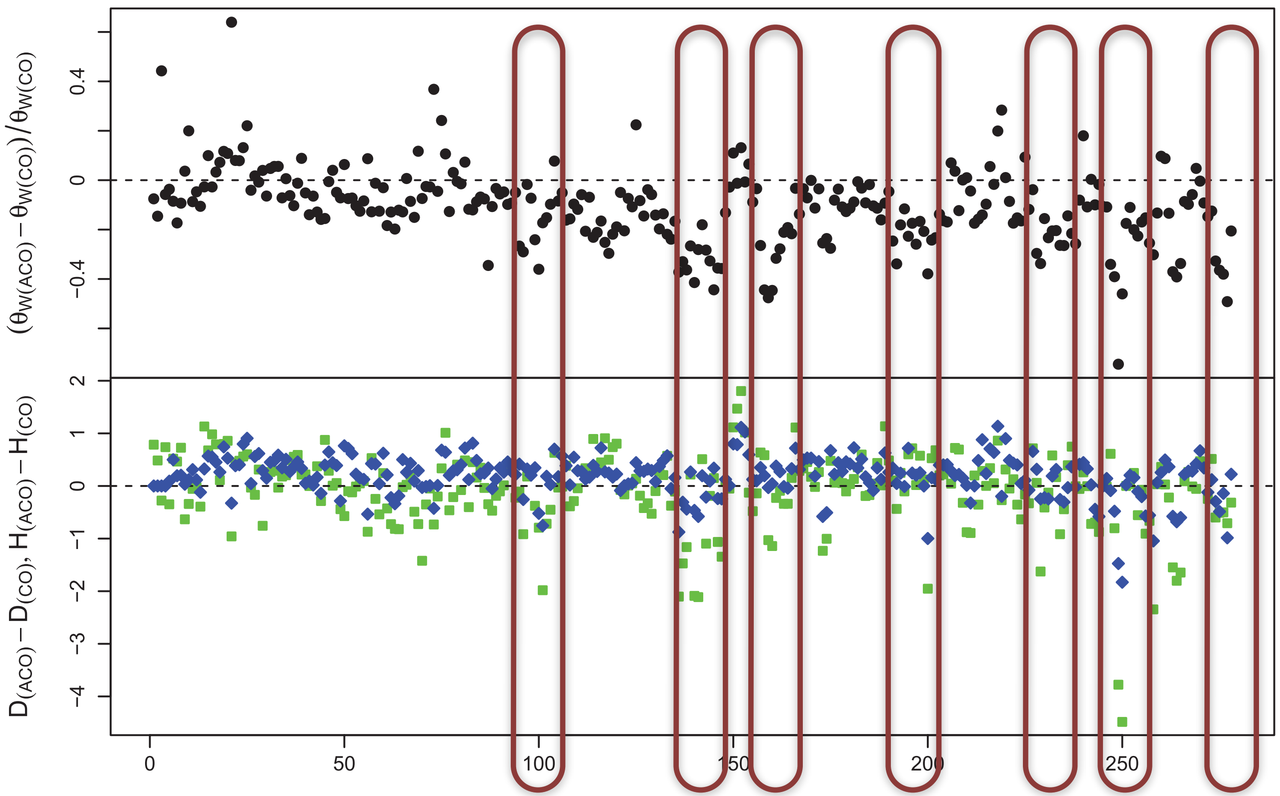
Example: experimental evolution for accelerated life cycle in Drosophila



Burke et al, Nature 2010 (analysis from Ferretti et al, Mol Ecol 2013)

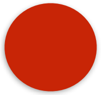
Signals of recent selection from sequences

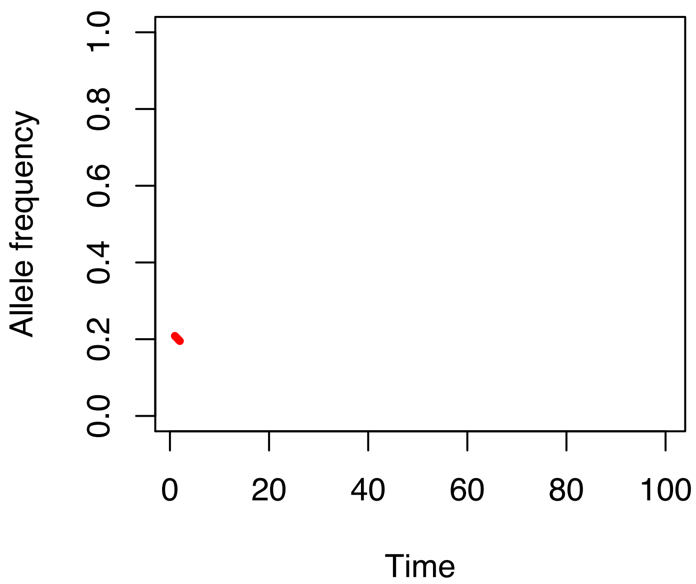
Example: experimental evolution for accelerated life cycle in Drosophila



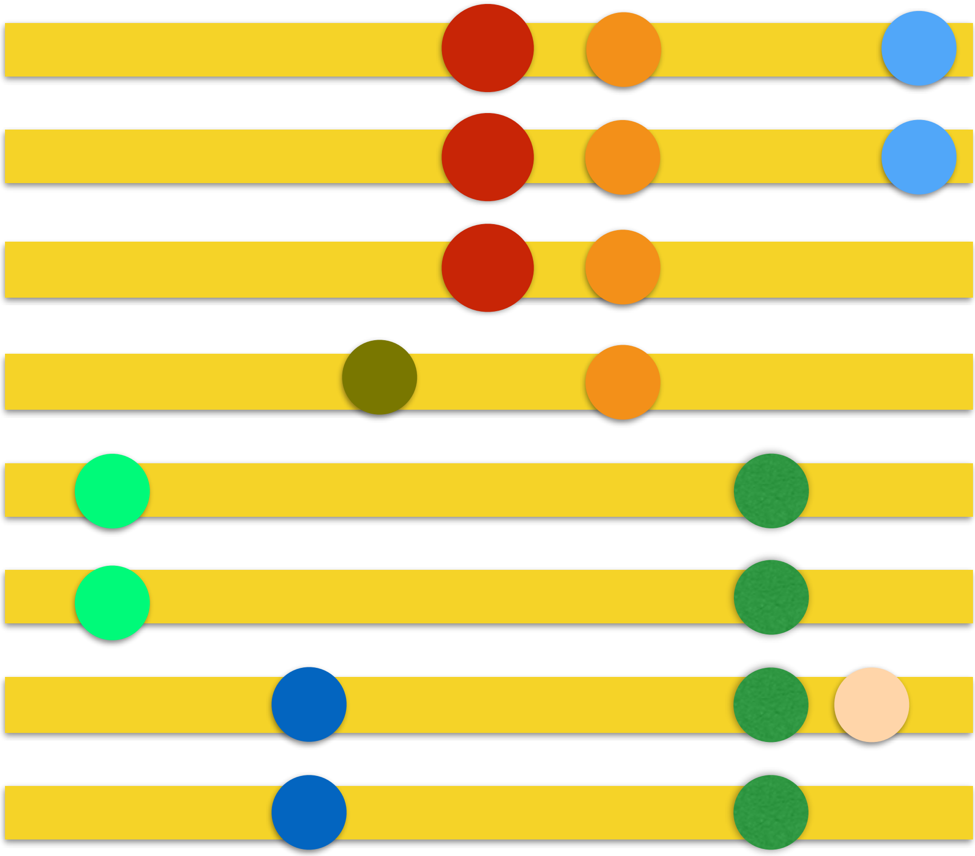
Partial selective sweeps?

