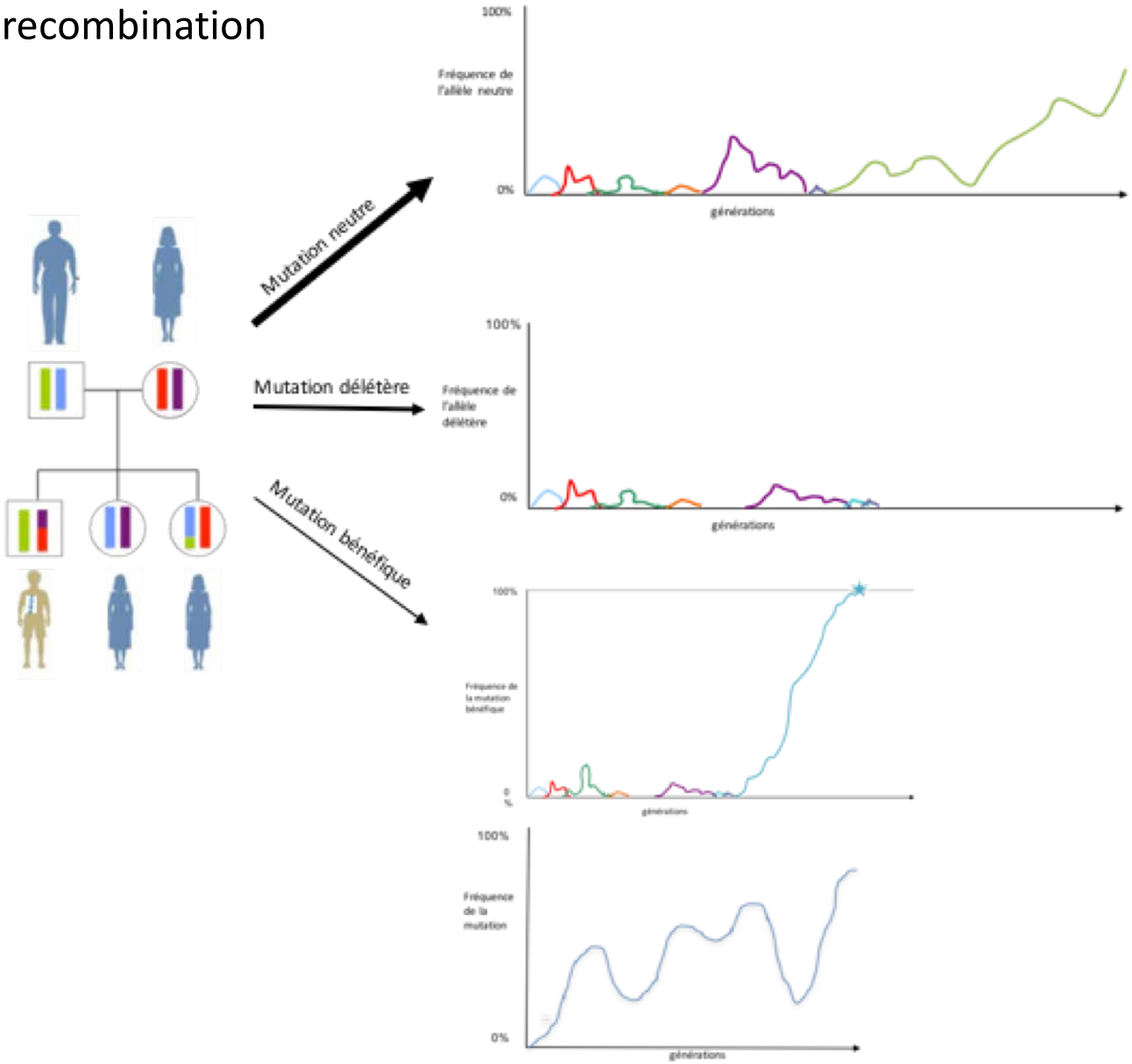
À la recherche de la base moléculaire des adaptations

Molly Przeworski

Cours #6

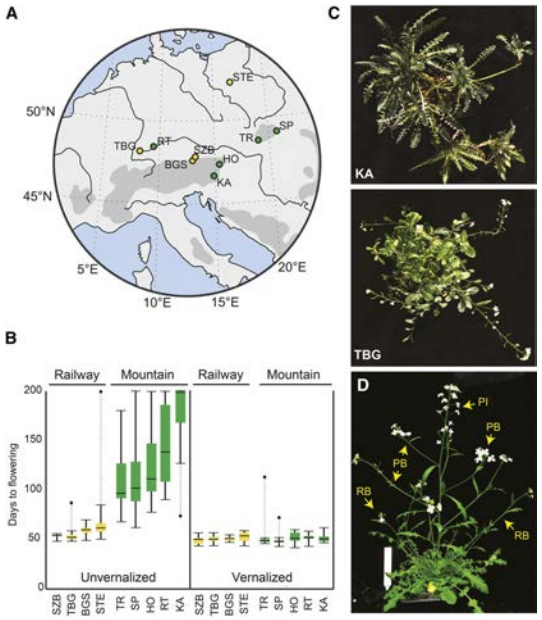
genetic drift & natural selection

mutation & recombination

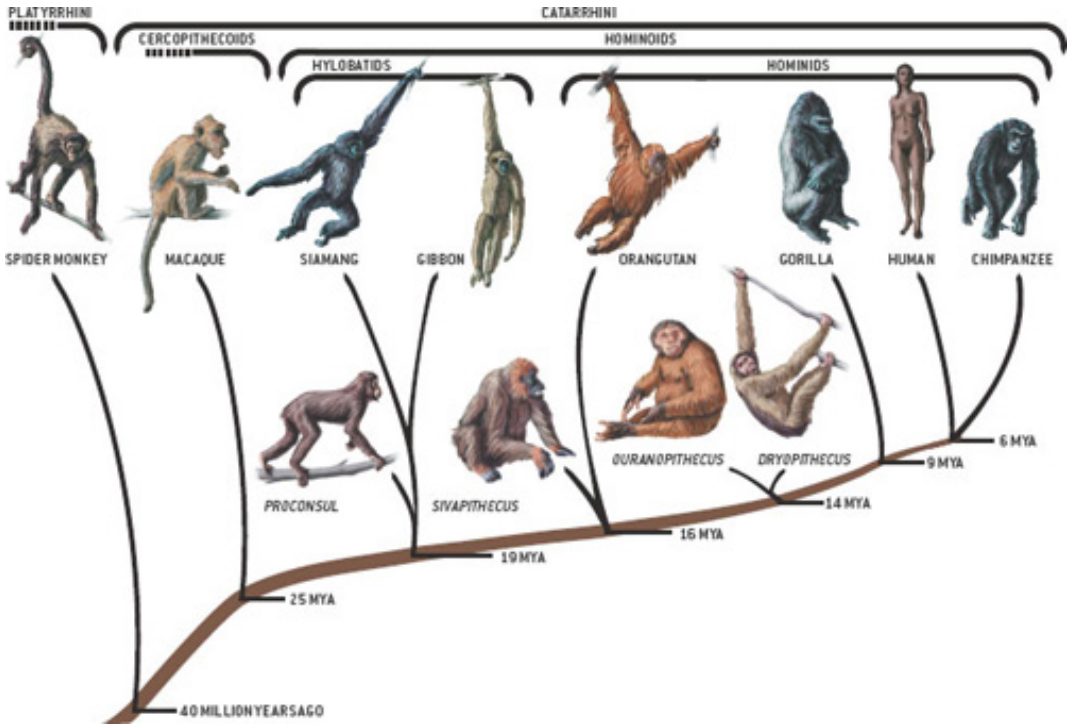


https://www.biotecnika.org/2018/04/in-a-moonshot-for-biology-earth-biogenome-project-to-sequence-all-of-the-planets-eukaryotic-biodiversity/

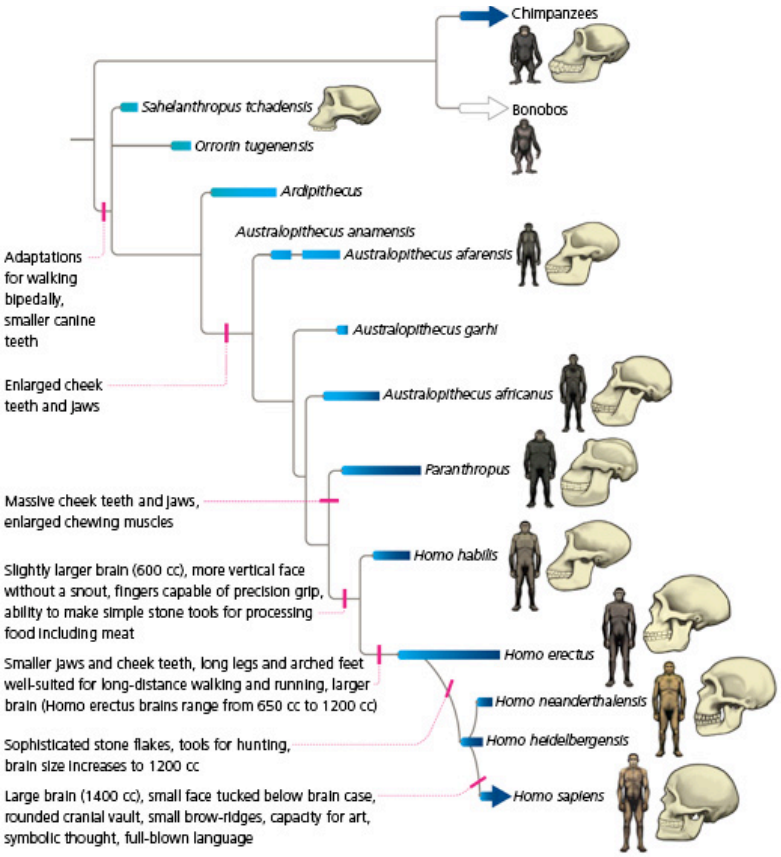
https://www.smithsonianmag.com/science-nature/becoming-human-the-evolution-of-walking-upright-13837658/



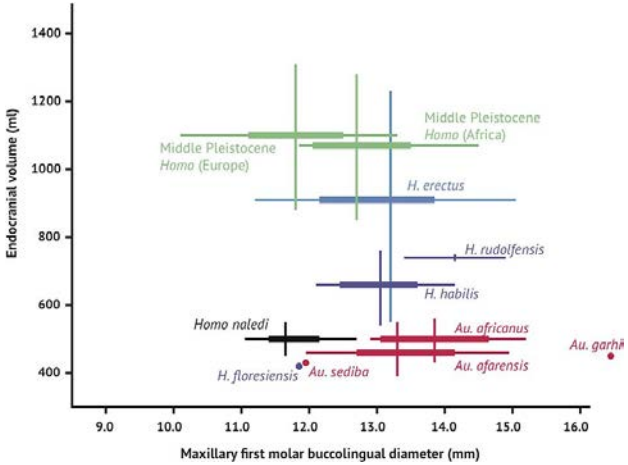
http://www.plantphysiol.org/content/171/1/437



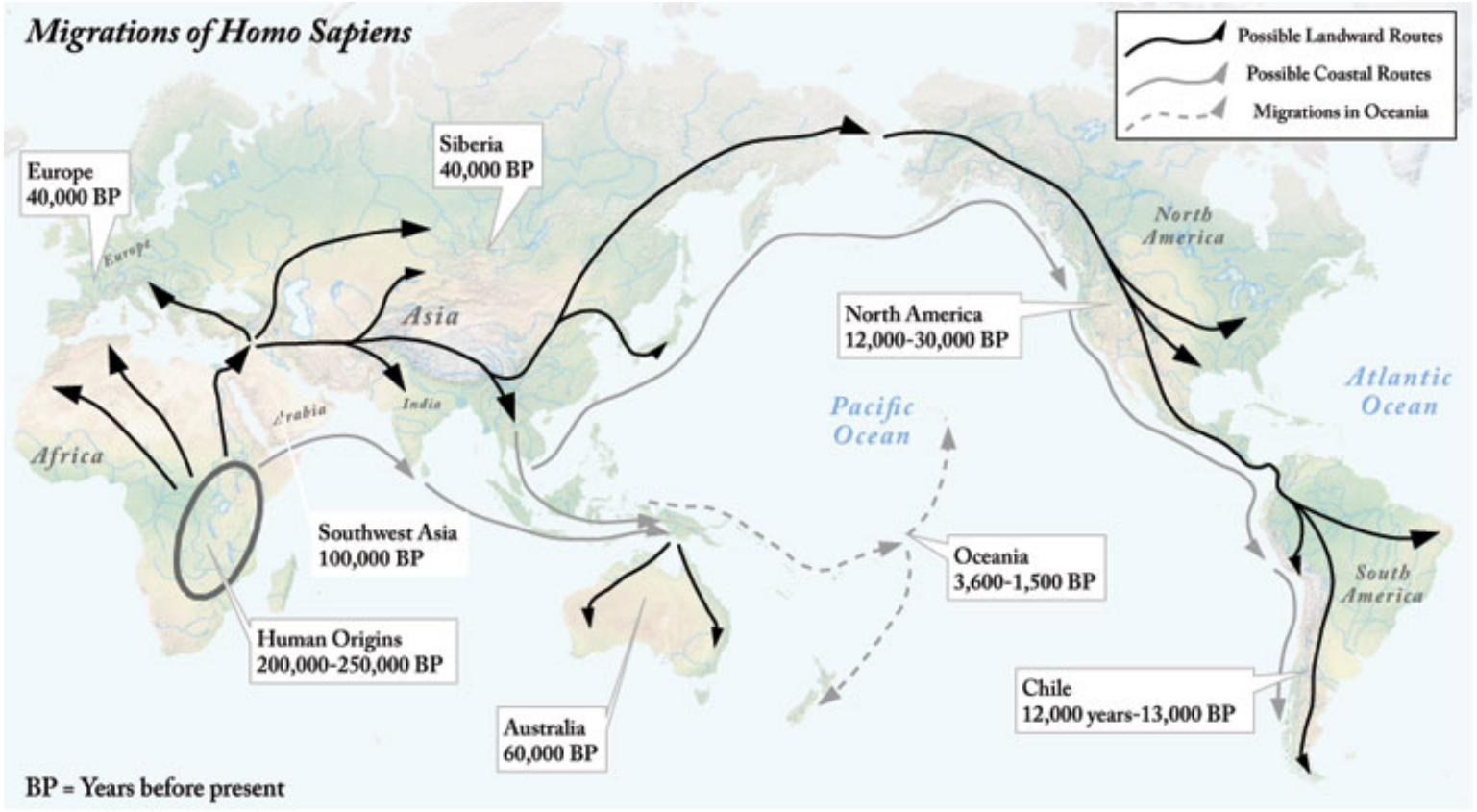
Borrowed from Begun & Gurche 2006 Scientific American



http://evolution.berkeley.edu/evolibrary/article/evograms_07

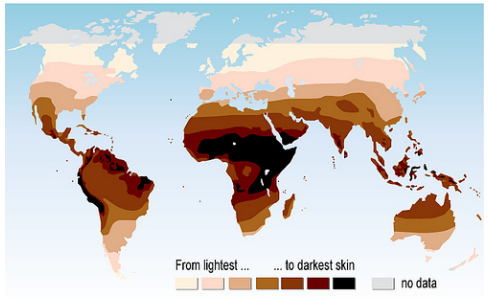


https://en.wikipedia.org/wiki/Human_evolution



http://worldhistoryforusall.sdsu.edu/eras/era2.php

Skin colour map for indigenous people
Predicted from multiple environmental factors

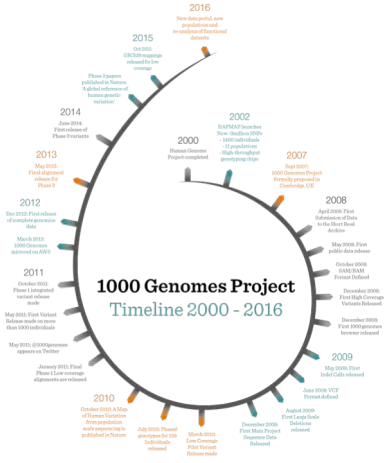


Source: Chaplin G.[®] , *Geographic Distribution of Environmental Factors Influencing Human Skin Coloration*, American Journal of Physical Anthropology 125:292–302, 2004; map updated in 2007.

