Are humans still evolving?

Lawrence Uricchio
Alison Feder

Course outline

Introduction to human genetics

Using the shell in 5 minutes

How do traits change in prevalence?

Break

Introduction to genomic data

Adaptation for lactase persistence? A single locus example.

Adaptation for height? A multi-locus example

Discussion/Discoveries

Pop QUIZ!!!!

- Roughly what percentage of the sites in two random human genomes will differ at the DNA level?
- Does the answer to the previous question depend on the ancestry (*i.e.,* continental level origin) of the two random individuals?
- How long has it been since humans diverged from Chimpanzees?
- Roughly how much does a human genome differ from a Chimpanzee genome? How about a mouse genome?
- Which ‘evolves’ faster, human language or human DNA?
- Is ‘natural selection’ synonymous with ‘evolution’?

Answers

- Roughly what percentage of the sites in two random human genomes will differ at the DNA level? ~***0.1%***
- Does the answer to the previous question depend on the ancestry (*i.e.,* continental level origin) of the two random individuals? ***Only very weakly***
- How long has it been since humans diverged from Chimpanzees? ***~5-6 million years***
- Roughly how much does a human genome differ from a Chimpanzee genome? How about a mouse genome? ***Chimp: ~1%, Mouse: ~15% (in protein coding regions)***
- Which ‘evolves’ faster, human language or human DNA? ***By almost any measure, language changes waaaaay faster!***
- ***Evolution proceeds even in the absence of natural selection***

Central question & approach

- If we are so similar at the DNA level, why is there such wide variation among people?

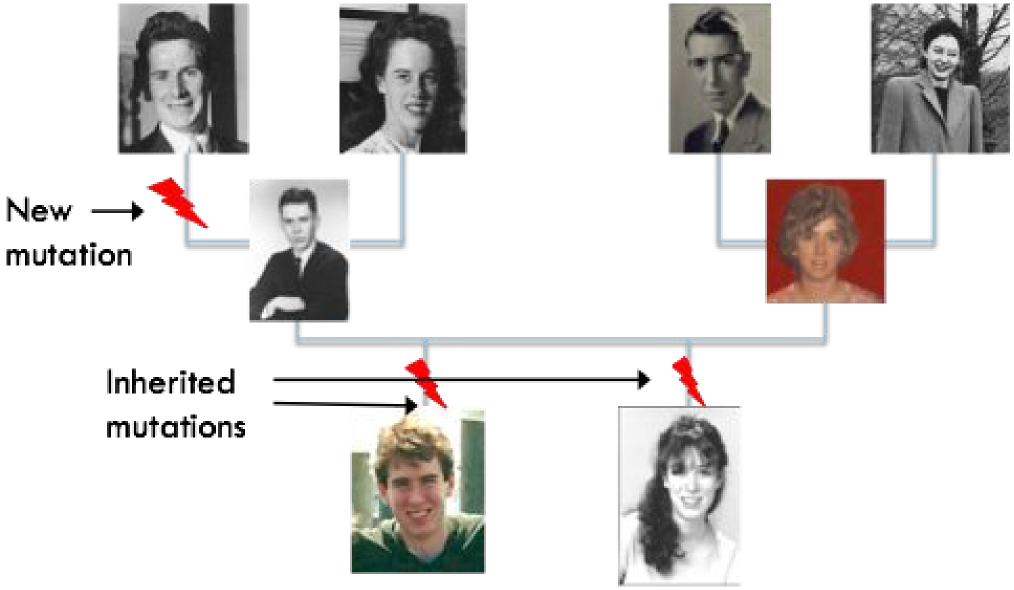
- We will use the techniques of population genetics to address this question, including the development of models, running computer simulations, and measuring patterns of variation in real human DNA sequences

Population genetics

- Population genetics is the study of how relatedness among individuals affects population level genetic variation and genetic traits

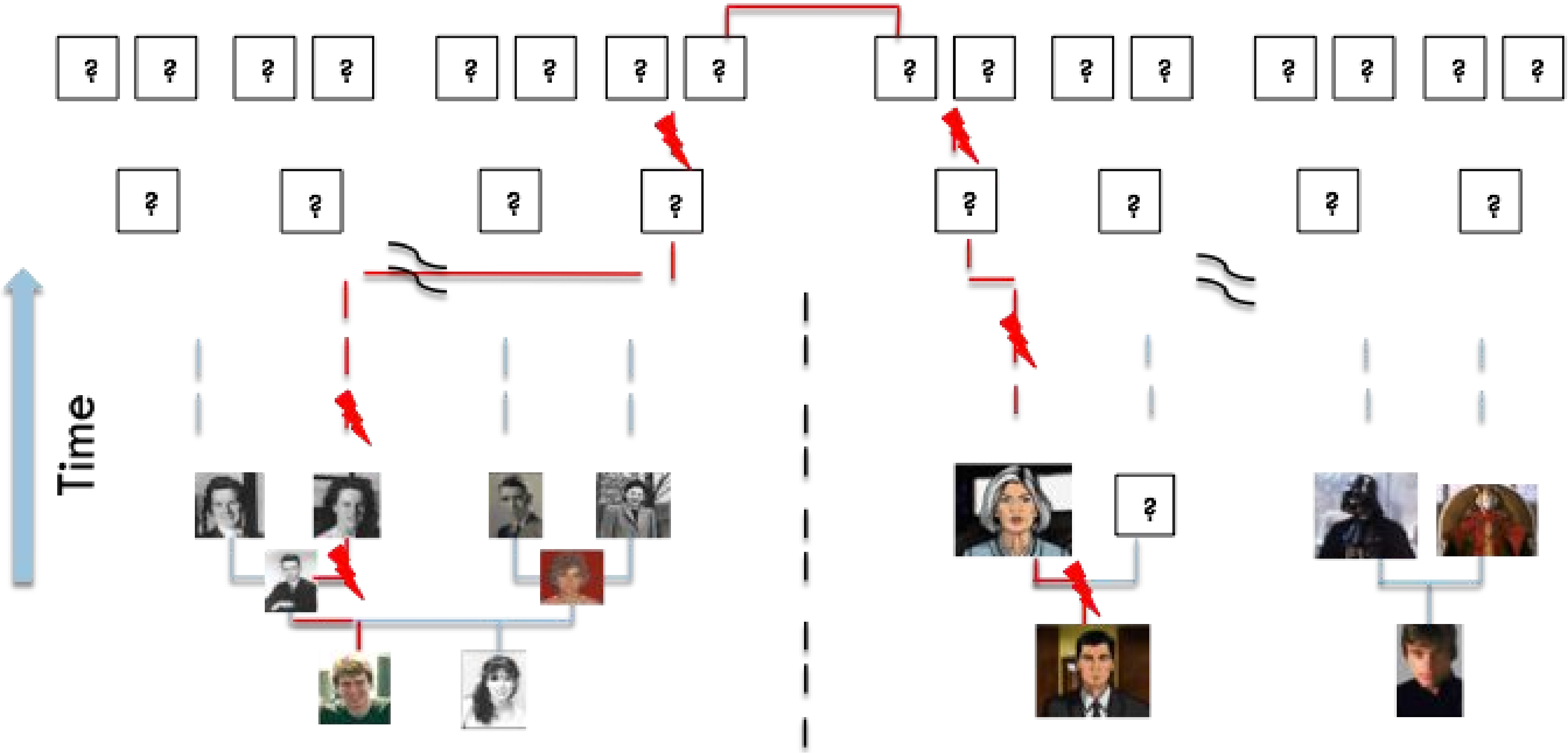
Population genetics

- Population genetics is the study of how relatedness among individuals affects genetic variation and genetic traits



