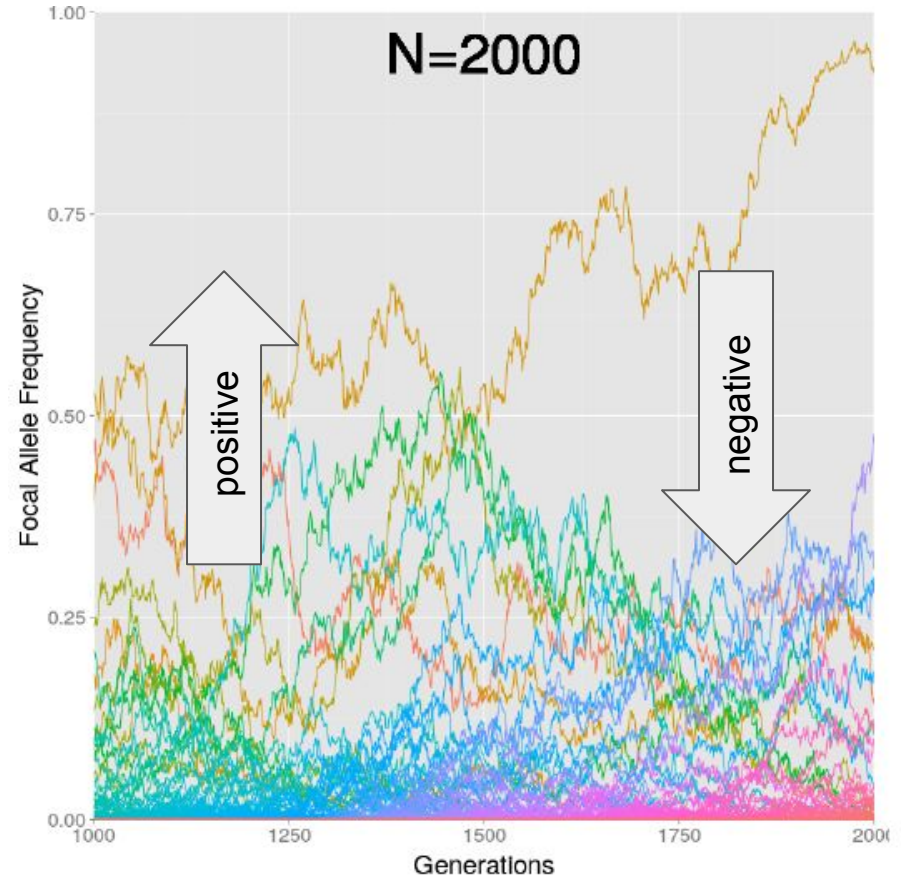


5 April, 2016:
Population genetics III

Site frequency spectrum (3)

Selection

- described with classical drift models, not with coalescence
- naive simulation, neutral expectation:
- selection increases/decreases fitness and thus allele frequency