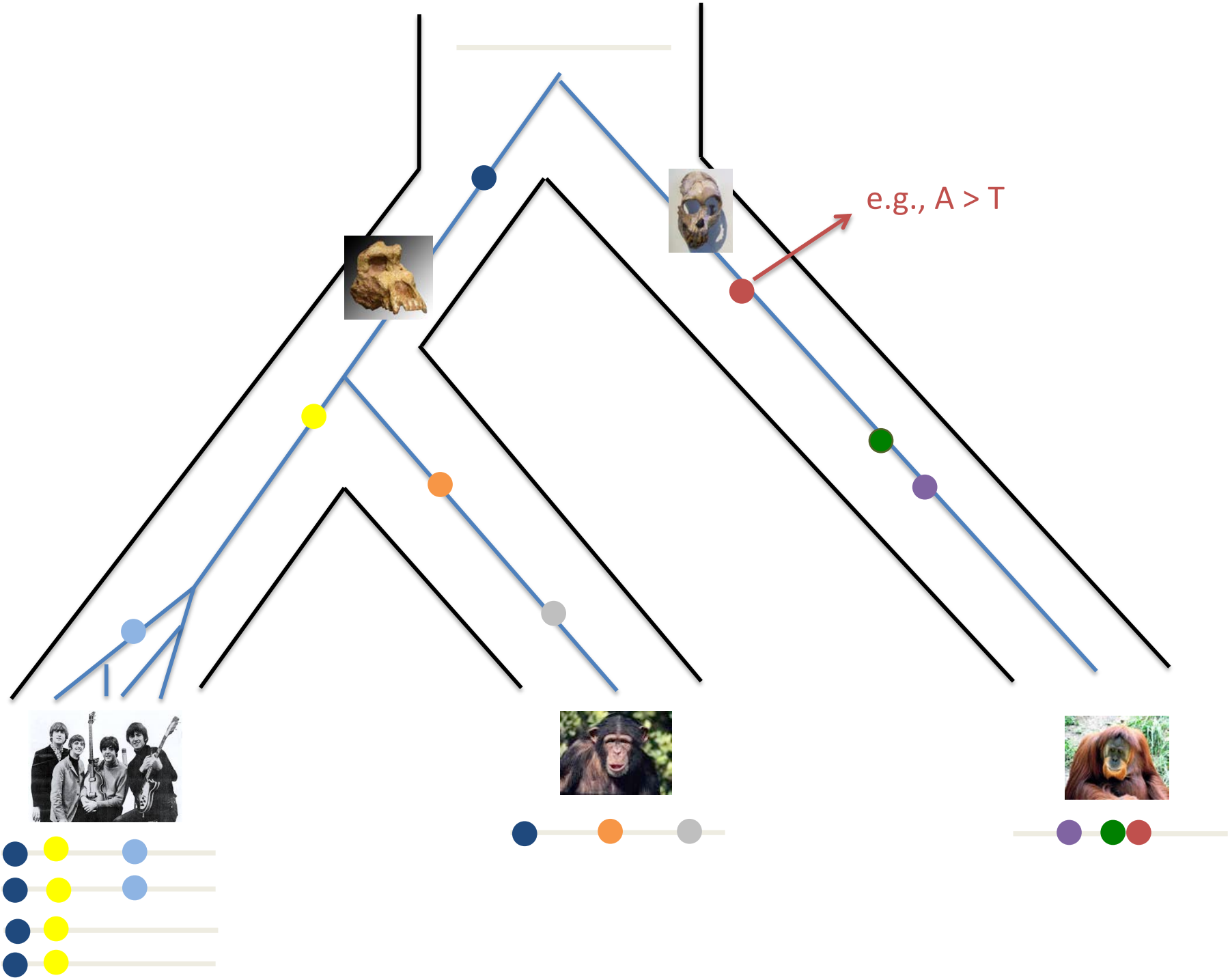
http://www.windows2universe.org/earth/polar/inuit_image_gallery.html

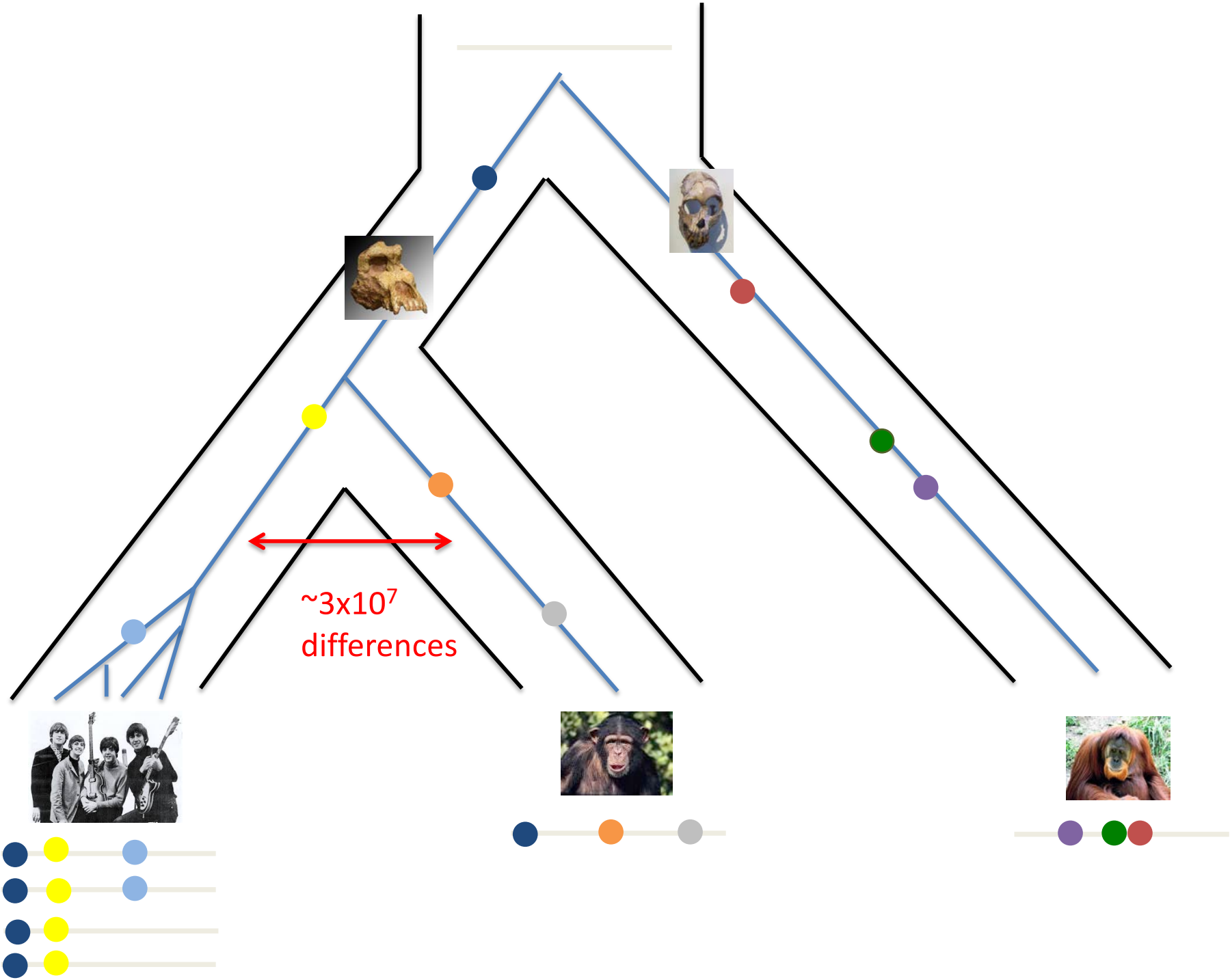
http://commons.wikimedia.org/wiki/File:Maasai_tribe.jpg

- How many changes were involved?
- What types of change do they involve (e.g., enhancers, gene duplications etc…)?
- Can we generate a comprehensive list of the changes?
- What were the typical fitness effects of beneficial changes?



CATGACGTCGCGGACAACCCAGAATTGTCTTGAGCGATGGTAAGATCTAACCTCACTGCCGGGGGAGGCTCATAC
CTGGGGCTTTACTGATGTCATACCGTCTTGCACGGGGATAGAATGACGGTGCCCGTGTCTGCTTGCCTCGAAGCA
ATTTTCTGAAAGTTACAGACTTCGATTAAAAAGATCGGACTGCGCGTGGGCCCCGGAGAGACATGCGTGGTAGTCA
TTTTTCGACGTGTCAAGGACTCAAGGGAATAGTTTGGCGGGAGCGTTACAGCTTCAATTCCCAAAGGTCGCAAGA
CGATAAAATTCAACTACTGGTTTCGGCCTAATAGGTCACGTTTTATGTGAAATAGAGGGGAACCGGCTCCCAAAT
CCCTGGGTGTTCTATGATAAGTCCTGCTTTATAACACGGGGCGGTTAGGTTAAATGACTCTTCTATCTTATGGTG
ATCCAAGCGCCCGCTAATTCTGTTCTGTTAATGTTCATACCAATACTCACATCACATTAGATCAAAGGATCCCCG
AGCCCAGTCGCAAGGGTCTGCTGCTGTTGTCGACGCCTCATGTTACTCCTGGAATCTACCTGCCCTCCCCTCACC
GGTTAAGGCGTGTGATCGACGATGCAGGTATACATCGGCTCGGACCTACAGTGGTCGATCGACTGGCTACTGGCT
TCGCGGTTCGGCGCGTAGTTGAGTGCGATAACCCAACCGGTGGCAAGTAGCAAGAAGACCTACCTGGGTCACCTT
AGACAACCTAACTAATAGTCTCTAACGGGGAATTACCTTTACCAGTCTCATGCCTCCAATATATCTGCACCGCTT
CAATGATATCGCCCACAGAAAGTAGGGTCTCAGGTATCGCATACGCCGCGCCCGGGTCCCAGCTACGCTCAGGAC
GACAGTAGAGAGCTATTGTGTAATTCAGGCTCAGCATTCATCGACCTTTCCTGTTGTGAATATTGTGCTAATGCA
TCTCGTCCGTAACGATCTGGGGGGCAAAACCGAATATCCGTATTCTCGTCCTACGGGTCCACAATGAGAAAGTCC
TGCGCGTGATCGTCAGTTAAGTTAAATTAATTCAGGCTACGGTAAACTTGTAGTGAGCTAAGAATCACGGGAATC
ACGGGTTCGCTACAGATGAACTGAATTTATACACGGACAACTCATCGCCCATTTGGGCGTGGGCACCGCAGATCA
AAAGTGGCAGATTAGGAGTGCTTGATCAGGTTAGCAGGTGGACTGTATCCAACAGCGCATCAAACTTCAATAAAT
CCAAAGCGTTGTAGTGGTCTAAGCACCCCTGAACAGTGGCGCCCATCGTTAGCGTAGTACAACCCTTCCCCCTTG
AGGTGCGACATGGGGCCAGTTAGCCTGCCCTATATCCCTTGCACACGTTCAATAAGAGGGGCTCTACAGCGCCGC
TTTTTAAATTAGGATGCCGACCCCATCATTGGTAACTGTATGTTCATAGATATTTCTTCAGGAGTAATAGCGACA
AGCTGACACGCAAGGGTCAACAATAATTTCTACTATCACCCCGCTGAACGACTGTCTTTGCAAGAACCAACTGGG
CTTAGATTCGCGTCCTAACGTAGTGAGGGCCGAGTCATATCATAGATCAGGCATGAGAAACCGACGTCGAGTCTA
CACACGAGTTGTAAACAACTTGATTGCTATACTGTAGCTACCGCAAGGATCTCCTACATCAAAGACTACTGGGCG
ATCTGGATCCGAGTCAGAAATACGAGTTAATGCAAATTTACGTAGACCGGTGAAAACACGTGCCATGGGTTGCGT
AGACCGTAGTCAGAAGTGTGGCGCGCTATTCGTACCGAACCGGTGGAGTATACAGAATTGCTCTTCTACGACGTA
AGGAGCTCGGTCCCCAATGCACGCCAAAAAAGGGAATAAAGTATTCAAACTGCGCATGGTCCCTCCGCCGGTGGCA
CTATTATCCATCCGAACGTTGAACCTACTTCCTCGGCTTATGCTGTCCTCAACAGTATCGCTTATGAATCGCATG
CGGCTGTGGATCTTAACGGCCACATTCTTAATTCCGACCGATCACCGATCGCCTTTCCTCGCTGGTACAATGAGT
ACTAAGTTATCCAGATCAAGGTTTGAACGGACTCGTATGACATGTGTGACTGAACCCGGGAGGAAATGCAGAGAA
CTGTTTCAAGGCCTCTGCTTTGGTATCACTCAATATATTCAGACCAGACAAGTGGCAAAATTTCGTGCGCCTCTC
CTAGGTATTCACGCAACCGTCGTAACATGCACTAAGGATAACTAGCGCCAGGGGGGCATACTAGGTCCCGGAGCT
AAAGACTACCCTATGGATTCCTTGGAGCGGGGACAATGCAGACCGGTTACGACACAATTATCGGGATCGTCTAGA
GGTATTATTAGCAAGACAATAAAGGACATTGCACAGAGACTTATTAGAATTCAACAAACAGGATCATATCATGCG
GTGTTGGGTCGGGCAAGTCCCCGAAGCTCGGCCAAAAGATTCGCCATGGAACCGTCTGGTCCTGTTAGCGTGTAC
GCCTGCTCCTGTTCCGGGTACCATAGATAGACTGAGATTGCGTCAAAAAATTGCGGCGAAAATAGAGGGGCTCCT
TGTAGAAATACCAGACTGGGGAATTTAAGCGCTTTCCACTATCTGAGCGACTAAACATCAACAAATGCGTCTACT
CGAATCCGCAGTAGGCAATTACAACCTGGTTCAGATCACTGGTTAATCAGGGATGTCTTCATAAGATTATACTTG
CCCCGACGCGACAGCTCTTCAAGGGGCCGATTTTTGGACTTCAGATACGCTAGAATTTAAAGGGTCTCTTACACC
TGCTGCGGCCTGCAGGGACCCCTAGAACTTGCCGCCTACTTGTCTCAGTCTAATAACGCGCGAAGCCGTGGGGCA
CGTGACCTTAAGTCGCAGAGCGAGTGATGAATTTGGGACGCTAATATGGGTGAATAGAGACTTATATCATCAGGG