Evolution of a neutral sequence

Frequency of the  mutation in the population

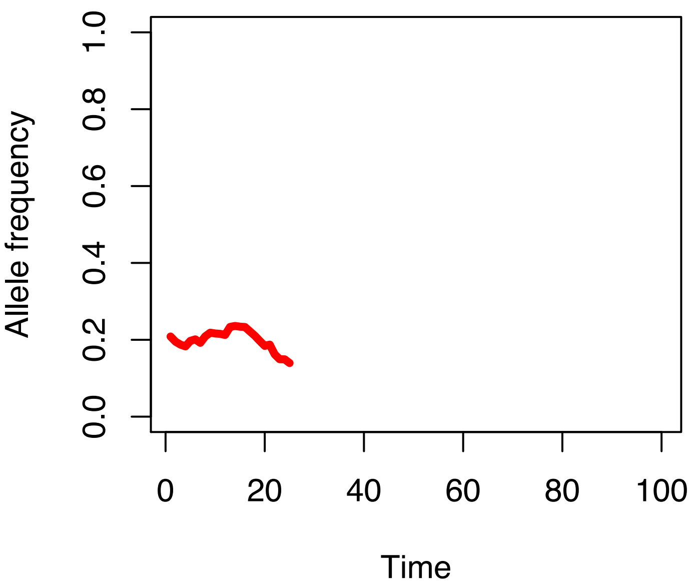


Mutations in a sample:

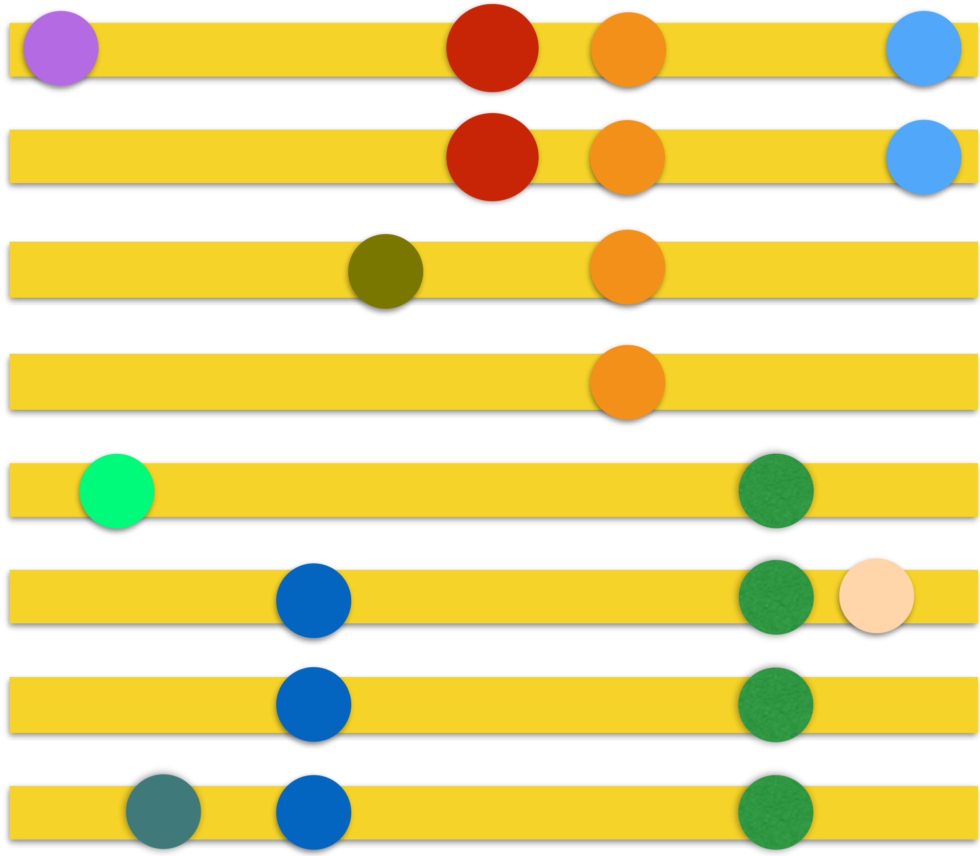


Evolution of a neutral sequence

Frequency of the
mutation in the population

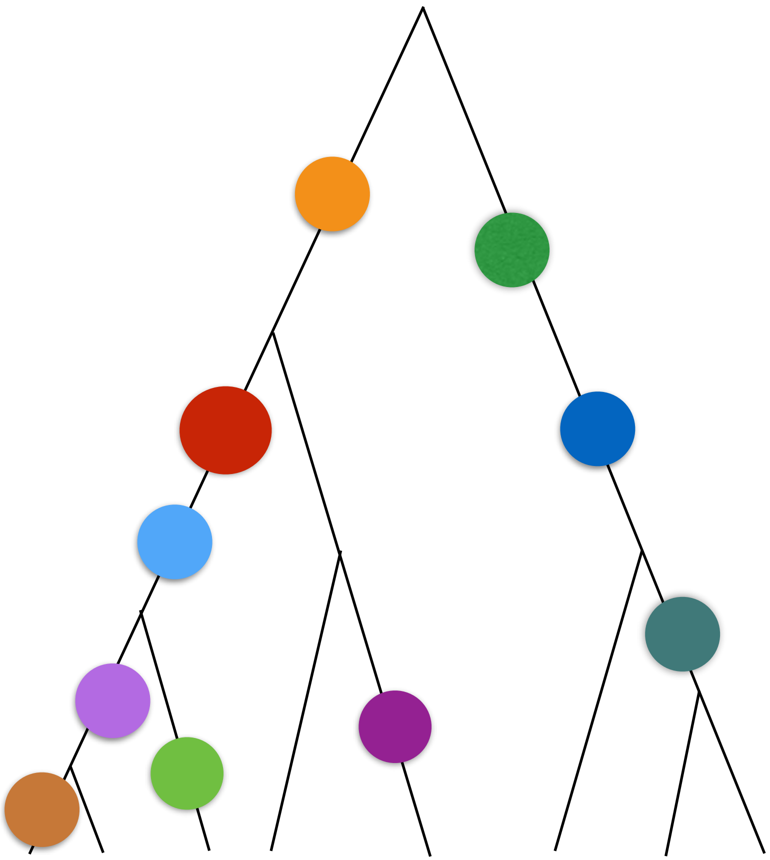


Mutations in a sample:

