Detecting Adaptive Evolution

Alan R. Rogers

September 14, 2021

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.

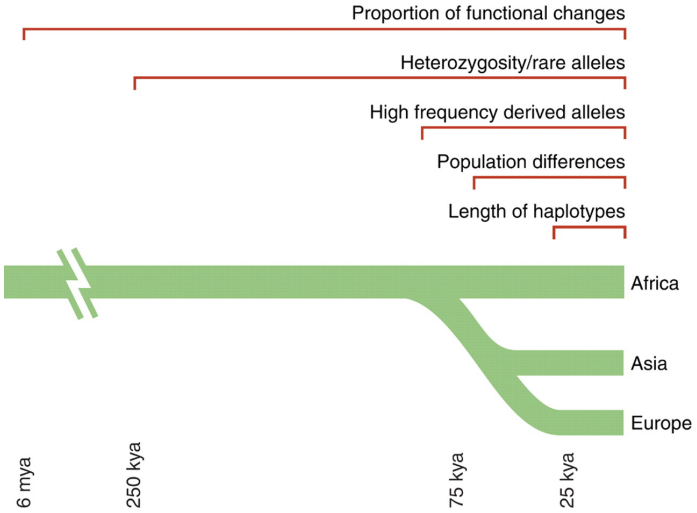
—Helena Cronin 2000

Is this really true? How could we know?

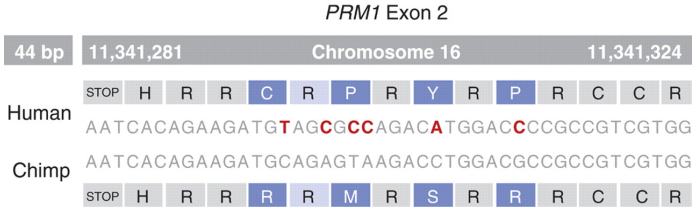
Signatures of selection

- High proportion of functional changes
- Reduction of gene diversity
- Population differences
- Excess of rare derived alleles.
- Common allele on long LD block

Time scale for signatures of selection



Protamine 1 (PRM1) gene



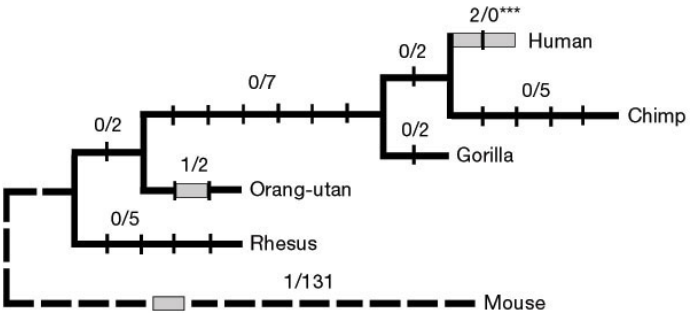
- compacts sperm DNA
- 13/14 human-chimp diffs are non-synonymous (6 shown here)

Proportion of functional changes at FOXP2 locus

- Mutations at FOXP2 cause problems with language.
- Few amino acid changes within mammals ⇒ strong selective constraint.
- Two mutations on human lineage, neither synonymous ⇒ selection

(Enard et al 2002)

Evolution of FOXP2

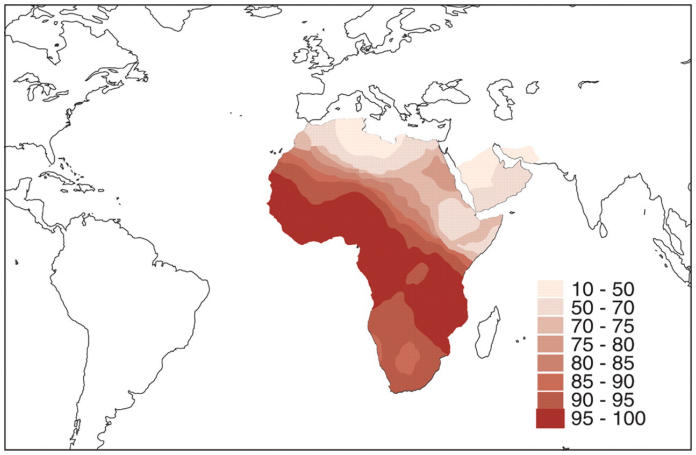


Vertical bars **separate** nucleotide substitutions; gray indicates amino-acid changes. (Enard et al 2002)