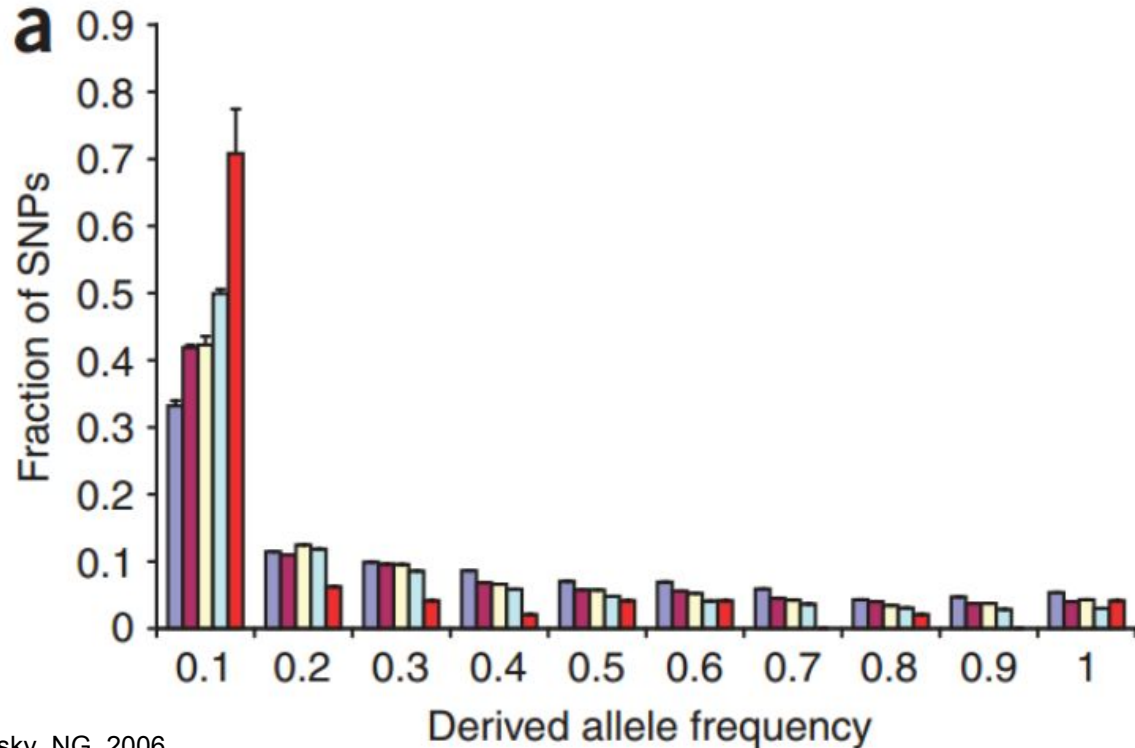


Site frequency spectrum (3)

- miRNA conservation:

- syn sites
- 3'UTR
- cons 3'UTR
- nonsyn sites
- cons miRNA

- increases in low frequencies due to negative selection

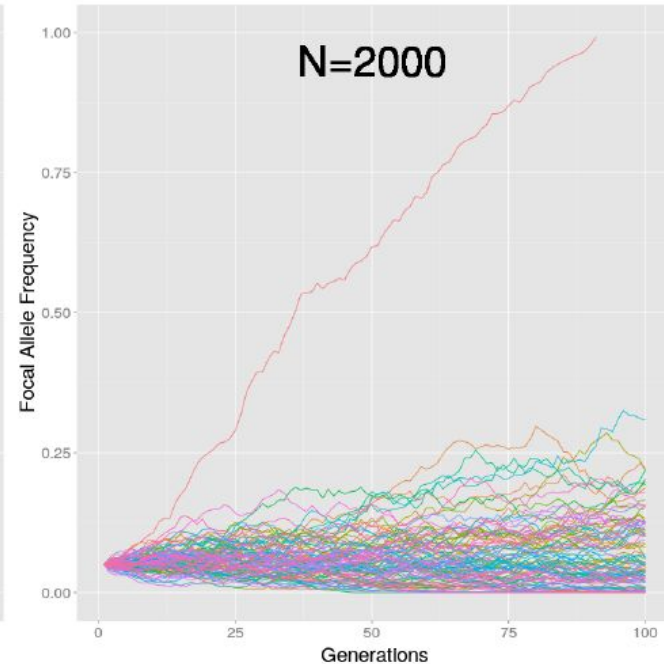
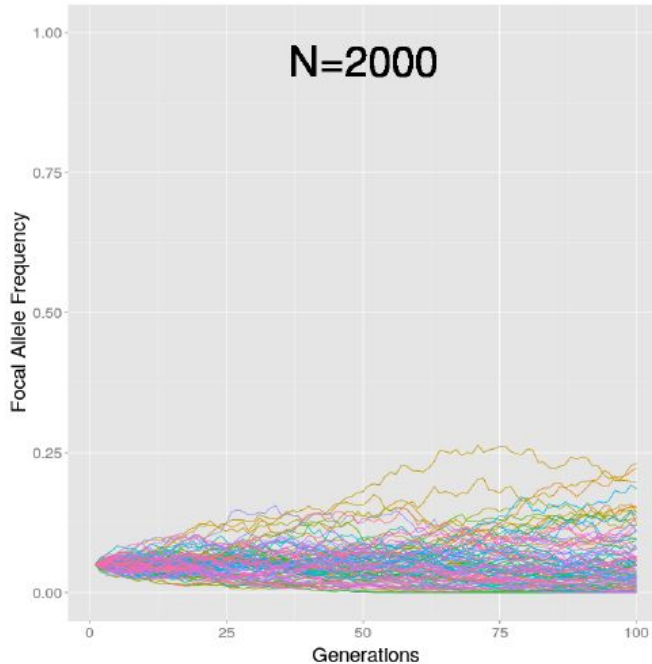