e.g., A > T


~3x10^7
differences

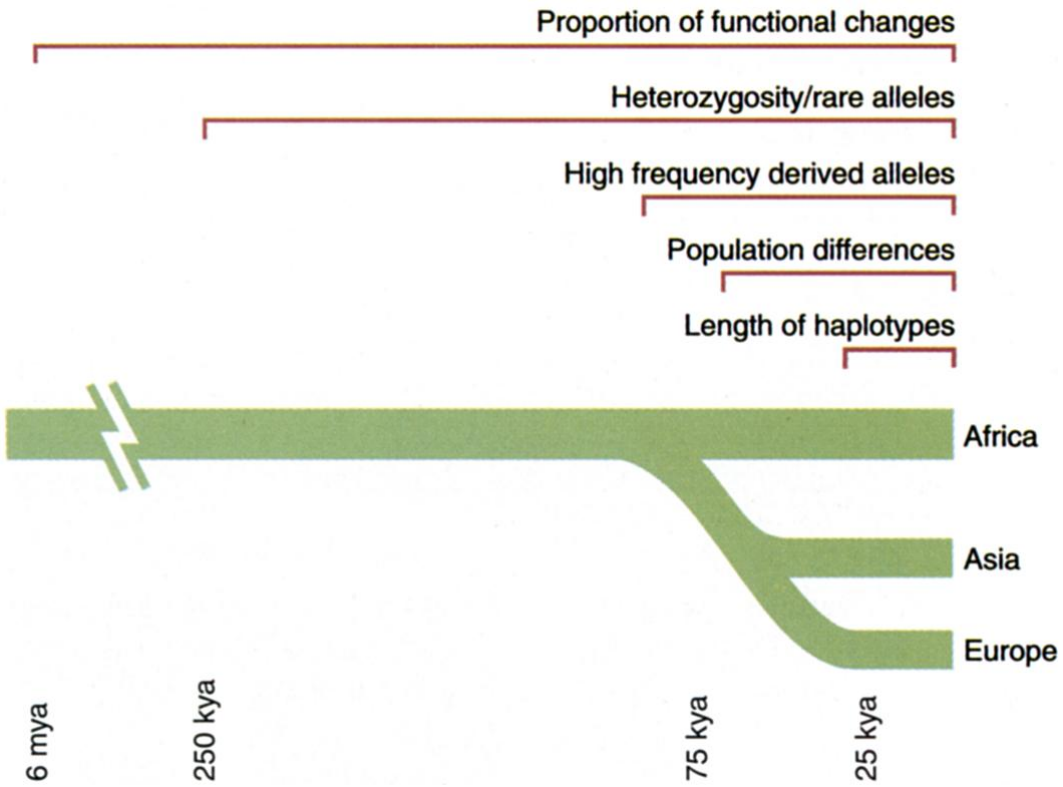
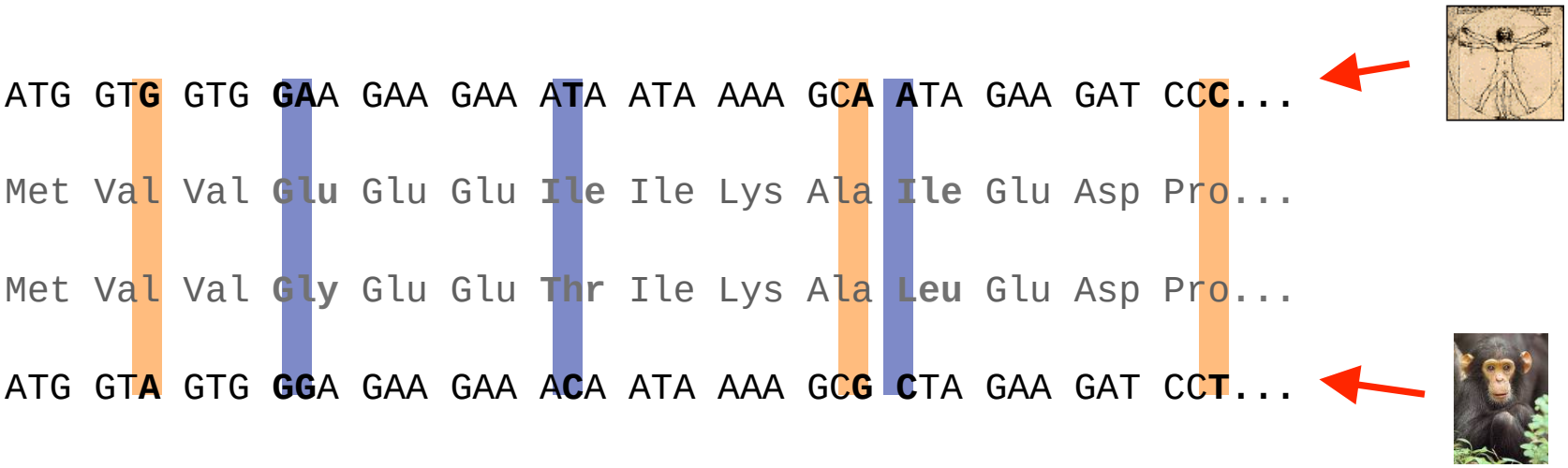


Fig. 1. Time scales for the signatures of selection. The five signatures of selection persist over varying time scales. A rough estimate is shown of how long each is useful for detecting selection in humans. (See fig. S1 for details on how the approximate time scales were estimated).

Sabeti et al. 2006 Science



synonymous divergence	$D_s = 3$
non-synonymous divergence	$D_n = 3$

synonymous sites	$L_s = 8$
non-synonymous sites	$L_n = 34$

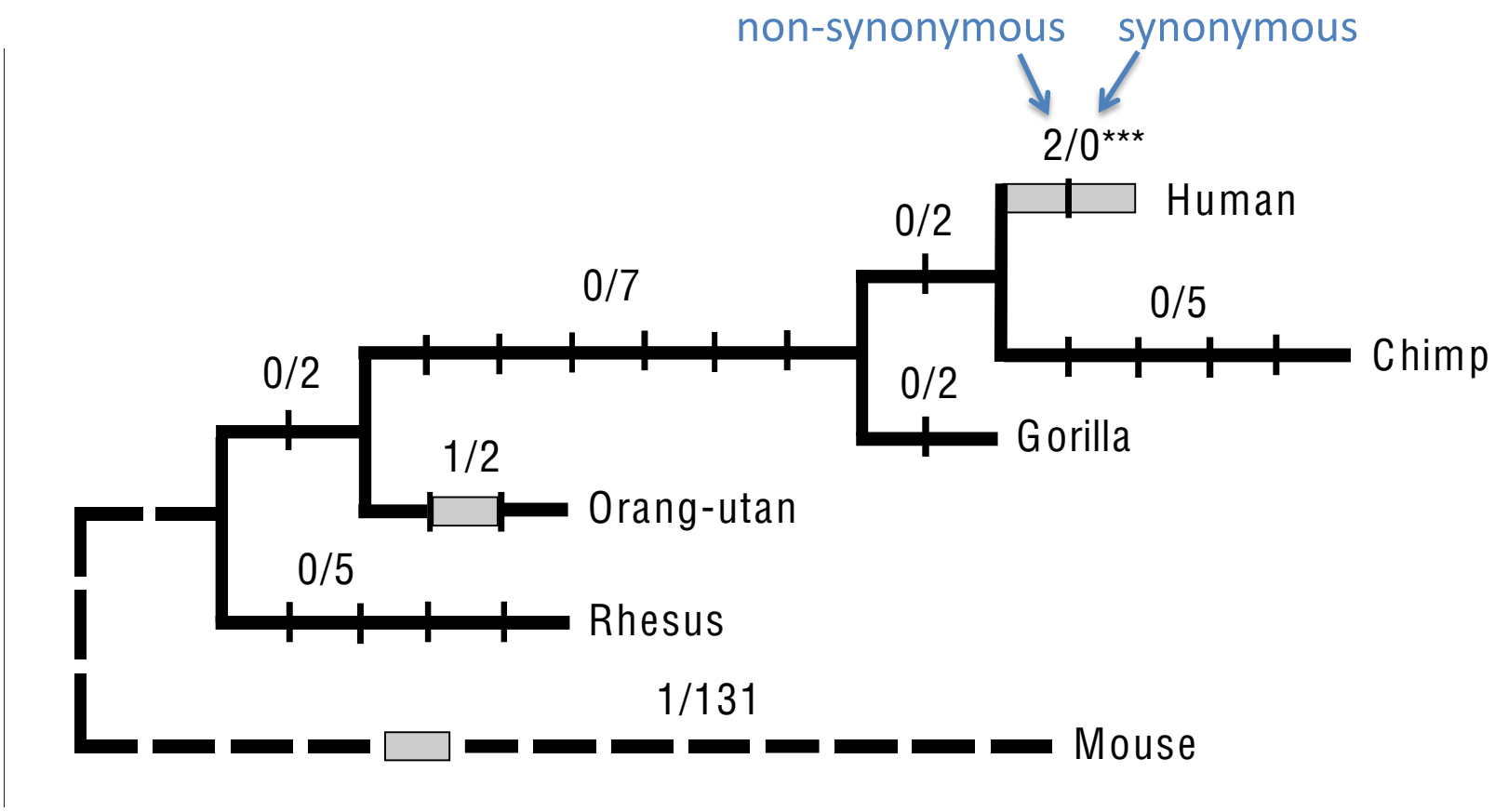
synonymous divergence per-site	$d_s = 3/8 \approx .38$
non-synonymous divergence per-site	$d_n = 3/34 \approx .09$