Signatures of selection

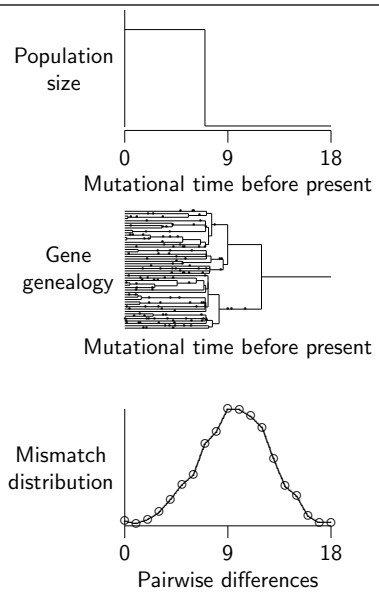
- ◦ High proportion of functional changes
- ⊗ Reduction of gene diversity
- ▶ Population differences
- ▶ Excess of rare derived alleles.
- ▶ Common allele on long LD block

Duffy map: population differences ⇒ selection



Signatures of selection

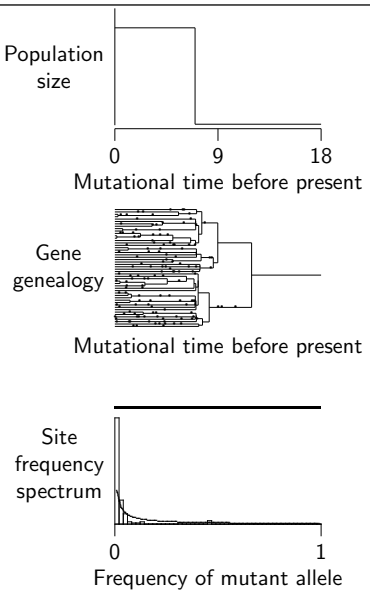
- ◦ High proportion of functional changes
- ⊗ Reduction of gene diversity
- ◦ Population differences
- ▶ Excess of rare derived alleles.
- ▶ Common allele on long LD block



Mismatch distribution of an expanded population

Selection has same effect.

But we can't use the mismatch distribution with nuclear DNA.



Spectrum of same population

*i*th bar: number of sites at which derived allele is present in *i* copies.

Selection: an excess of rare derived alleles.

Signatures of selection

- ○ High proportion of functional changes
- ⊗ Reduction of gene diversity
- ○ Population differences
- ○ Excess of rare derived alleles.
- ▶ Common allele on long LD block

Look at some DNA sequence data

```
Ref   T G C A T G T A A T G C T
------------------------------
A     . . . . A . . . . G A A .
B     . . . . A . . . . G A . .
C     . . . . . T . G . . . . C
D     C . . . . . . G G . . . C
E     C C . . . . . G G . . . C
F     C . . . . . C G . . . . C
G     C . T . . . C G . . . . C
H     C . T G . . C G . . . . C
------------------------------
    (Garrigan et al 2004)
```

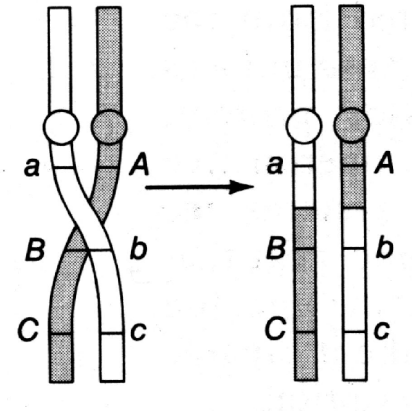
- ▶ A–H: human DNA sequences
- ▶ Dots mean identical to ref.
- ▶ Note similarity between A and B.
- ▶ Called "linkage disequilibrium" (LD)
- ▶ Natural selection generates LD.

