

Detecting Adaptive Evolution

Alan R. Rogers

March 28, 2023

Conventional wisdom

Something must have happened to weaken the selective pressure drastically. We cannot escape the conclusion that man's evolution towards manness suddenly came to a halt.

—Ernst Mayr 1963

Natural selection has almost become irrelevant in human evolution. There's been no biological change in humans in 40,000 or 50,000 years. Everything we call culture and civilization we've built with the same body and brain.

—Stephen Jay Gould 2000

Certainly, human nature is fixed. It's universal and unchanging —common to every baby that's born, down through the history of our species.

—Helena Cronin 2000

Conventional wisdom

Something must have happened to weaken the selective pressure drastically. We cannot escape the conclusion that man's evolution towards manness suddenly came to a halt.

—Ernst Mayr 1963

Natural selection has almost become irrelevant in human evolution. There's been no biological change in humans in 40,000 or 50,000 years. Everything we call culture and civilization we've built with the same body and brain.

—Stephen Jay Gould 2000

Certainly, human nature is fixed. It's universal and unchanging —common to every baby that's born, down through the history of our species.

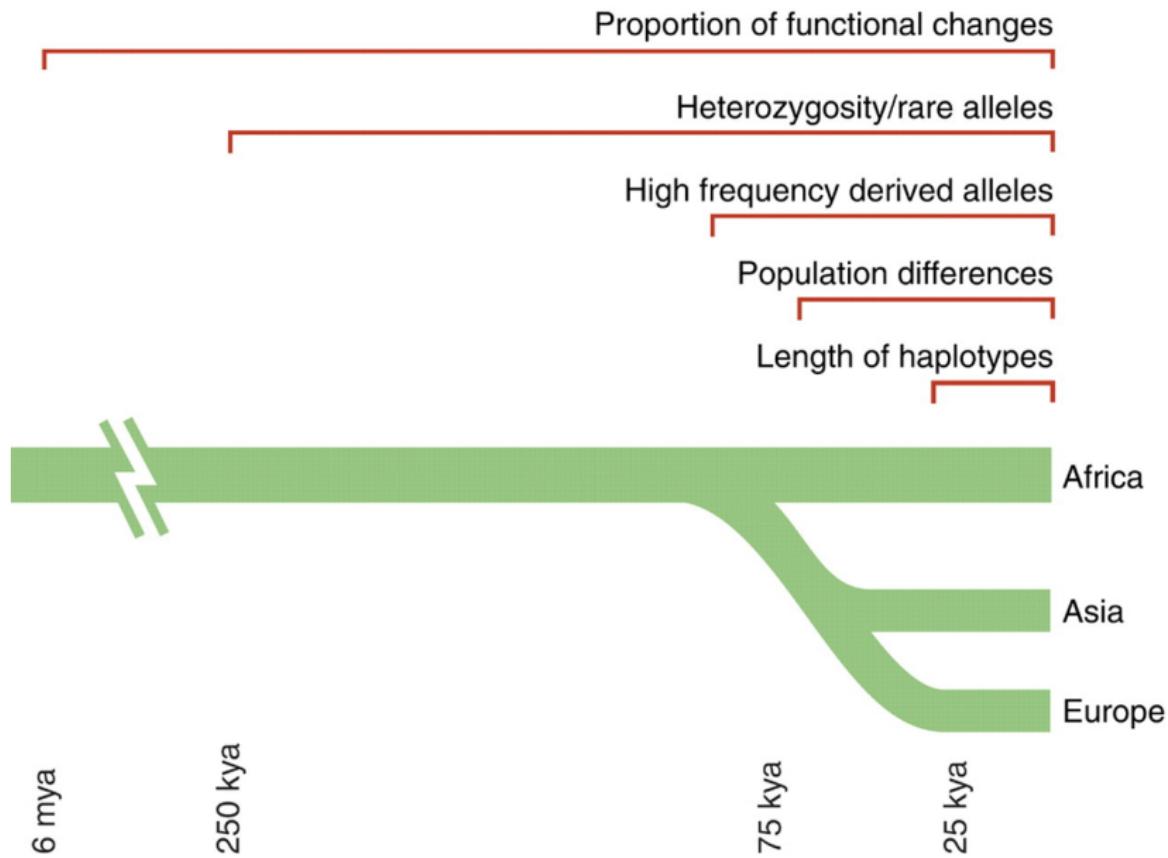
—Helena Cronin 2000

Is this really true? How could we know?

Signatures of selection

- ▶ High proportion of functional changes
- ▶ Reduction of gene diversity
- ▶ Population differences
- ▶ Excess of singletons
- ▶ Common allele on long LD block
- ▶ Singleton density

Time scale for signatures of selection



PRM1 gene

PRM1 Exon 2

	44 bp	11,341,281	Chromosome 16	11,341,324
	STOP H R R C R P R Y R P R C C R			
Human	AATCACAGAAGATGTAG CGCC AGAC A TGGACC CCCGCCGTCGTGG			
Chimp	AATCACAGAAGATGCAGAGTAAGACCTGGACGCCGCCGTCGTGG	STOP H R R R R M R S R R R C C R		

- ▶ compacts sperm DNA
- ▶ 13/14 human-chimp diffs are non-synonymous (6 shown here)

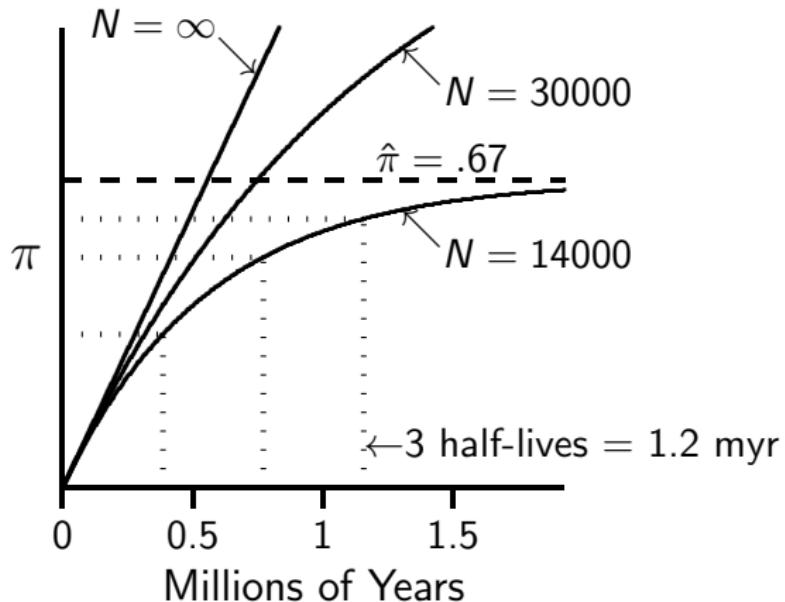
Melanocortin 1 receptor (MC1R) locus

- ▶ Affects color of skin and hair.
- ▶ human-chimp $K_a/K_s = 0.63$: large for a functional protein \Rightarrow weak selective constraint
- ▶ Yet $K_a/K_s = 0$ among Africans.
- ▶ On the other hand, $K_a/K_s \gg 0$ among Europeans

Hypothesis

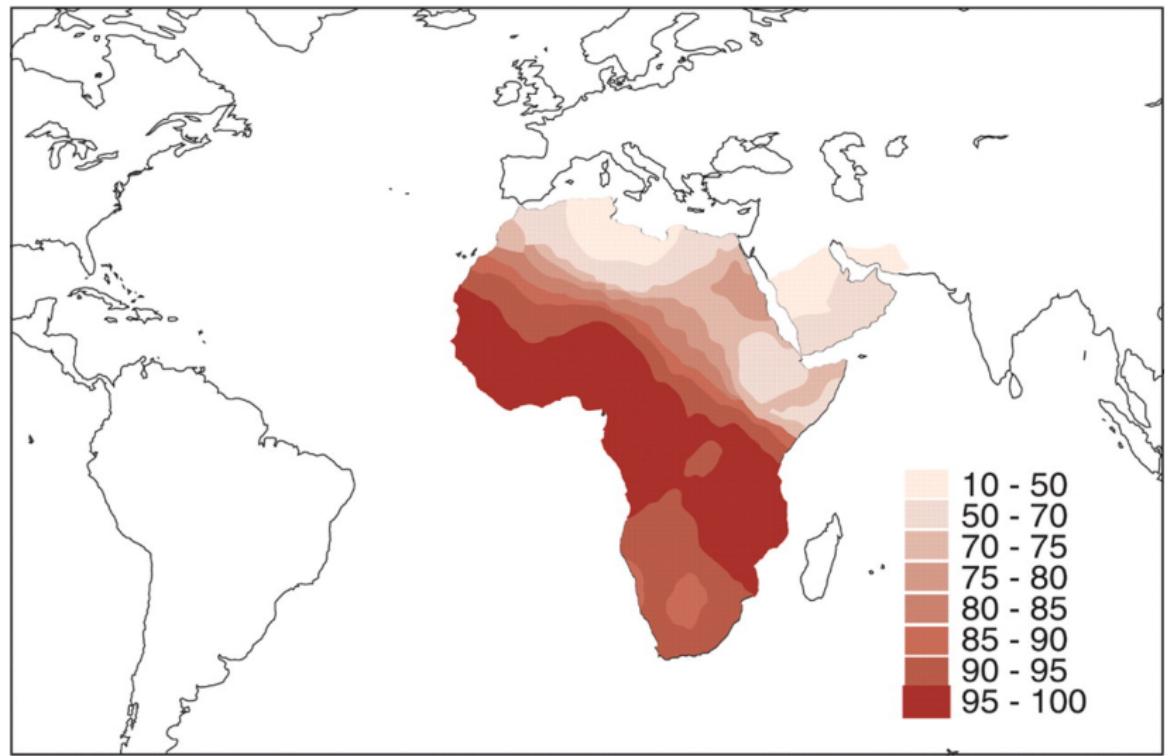
- ▶ Before loss of body hair: weak selective constraint on MC1R
- ▶ After loss: constraint strong w/i Africa; weak w/i Europe
- ▶ Selective sweep at loss of body hair
- ▶ Neutral diversity within Africa accumulated since the sweep.