On average $d_n/d_s \approx 1/6$ => selection to maintain the protein

dn>ds => selection favors protein changes

Muse & Gaut 1994; Goldman & Yang 1994

FOXP2

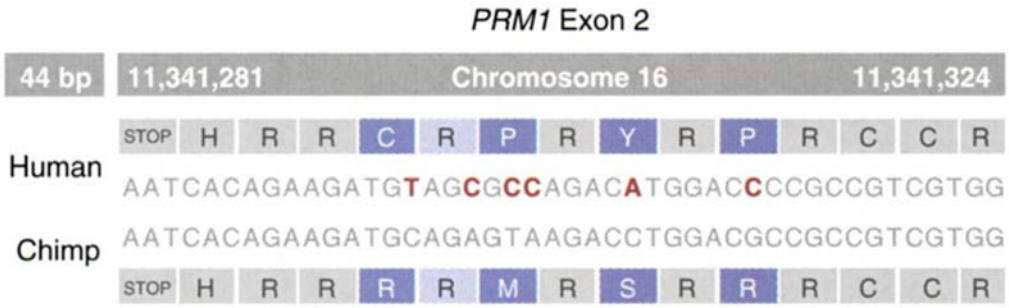


Enard et al. 2002 Nature

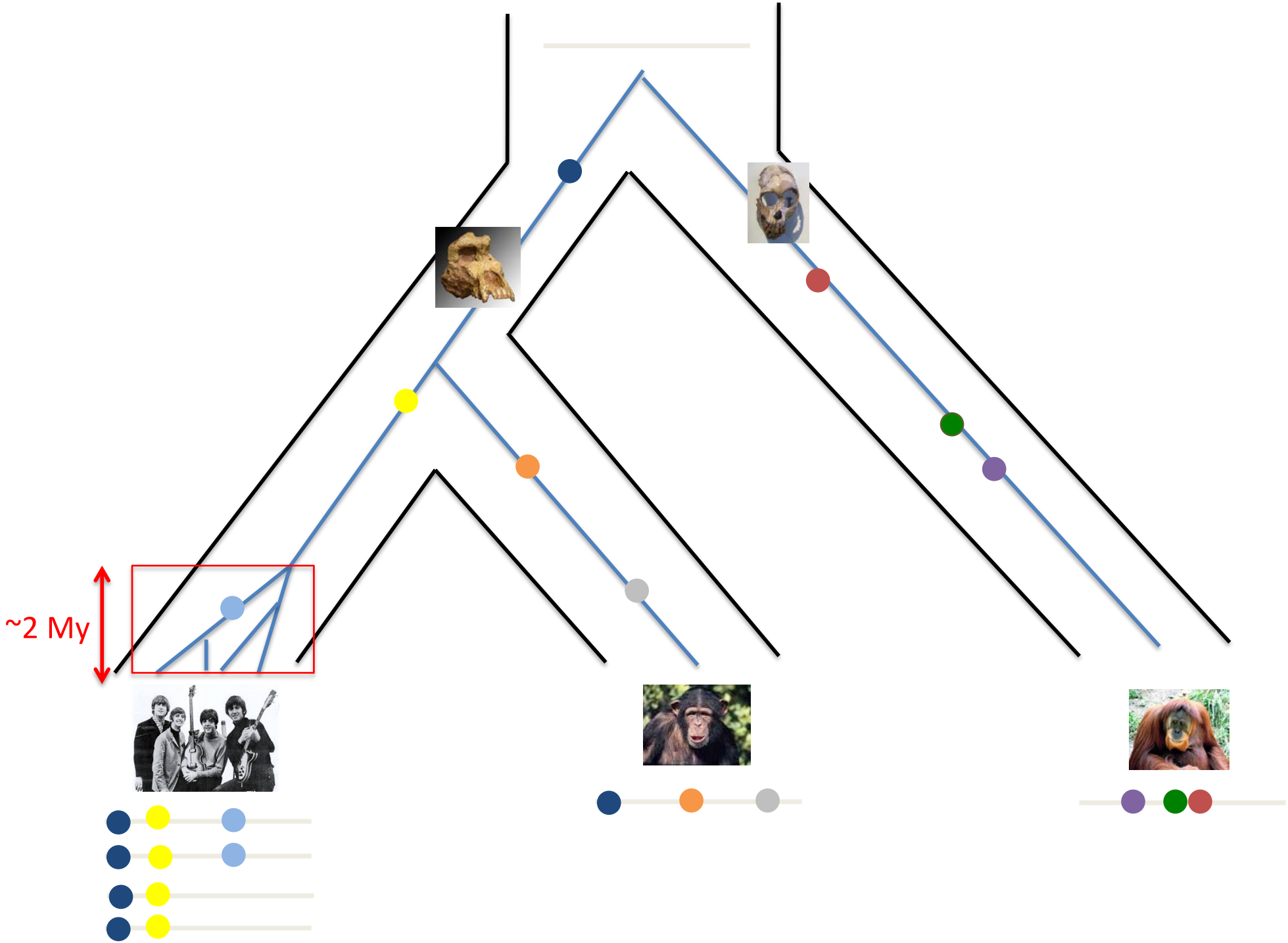
Table 1. Biological Process Categories with an Excess of Putatively Positively Selected Genes (Nominal *p* less than 0.05; MWU) among a Total of 133 Biological Process Categories

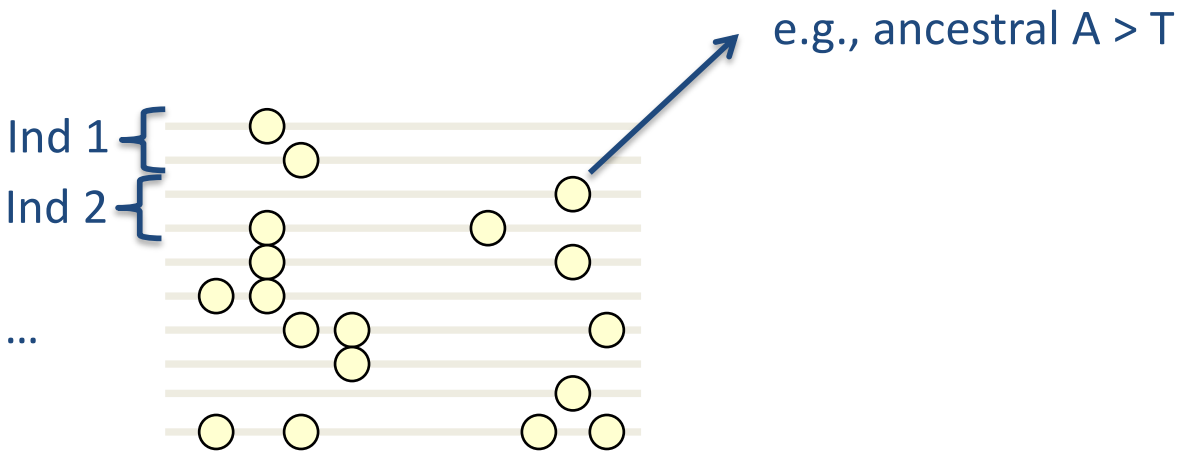
Biological Process	Number of Genes	*p*-Value
Immunity and defense	417	0.0000
T-cell-mediated immunity	82	0.0000
Chemosensory perception	45	0.0000
Biological process unclassified	3,069	0.0000
Olfaction	28	0.0004
Gametogenesis	51	0.0005
Natural killer-cell-mediated immunity	30	0.0018
Spermatogenesis and motility	20	0.0037
Inhibition of apoptosis	40	0.0047
Interferon-mediated immunity	23	0.0080
Sensory perception	133	0.0160
B-cell- and antibody-mediated immunity	57	0.0298