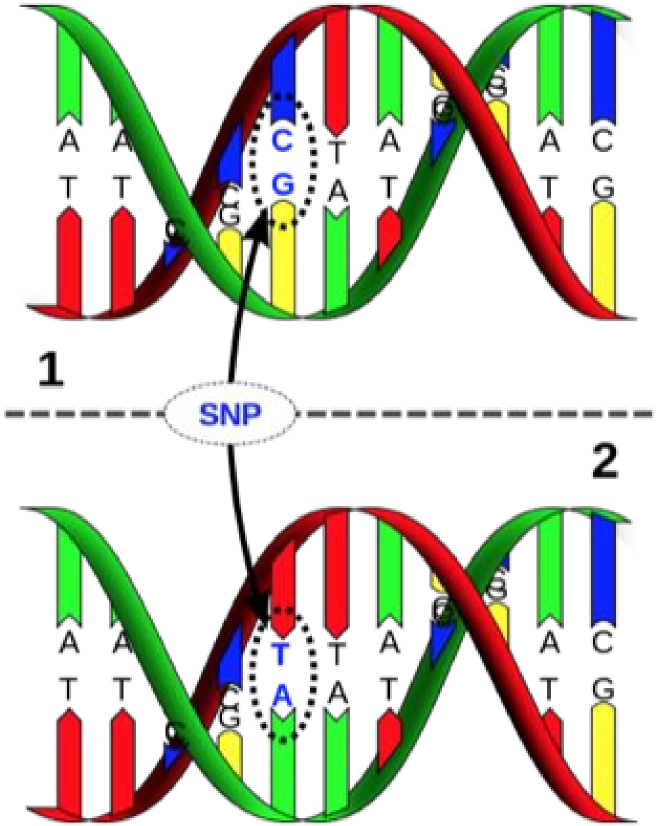
- Mutation rate in humans is ~ 10^{-8}
- With 3 billion base pairs, that means we expect around 30 mutations per generation!

- What if we don't actually know the full pedigree?

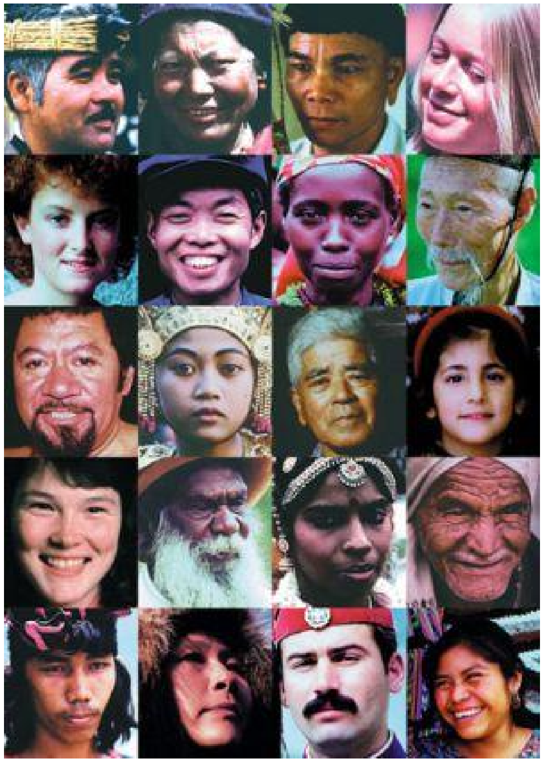


Where does variation come from?

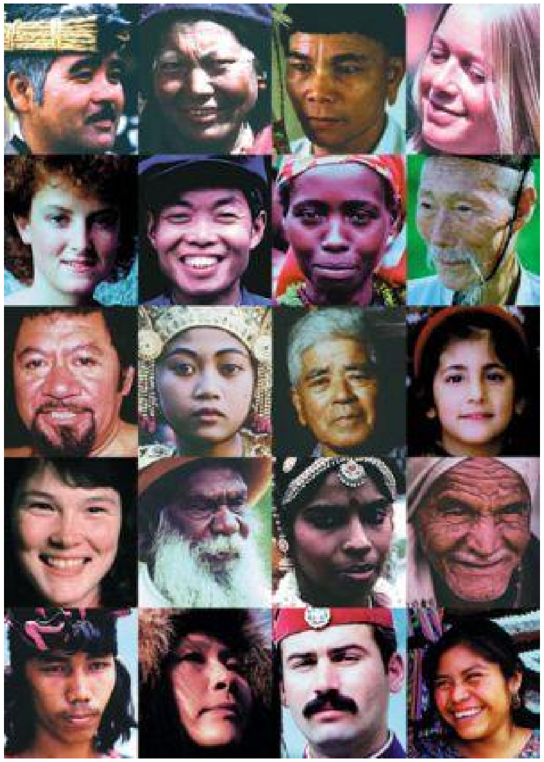
- A SNP is a DNA sequence variant occurring when a single nucleotide in the genome differs between members of a species
- SNPs represent the most fundamental level of variation