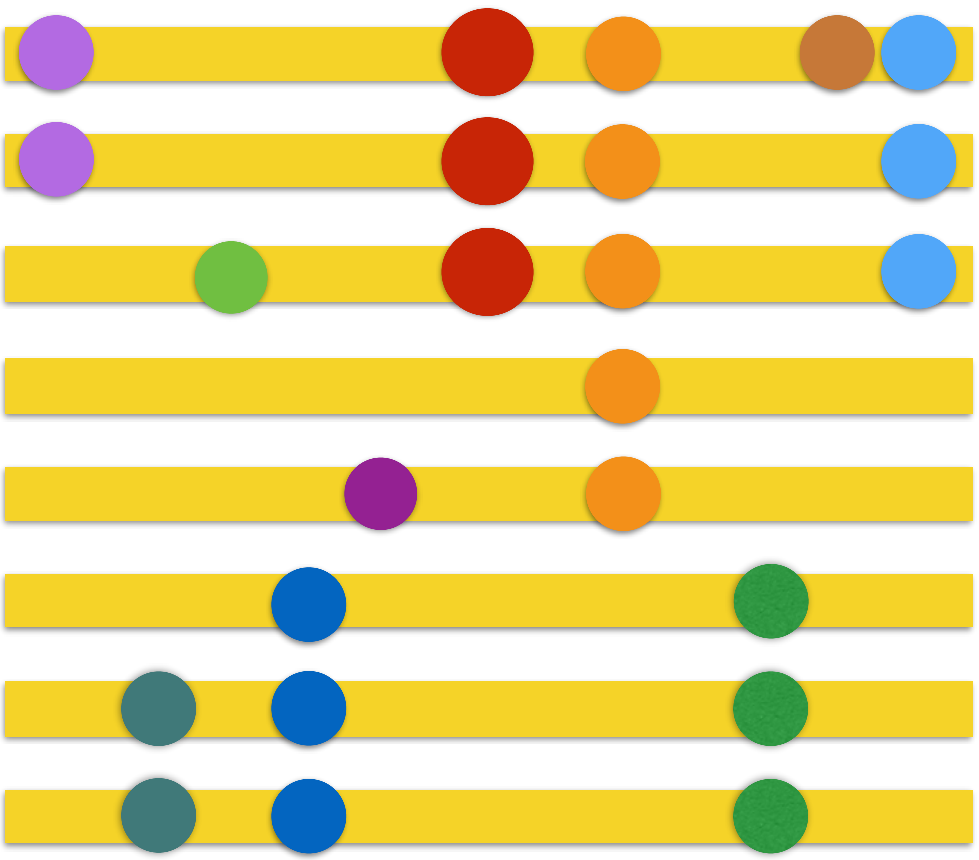


Evolution of a neutral sequence

Genealogy of the sample



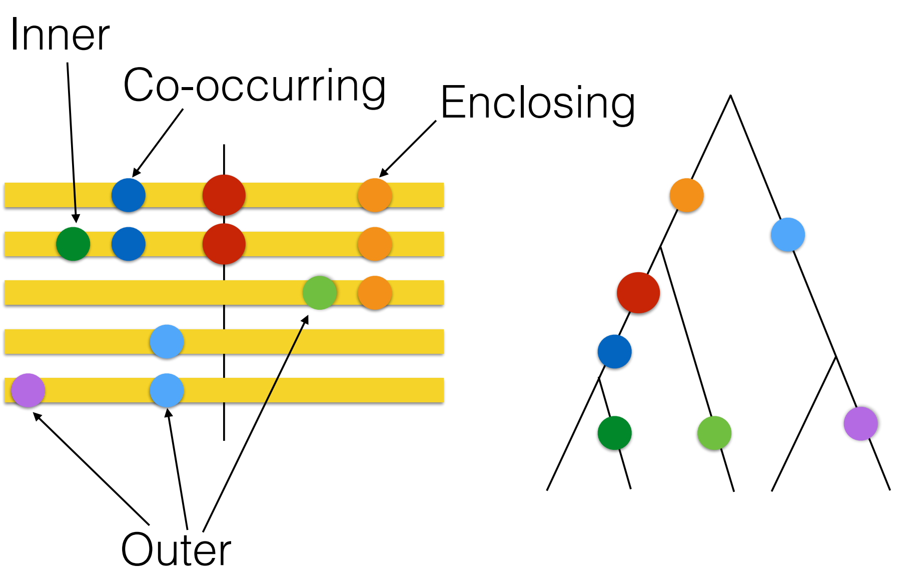
Mutations in a sample:

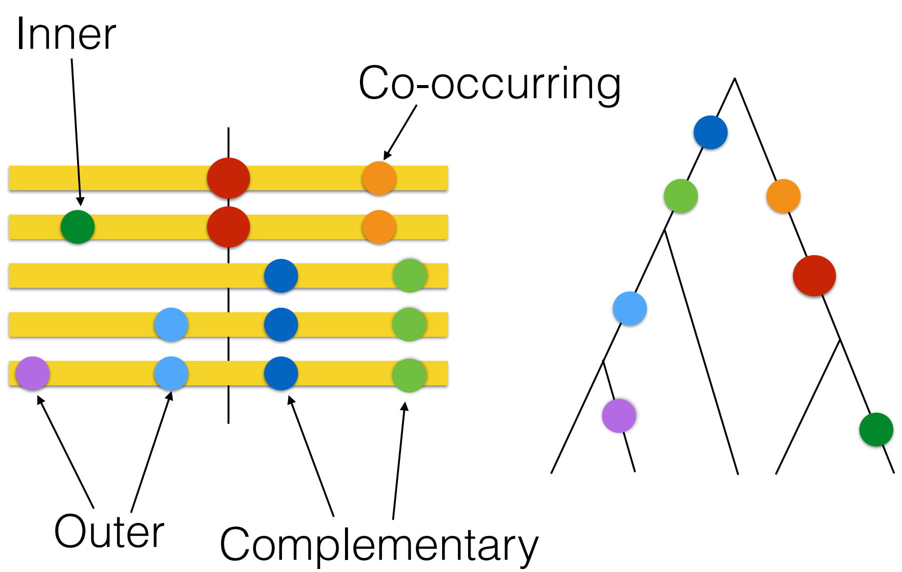


How? Coalescent theory of pairs of mutations

no recombination: 2 types of pairs, *nested* or *disjoint*

and 5 types of mutations relative to a focal mutation…

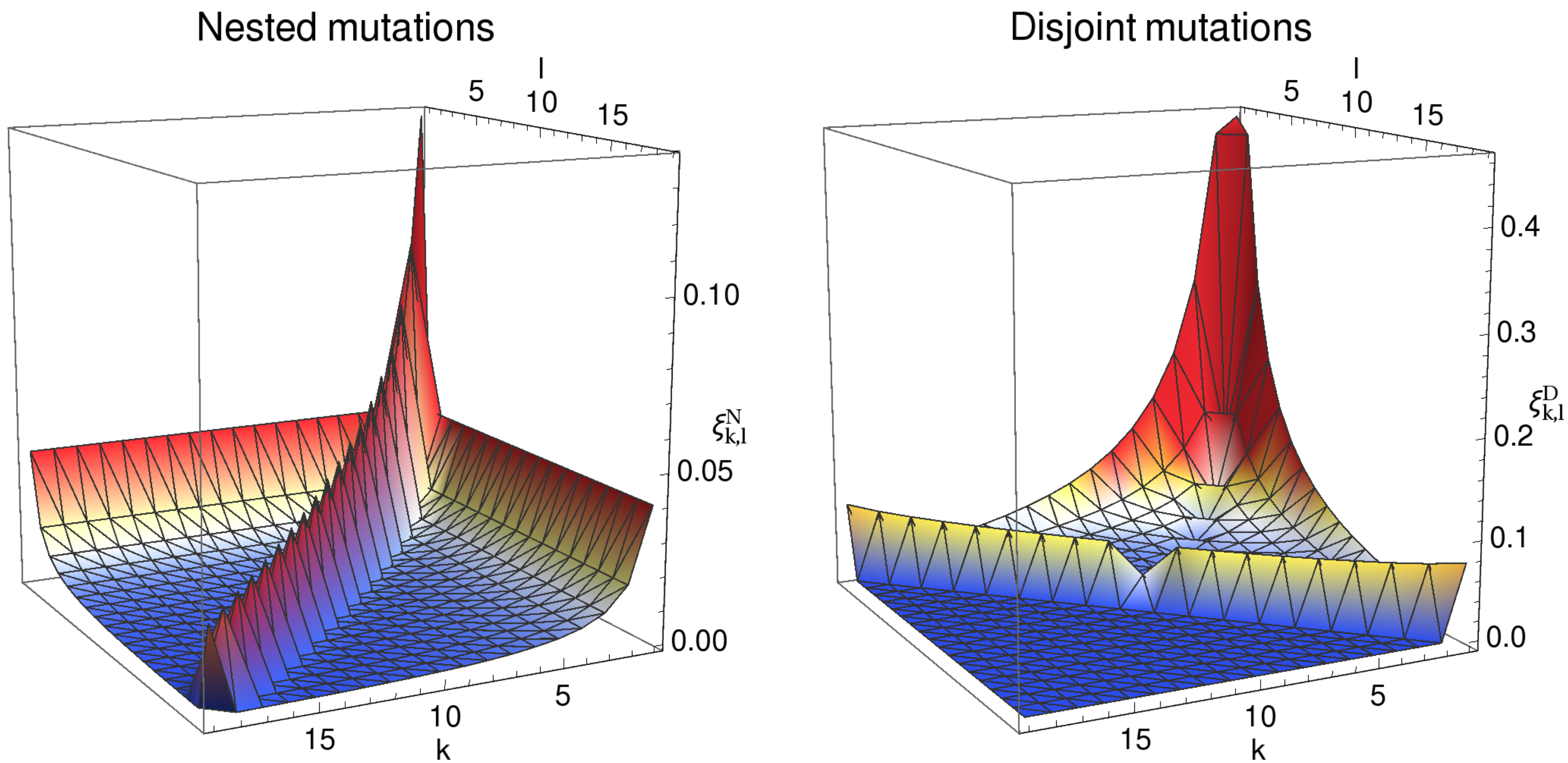




nested = inner + cooccurring + enclosing

disjoint = outer + complementary

2-SFS



for neutral model (from Kingman's coalescent)

2-SFS and spectrum of linked mutations

From *2-SFS of pairs of mutations* f,f_0
to *SFS of mutations* f *linked to a focal one* f_0

"joint" 2-SFS : "conditional" SFS of linked sites

=

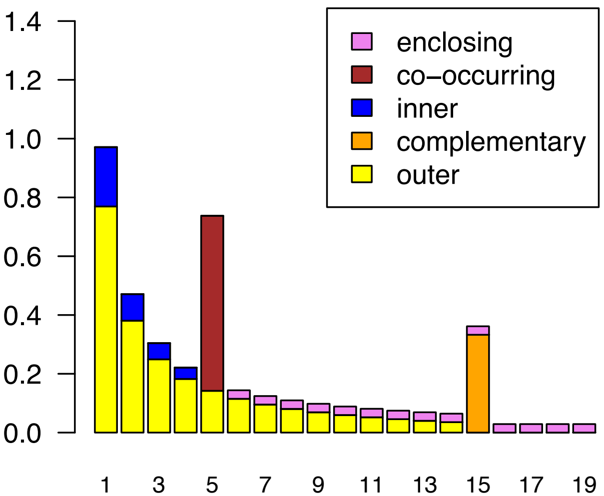
joint probability $p(f,f_0)$: conditional probability $p(f|f_0)$

