

Detecting Adaptive Evolution

Alan R. Rogers

March 26, 2020

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.

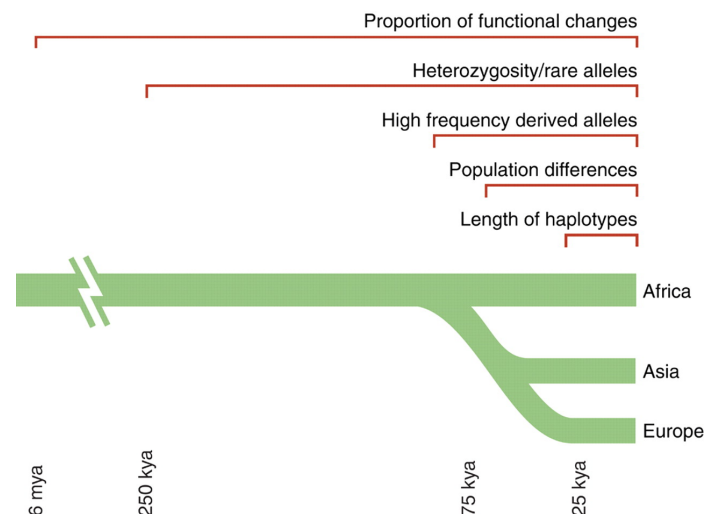
—Helena Cronin 2000

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			
Human		STOP	H	R	R	C	R	P	R	Y	R	P	R	C	C	R
		AATCACAGAAGATGTAGCGCCAGACATGGACCCTCCGCGTCGTGG														
Chimp		AATCACAGAAGATGTCAGAGTAAGACCTGGACGCCGCGTCGTGG														
		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)

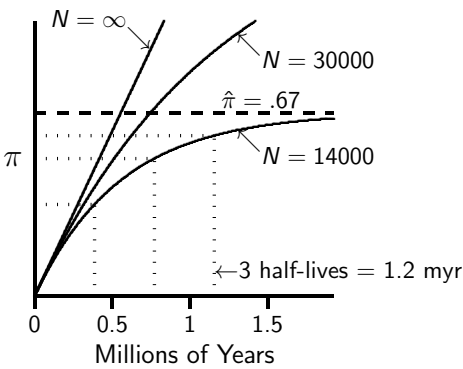
Melanocortin 1 receptor (MC1R) locus

- ▶ 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

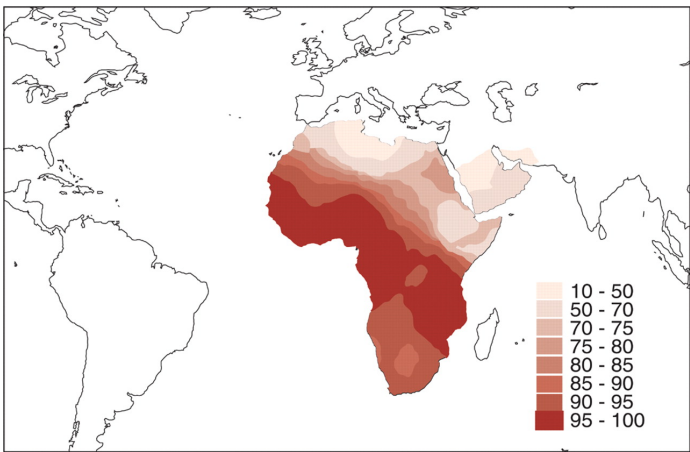
- ▶ Before loss of body hair: weak selective constraint on MC1R
- ▶ After loss: constraint strong w/i Africa; weak w/i Europe
- ▶ Selective sweep at loss of body hair
- ▶ Neutral diversity within Africa accumulated since the sweep.