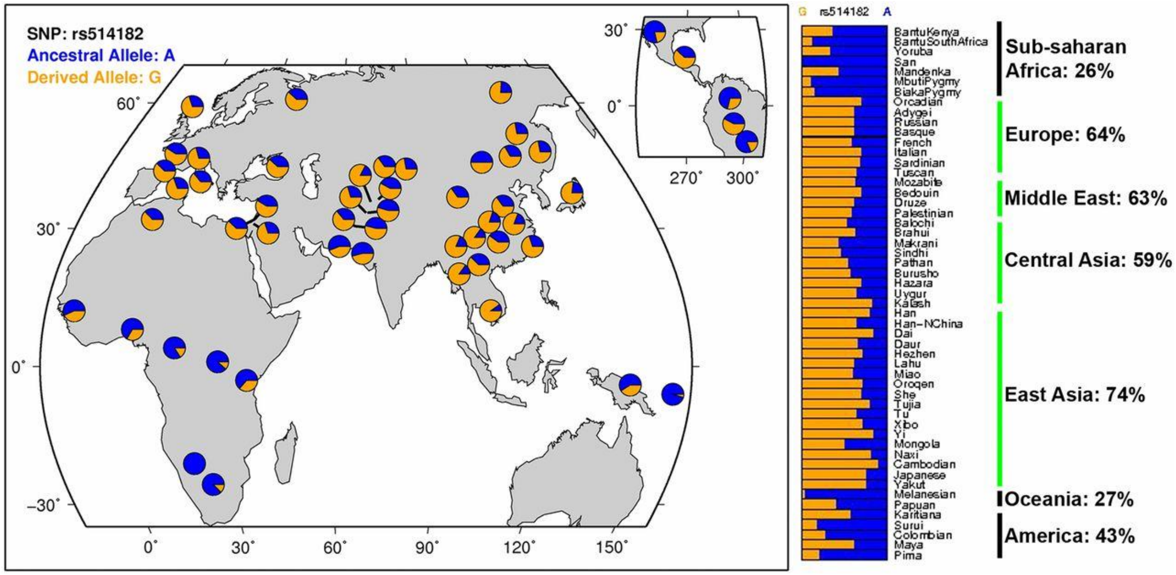


Wide variation in human phenotypes



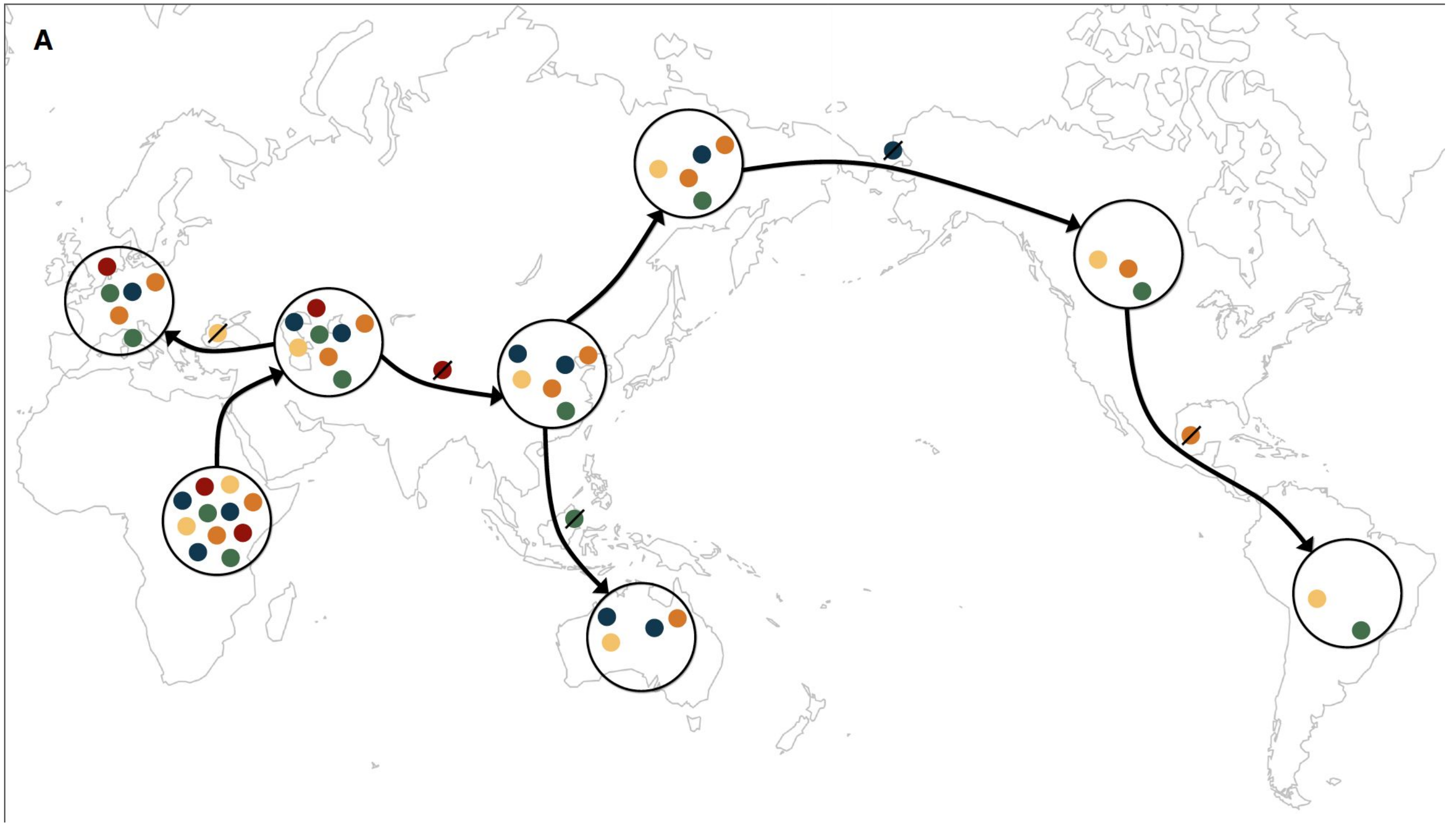
Wide variation in human phenotypes… and genotypes





roughly 0.1% of sites differ between a pair of random human chromosomes!

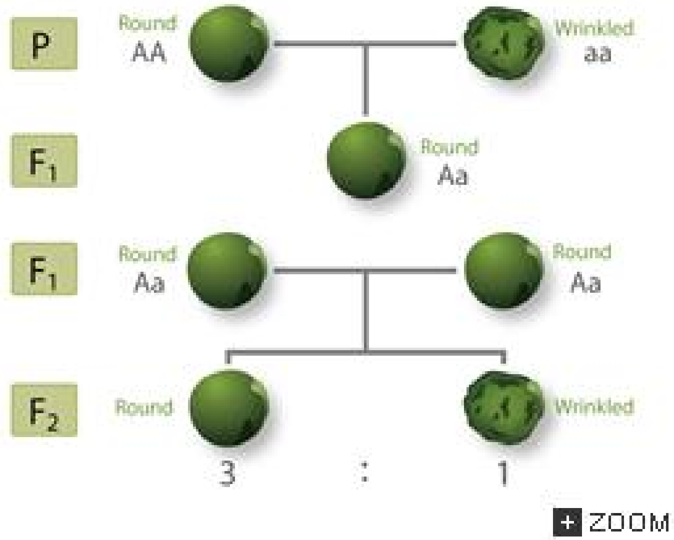
Wide variation in human phenotypes… and genotypes



Kang, 2015

How does genetic variation affect traits?

1) Mendelian traits: Single locus, large effect



1) Complex traits: Many loci, small effect

Why do traits vary between populations?

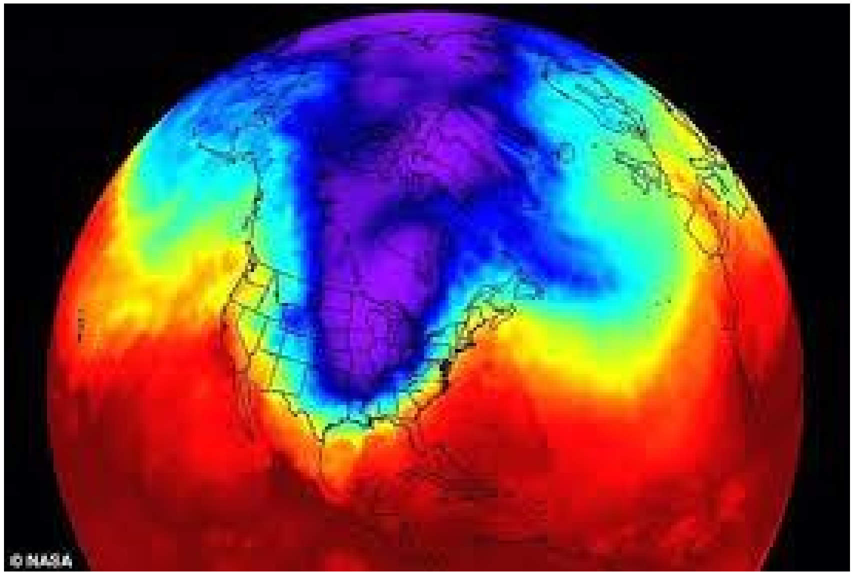
Why do genotypes/traits vary between populations?

- Mutation
- Migration

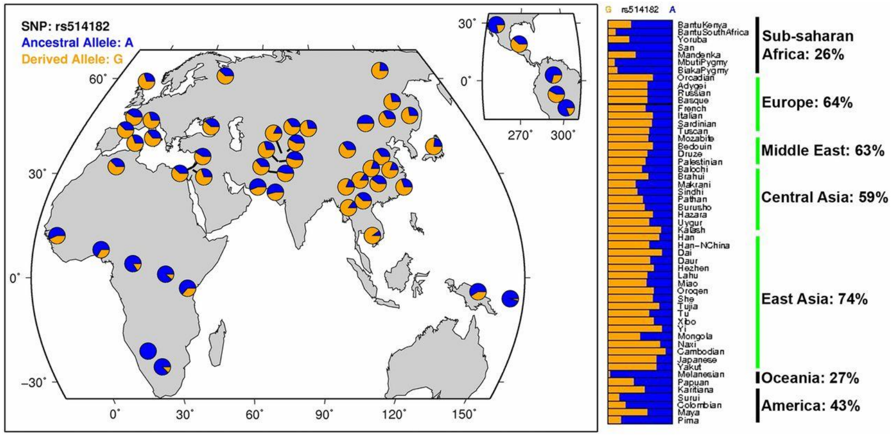
http://cache.gawkerassets.com/assets/images/4/2010/12/spidermanmovieposter.jpg

Why do traits vary between populations?