Time required to generate African neutral $\hat{\pi}$

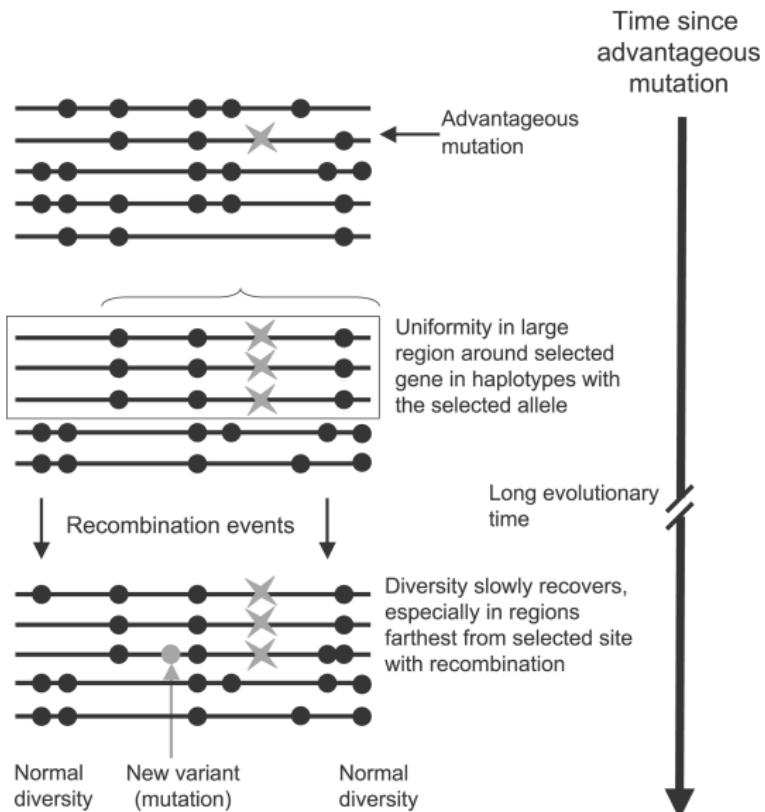


Trajectories of π following a selective sweep. Dotted lines show the half-lives when $N = 14,000$; the dashed line shows $\hat{\pi}$, the observed value of π in Africa.

Duffy map



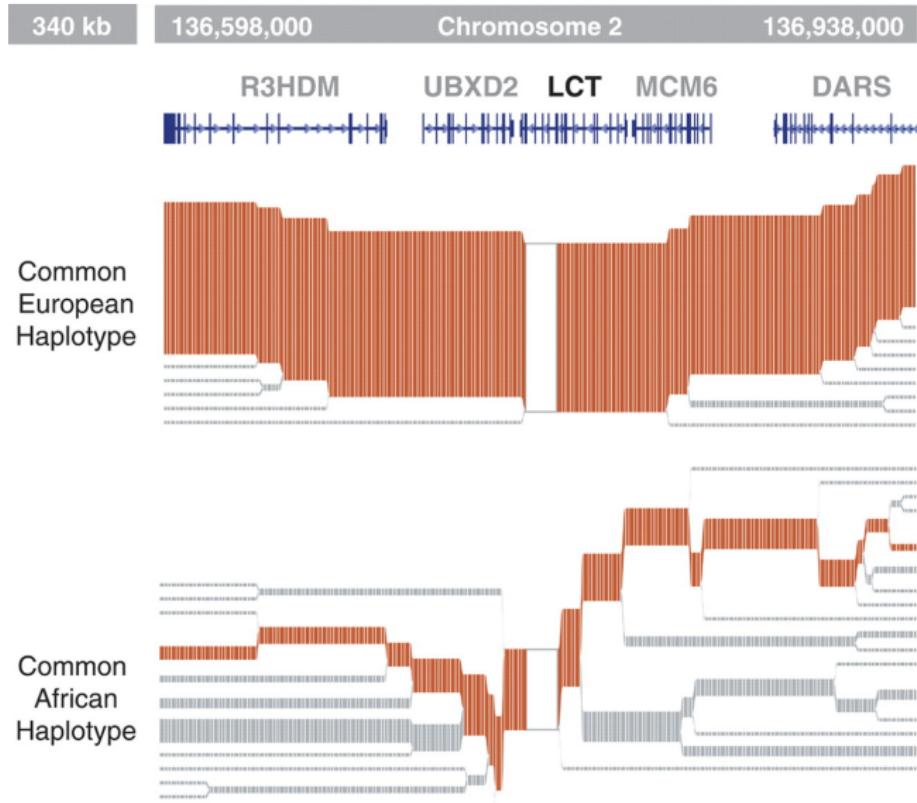
How a “hard” selective sweep generates LD



DNA sequences from region of human lactase gene

```
cgcttcaggcatttatctaaacagaccaacgtAgggtacaatgcctaaccagacgttcaactct
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.gttccgG..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....ggt....cg.gt.t..c
51 ta.ccggta....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt....cg.gt.t..c
52 ta.ccggta....g.t..atc.g.tc.g.tg.tc.gG..a.g.g....tg....ggt....cg.gt.t..c
53 ta.ccggta....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt....cg.gt.t..c
```

Huge block of LD around lactase allele in Europe



Idea behind Extended Haplotype Homozygosity (EHH)

```
ctaaacagaccaacgtAgggtacaatgcctaaccagacgttt  
20 .....  
21 .....  
22 .....  
23 .....  
24 .....  
25 .....  
26 .....  
27 t.....  
28 t.....  
29 .....c.....  
37 .....G..a.gt.....t.....gac.c  
38 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
39 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
40 t..g.....t..ttccgG..a.gt.....t.....gac.c  
41 t..g.....t..gttccgG..a.gt.....t.....gac.c  
44 t..g.....t..ttc.gG..acgt.....t.....gac.c  
45 t..g.....t..gttc.gG..a.gt.....t.....gac.c  
46 t..gg..c.....tc.gGaaa.g..ccttt...tg.....cg.  
47 t..g.....t..gttccgG..a.gt.....t.....gac.c  
48 t..g.....t..gttccgG..a.gt.....t.....gac.c  
49 t..g.....t..gttccgG..a.gt.....t.....gac.c  
50 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.  
51 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.
```

Upper half: pairs of chromosomes
are identical at most sites

Idea behind Extended Haplotype Homozygosity (EHH)

```
ctaaacagaccaacgtAgggtacaatgcctaaccagacgttt  
20 .....  
21 .....  
22 .....  
23 .....  
24 .....  
25 .....  
26 .....  
27 t.....  
28 t.....  
29 .....c.....  
37 .....G..a.gt.....t.....gac.c  
38 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
39 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
40 t..g.....t.tttcgG..a.gt.....t.....gac.c  
41 t..g.....t.gttccgG..a.gt.....t.....gac.c  
44 t..g.....t.ttc.gG..acgt.....t.....gac.c  
45 t..g.....t.gttc.gG..a.gt.....t.....gac.c  
46 t..gg..c.....tc.gGaaa.g..ccttt...tg.....cg..  
47 t..g.....t.gttccgG..a.gt.....t.....gac.c  
48 t..g.....t.gttccgG..a.gt.....t.....gac.c  
49 t..g.....t.gttccgG..a.gt.....t.....gac.c  
50 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.  
51 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.
```

Upper half: pairs of chromosomes
are identical at most sites

Lower half: pairs identical at fewer
sites

Idea behind Extended Haplotype Homozygosity (EHH)

```
ctaaacagaccaacgtAgggtacaatgcctaaccagacgttt  
20 .....  
21 .....  
22 .....  
23 .....  
24 .....  
25 .....  
26 .....  
27 t.....  
28 t.....  
29 .....c.....  
37 .....G..a.gt.....t.....gac.c  
38 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
39 t..gg..c.....tc.gGaaa.g..ccttt...tg.....c..  
40 t..g.....t..ttccgG..a.gt.....t.....gac.c  
41 t..g.....t..gttccgG..a.gt.....t.....gac.c  
44 t..g.....t..ttc.gG..acgt.....t.....gac.c  
45 t..g.....t..gttc.gG..a.gt.....t.....gac.c  
46 t..gg..c.....tc.gGaaa.g..ccttt...tg.....cg..  
47 t..g.....t..gttccgG..a.gt.....t.....gac.c  
48 t..g.....t..gttccgG..a.gt.....t.....gac.c  
49 t..g.....t..gttccgG..a.gt.....t.....gac.c  
50 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.  
51 tcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.
```