What happens when a favorable mutation strikes some lucky chromosome?

- ▶ Better-than-average chance of spreading through population.
- ▶ What is it that spreads?
  - ▶ The entire chromosome?
  - ▶ Just the mutated nucleotide site?

Neither: the answer is in between.

Cross-overs shuffle DNA



- ▶ occur during reproduction.
- ▶ shuffle parental chromosomes.
- ▶ sites far apart shuffled more.
- ▶ result: "recombinant" chromosomes

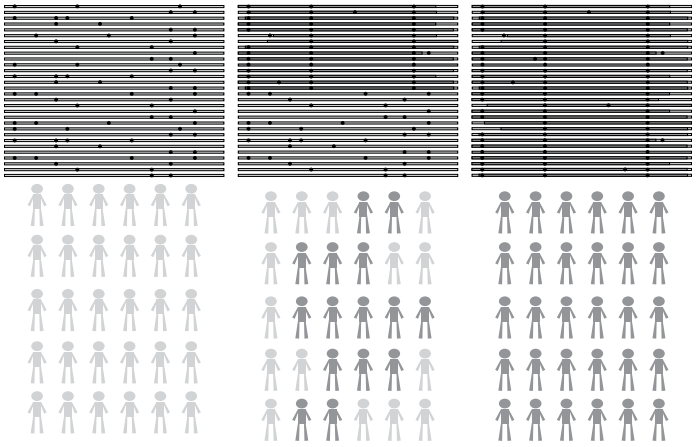
The chromosome you pass on to your child

- ▶ Not like either of those you inherited from Mom or Dad.
- ▶ A mosaic, made of pieces of Mom's and Dad's.
- ▶ Technical terms: *recombination*, *recombinant chromosomes*

What happens to lucky chromosome?

- ▶ Favorable mutation surrounded by region from original chromosome.
- ▶ Recombination gradually reduces size of that region.
- ▶ The older the mutation, the smaller the region.
- ▶ Technical term: *linkage disequilibrium* (LD)

How a selective sweep generates LD



Common neutral mutations

- May accidentally drift to high frequency, but this takes a long time.
- Plenty of time for recombination.
- Sit on *short* stretches of original chromosome.

Common favorable mutations

- Increase rapidly in frequency
- Little time for recombination.
- Sit on *long* stretches of original chromosome.

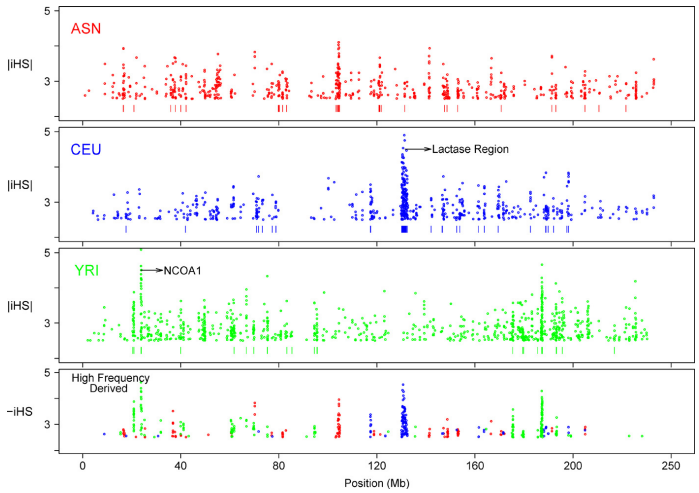
DNA sequences from region of human lactase gene