Environmental variation

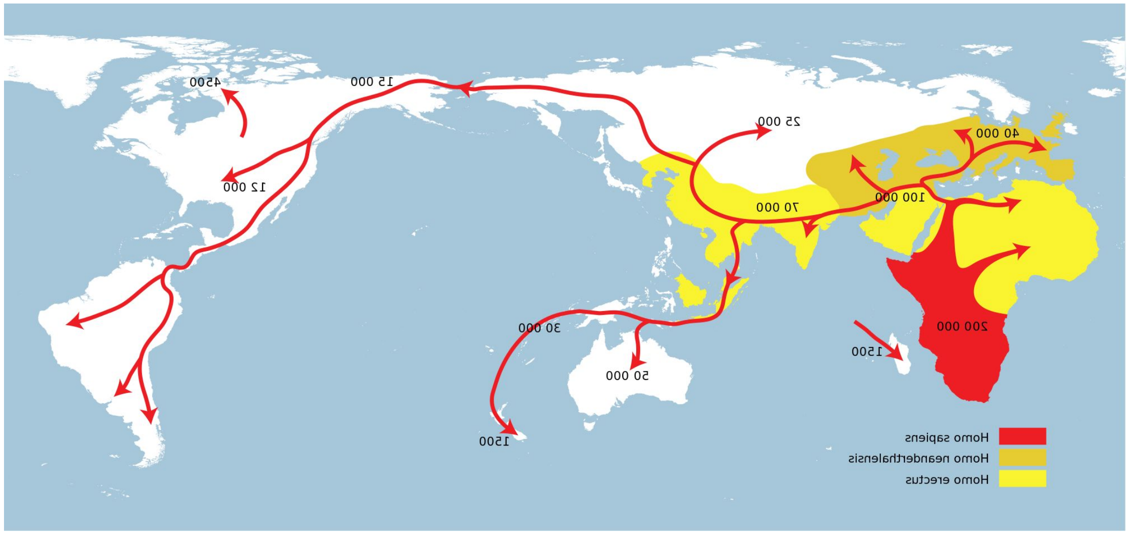


Genetic variation

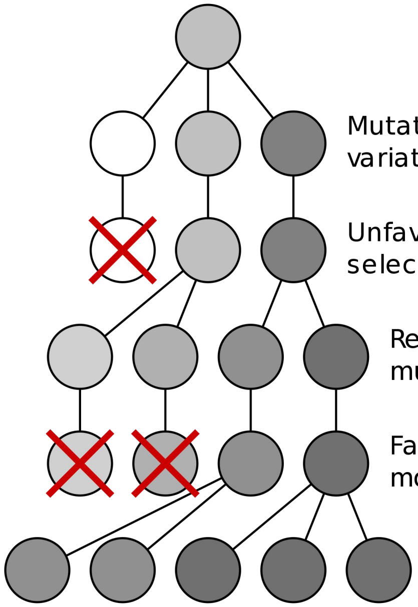


What factors affect genetic variation?

Demography/Drift



Selection



Mutation creates variation

Unfavorable mutations selected against

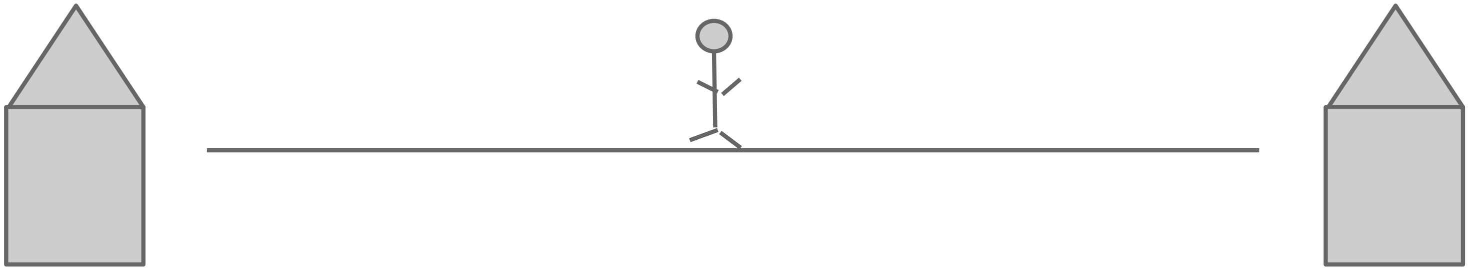
Reproduction and mutation occur

Favorable mutations more likely to survive

... and reproduce

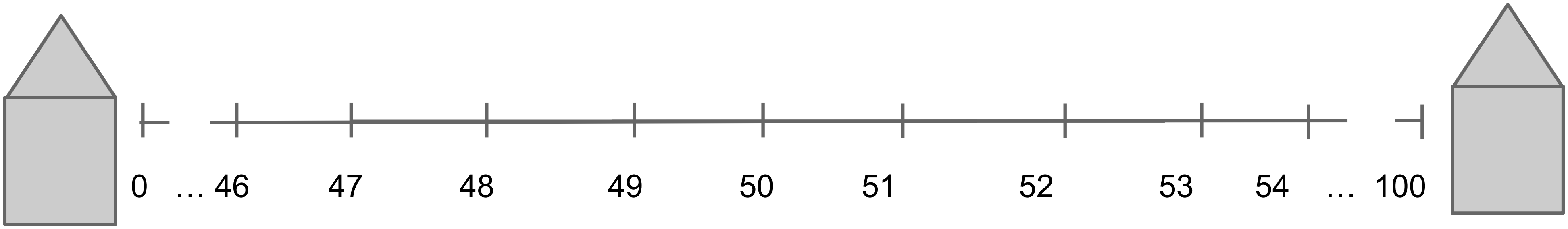
Intuition: A drunkard’s walk

Imagine a drunkard at a location halfway between two bars. Every step he takes has an equal chance of going towards either bar. If he stumbles onto either bar, he will walk into it.



Translation to demonstration cont.

Now imagine if instead of 6 individuals, you had 100 individuals:

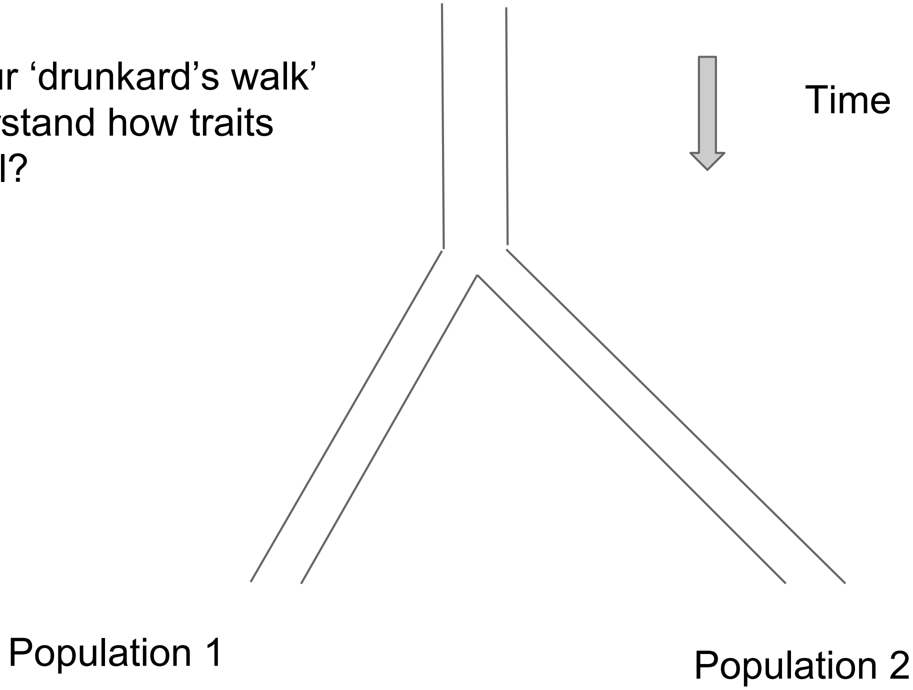


Genetic drift: an experiment

- Break into small group
- Use python (or R/perl/whatever software you like) to simulate the following process:
    - Heads (>= *0.5*): go to right
    - Tails (< *0.5*) : go to left
    - Stop if you hit the end. Record number of steps and side
    - Move barrier/repeat
- Can we say anything about the number of steps before reaching the barrier/which barrier we hit?

Simple model with 2 populations

How can we use our ‘drunkard’s walk’ framework to understand how traits evolve in this model?



Using the shell to run programs

Getting to the command line in osx

- open up the ‘terminal’ application
- you can use the spotlight to find ‘terminal’
- type ‘python’ to enter an interactive session

Getting to the command line in windows

- click ‘start’
- search for ‘command prompt’
- double click
- type ‘python’ in the window

But this is just a toy model

- What other factors should we consider if we were to try to accommodate more complexity?
- How do we hypothesize that these results will change based on these factors?

How do traits change in frequency?

Let’s do some simulations of gentoypes and traits to find out!

```
--[harrisk: sim_code]$ python traj.py

This software simulates trajectories of alles under selection in divergent populations

usage: python traj.py <Size of population 1> <Size of population 2> <Number of simulated alleles> <total time in coalescent units> <fraction_selected>
```

Course outline

Introduction to human genetics

Using the shell in 5 minutes

How do traits change in frequency?