Frequency of linked mutations

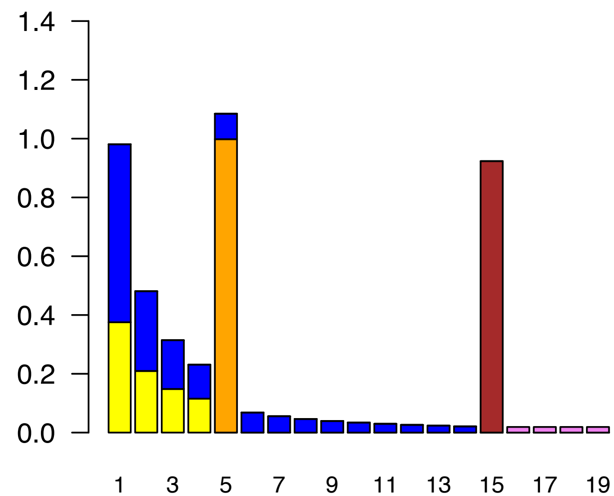
$$
\begin{aligned}
\mathrm{E}[\xi^{(i)}(f|f_0)] =& \theta L \cdot \frac{f_0}{(1-f)^2}\left(1+\frac{1}{f}+\frac{2\ln(f)}{1-f}\right), \quad f < f_0 \\
\mathrm{E}[\xi^{(co)}(f|f_0)] =& \theta L \cdot \delta(f-f_0)\frac{2f_0}{1-f_0}\left(-\frac{\ln(f_0)}{1-f_0}-1\right) \\
\mathrm{E}[\xi^{(e)}(f|f_0)] =& \theta L \cdot \frac{f_0}{(1-f_0)^2}\left(1+\frac{1}{f_0}+\frac{2\ln(f_0)}{1-f_0}\right), \quad f > f_0 \qquad (14) \\
\mathrm{E}[\xi^{(cm)}(f|f_0)] =& \theta L \cdot \delta(f-1+f_0)\left[\frac{1-f_0}{f_0}\log(1-f_0)+\left(\frac{f_0}{1-f_0}\right)^2\log(f_0)+\frac{1}{1-f_0}\right] \\
\mathrm{E}[\xi^{(o)}(f|f_0)] =& \theta L \cdot \left[\frac{1}{f}-\frac{f_0}{(1-f)^2}\left(1+\frac{1}{f}+\frac{2\ln(f)}{1-f}\right)\right. \\
& \left. -\frac{f_0}{(1-f_0)^2}\left(1+\frac{1}{f_0}+\frac{2\ln(f_0)}{1-f_0}\right)\right], \quad f < 1-f_0
\end{aligned}
$$

Frequency of linked mutations

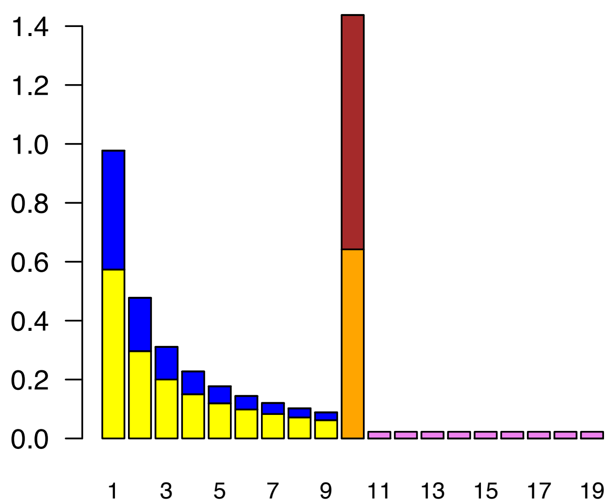
Linked mutation frequency (n=20,k=5)



Linked mutation frequency (n=20,k=15)

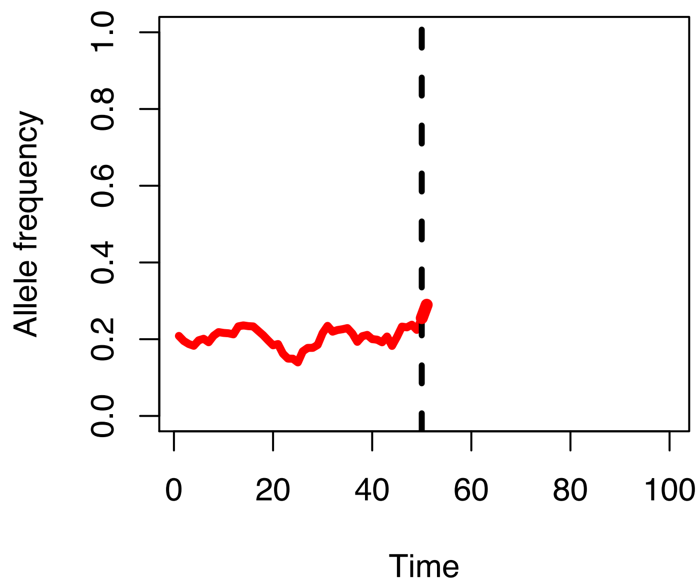


Linked mutation frequency (n=20,k=10)