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

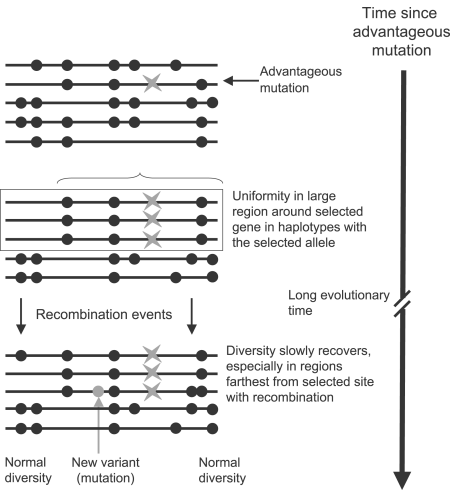


Trajectories of π following a selective sweep. Dotted lines show the half-lives when $N = 14,000$; the dashed line shows $\hat{\pi}$, the observed value of π in Africa.

Duffy map



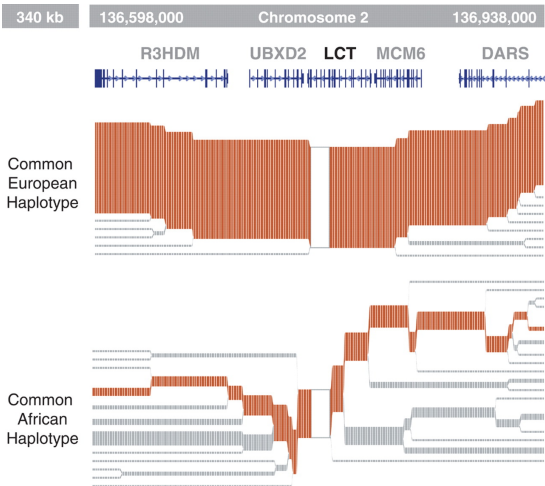
How a “hard” selective sweep generates LD



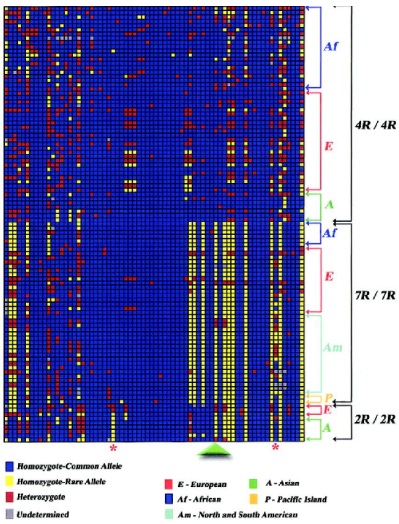
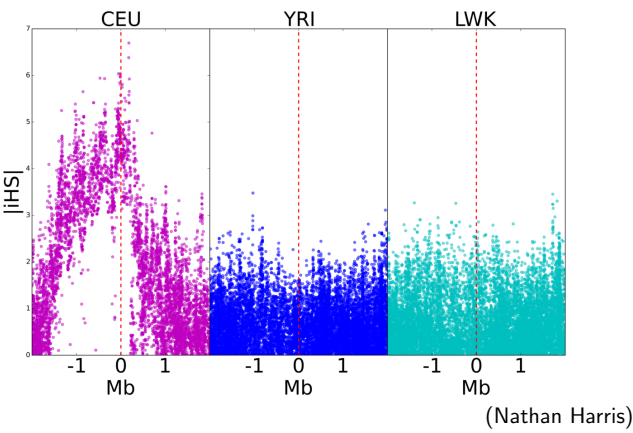
DNA sequences from region of human lactase gene

```
cgcttcaggcattcctatctaacaagaccaacgtaAgggtacaatgctaaccagacgtttcaactct
20 .....
21 .....
22 .....
23 .....
24 .....
25 .....
26 .....
27 .....t.....
28 .....t.....
29 .....c.....
37 .....G.a.gt.....gac.c.tgtct.
38 ...cogga...gat..at..gg..c.....tc.gGaaa.g..ccttt...tg.....c..t.t...
39 ...cogga...gat..at..gg..c.....tc.gGaaa.g..ccttt...tg.....c..t.t...
40 .tcc...agtag.t.cat..g.....t.ttcgG..a.gt.....t.....gac.c.tgtct.
41 .tcc...agtag.t.cat..g.....t.gttccG..a.gt.....t.....gac.c.tgtct.
42 .tcc...agtag.t.cat..g.....t.gttccG..a.gt.....t.....gac.c.tgtct.
43 .tcc...agtag.t.cat..g.....t.gtc.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 ...cogga...gat..at..gg..c.....tc.gGaaa.g..ccttt...tg.....cg.gt.t..c
47 .tcc...agtag.t.cat..g.....t.gttccG..a.gt.....t.....gac.c.tgtct.
48 .tcc...agtag.t.cat..g.....t.gttccG..a.gt.....t.....gac.c.tgtct.
49 .tcc...agtag.t.cat..g.....t.gttccG..a.gt.....t.....gac.c.tgtct.
50 tatcogga...g.tc.atcgg.tc.g.tg.tc.gG..a.g.....tg...ggt...cg.gt.t..c
51 ta.cogga...g.t.atcgg.tc.g.tg.tc.gG..a.g.....tg...ggt...cg.gt.t..c
52 ta.cogga...g.t.atc.g.tc.g.tg.tc.gG..a.g.....tg...ggt...cg.gt.t..c
53 ta.cogga...g.t.atcgg.tc.g.tg.tc.gG..a.g.....tg...ggt...cg.gt.t..c
```