Break

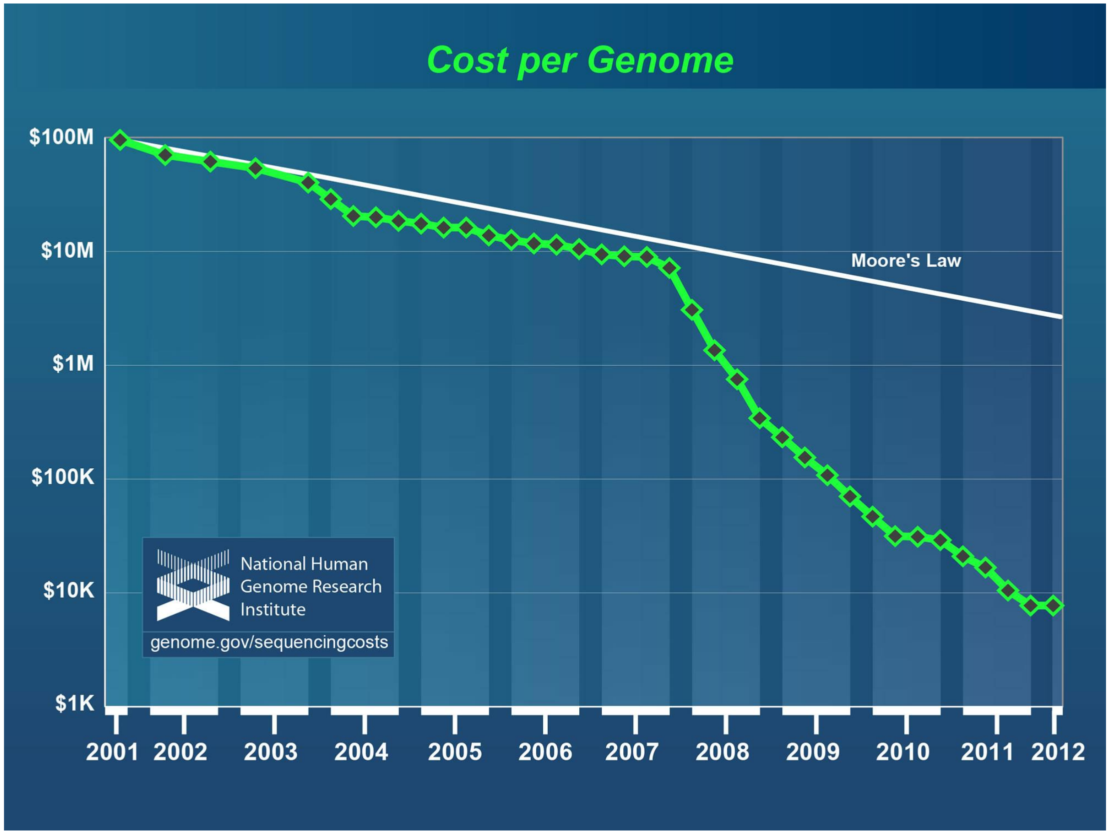
Introduction to genomic data

Adaptation for lactase persistence? A single locus example.