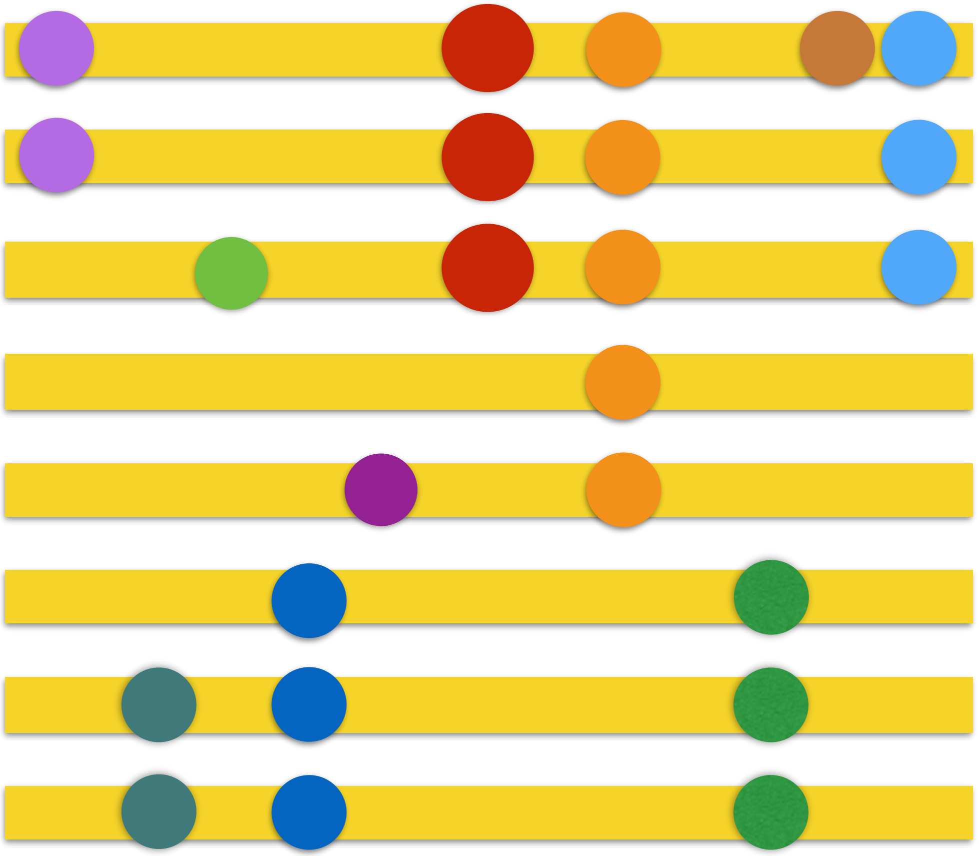


Sudden environmental change!
e.g. new viral strain

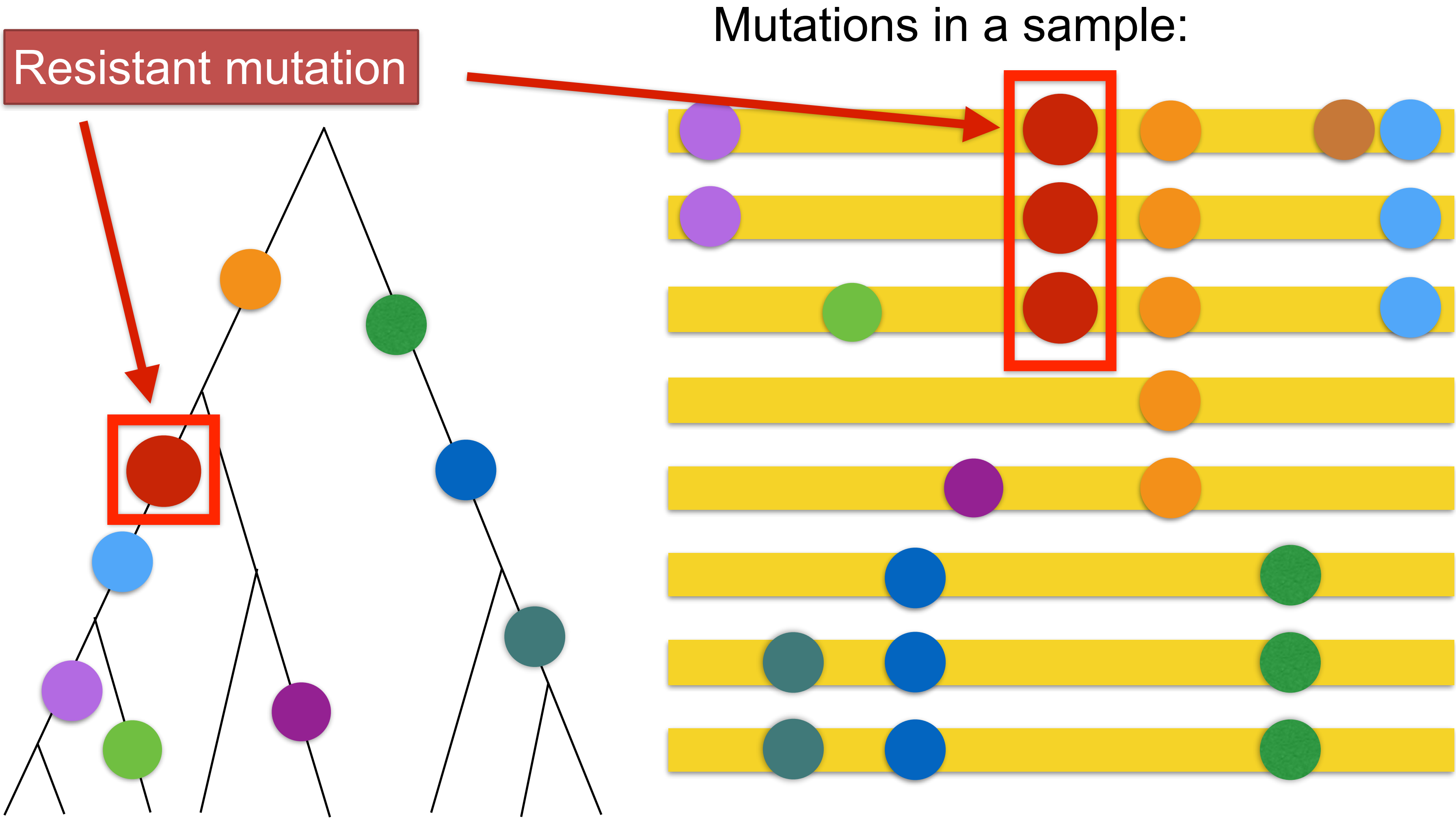
Frequency of the mutation in the population



Mutations in a sample:



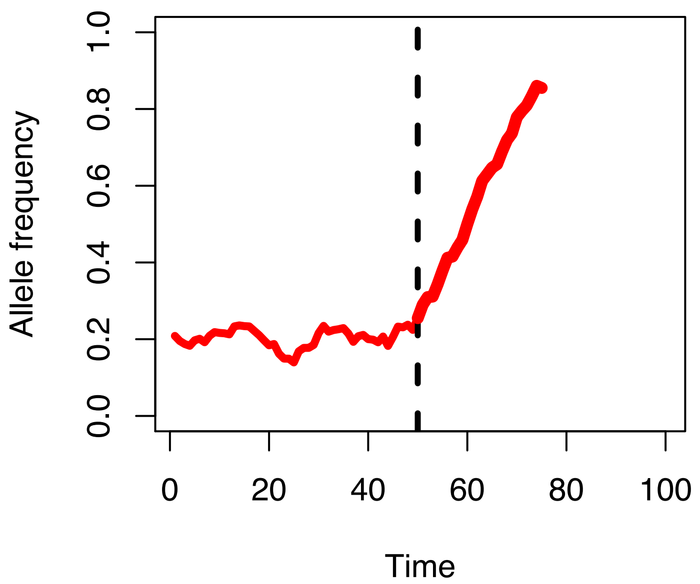
Sudden environmental change!
e.g. new viral strain