Huge block of LD around lactase allele in Europe



Huge block of LD around lactase allele in Europe

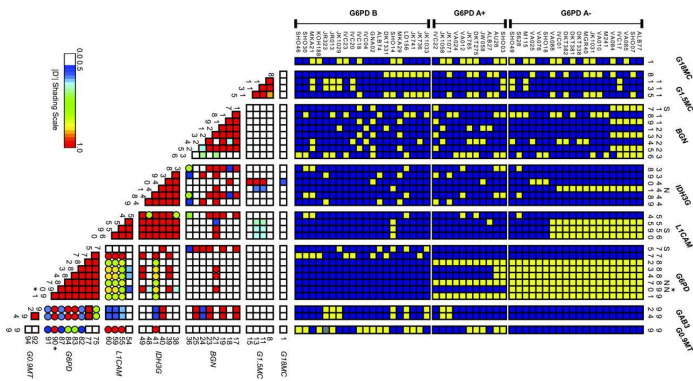


Signature of an ongoing selective sweep at DRD4

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

Linkage disequilibrium at G6PD

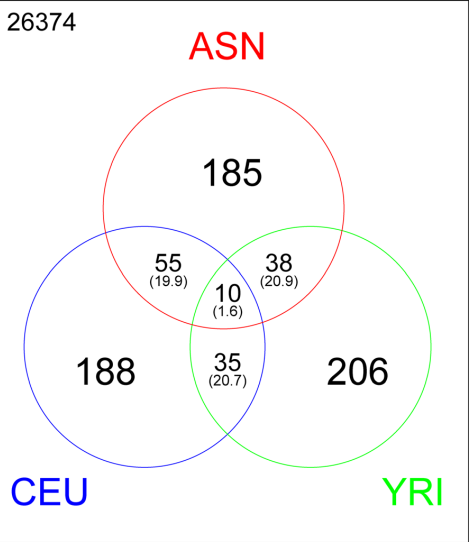
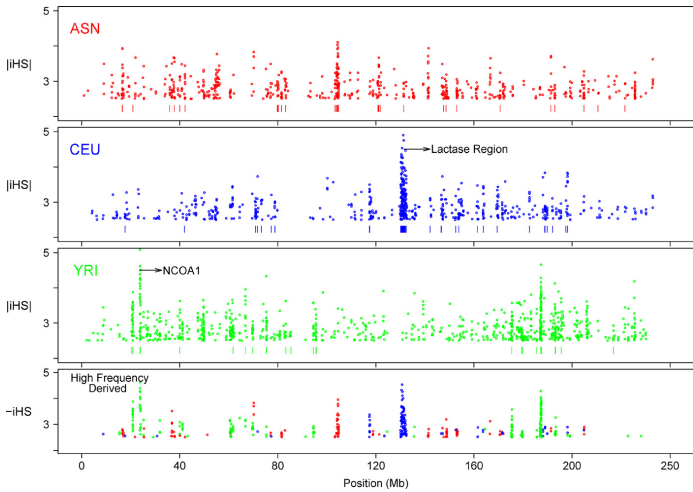
Left: LD plot; right: haplotype plot



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)

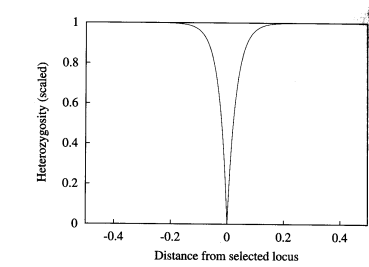


Voight et al (2006):
431 sweeping loci.