Chen and Rajewsky, NG, 2006

Selective sweeps

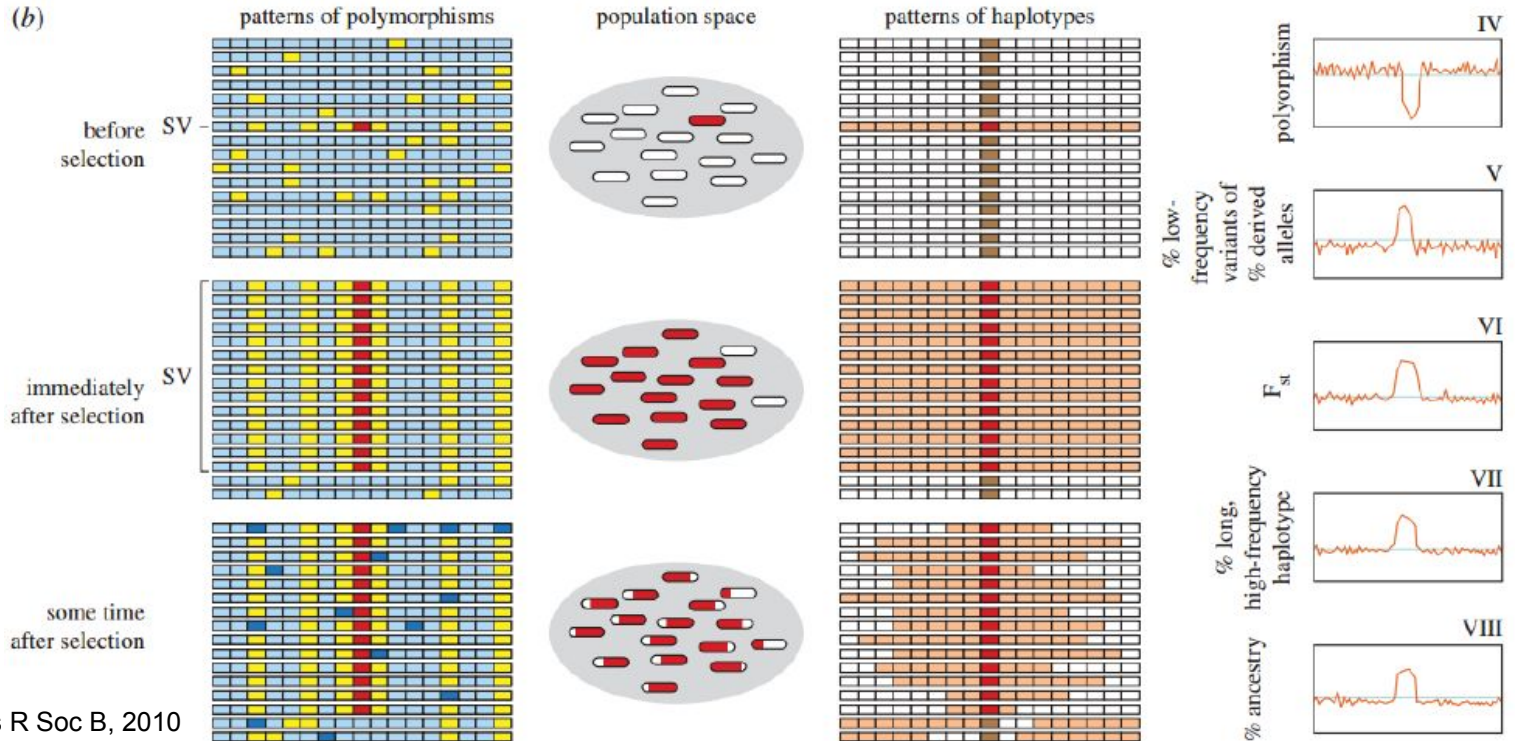
Selective (=fitness) advantage increases allele's chances of fixation