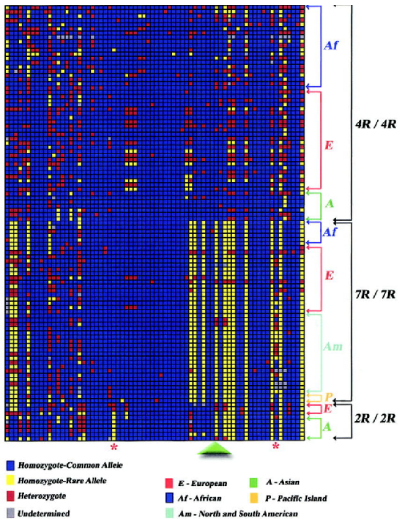
```
   cgcttcaggcattcctatctaaacagaccaacgtaAgggtacaatgcctaacccagacgtttcaactct
20 ....................................................................
21 ....................................................................
22 ....................................................................
23 ....................................................................
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.gttc.gG..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.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..c
52 ta.ccgga....g.t..atc.g.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..c
53 ta.ccgga....g.t..atcgg.tc.g.tg.tc.gG..a.g.g....tg....ggt...cg.gt.t..c
```


Evidence for natural selection at lactase gene

- Lucky chromosomes nearly identical.
- Unlucky chromosomes vary.
- Region of LD covers nearly a million nucleotides in European population.
- Absent in Africa and Asia.

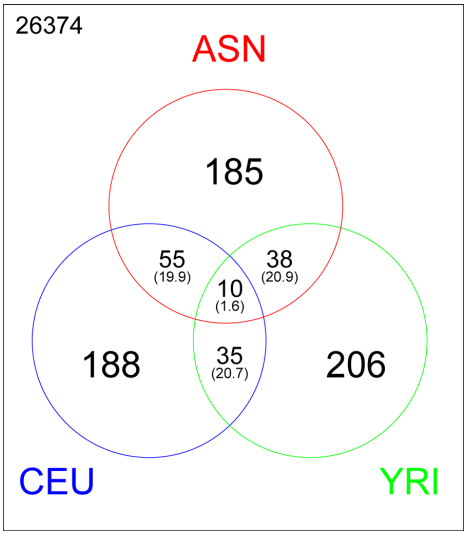
LD on human chromosome 2 (Voigt et al 2006)





Signature of an ongoing selective sweep at DRD4

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



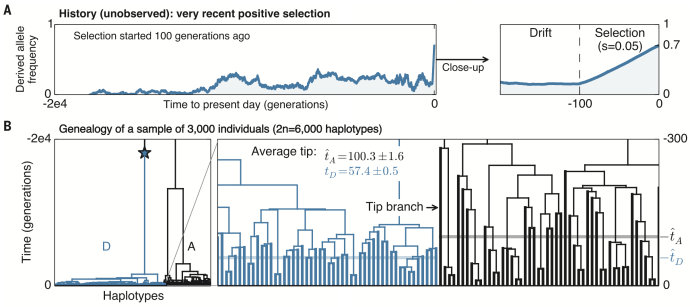
Voight et al (2006): 431 sweeping loci.

ASN: Asia
YRI: Africa
CEU: Europe.

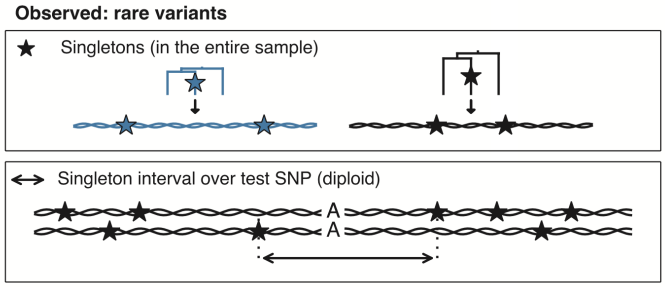
Most are sweeping w/i only one continent.

Singleton Density Score (SDS)

At a selected locus, haplotypes carrying the favored allele have shorter tip branches.