Borrowed from Nielsen et al. 2005



Borrowed from Sabeti et al. 2006

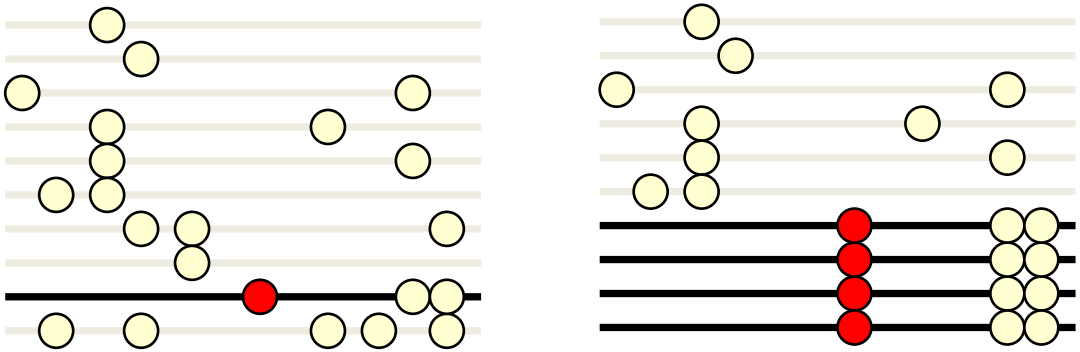

~2 My



The “classic” sweep model

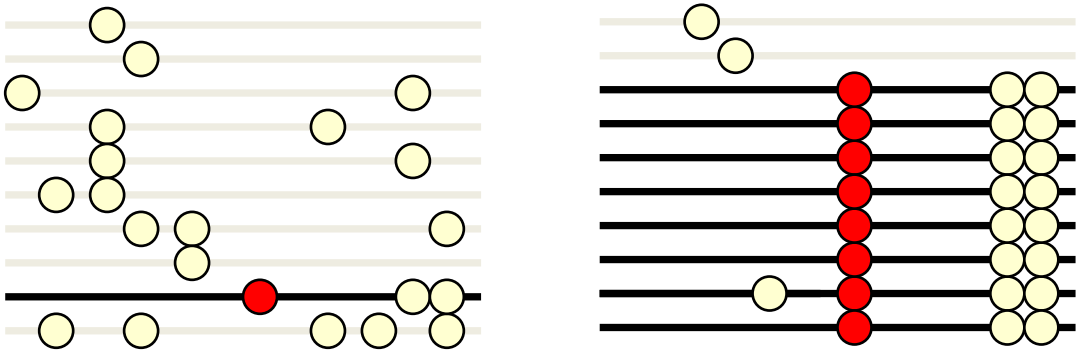
Maynard Smith & Haigh 1974

sweep



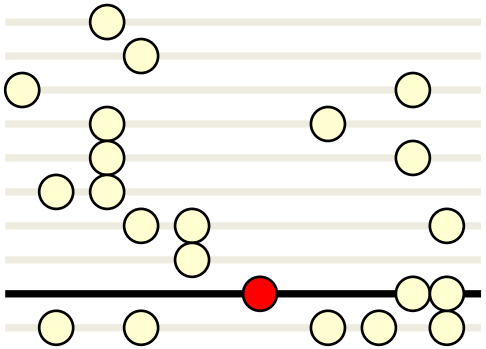
Maynard Smith & Haigh 1974

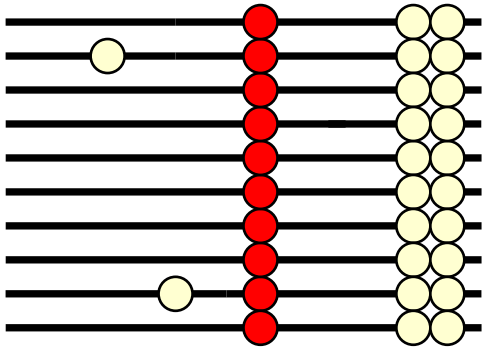
sweep



Maynard Smith & Haigh 1974

sweep



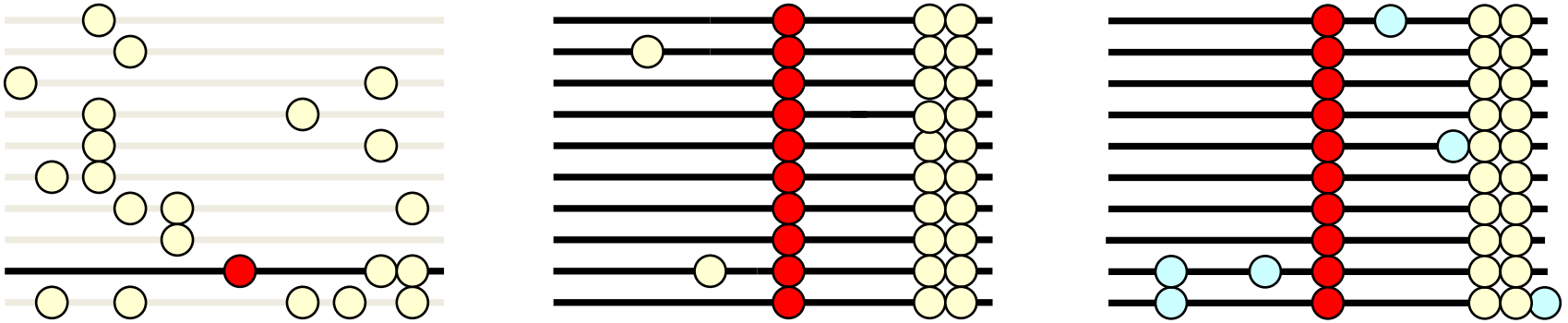


Takes on the order of tens of thousands of years in humans

Maynard Smith & Haigh 1974

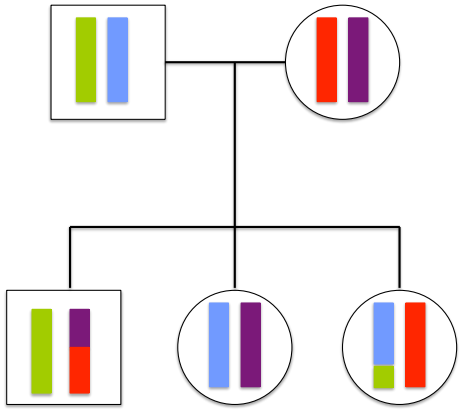
sweep

recovery



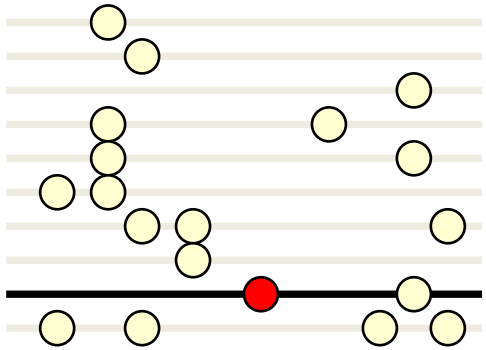
Recovery takes up to 2M years in humans

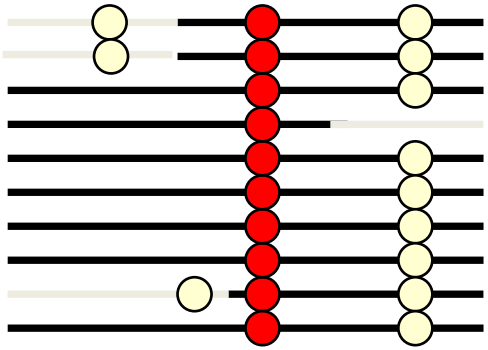
Maynard Smith & Haigh 1974

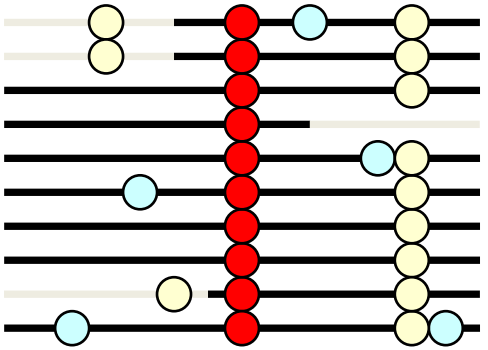


sweep

recovery

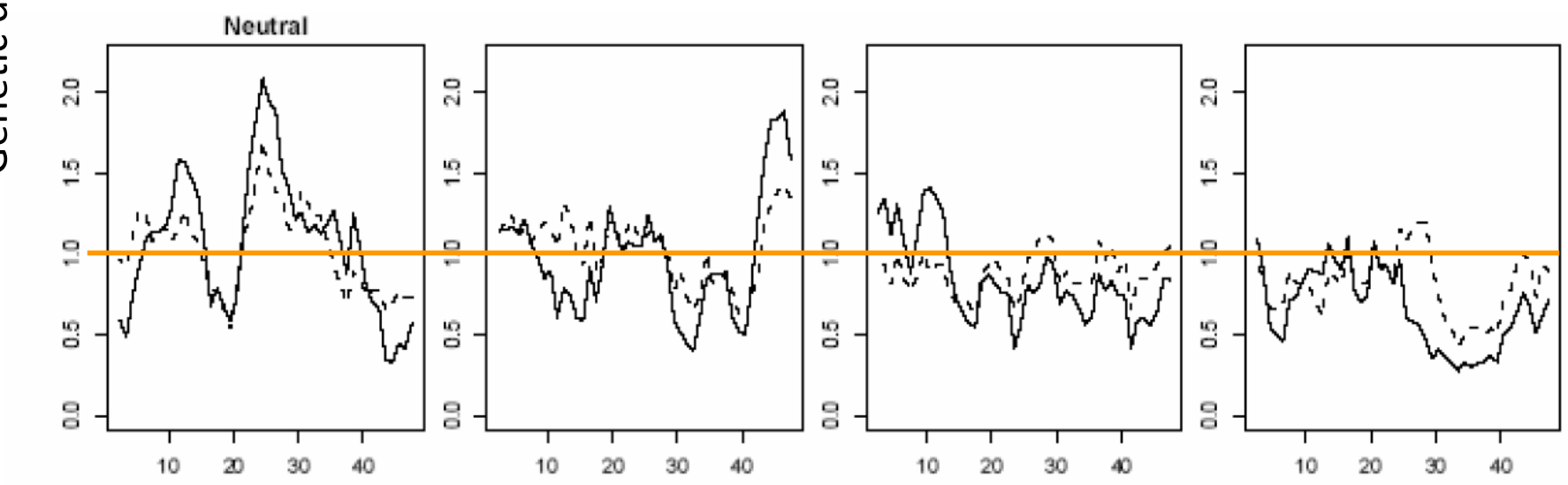




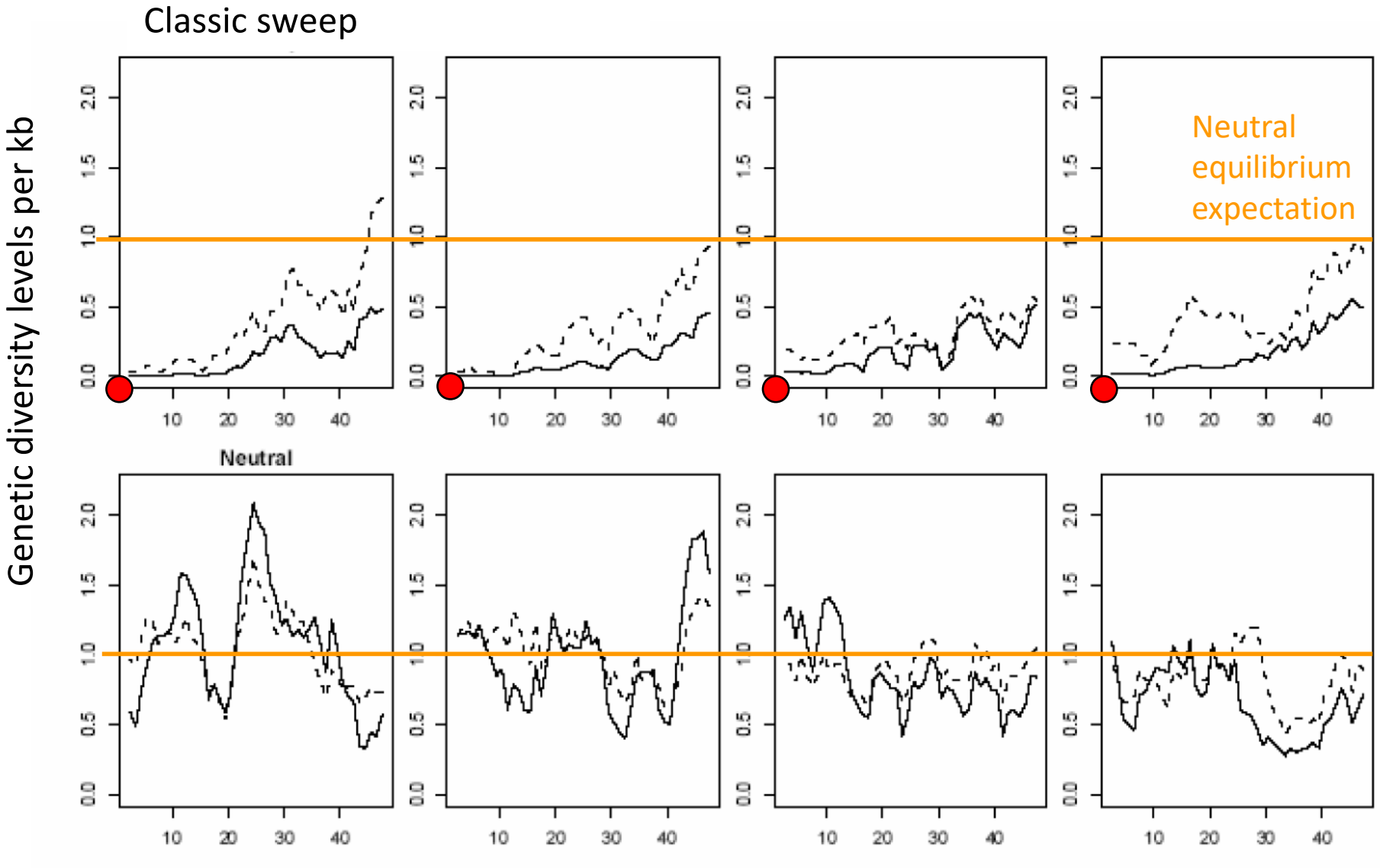


depends on s/r

Genetic diversity levels per kb

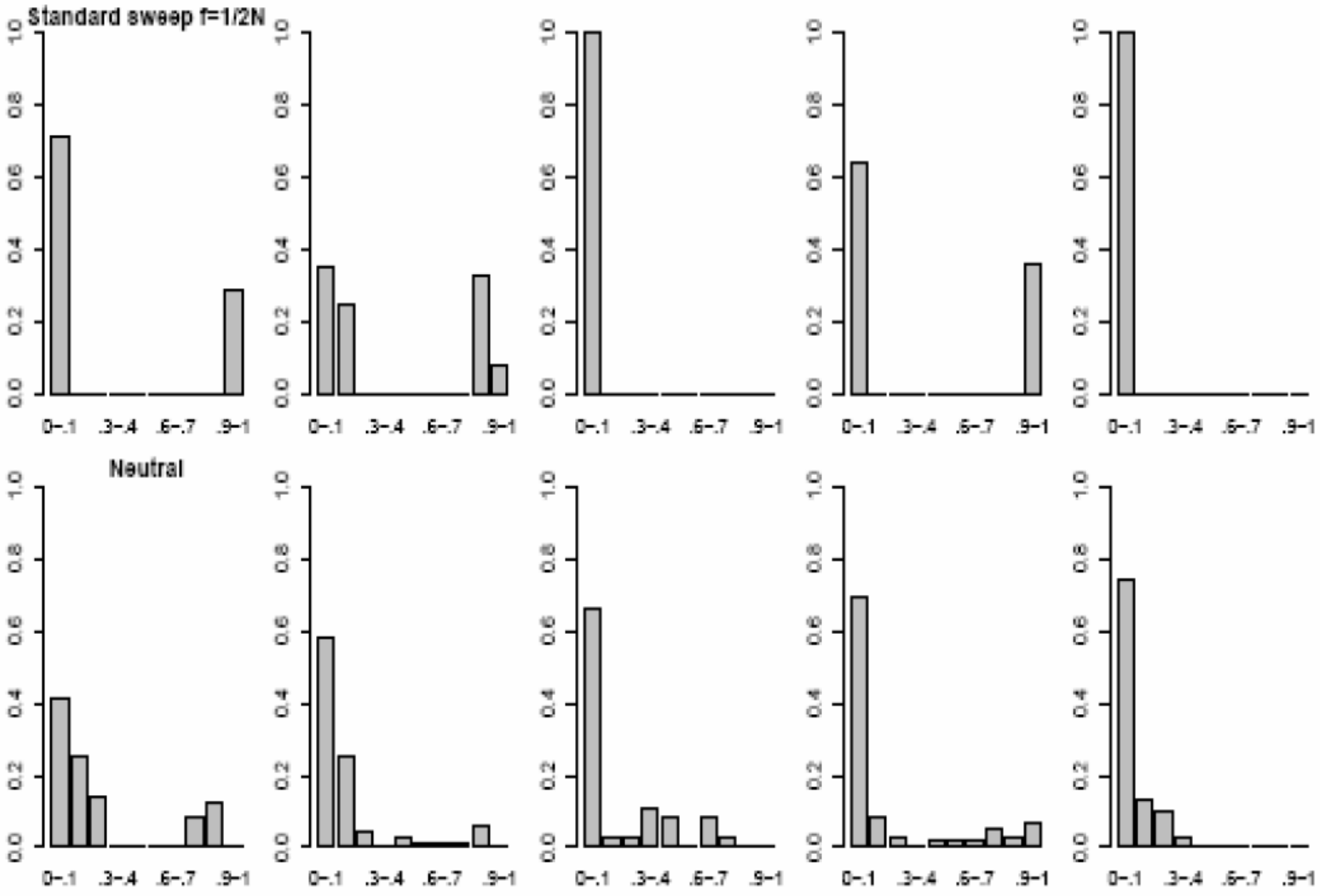


Przeworski et al. 2005



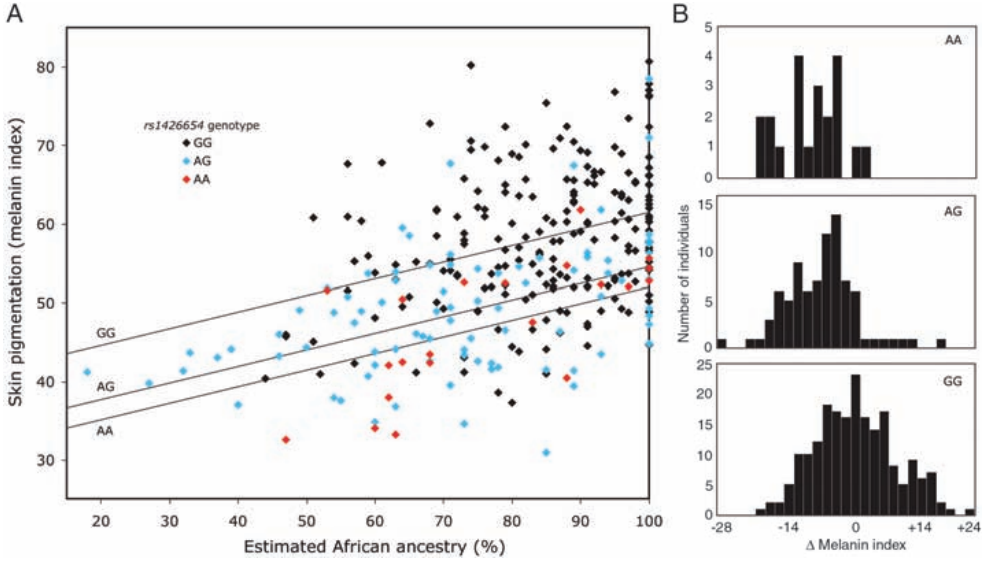
Przeworski et al. 2005

Five realizations of a sweep (that just ended)



Przeworski et al. 2005

A canonical classic sweep: *SLC24A5*



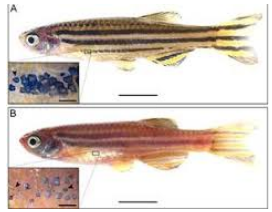
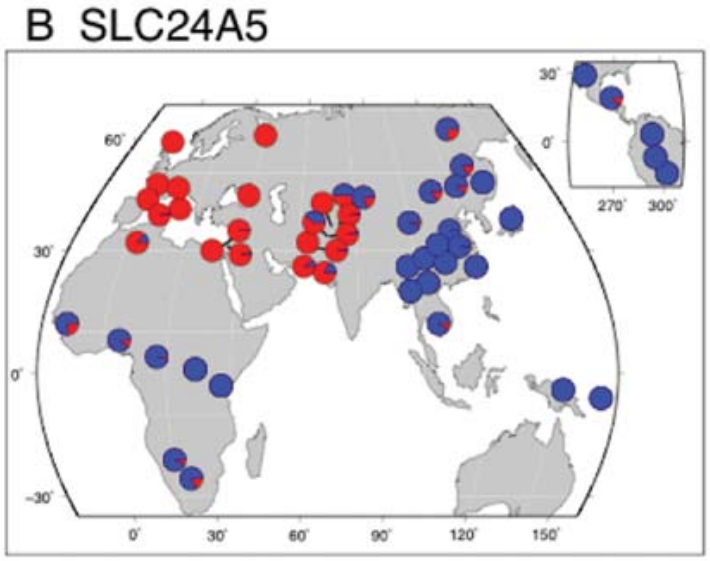
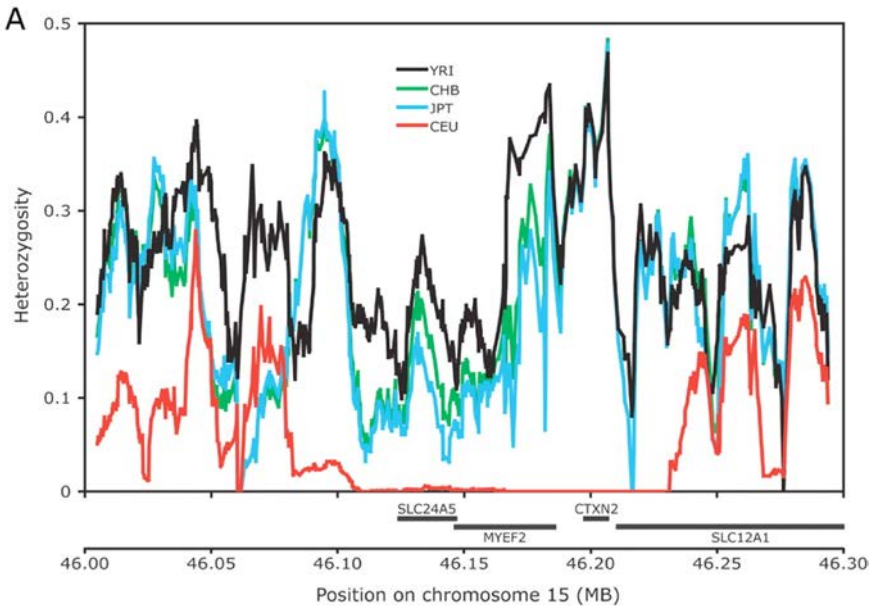