Upper half: pairs of chromosomes
are identical at most sites

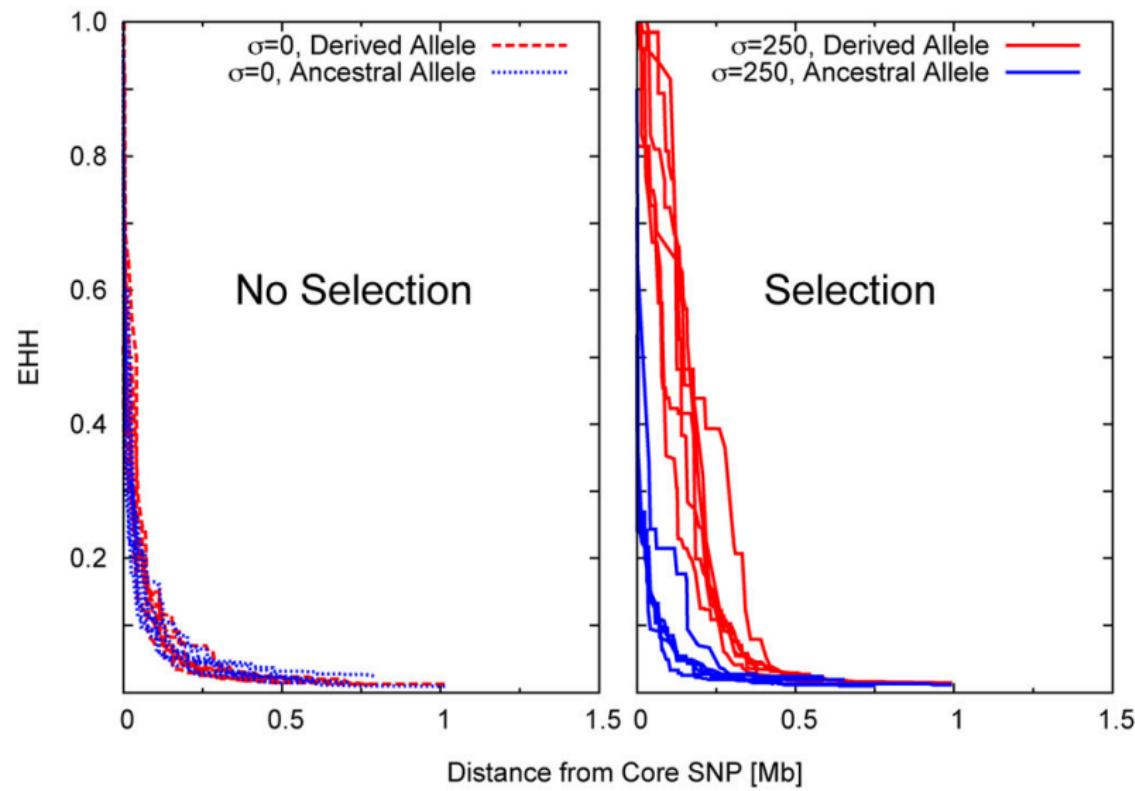
Lower half: pairs identical at fewer
sites

This idea underlies EHH.

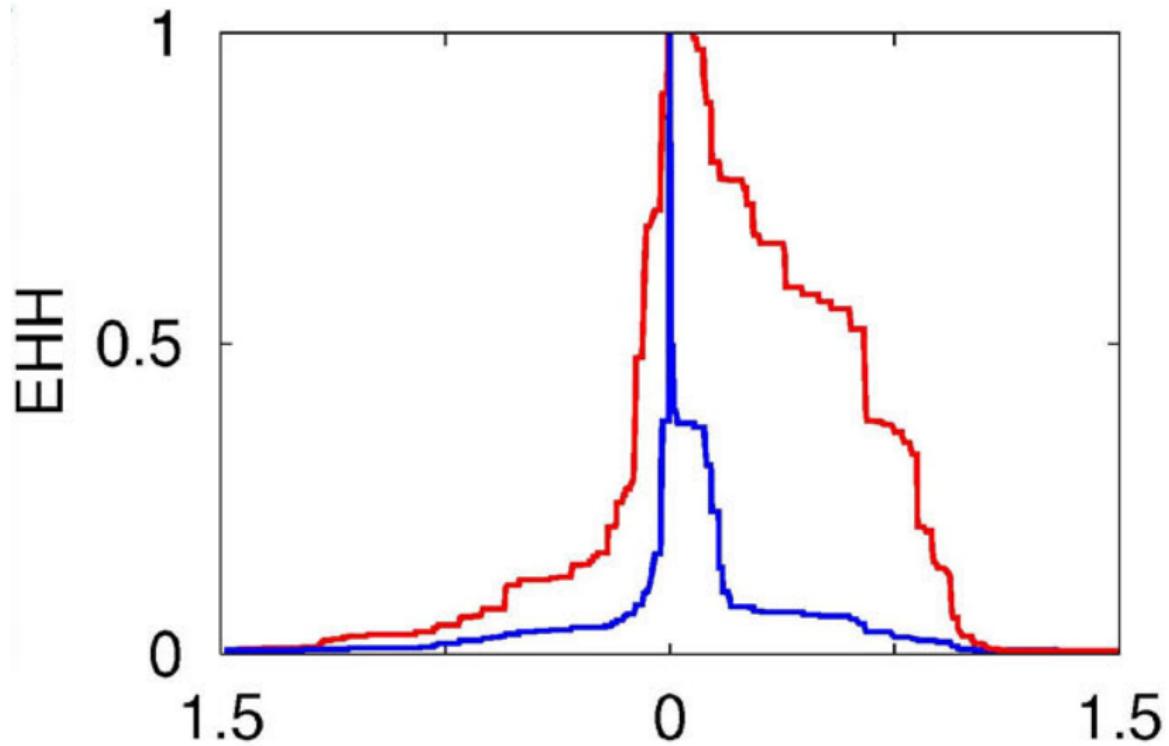
Extended haplotype homozygosity (EHH)

- ▶ Select chromosomes that carry allele A at focal site.
- ▶ Within this set, calculate the fraction of pairs that are identical at another site, x base-pairs away. This is $\text{EHH}(x)$.
- ▶ Do the same for chromosomes that *don't* carry A .
- ▶ **relative EHH** is the ratio of the two.

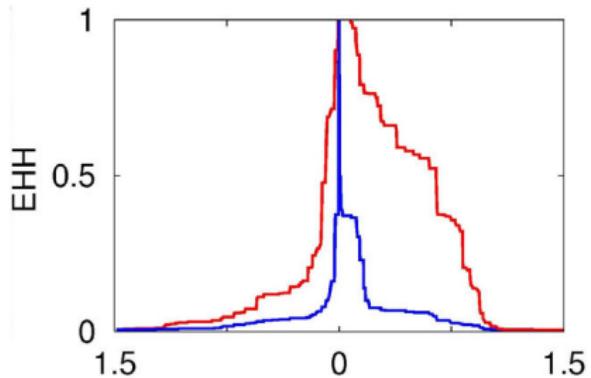
Decay of EHH with distance along chromosome