- rapid increase in frequency is called *sweep*



Methods for detecting adaptive evolution

Selective sweep affects neighbouring loci, removes variation



Oleksyk et al., Phil Trans R Soc B, 2010

Selective sweeps

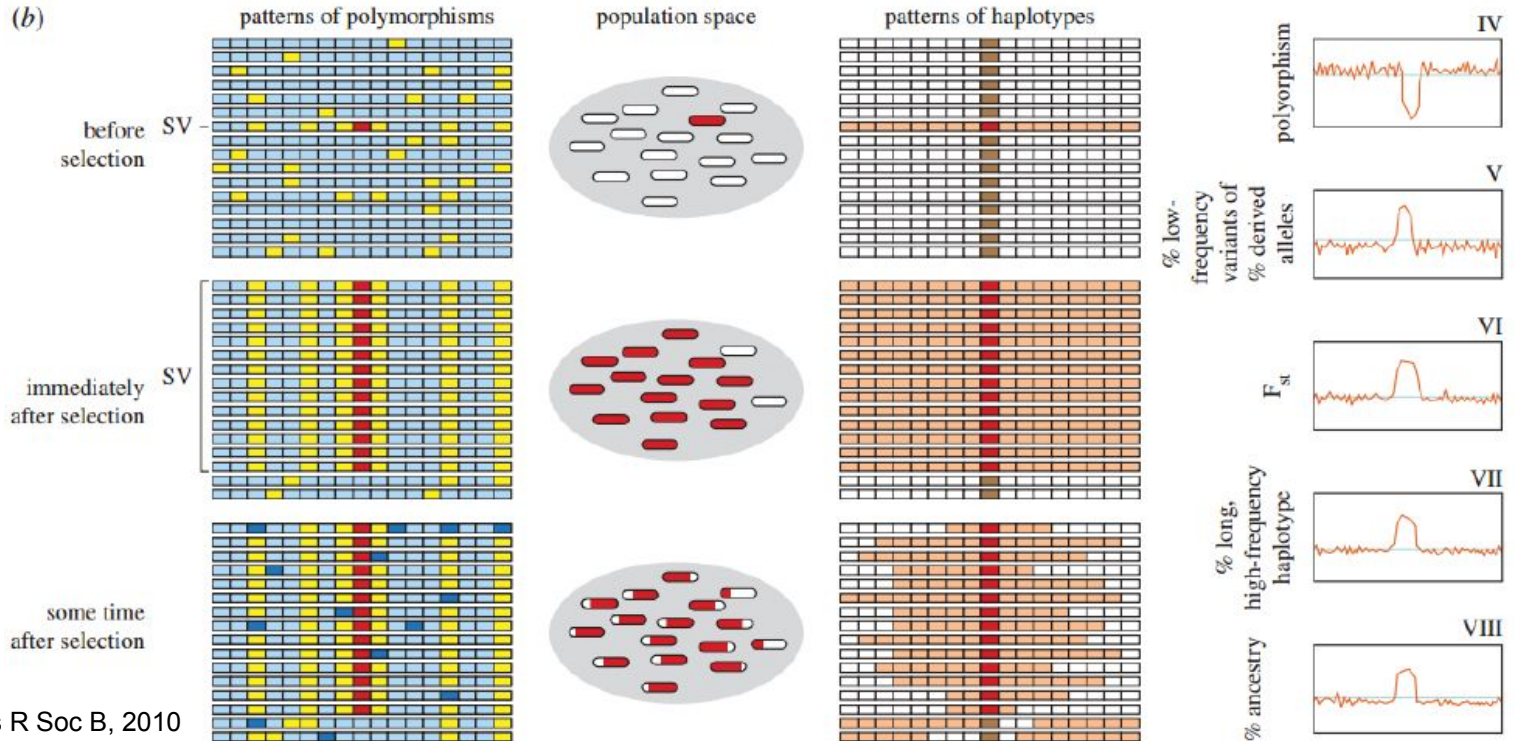
Why interesting?