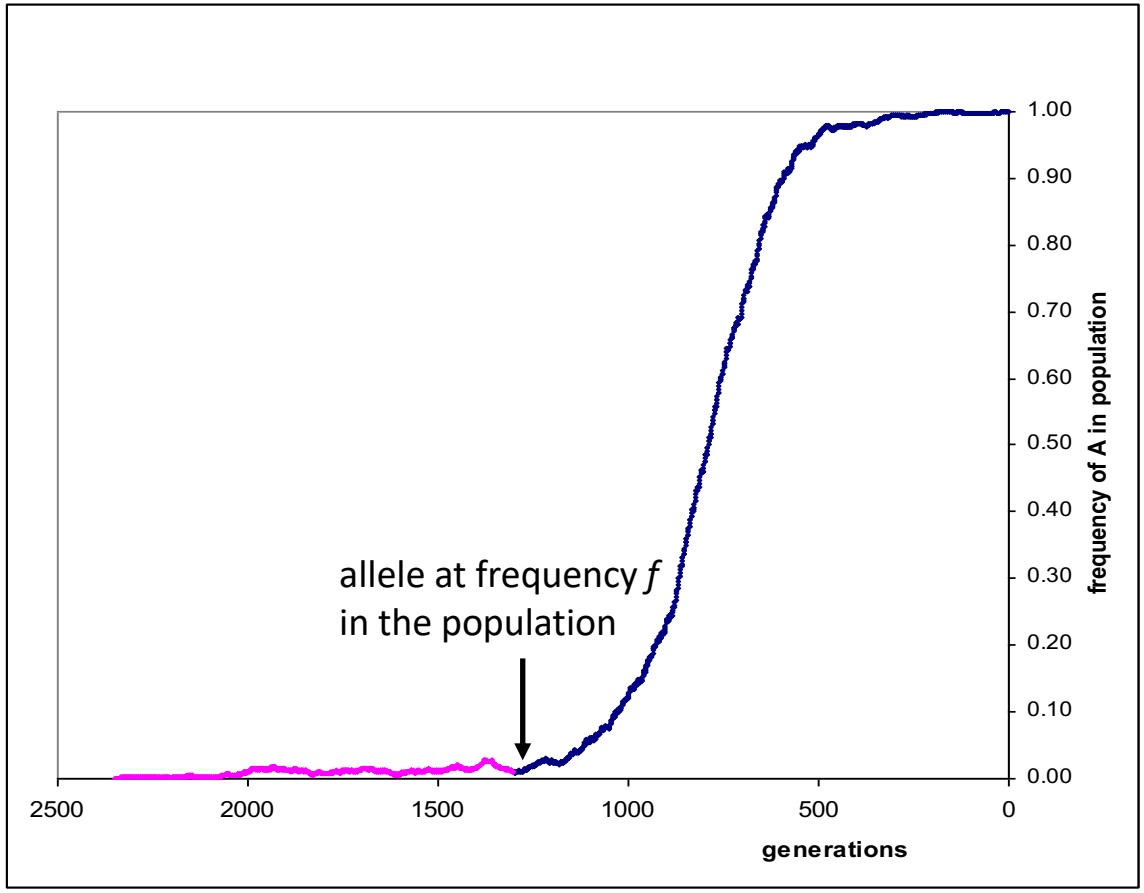
Fig. 6. Effect of *SLC24A5* genotype on pigmentation in admixed populations. (**A**) Variation of

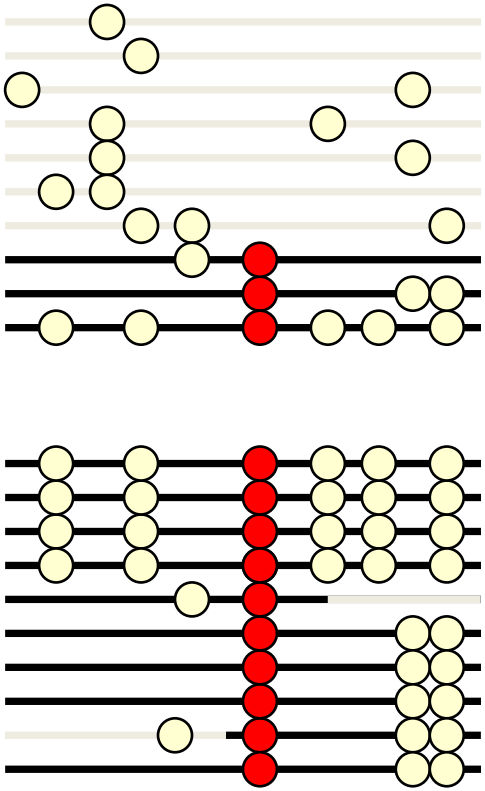




Borrowed from Lamason et al. 2005, Coop et al. 2009

Selection on standing variation

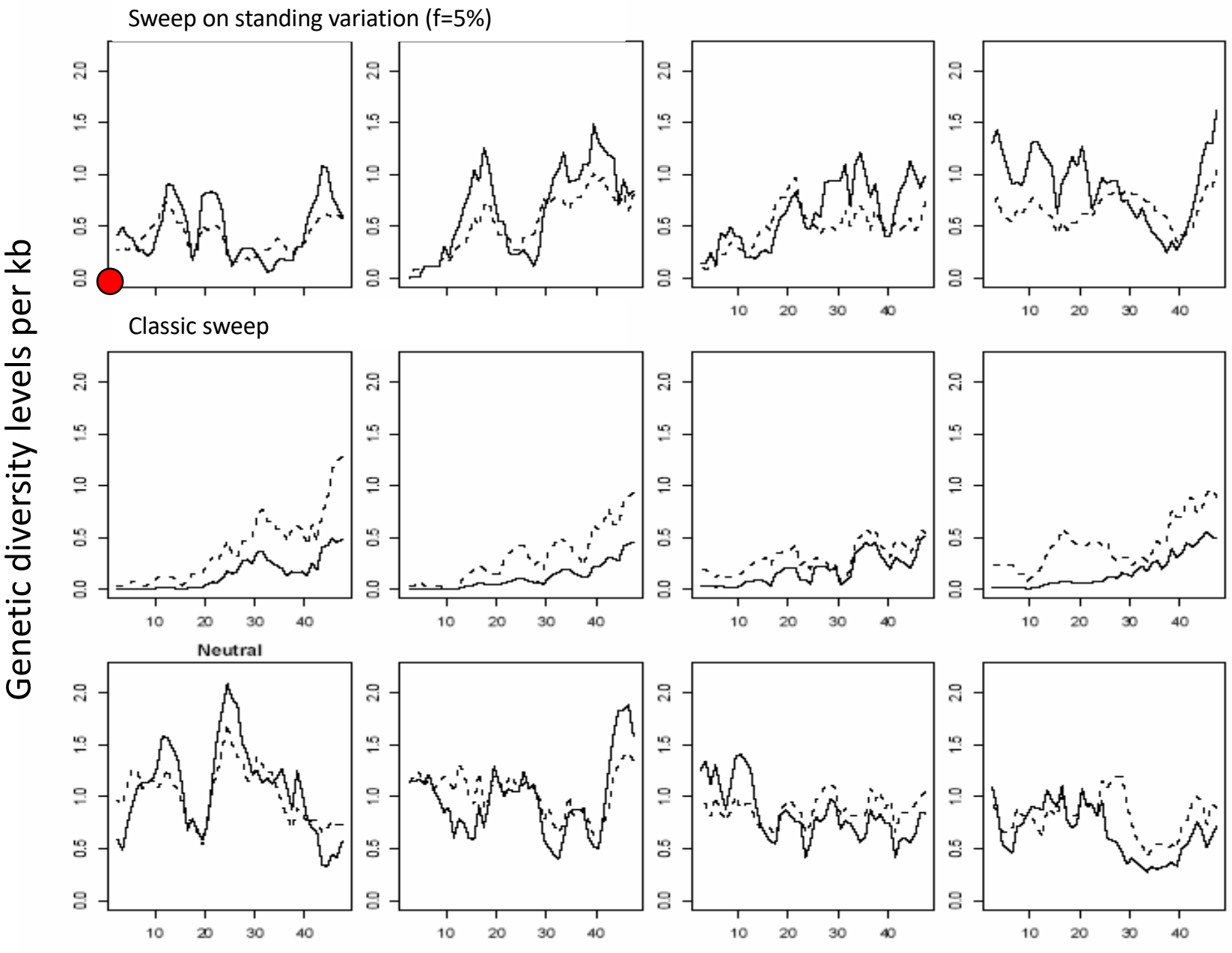




Allele is neutrally evolving

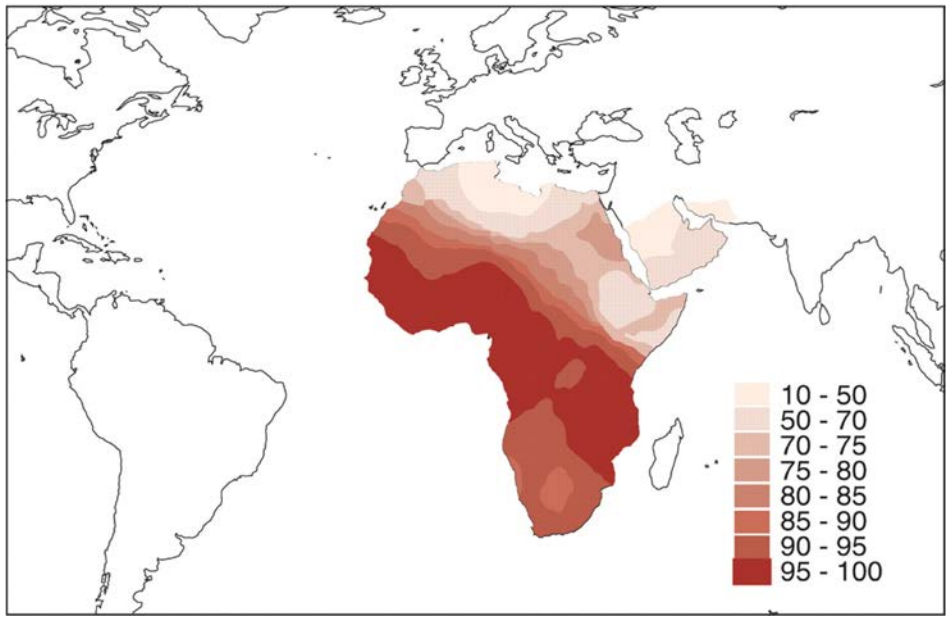
Allele is favored

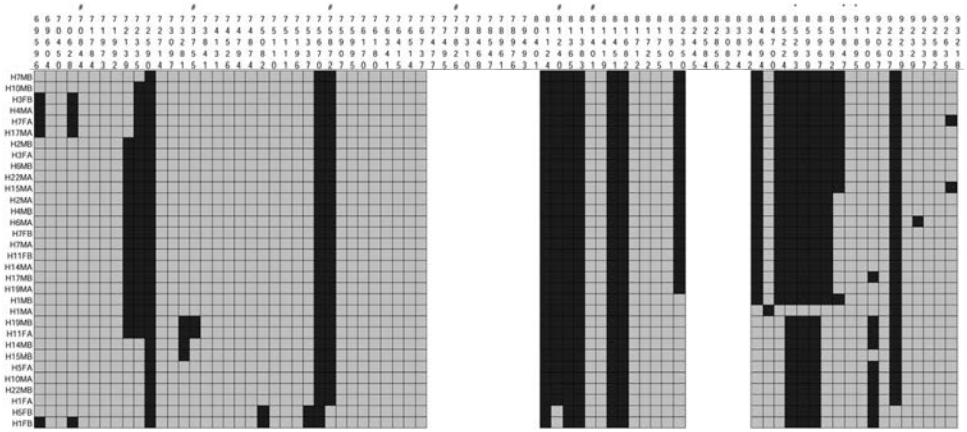
Przeworski et al. 2005, see also Hermission and Pennings 2005, 2006