EHH at candidate locus: SPAG4

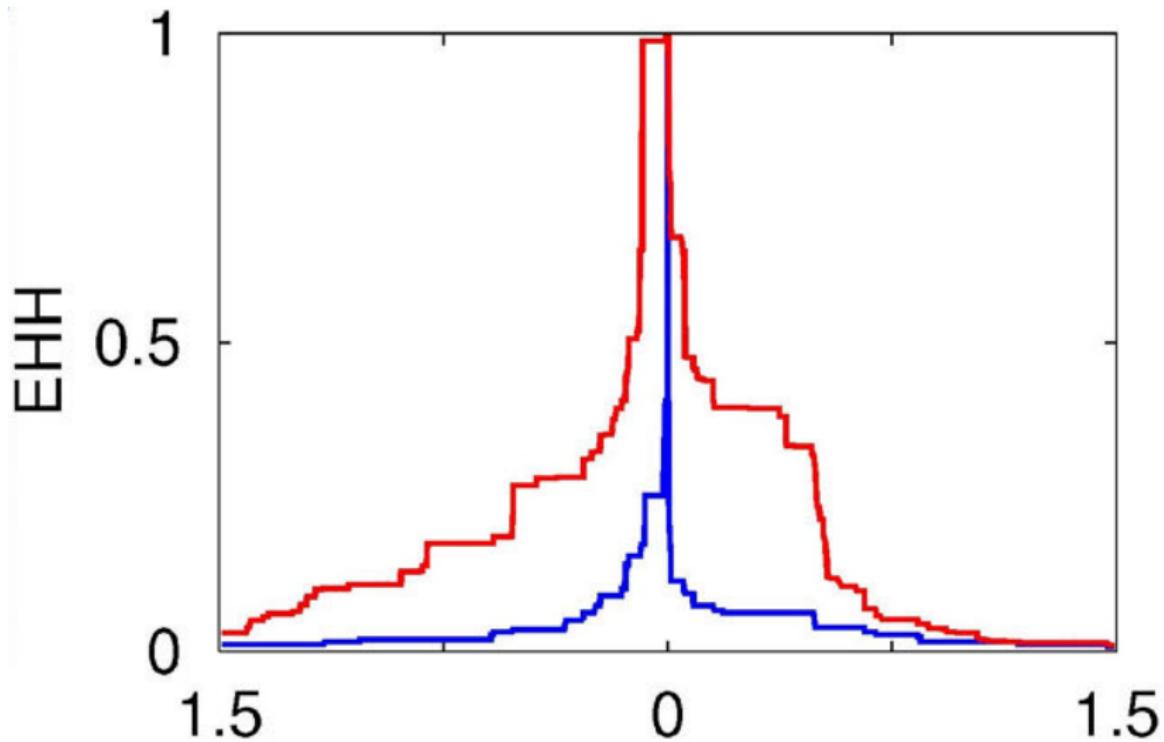


iHS, the integrated haplotype score

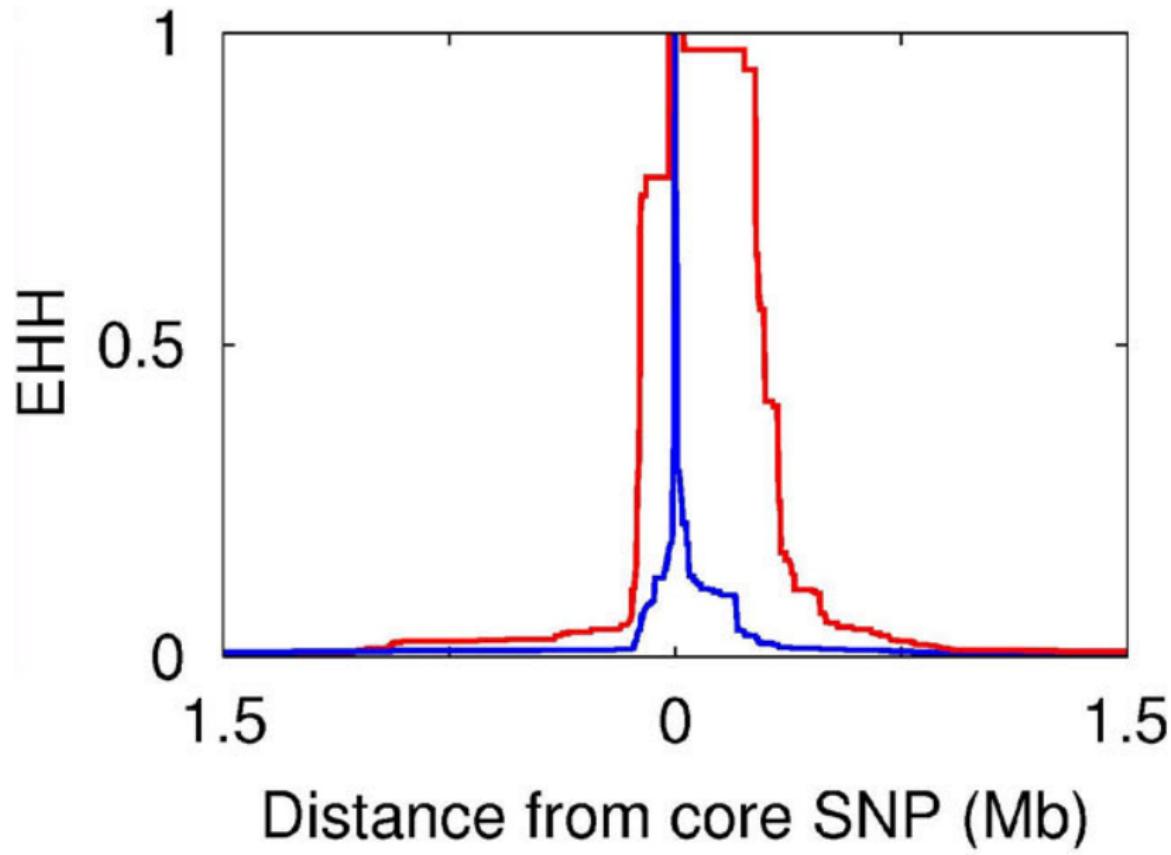


iHS is a normalized measure of the area between the two curves.

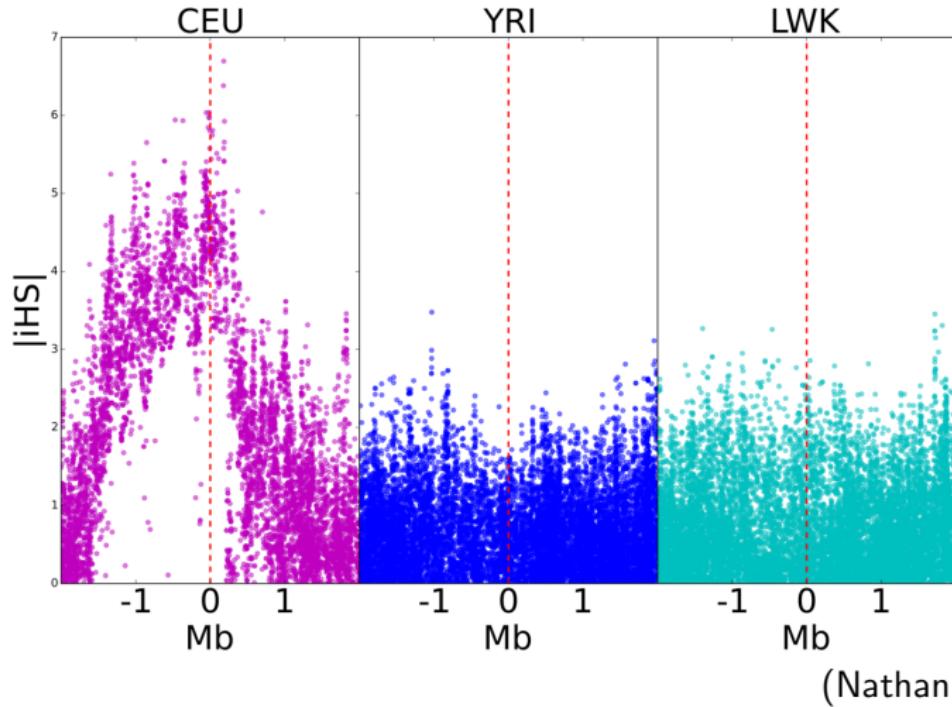
EHH at candidate locus: SNTG1

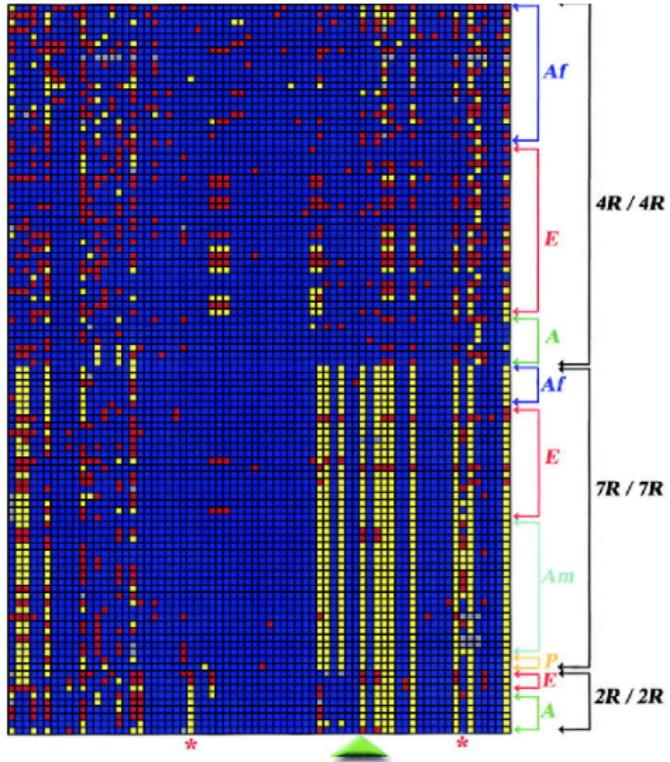


EHH at candidate locus: NCOA1



Huge block of LD around lactase allele in Europe



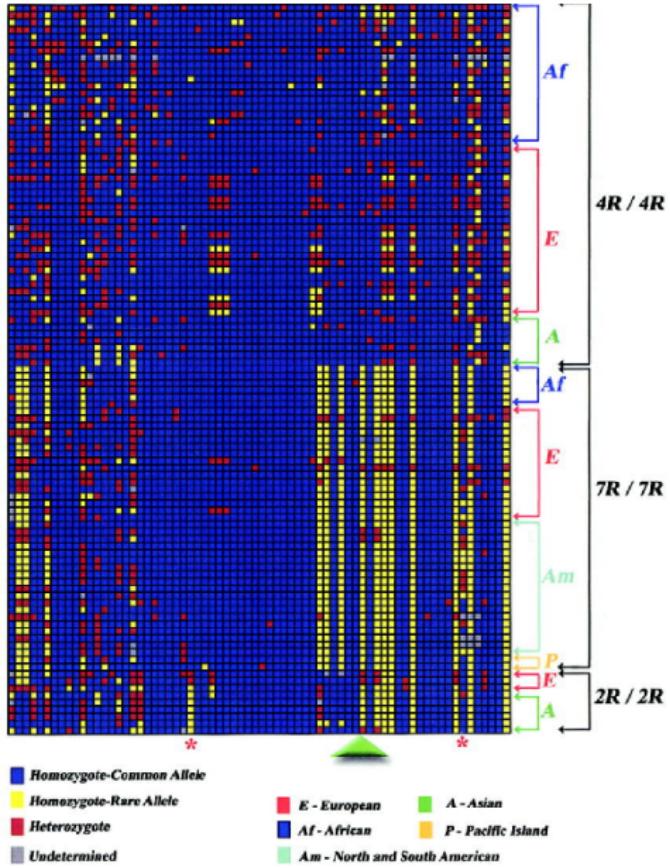


█ Homozygote-Common Allele
█ Homozygote-Rare Allele
█ Heterozygote
█ Undetermined