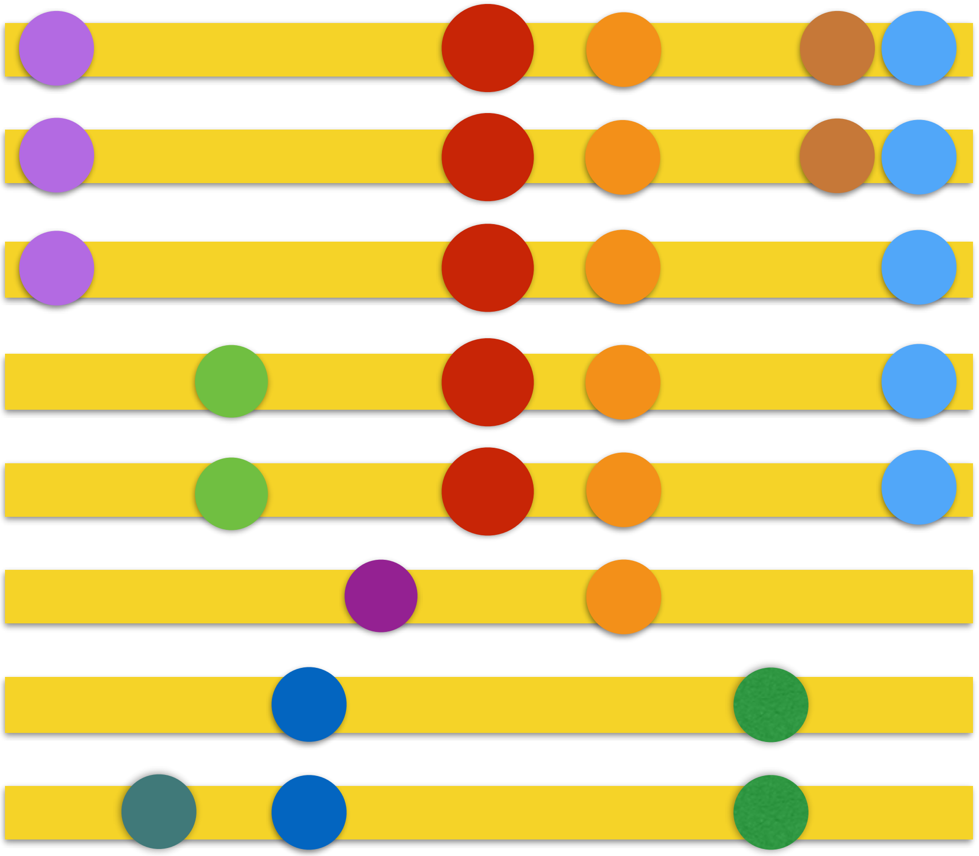
Selection on standing variation!

Selection for resistance

Frequency of the mutation in the population



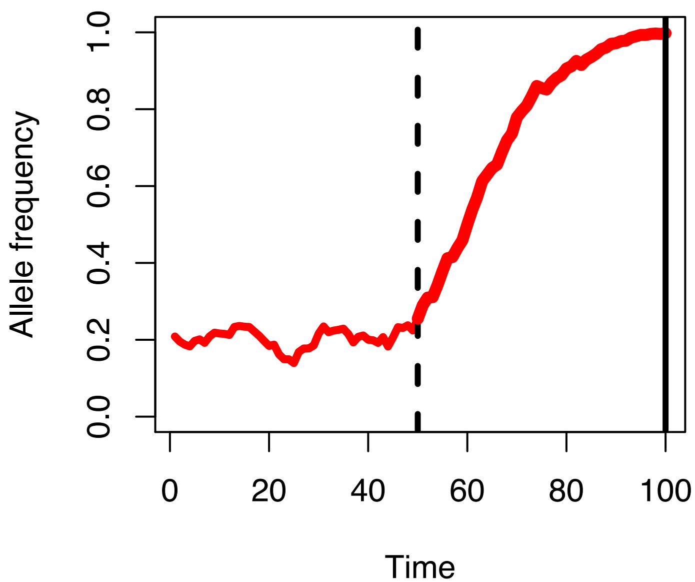
Mutations in a sample:



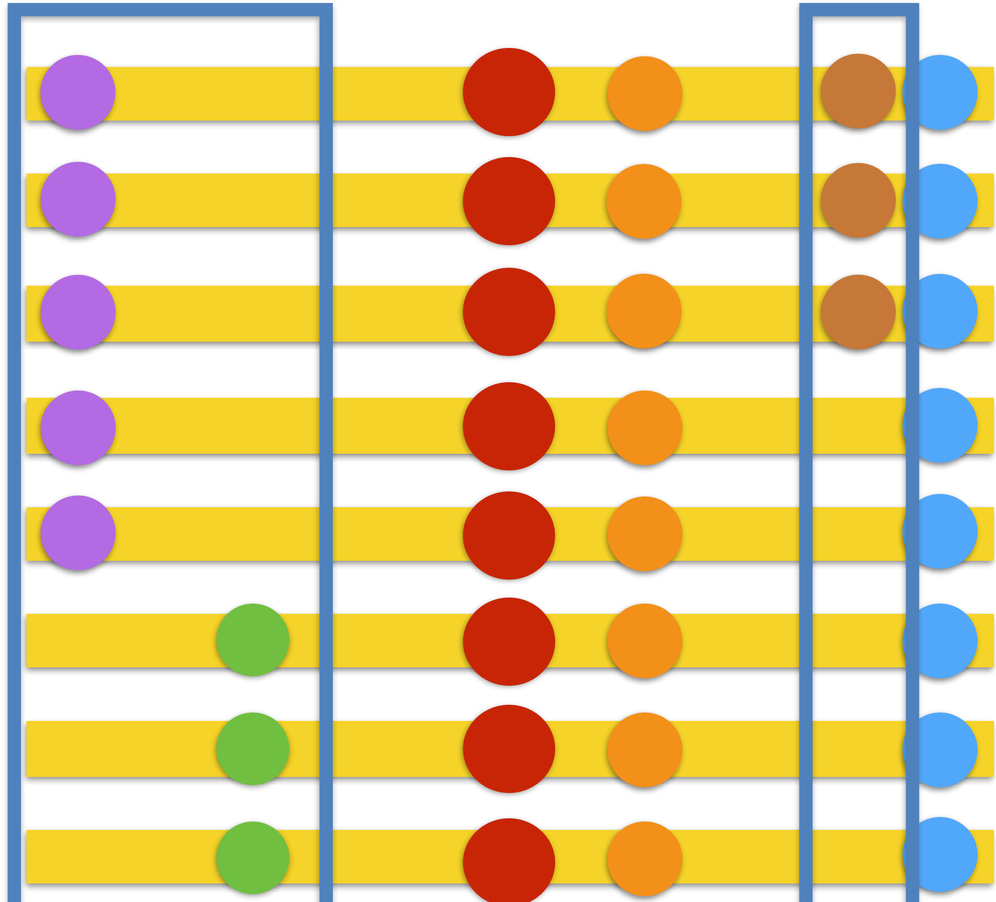
Selection on standing variation!