For example, human was an ape eating fruits and living in Africa

- quick expansion to new environments
 - climate, sun's radiation, altitude
- changing diet
 - agriculture, dairy farming, sugar(!)
- population structure and bottlenecks
 - locally high frequencies of harmful/disease alleles

Methods for detecting adaptive evolution

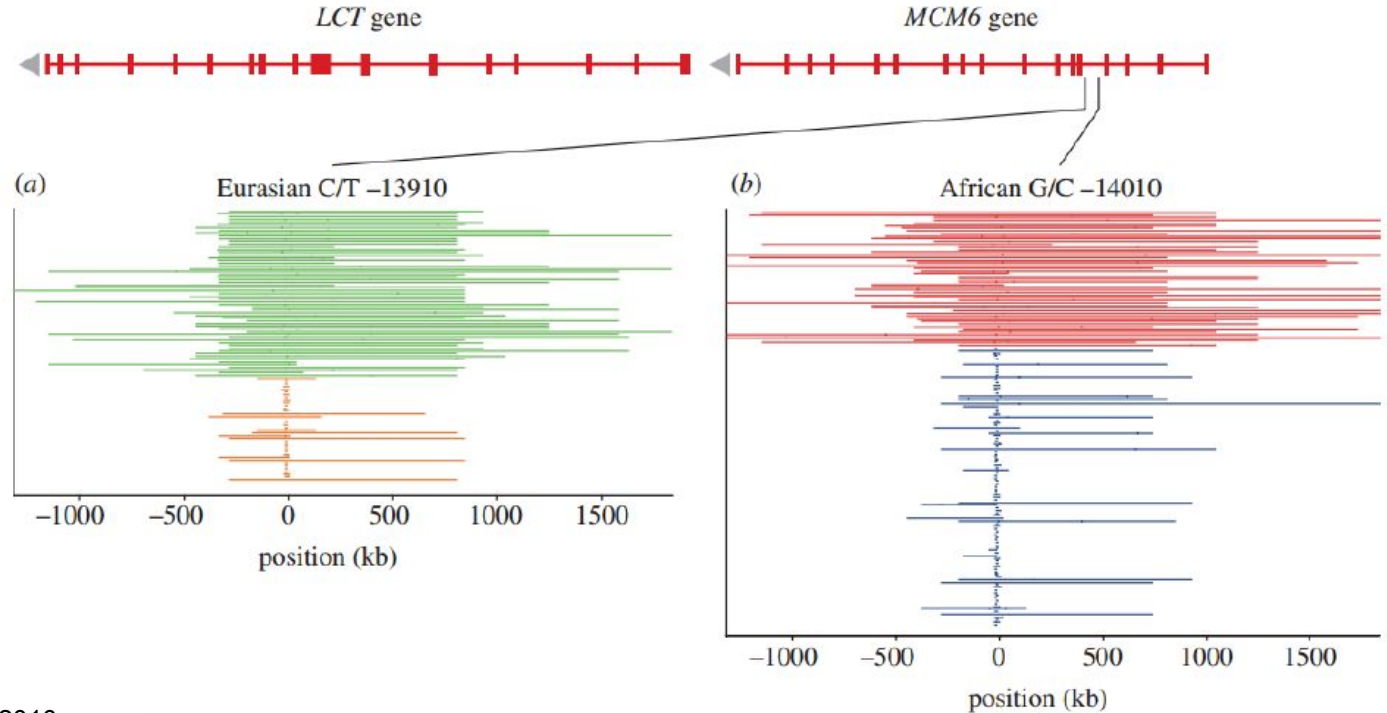
Selective sweep affects neighbouring loci, removes variation