Adaptation for height? A multi-locus example

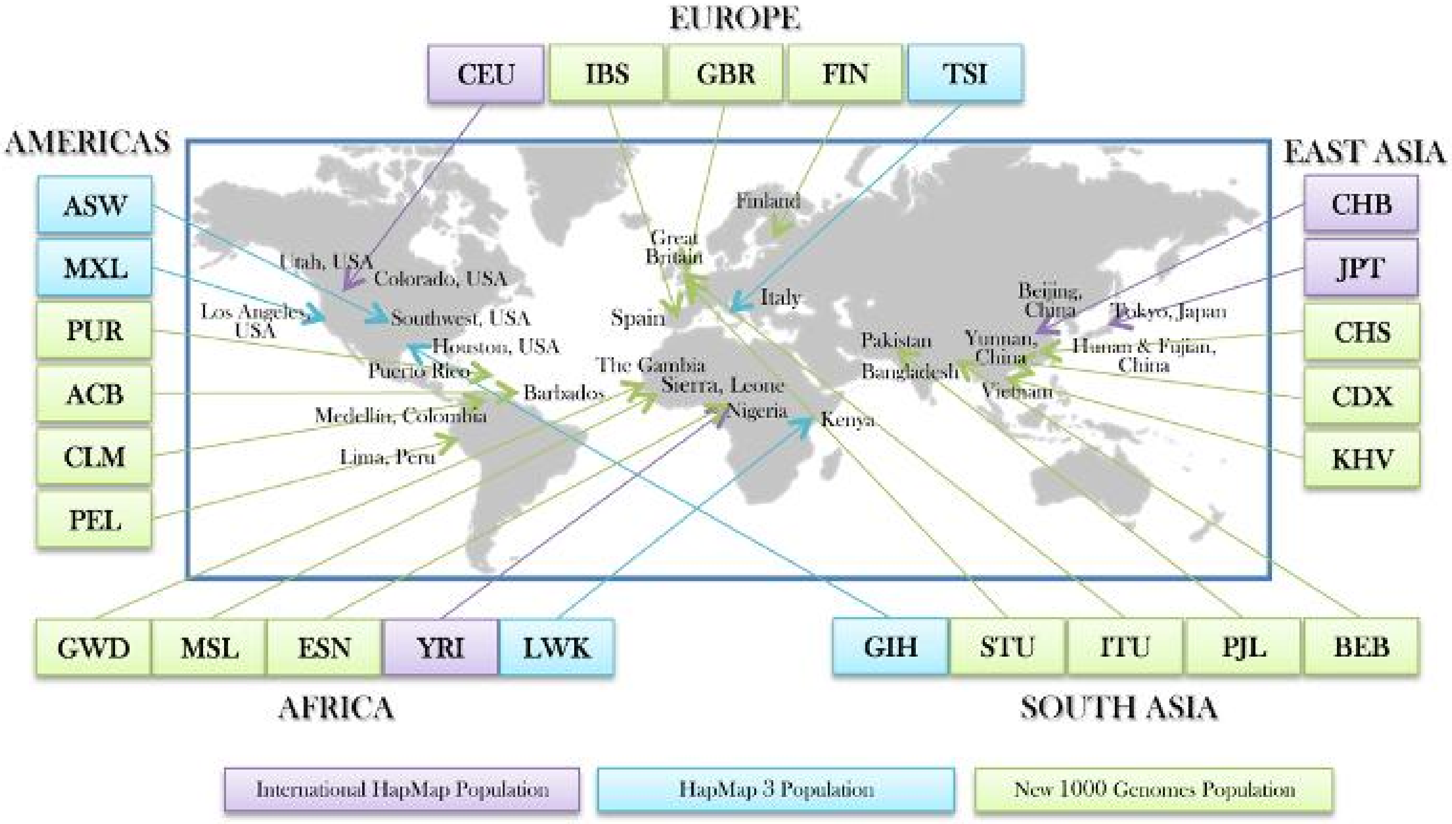
Discussion/Discoveries

Cost of sequencing

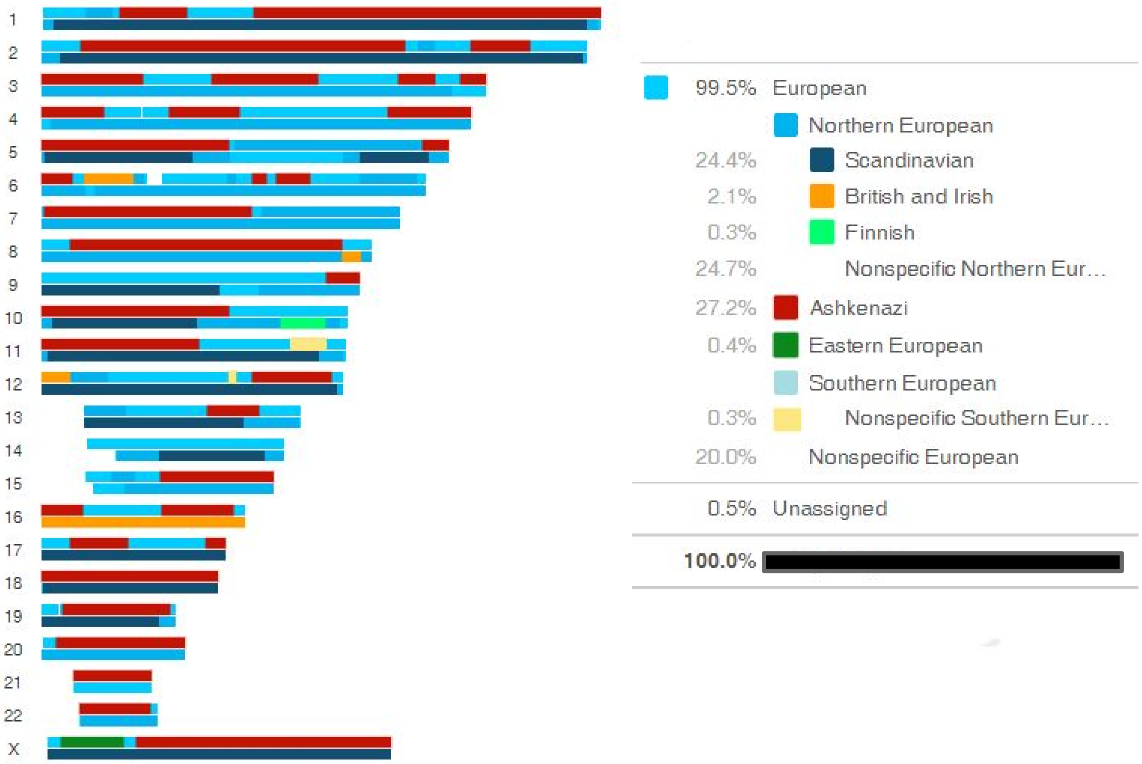


https://en.wikipedia.org/wiki/File:Genome_sequencing_costs,_May_2012.jpg

Introduction to genomic data

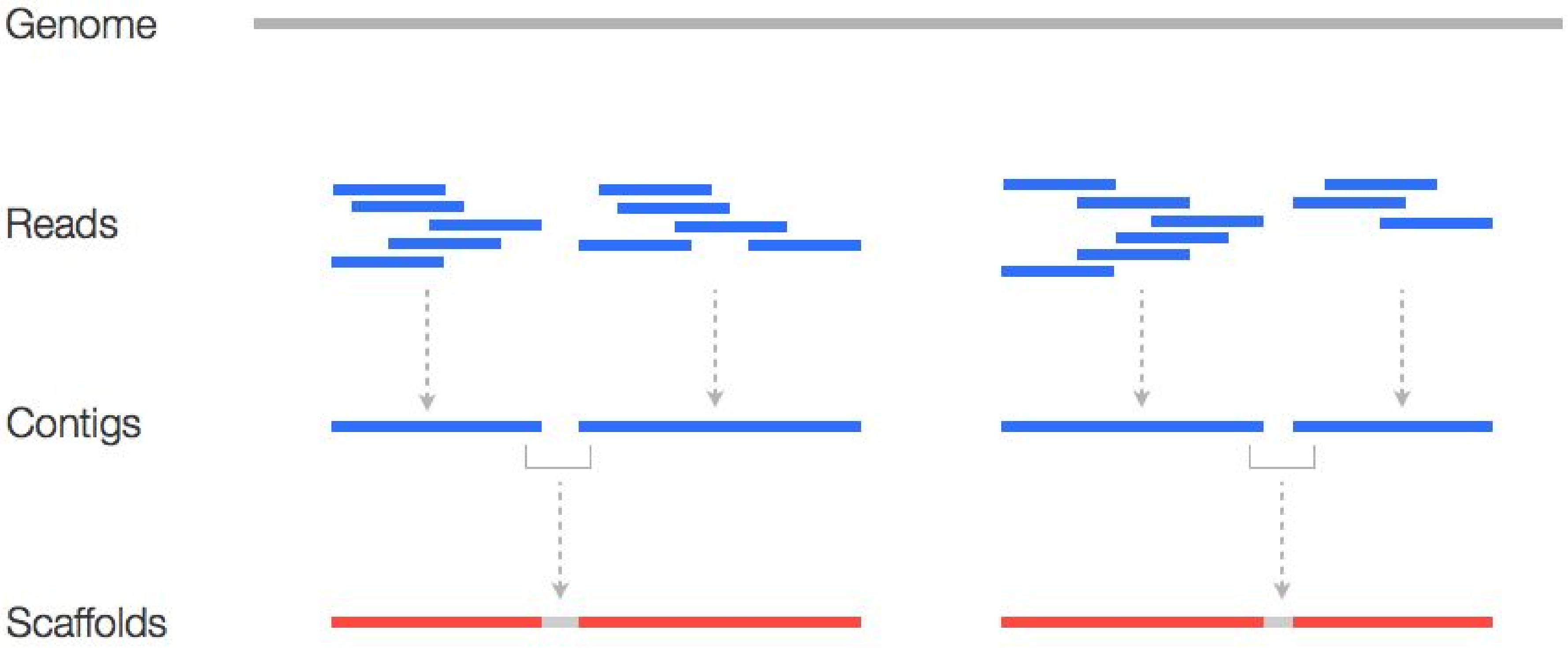


23 & Me



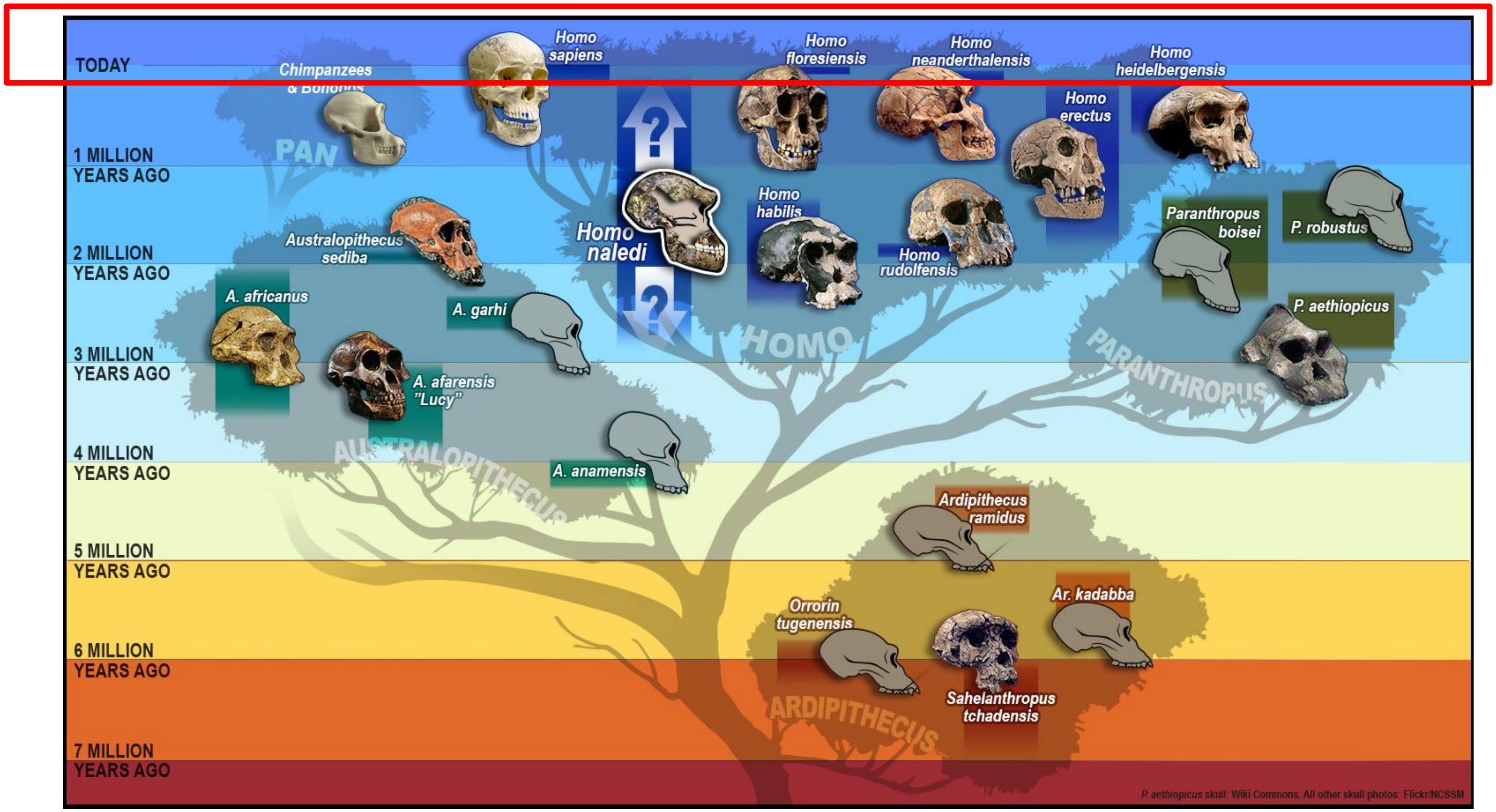
🟦	99.5%	European
	🟦	Northern European
24.4%	🟦	Scandinavian
2.1%	🟧	British and Irish
0.3%	🟩	Finnish
24.7%		Nonspecific Northern Eur…
27.2%	🟥	Ashkenazi
0.4%	🟩	Eastern European
	🟦	Southern European
0.3%	🟨	Nonspecific Southern Eur…
20.0%		Nonspecific European
0.5%		Unassigned
100.0%		