Selection for resistance: *soft sweep*

Frequency of the
mutation in the population



Mutations in a sample:



Detected by pattern of sequence diversity

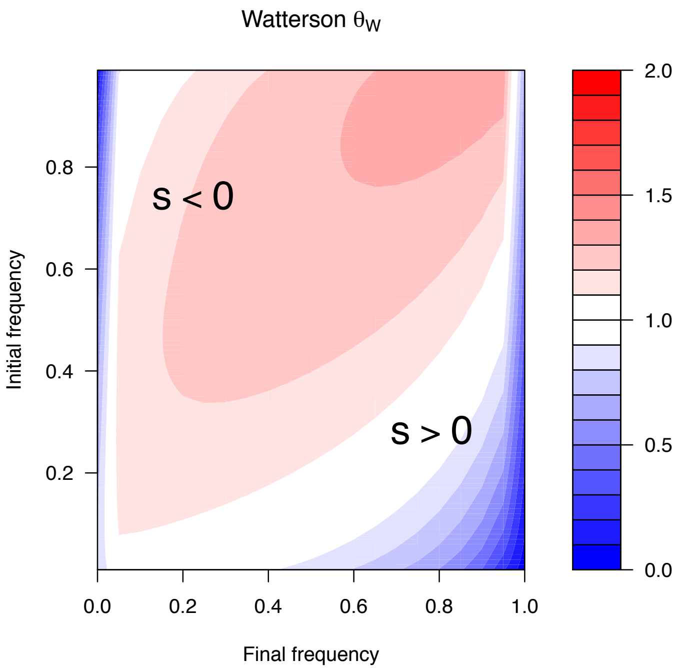
From neutral spectrum to fast sweep

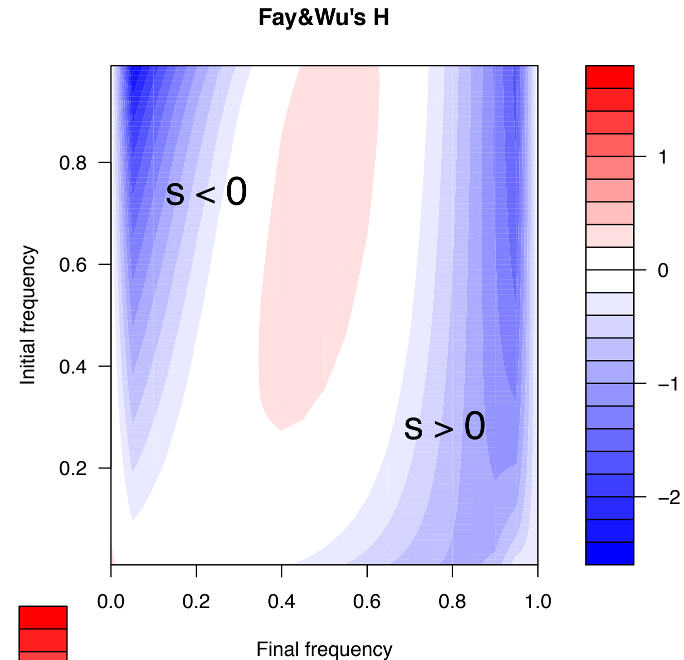
Selected allele changes frequency from f_0 to f_s

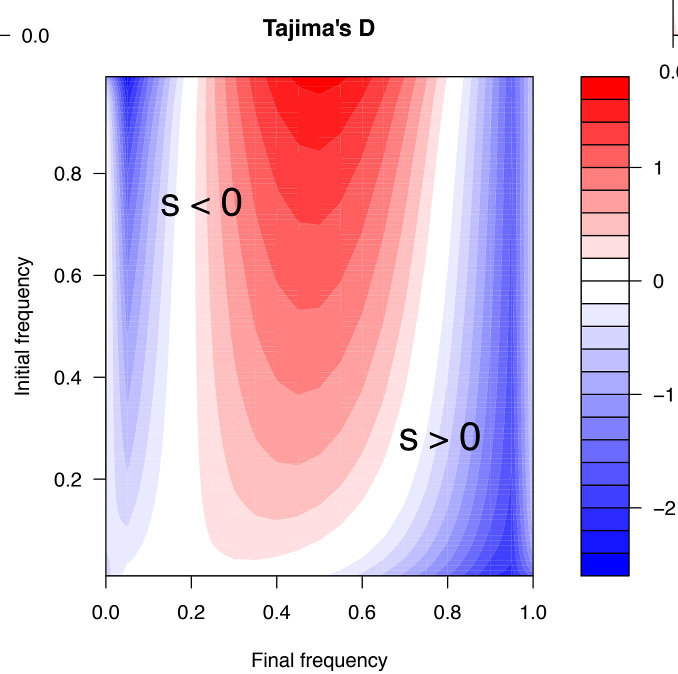
Rescale frequencies according to the component of the spectrum and the change in frequency of the selected allele:

Component	Freq. 1st subpop.	Freq. 2nd subpop.	Final frequency
nested	f/f_0	0	$f_s f/f_0$
co-occurring	1	0	f_s
containing	1	$(f-f_0)/(1-f_0)$	$f_s+(1-f_s)(f-f_0)/(1-f_0)$
complementary	0	1	$1-f_s$
exclusive	0	$f/(1-f_0)$	$(1-f_s)f/(1-f_0)$

Signatures of ongoing selection







Perspectives

- Composite Likelihood-Ratio Test for partial sweeps
- Maximum Composite Likelihood inference of the initial and final frequency of the sweep
- Similar methods applied to balancing selection
- Include recombination (approximate) and joint sweep finder with haplotype-based methods

Acknowledgments

- Emanuele Raineri
- Thomas Wiehe
- Paolo Ribeca
- Adrià Madico Ferrer

AGENCE NATIONALE DE LA RECHERCHE ANR

THE Pirbright INSTITUTE

BBSRC bioscience for the future

Thanks for your attention!

Frequency spectrum linked to a soft selective sweep

$$\xi^{soft}(f|f_0) = \theta L \cdot \frac{f_0^2}{(1 - f f_0)^2} \left(1 + \frac{1}{f f_0} + \frac{2\ln(f f_0)}{1 - f f_0}\right)$$

Applicable to selection on standing variation