Oleksyk et al., Phil Trans R Soc B, 2010

Methods for detecting adaptive evolution

Young alleles sit in long haplotypes

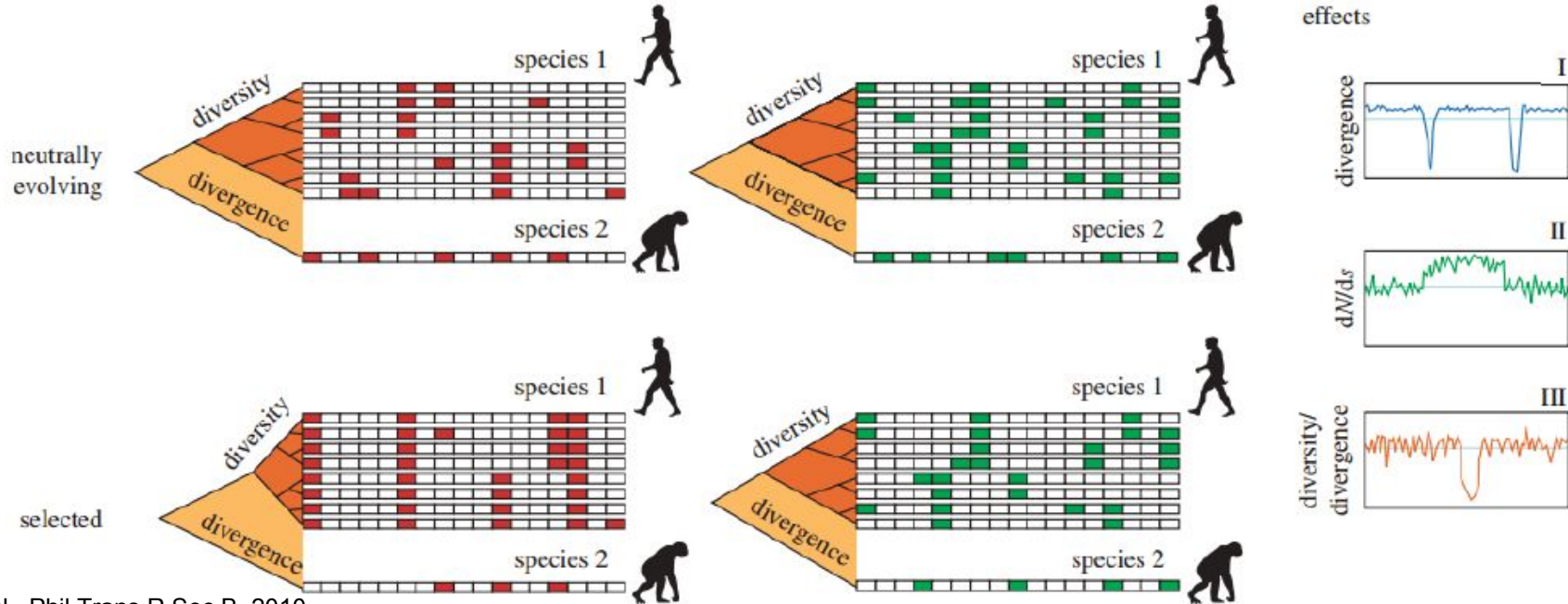


Oleksyk et al., Phil Trans R Soc B, 2010

Methods for detecting adaptive evolution

Outgroup-based approaches

- intra- and inter-species variations are different

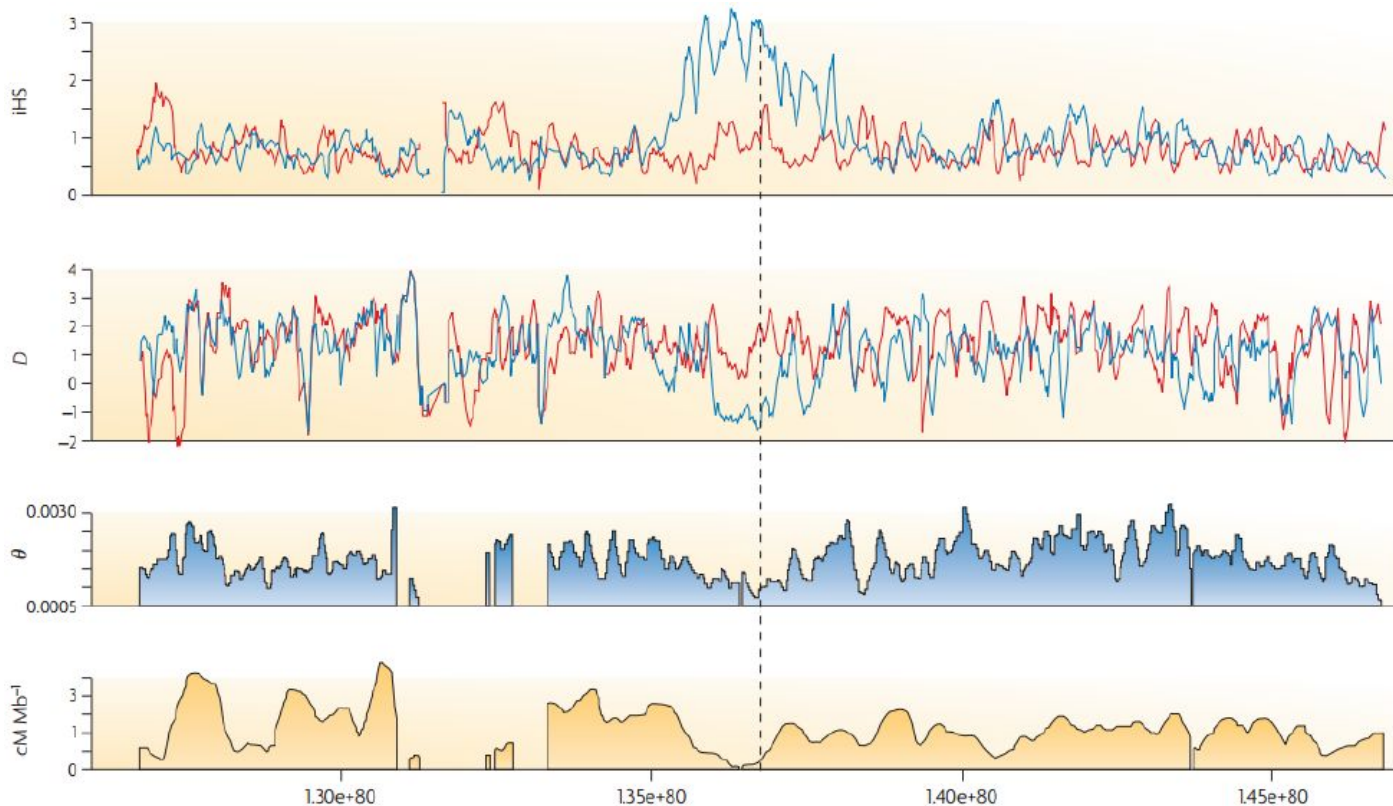


Olekyk et al., Phil Trans R Soc B, 2010

Methods for detecting adaptive evolution

Signal in summary statistics is often noisy

- LCT locus
- **asian**
- **caucasian**



Nielsen et al., NRG, 2007

Methods for detecting adaptive evolution: F_{ST}

F_{ST} , fixation index

- measures how much populations differ genetically
- $F_{ST} = (H_T - H_S) / H_T$ where H_T is heterozygosity in total population and H_S is heterozygosity in subpopulations

Interpretation:

- low F_{ST} : variation within subpopulations, i.e. no differences between populations