█ E - European █ A - Asian
█ Af - African █ P - Pacific Island
█ Am - North and South American

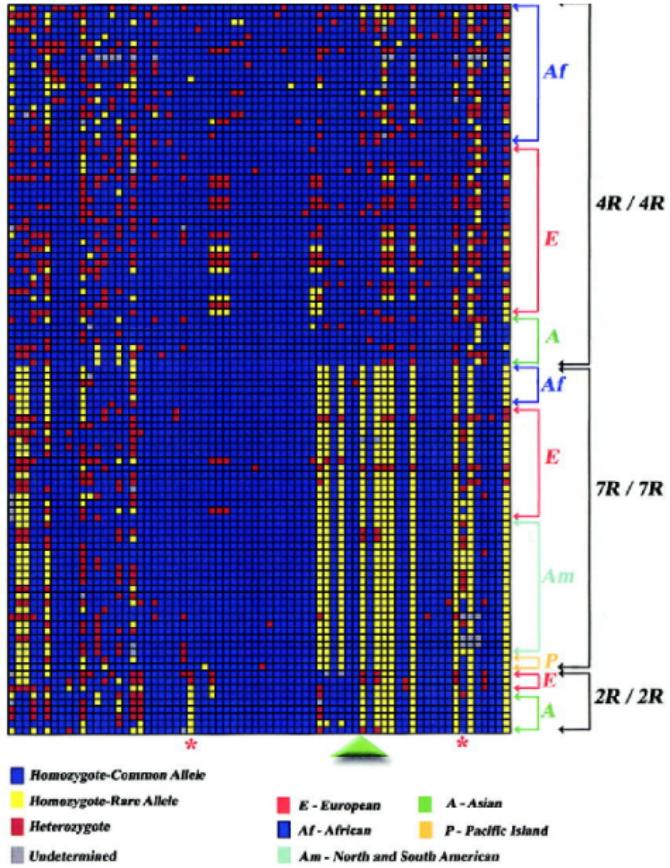
Signature of an ongoing selective sweep at DRD4

- ▶ Sweeping allele is common



Signature of an ongoing selective sweep at DRD4

- ▶ Sweeping allele is
 - ▶ common
 - ▶ has low diversity over large region

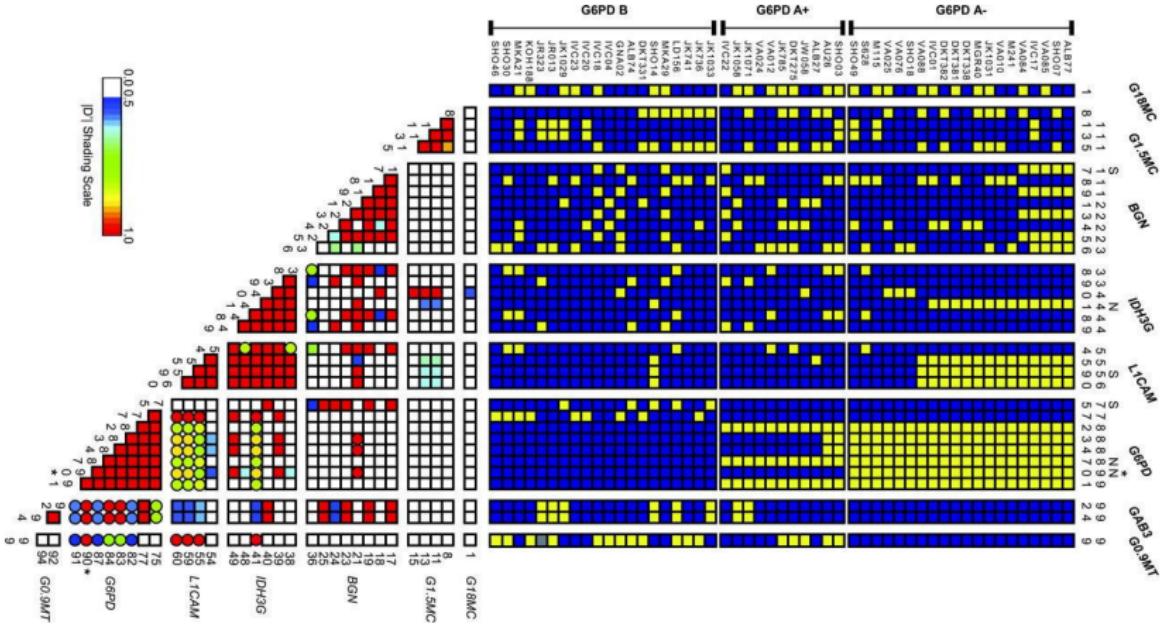


Signature of an ongoing selective sweep at DRD4

- ▶ Sweeping allele is
 - ▶ common
 - ▶ has low diversity over large region
- ▶ High LD over large region

Linkage disequilibrium at G6PD

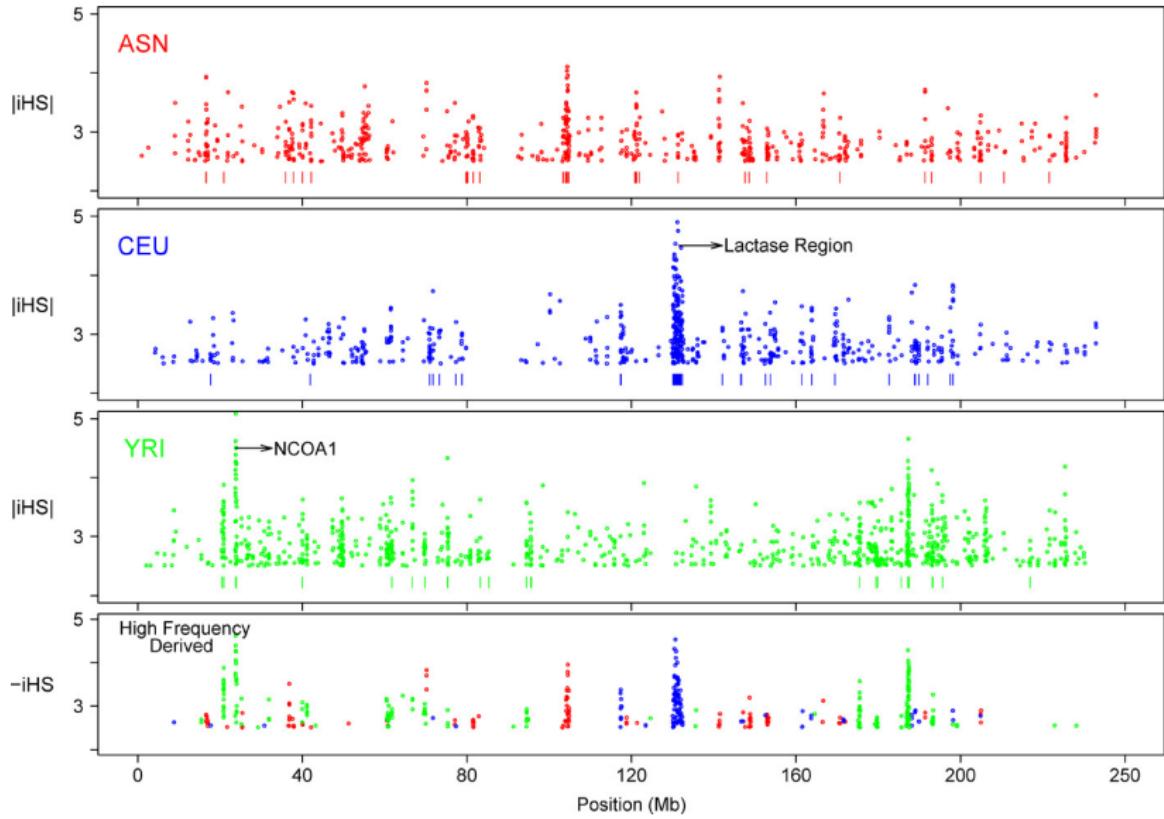
Left: LD plot; right: haplotype plot



Study of Voight et al (2006)