What does sequencing data look like?



http://ecoevo.unit.oist.jp/lab/wp-content/uploads/2013/08/GenomeAssembly.png

Snapshot of data



P. aethiopicus skull: Wiki Commons. All other skull photos: Flickr/NCSSM

How can you test for selection if you only have a snapshot?

What is the genetic basis of traits?

		SNP1	SNP2	…
People with disease	ATCCCA... ATCCCG... ATCGCA... ATCCCG... ...	750/1000 have C 250/1000 have G C has frequency 75%	493/1000 have A 507/1000 have G Both A and G have frequency approximately 50%	
Healthy People	ATCGCG... ATCGCG... ATCGCA... ATCGCA... ...	10/1000 have C 990/1000 have G C has frequency 1%	505/1000 have A 495/1000 have G Both A and G have frequency approximately 50%	

GWAS for complex diseases

Disease	Number of Genes	% of variation that is genetic	% of genetic variation explained by known genes
Alzheimer's	4	79	23.2
Bipolar disorder	5	77	2.8
Breast Cancer	13	53	12.5
Coronary Artery Disease	12	49	25.2
Crohn’s	32	55	13.4
Prostate Cancer	27	50	31.2
Schizophrenia	4	81	0.4
Systemic lupus erythematosus	23	66	13.2
Type I Diabetes	45	80	13.6
Type II Diabetes	25	42	27.9

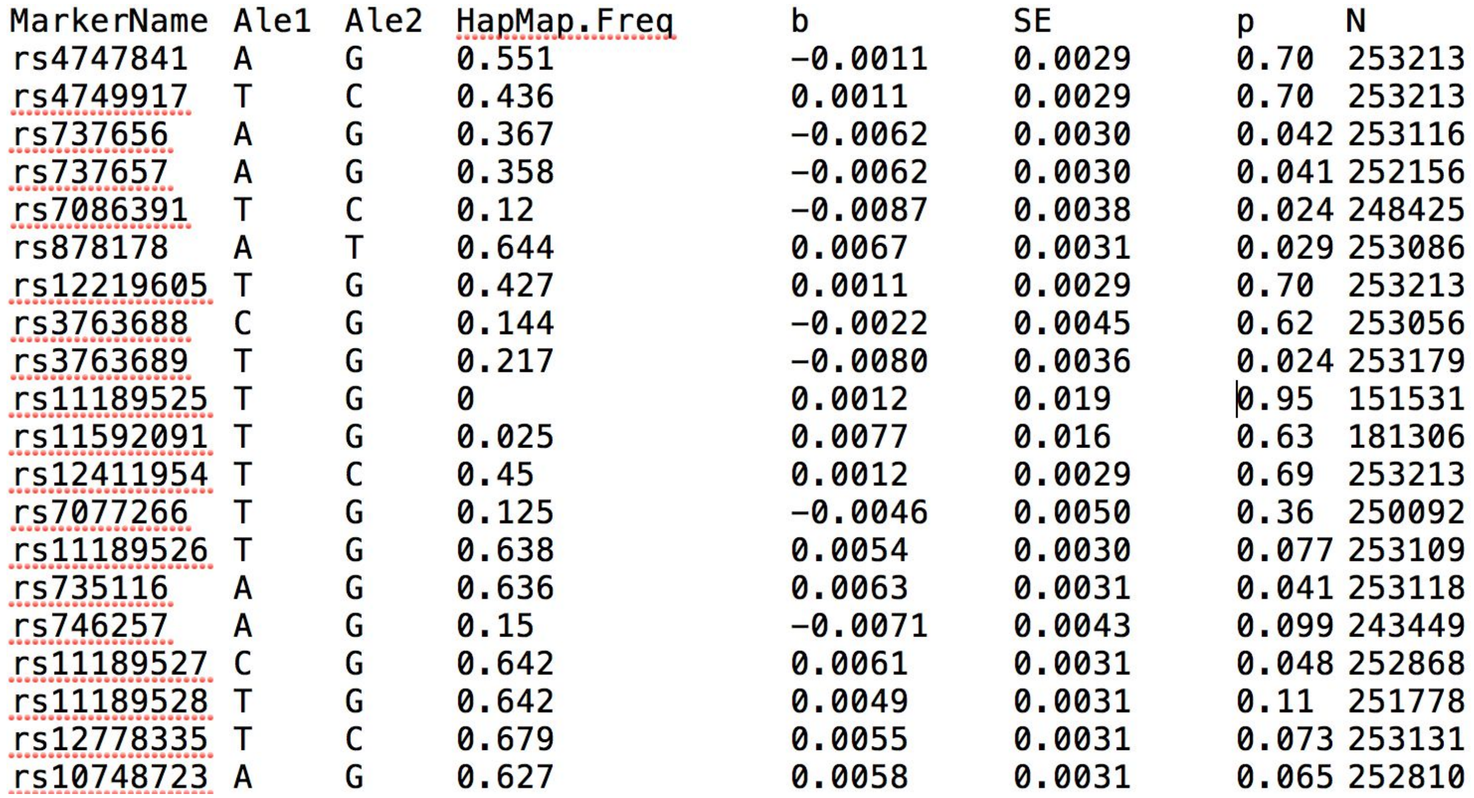
Selection on human height

Real human sequence data!

GBR

TSI

GIANT: Genetic Investigation of ANthropometric Traits



MarkerName	Ale1	Ale2	HapMap.Freq	b	SE	p	N
rs4747841	A	G	0.551	-0.0011	0.0029	0.70	253213
rs4749917	T	C	0.436	0.0011	0.0029	0.70	253213
rs737656	A	G	0.367	-0.0062	0.0030	0.042	253116
rs737657	A	G	0.358	-0.0062	0.0030	0.041	252156
rs7086391	T	C	0.12	-0.0087	0.0038	0.024	248425
rs878178	A	T	0.644	0.0067	0.0031	0.029	253086
rs12219605	T	G	0.427	0.0011	0.0029	0.70	253213
rs3763688	C	G	0.144	-0.0022	0.0045	0.62	253056
rs3763689	T	G	0.217	-0.0080	0.0036	0.024	253179
rs11189525	T	G	0	0.0012	0.019	0.95	151531
rs11592091	T	G	0.025	0.0077	0.016	0.63	181306
rs12411954	T	C	0.45	0.0012	0.0029	0.69	253213
rs7077266	T	G	0.125	-0.0046	0.0050	0.36	250092
rs11189526	T	G	0.638	0.0054	0.0030	0.077	253109
rs735116	A	G	0.636	0.0063	0.0031	0.041	253118
rs746257	A	G	0.15	-0.0071	0.0043	0.099	243449
rs11189527	C	G	0.642	0.0061	0.0031	0.048	252868
rs11189528	T	G	0.642	0.0049	0.0031	0.11	251778
rs12778335	T	C	0.679	0.0055	0.0031	0.073	253131
rs10748723	A	G	0.627	0.0058	0.0031	0.065	252810

Evidence for selection on height?

```
> python rsid.browse.py [options] rsidfile.txt

Check the rsidfiles/ directory (i.e.)
rsidfiles/rsidfile.set.1.1000.txt

-hist
Prints the mean of GBR frequency - TSI frequency
Plots a histogram of GBR frequency - TSI frequency for all SNPs
```

How big do frequency differences need to be in order to be considered “real”?