Shorter tips imply fewer singletons



SDS, the singleton density score

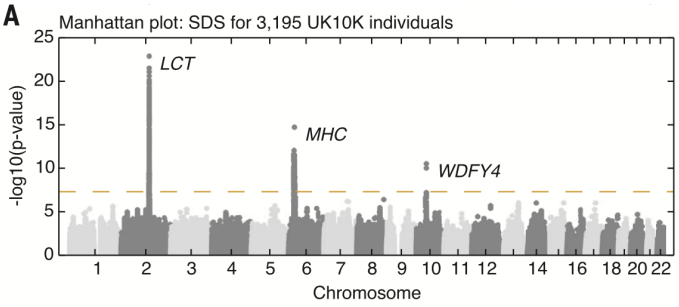
At each nucleotide site the varies, separate the sample of haplotypes into two groups: those carrying one allele and those carrying the other.

Within a region around the focal nucleotide site, count the number of singleton sites within each group.

SDS is a function of the ratio of these numbers.

Sensitive to recent selection: last 2000 years.

Genome-wide SDS scan



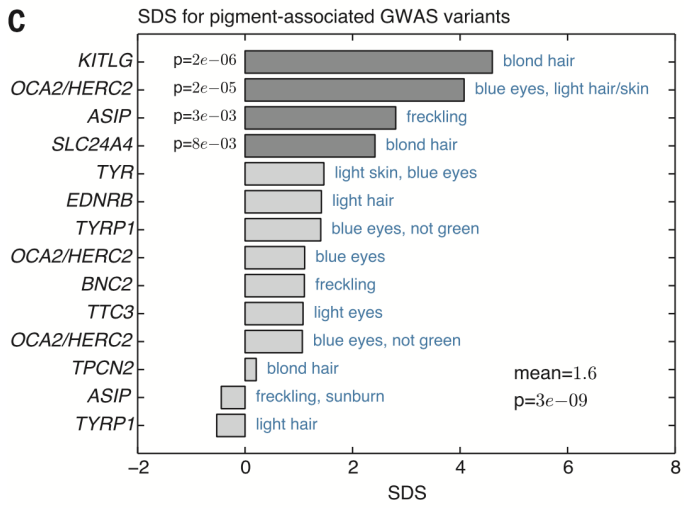
LCT, lactase persistance; MHC, adaptive immune system; WDFY4, defence against viruses and tumors.

SDS at GWAS loci

Genome-wide association studies (GWAS) find loci associated with variation in quantitative characters, such as height.

Field et al (2016) looked at SDS scores of GWAS-related loci.

Several GWAS loci have inflated SDS scores



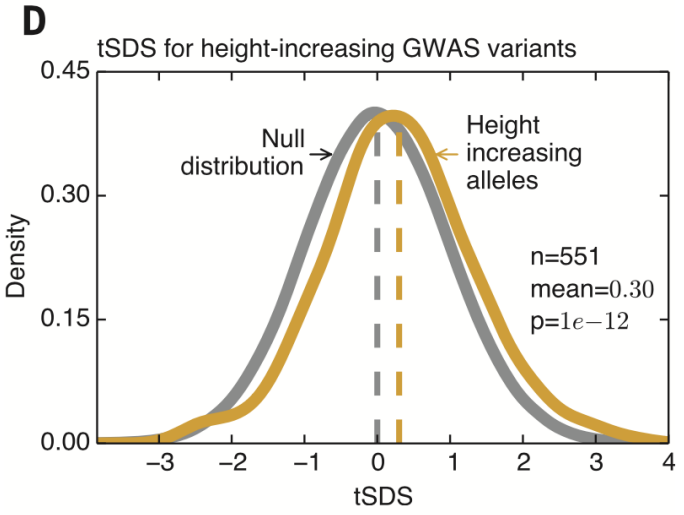
tSDS: SDS at quantitative traits

Calculate SDS at each site.

Adjust sign at each site so that SDS is positive if selection tends to increase trait value.

Sum these adjusted SDS values to obtain tSDS.

tSDS: selection for increased stature in Europe



Summary

- Hundreds human genes are under positive selection.
- K_a/K_s detects ancient selection
- Population differences: intermediated time depth
- iHS: sensitive to selection in past 20,000 years
- SDS: selsitive to selection in past 2000 years