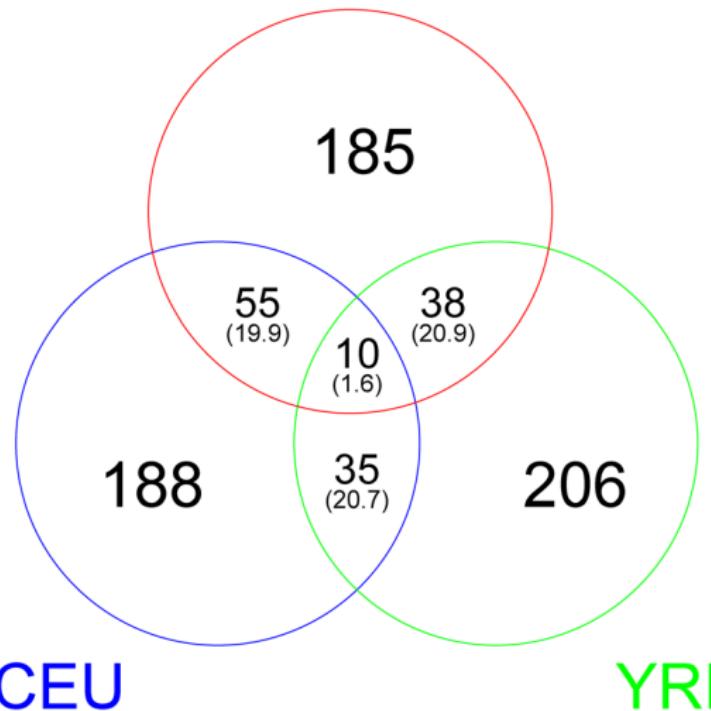
- ▶ 800,000 SNPs in 309 people
- ▶ 431 sweeping loci
- ▶ Most sweeps started w/i past 10,000 years

LD on human chromosome 2 (Voight et al 2006)



26374

ASN



Voight et al (2006):
431 sweeping loci.

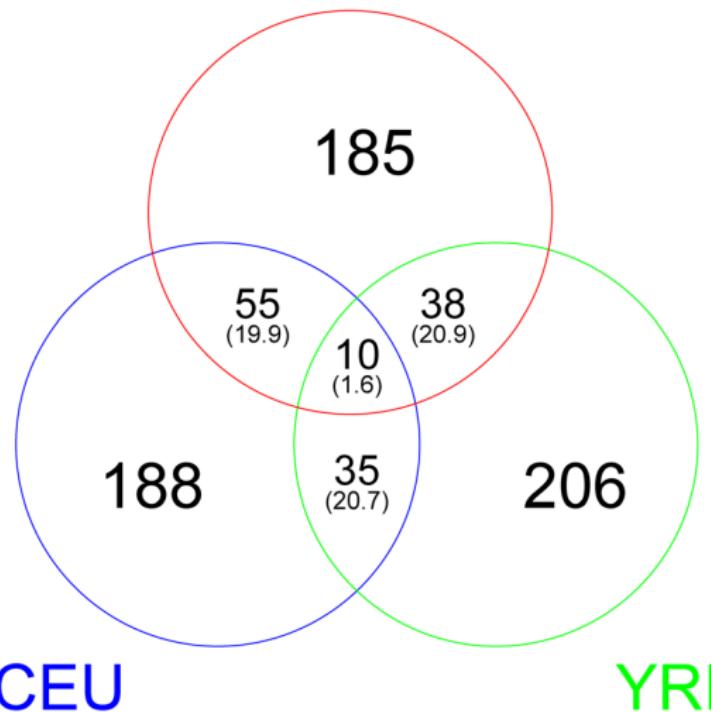
ASN: Asia

YRI: Africa

CEU: Europe.

26374

ASN



Voight et al (2006):
431 sweeping loci.

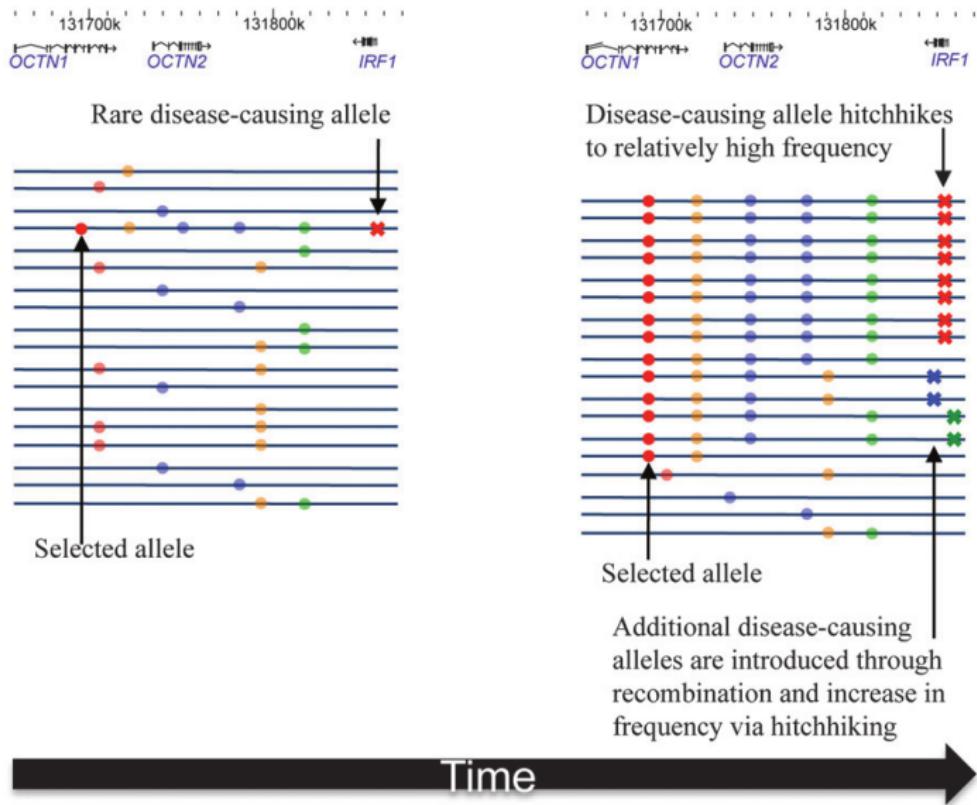
ASN: Asia

YRI: Africa

CEU: Europe.

Most are sweeping w/i
only one continent.

IBD5



(Huff et al. 2012)

IBD5 and inflammatory bowel disease

- ▶ IBD5 is a 250 kb haplotype associated with Crohn's disease.
- ▶ Within this region, variant 503F of the OCTN1 gene covaries with Crohn's.
- ▶ OCTN1 transports the antioxidant ergothioneine.
- ▶ Why should such a gene cause Crohn's disease?

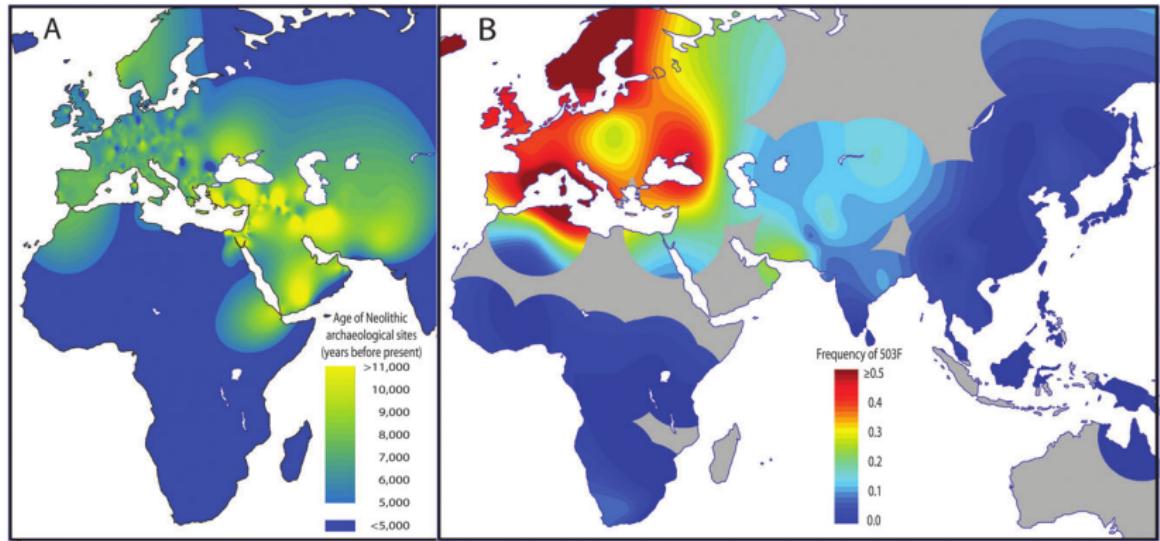
Ergothioneine (ET)

- ▶ An antioxidant, synthesized by fungi, and present in most plants and animals.
- ▶ Low in wheat, barley, lentils, and peas—foods domesticated early in Middle East.
- ▶ Early farmers would have lacked ET.