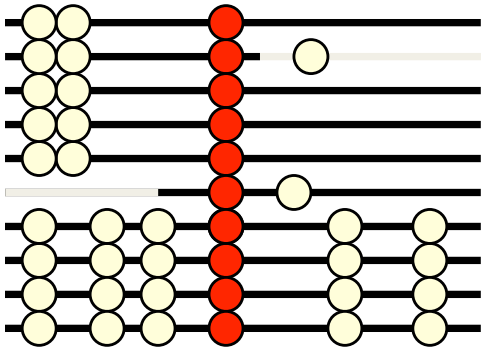


Przeworski et al. 2005

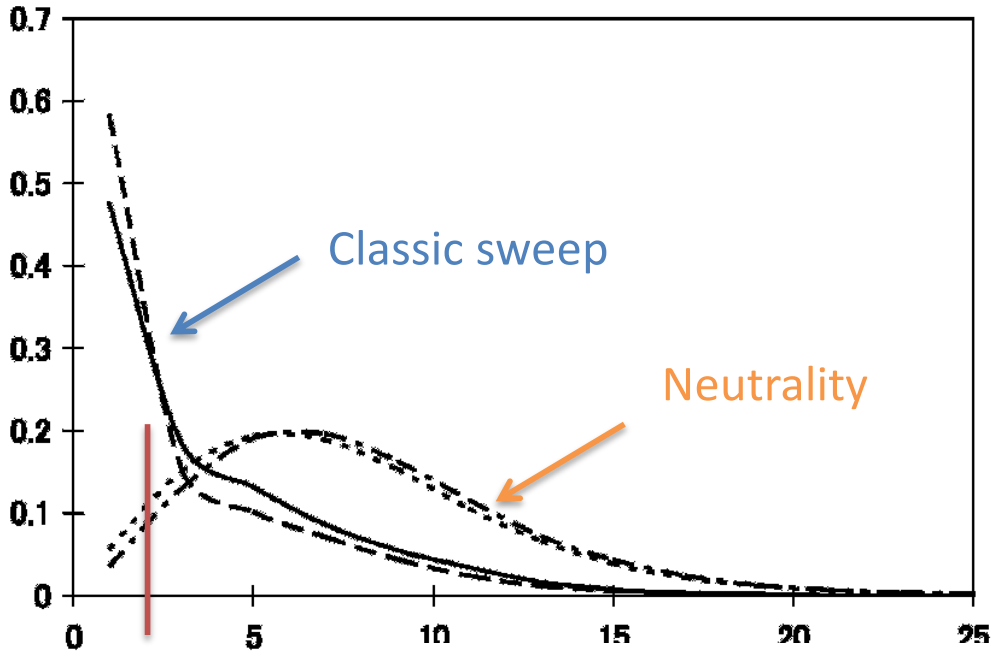
Duffy blood group locus







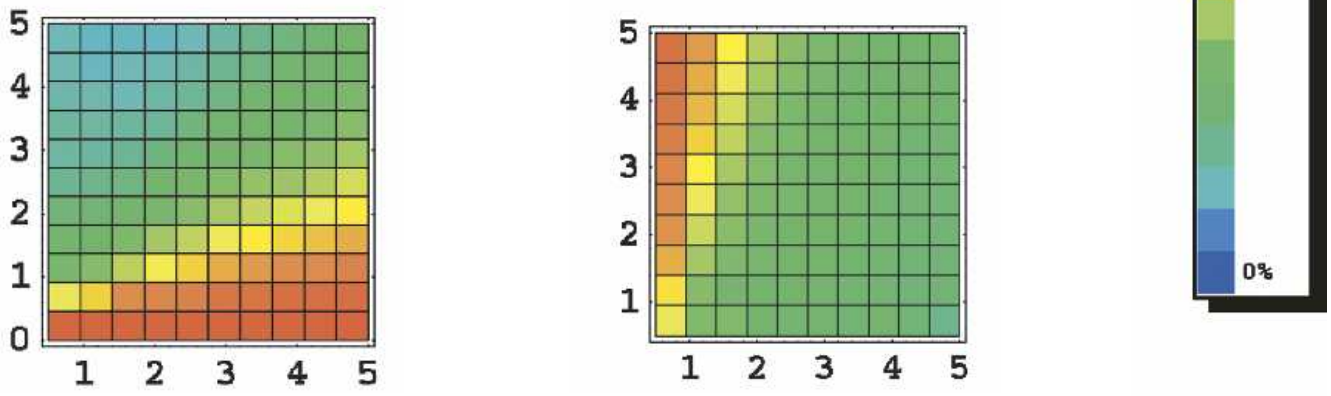
Borrowed from Hamblin & Di Rienzo 2002



Proportion of loci that were recent targets of a classic sweep

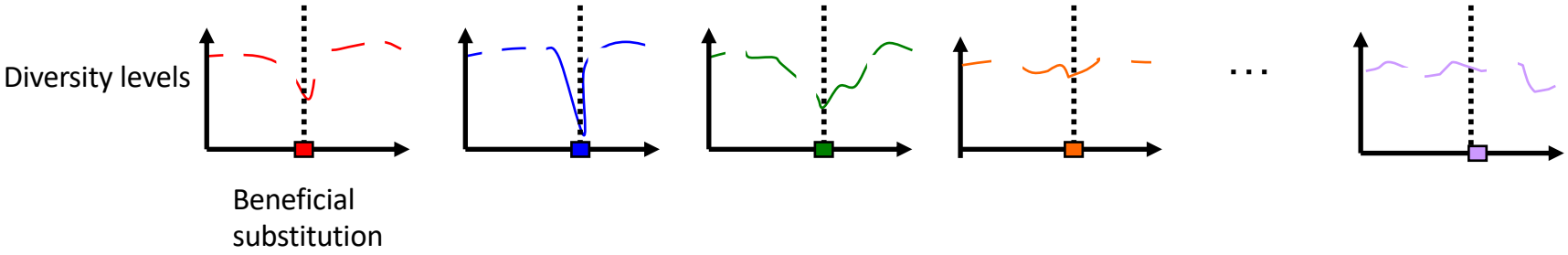
FDR

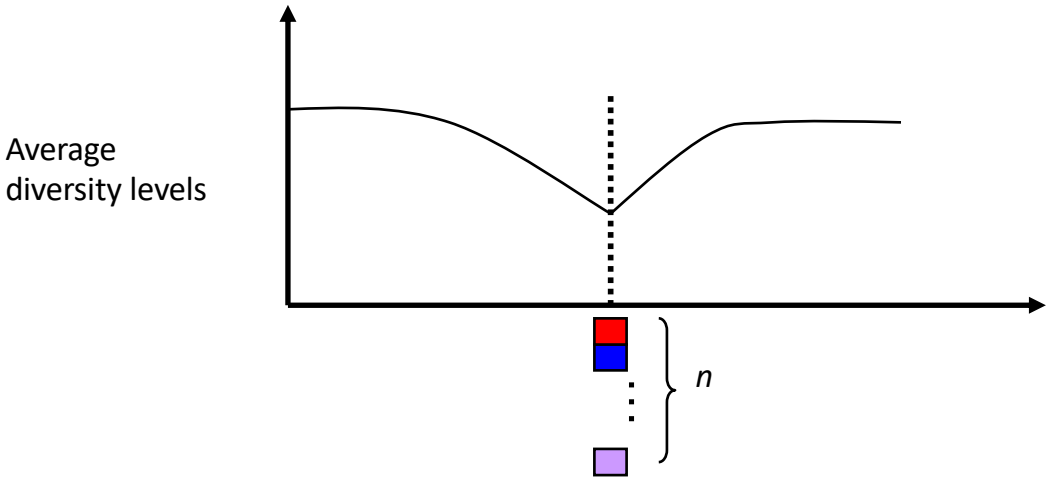
False negative rate

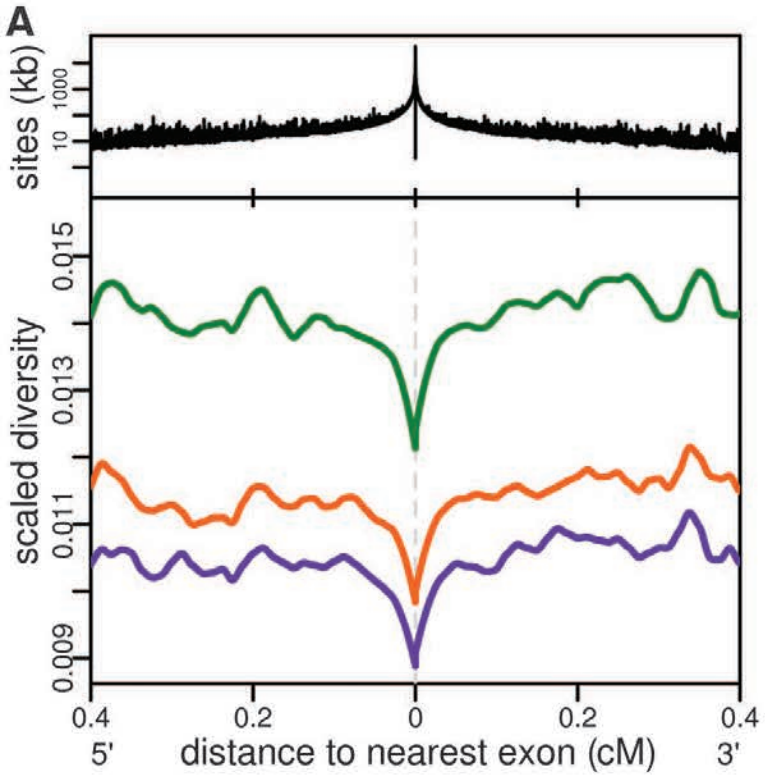


Distribution cut-off (tail %-tile)