- high F_{ST} : variation between subpopulations, i.e. populations are differences
- typically functional sites have low F_{ST} ; high values suggest selection

Methods for detecting adaptive evolution: F_{ST}

An example with real data

SNP	Africa	Europe	East Asia
FTO rs9939609	0.471	0.426	0.157
TCF7L2 rs7901695	0.629	0.325	0.044

For rs9939609 and AFR vs. EUR:

$$f = (0.471 + 0.426) / 2 = 0.449$$

$$H_T = 2 \times 0.449 \times (1 - 0.449) = 0.495$$

$$H_S = [2 \times 0.471 \times 0.529 + 2 \times 0.426 \times 0.574] / 2 = 0.494$$

$$F_{ST} = (0.495 - 0.494) / 0.495 = 0.002$$

Methods for detecting adaptive evolution: F_{ST}

An example with real data

SNP	Africa	Europe	East Asia
FTO rs9939609	0.471	0.426	0.157
TCF7L2 rs7901695	0.629	0.325	0.044

For rs9939609 and AFR vs. EAS:

$$f = (0.471 + 0.157) / 2 = 0.314$$

$$H_T = 2 \times 0.314 \times (1 - 0.314) = 0.431$$

$$H_S = [2 \times 0.471 \times 0.529 + 2 \times 0.157 \times 0.843] / 2 = 0.382$$

$$F_{ST} = (0.431 - 0.382) / 0.431 = 0.114$$