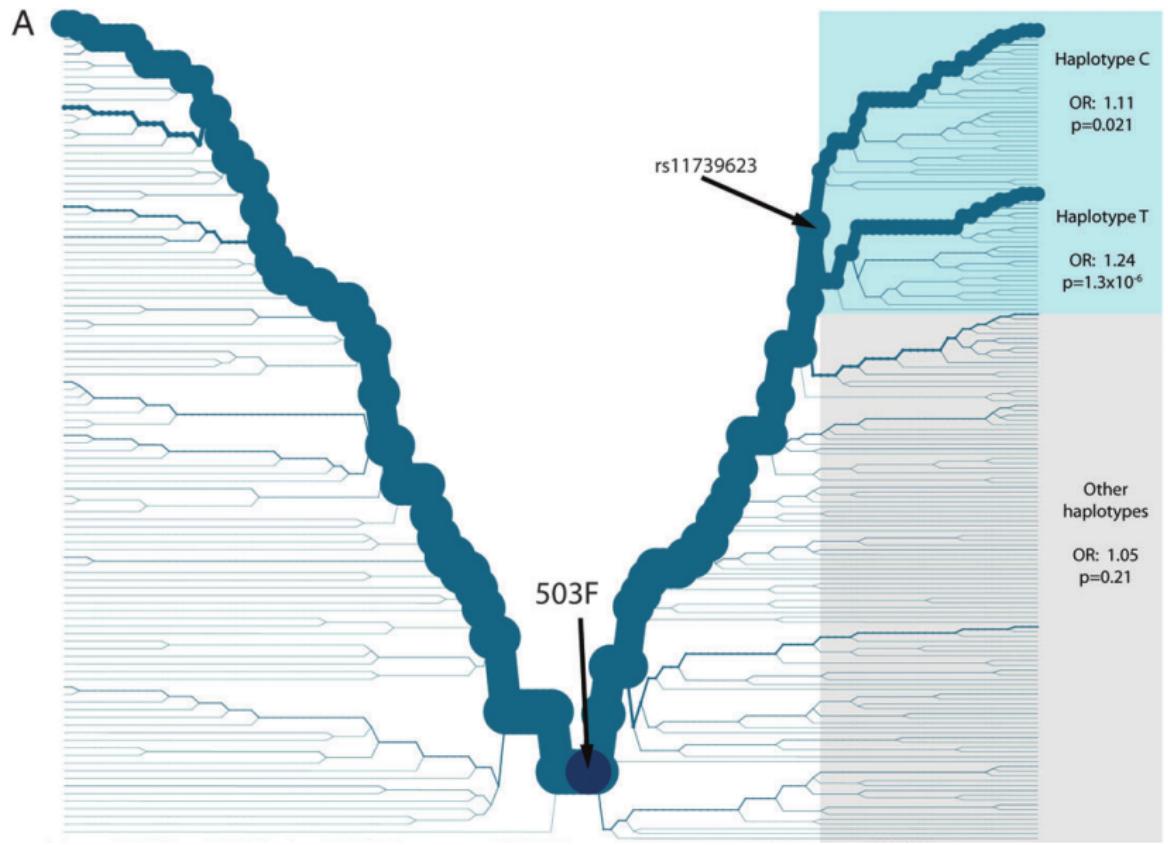
The OCTN1 protein

- ▶ Transports ergothioneine (ET).
- ▶ The 503F mutation increases rate of ET transport by 50%.
- ▶ Highly conserved in evolution, suggesting strong selection.
- ▶ Highly specific to ET.
- ▶ Common in Middle East and Europe; rare elsewhere.
- ▶ LD extent suggests that 503F mutated 7,750–19,025 y ago.

Distribution of early farming (A) and 503F (B)

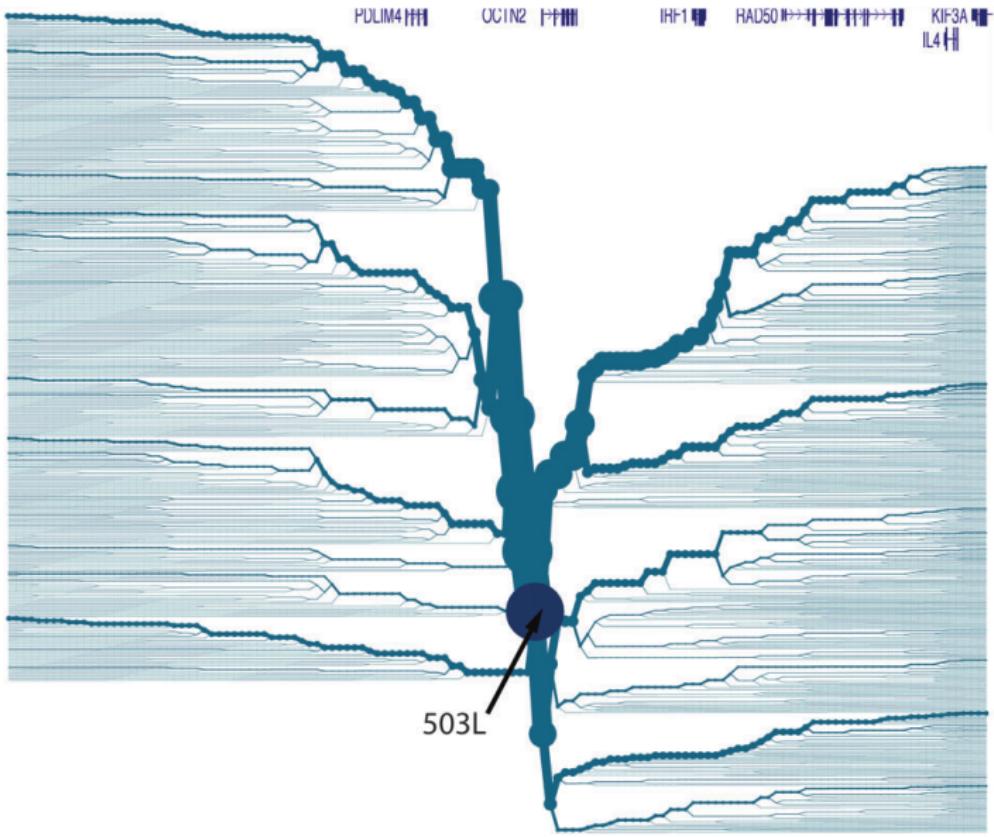


503F sits on a long LD block

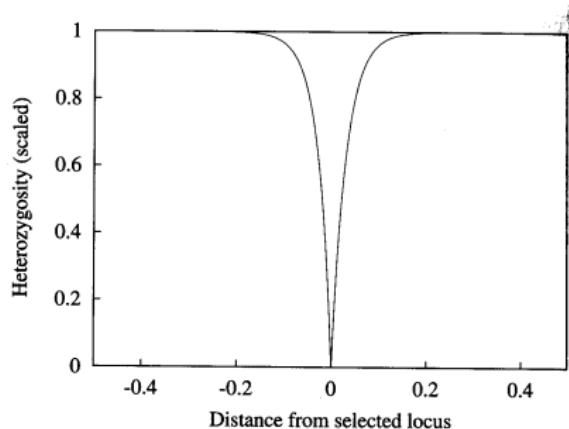


503L sits on a short LD block

B



Selective sweep reduces diversity at linked loci



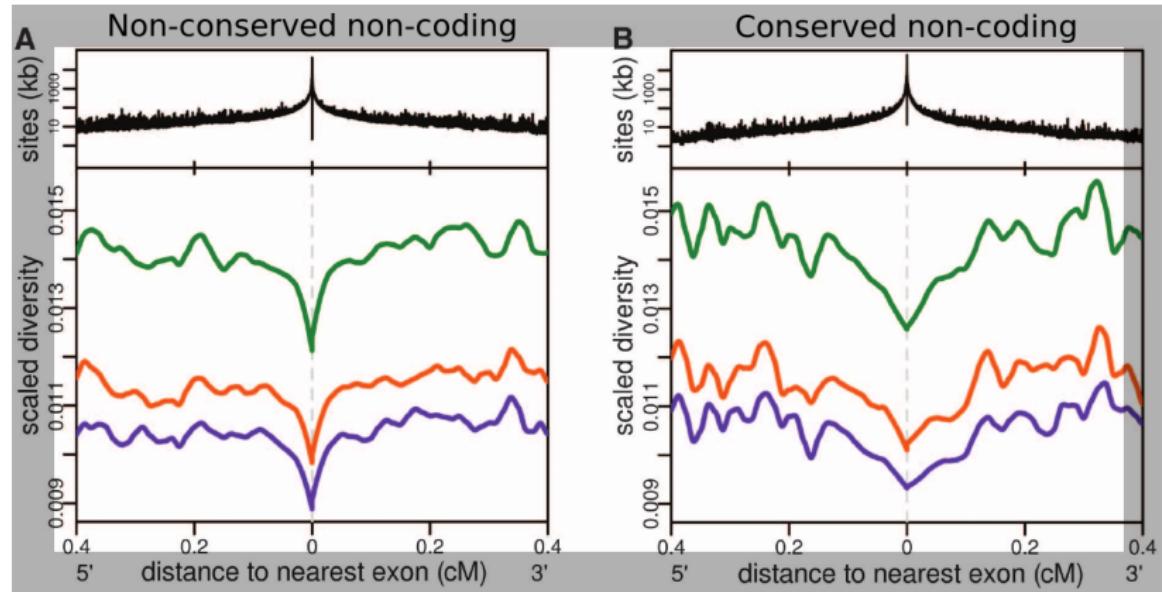
Horizontal axis: distance from selected locus in units of c/s , where c is recombination rate and s is selection coefficient.