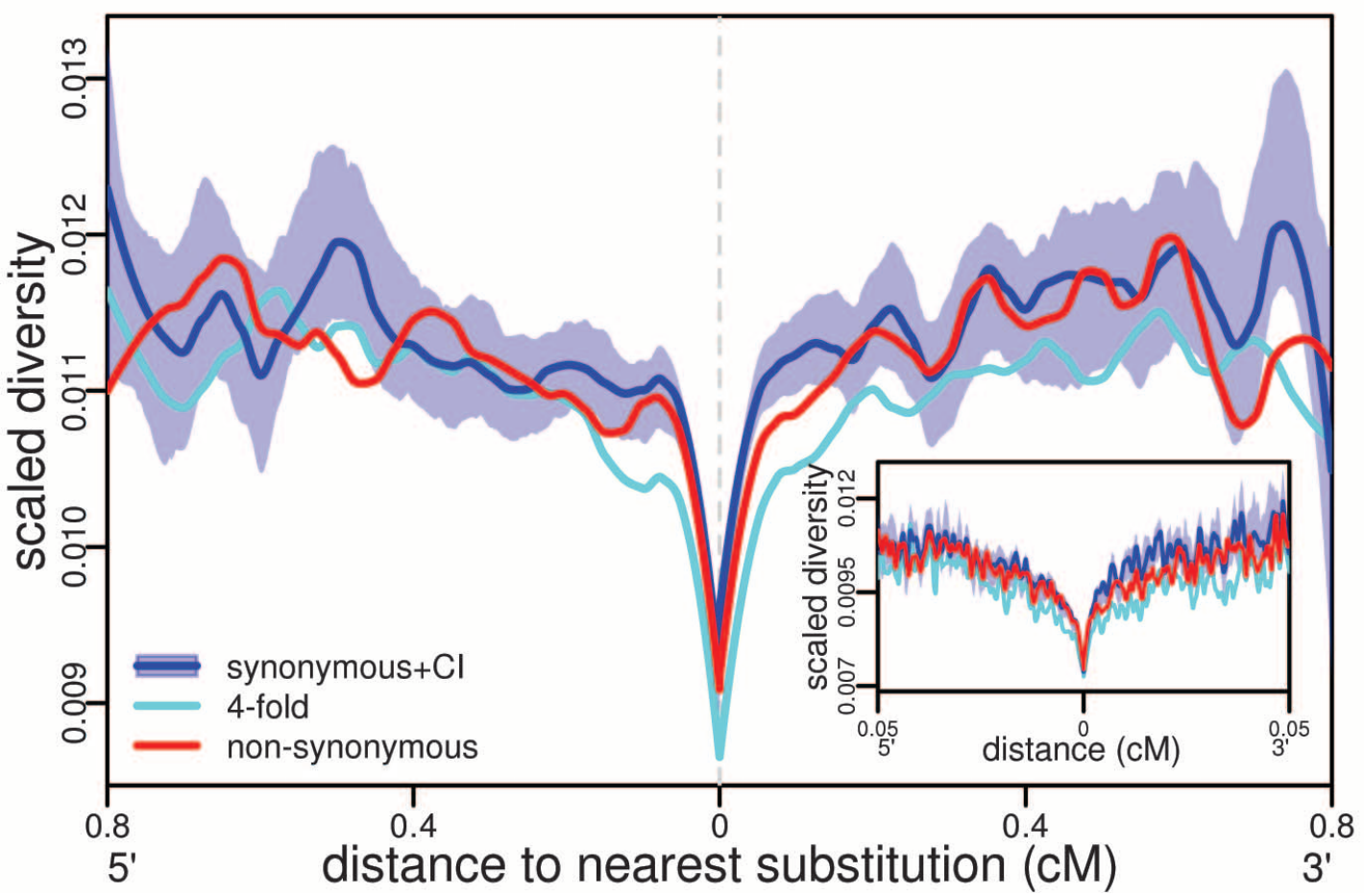
Teshima et al. 2006


Diversity levels
Beneficial substitution

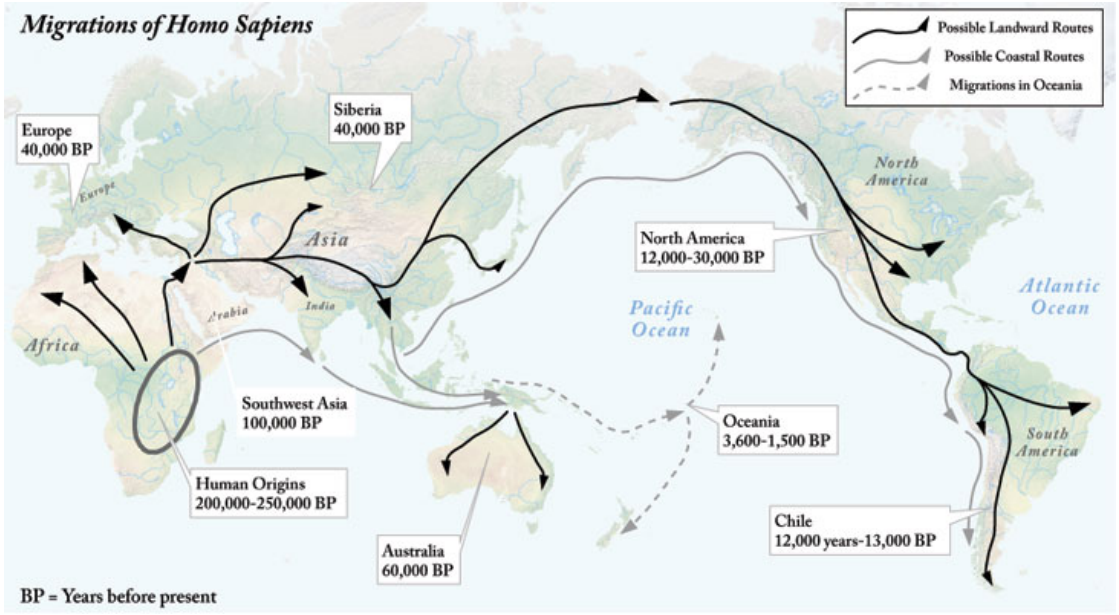

Average diversity levels
n



Hernandez et al. 2011

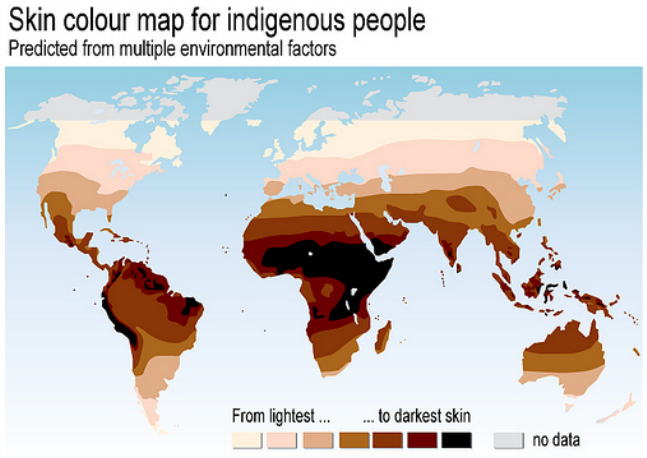


Hernandez et al. 2011



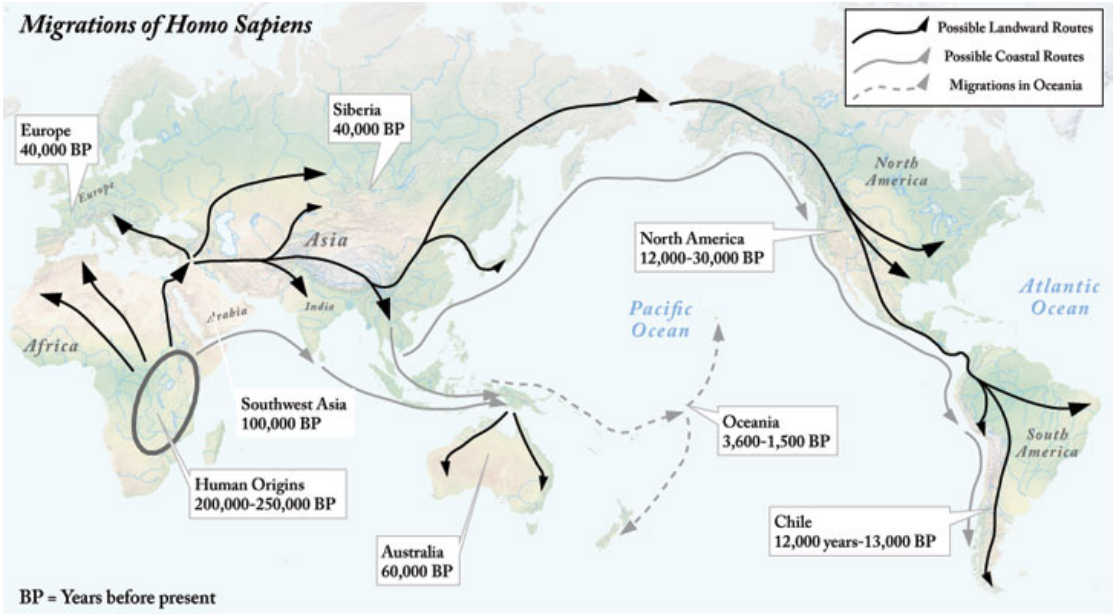
http://worldhistoryforusall.sdsu.edu/eras/era2.php

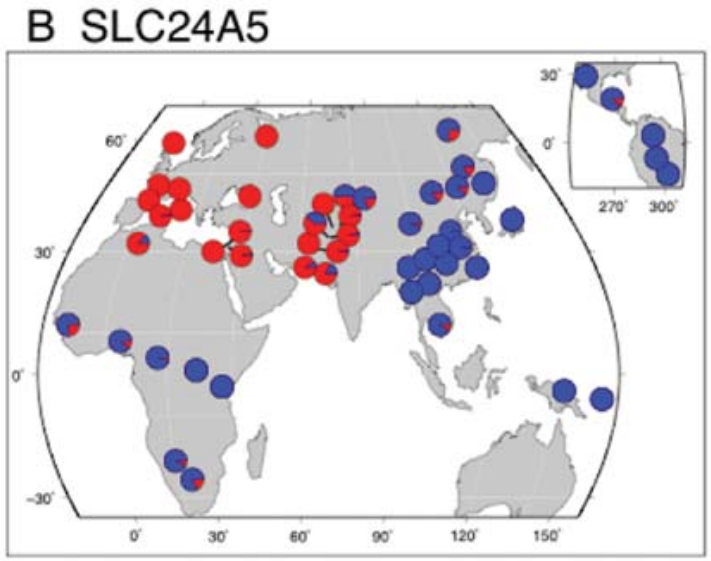
https://en.wikipedia.org/wiki/Allen%27s_rule

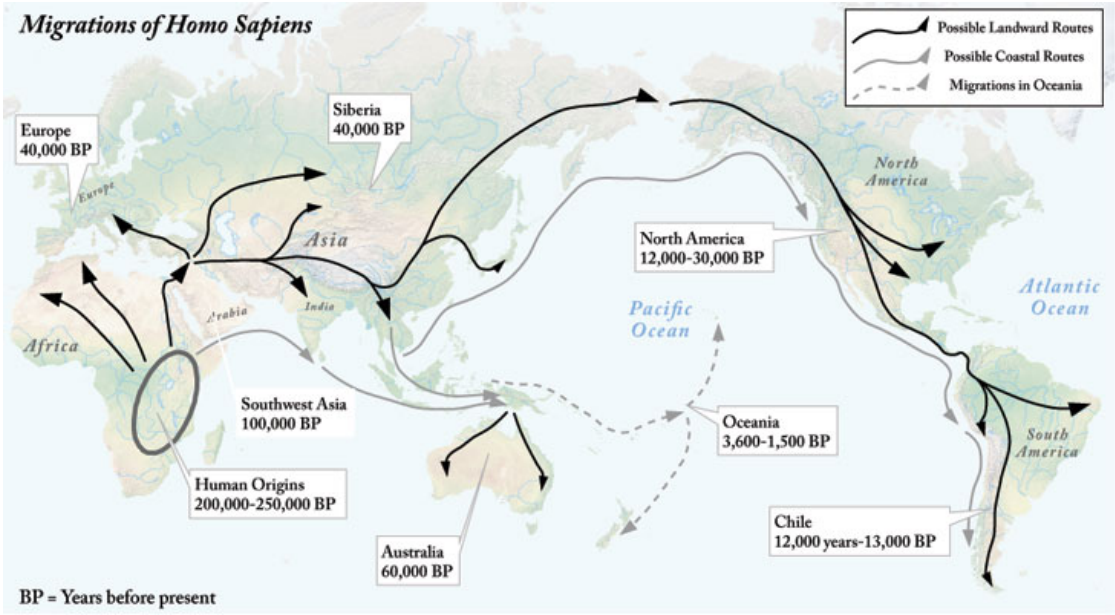


Source: Chaplin G.[©] , *Geographic Distribution of Environmental Factors Influencing Human Skin Coloration*, American Journal of Physical Anthropology 125:292–302, 2004; map updated in 2007.

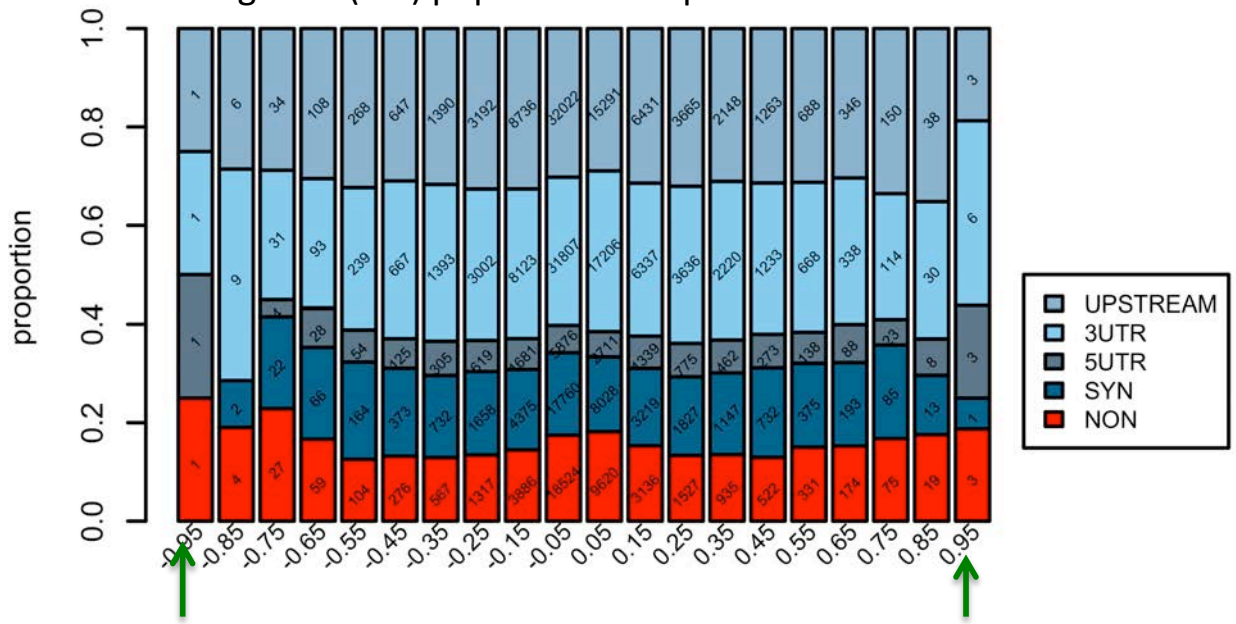
https://en.wikipedia.org/wiki/Maasai_people


Migrations of Homo Sapiens
Possible Landward Routes
Possible Coastal Routes
Migrations in Oceania
Europe
40,000 BP
Siberia
40,000 BP
North America
12,000-30,000 BP
Asia
Europe
Arabia
India
Africa
Pacific
Ocean
Atlantic
Ocean
North
America
South
America
Southwest Asia
100,000 BP
Oceania
3,600-1,500 BP
Human Origins
200,000-250,000 BP
Chile
12,000 years-13,000 BP
Australia
60,000 BP
BP = Years before present


B SLC24A5
60°
30°
0°
-30°
0°
30°
60°
90°
120°
150°
30°
0°
270°
300°

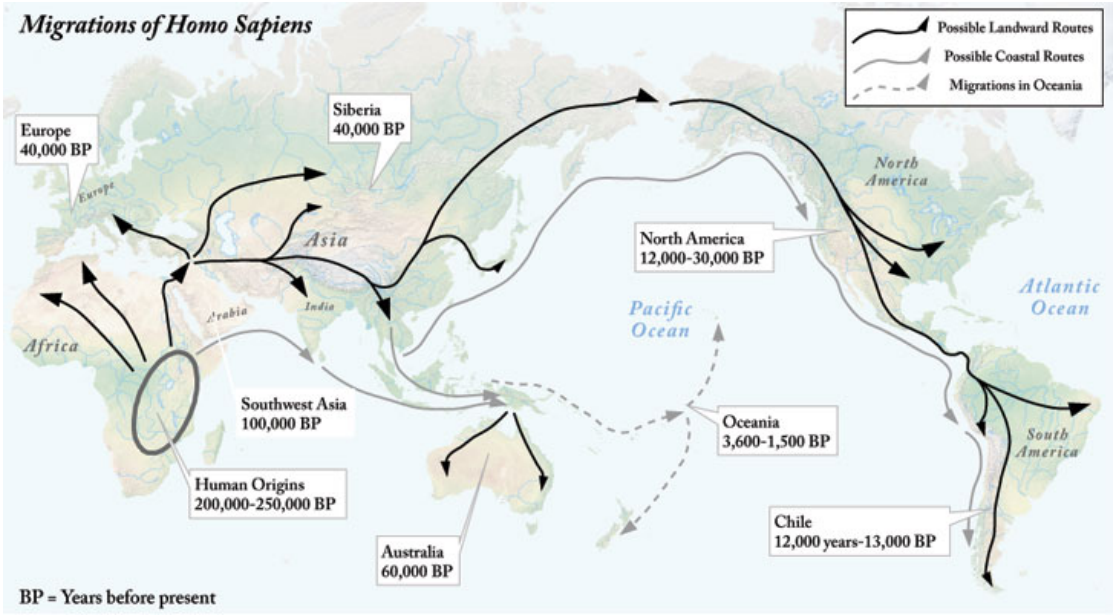


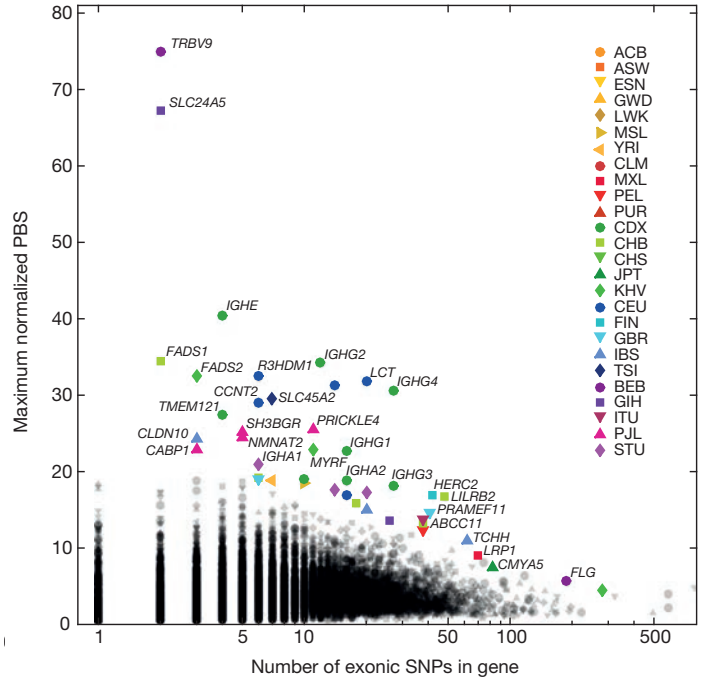
Comparing European-American (CEU) and Nigerian (YRI) population samples



~ fixed differences between populations

Coop et al. 2009; Hernandez et al. 2011





Borrowed from 1000 Genomes paper, 2015

Summary

- Humans have obviously adapted in numerous ways in the last 10 million years or so, as have other primates
- Adaptations sometimes leave a distortionary effect on patterns of genetic variation between species and among individuals. When we look for these distortionary effects, we (implicitly or explicitly) assume a specific model of how natural selection works
- The classic selective sweep model helps us identify a number of interesting cases, such as SLC24A5 and LCT. However, multiple lines of evidence suggest it was not a common mode of human adaptation, at least in recent human evolution
- An alternative is strong selection on individual alleles already present in the population, but there too, we see very few cases of highly differentiated alleles among human populations.
- So where are human adaptations hiding in the genome?