ASN: Asia
YRI: Africa
CEU: Europe.

Most are sweeping w/i
only one continent.

Selective sweep reduces diversity at linked loci



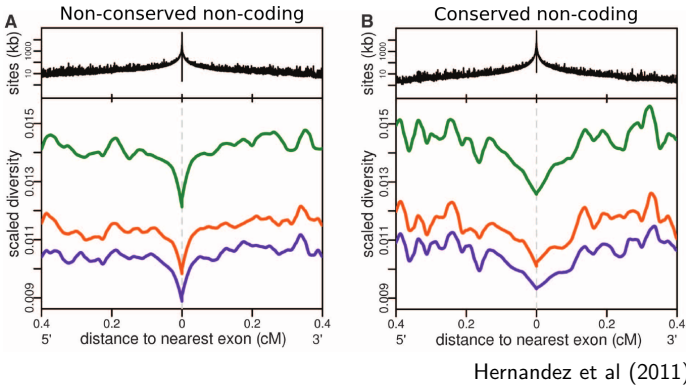
Vertical: surviving fraction of neutral heterozygosity at linked loci.

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

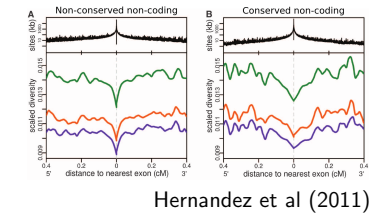
Purifying selection (against deleterious mutations) also reduces linked variation.

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

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



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



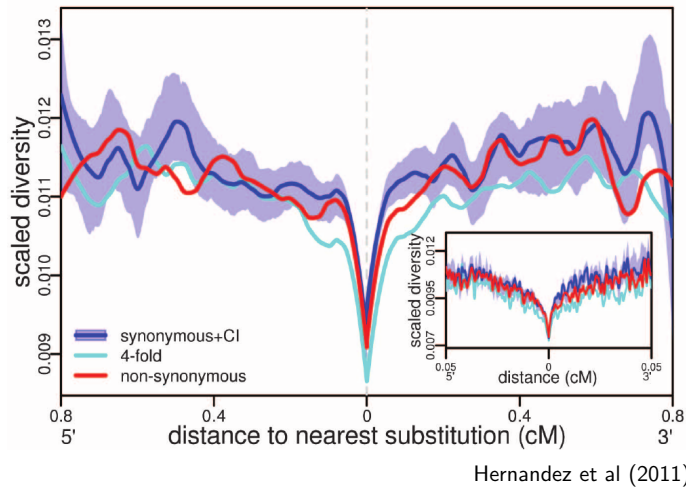
Diversity is low near exons. Why?

Old selective sweeps?

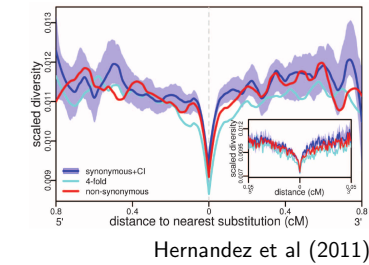
Background selection (against deleterious mutations)?

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



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



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.

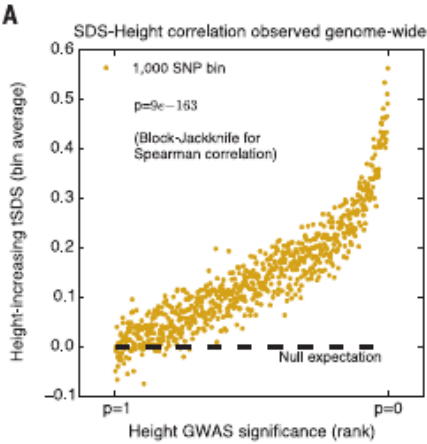
The Singleton Density Score (SDS, Field et al 2016)

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



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.