Vertical: surviving fraction of neutral heterozygosity at linked loci.

Reduction is appreciable at loci for which $c < s/10$. (Gillespie, 2004, Fig. 4.4)

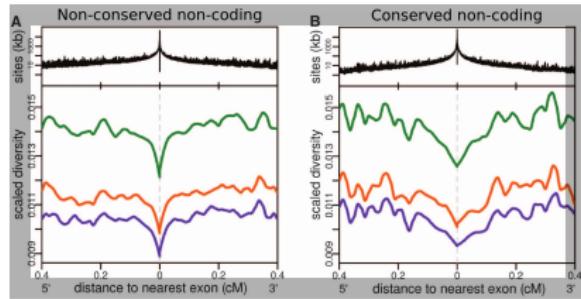
Purifying selection (against deleterious mutations) also reduces linked variation.

Gene diversity vs. distance from exons, scaled by human-rhesus divergence



Hernandez et al (2011)

Gene diversity vs. distance from exons, scaled by human-rhesus divergence



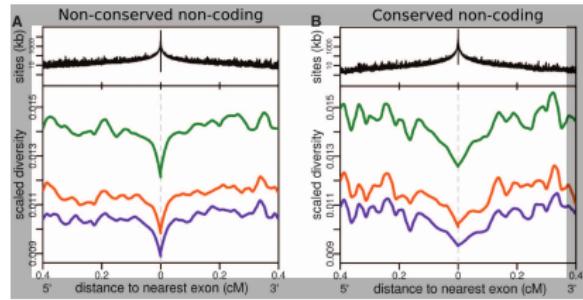
Hernandez et al (2011)

Diversity is low near exons. Why?

Old selective sweeps?

Background selection (against deleterious mutations)?

Gene diversity vs. distance from exons, scaled by human-rhesus divergence



Hernandez et al (2011)

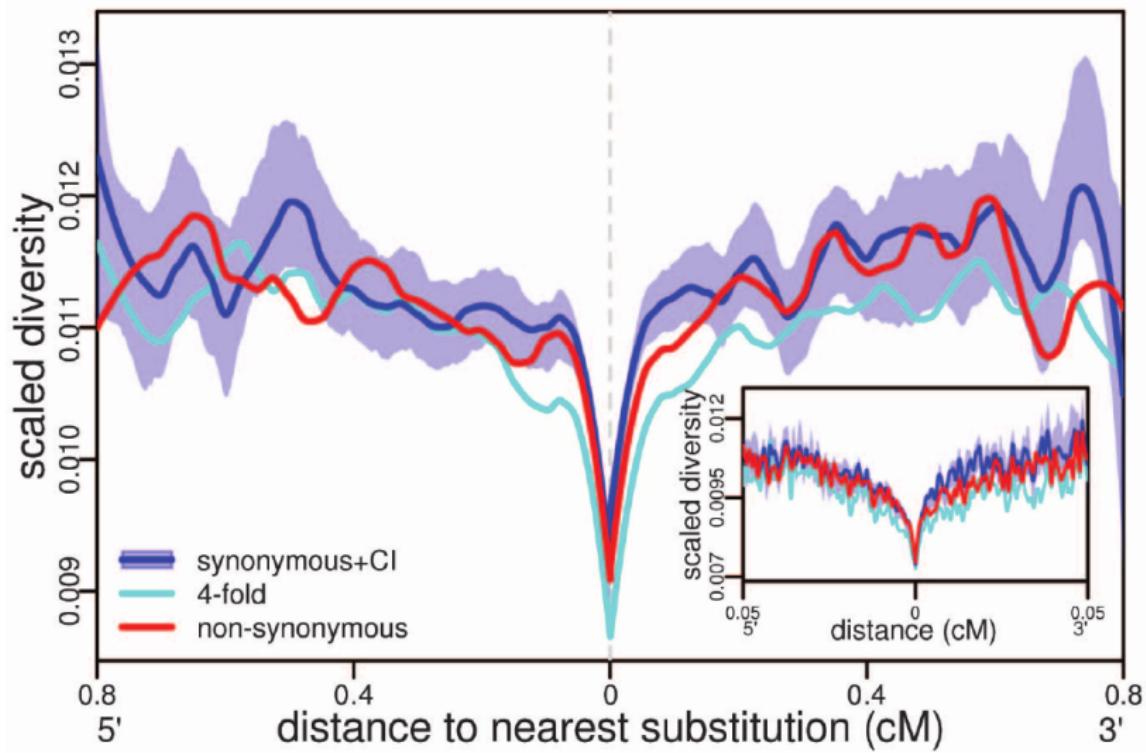
Diversity is low near exons. Why?

Old selective sweeps?

Background selection (against deleterious mutations)?

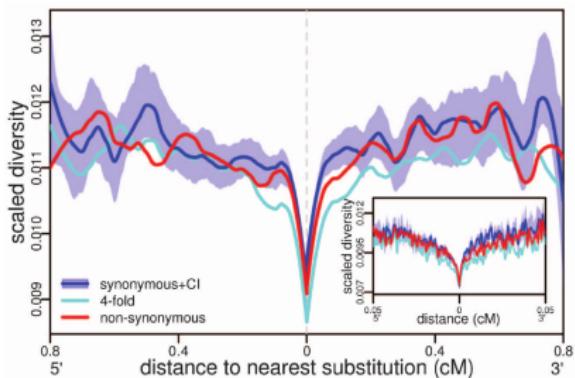
If it was sweeps, effect should be largest near exons with human-specific amino-acid substitutions.

But the patterns around synonymous and non-synonymous substitutions are the same.



Hernandez et al (2011)

But the patterns around synonymous and non-synonymous substitutions are the same.



Hernandez et al (2011)

If selective sweeps caused the dip in diversity, we would expect the dip surrounding non-synonymous substitutions to be wider.

This pattern is seen in *Drosophila simulans*, but not in humans.

Suggests that classical selective sweeps play only a minor role in adaptive evolution among humans.

Alternatives: selection from standing variation, and selection on quantitative variation. We need tools to study these effects.

The Singleton Density Score (SDS, Field et al 2016)

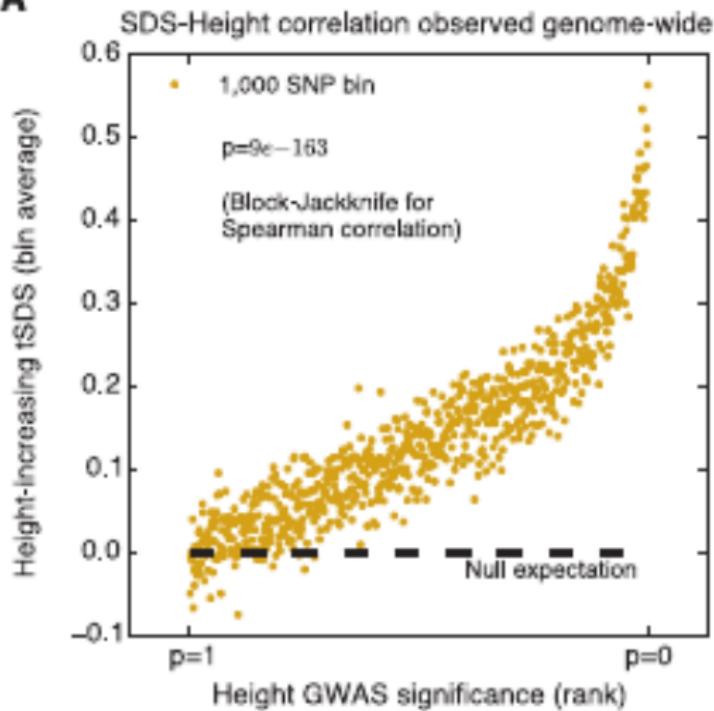
1. The environment changes, some class of haplotypes becomes advantageous, and begins to grow.
2. Growth within this class \Rightarrow recent coalescent events \Rightarrow short terminal branches \Rightarrow few singleton mutations.
3. Class of disfavored haplotypes: long terminal branches \Rightarrow more singleton mutations.

SDS is a normalized estimate of the difference in tip length between the haplotypes linked to two alleles at a nucleotide site.

Large samples provide sensitivity to recent selection. A sample of 3000 provides sensitivity to selection during ~ 75 generations, ~ 2000 y.

Selection on height alleles in UK

A



X axis: strength of GWAS evidence that locus affects stature

Y axis: strength of selection for "tall" allele.

Many of these GWAS associations are not statistically significant. Yet in aggregate, they demonstrate genome-wide selection increasing stature.

Summary

- ▶ Signature of a recent classic selective sweep: common derived allele surrounded by extensive LD.
- ▶ Many such sweeps have been discovered.
- ▶ Tend to be population-specific.
- ▶ In humans, classic sweeps account only for a minority of adaptive evolution.
- ▶ A new method (SDS) makes it possible to study recent selection on standing variation and on polygenic characters.