Evidence for selection on height?

```
> python rsid.browse.py [options] rsidfile.txt
```

Check the rsidfiles/ directory (i.e.)
rsidfiles/rsidfile.set.1.1000.txt

-hist
Prints the mean of GBR frequency - TSI frequency
Plots a histogram of GBR frequency - TSI frequency for all SNPs

-n
Performs **x** controls of GBR frequency - TSI frequency where the up and downregulating alleles are randomized
Plots a histogram of the **x** randomized GBR frequency - TSI frequency means

-randnum [**x**]
sets **x** (the number of randomized means)

Evidence for selection on height

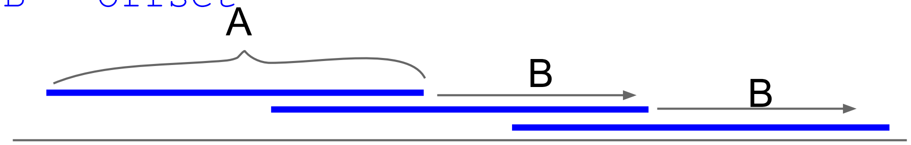
```
> python rsid.browse.py [options] rsidfile.txt
```

Check the rsidfiles/ directory (i.e.)
rsidfiles/rsidfile.set.1.1000.txt

-p

Plot the running mean of GBR-TSI differences

-windsize [A] [B]
Sets parameters for the running mean of window sizes
A = window size
B = offset



Other human traits under selection

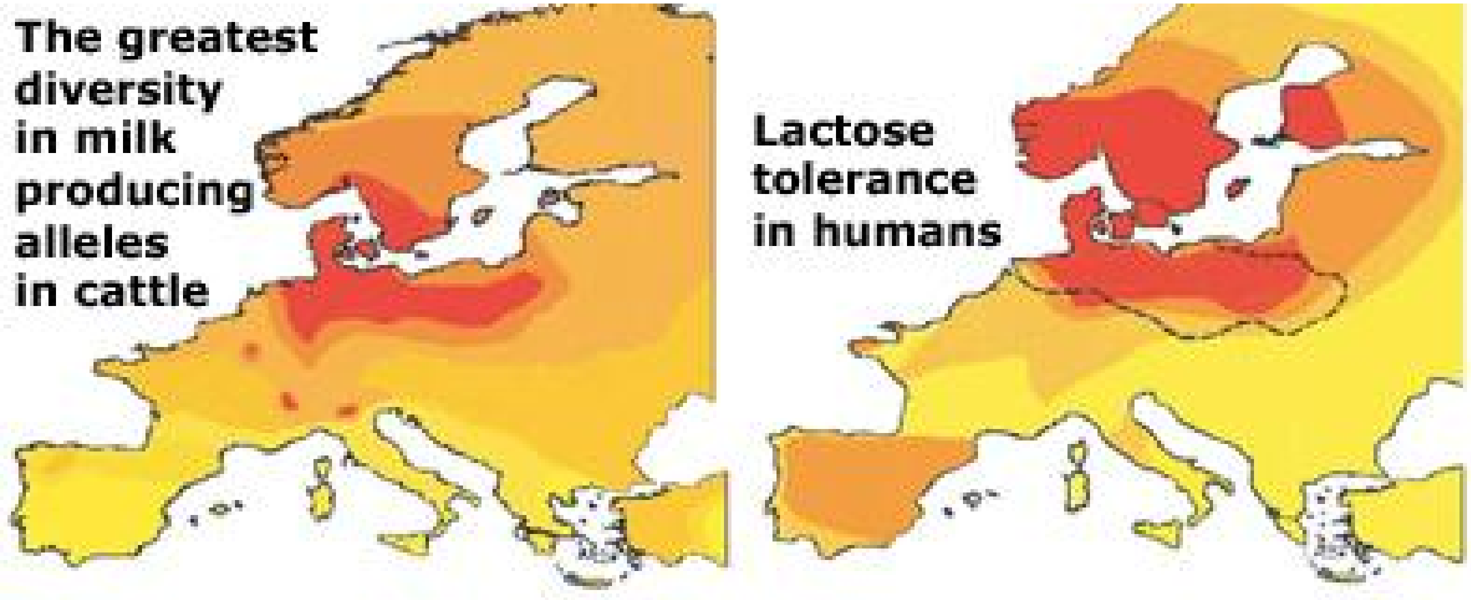
Lactose metabolism -- *strong evidence for selection on a small number of loci that confer the lactase persistence trait*

Immune function -- *immune genes are categorically enriched in many studies of the rate of evolution in human genes*

Olfaction -- *humans have a weaker sense of smell than many other closely related mammals, which may represent a relaxation of negative and/or positive selection on these genes*

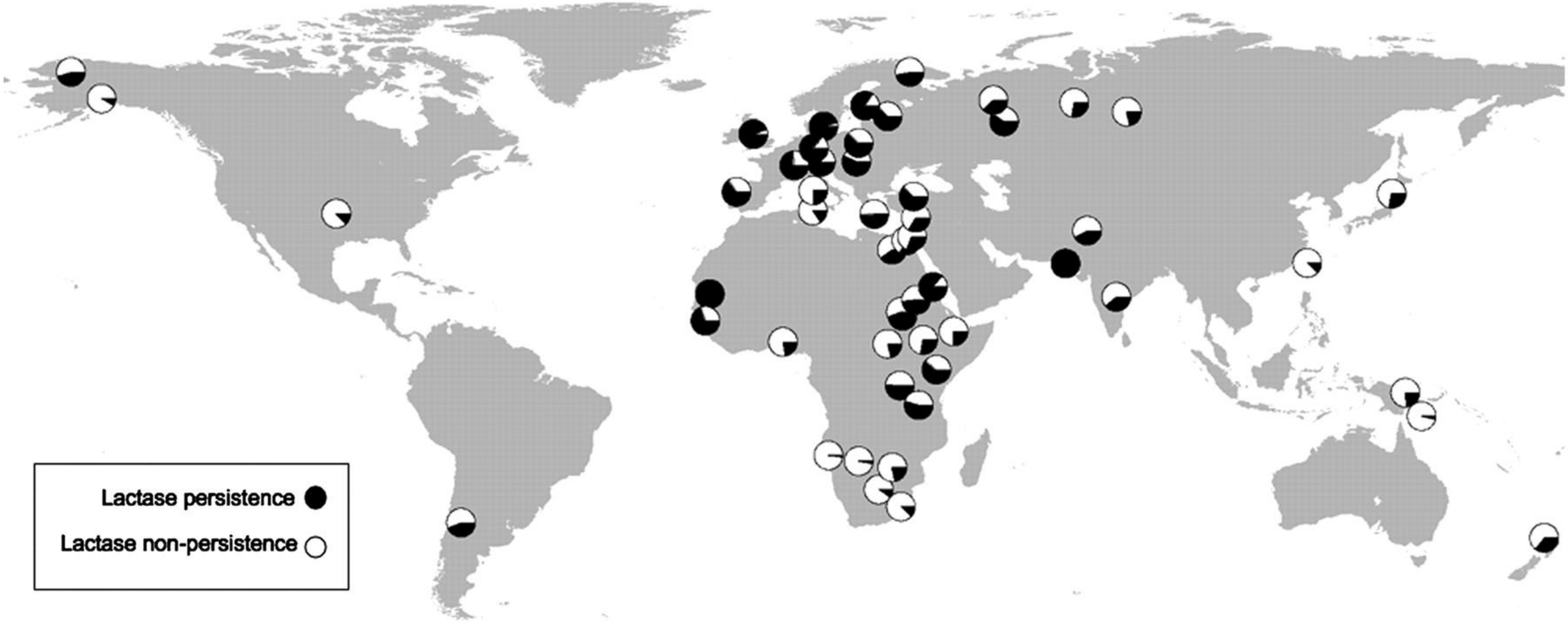
Transcription factors -- *many transcription factors have signatures of strong and recent selection* (Arbiza et al, Nature Genetics, 2013)

Human lactase persistence



Razib Khan’s modification of figure 1 from Beja-Pereira et al, Gene-culture coevolution between cattle milk protein genes and human lactase genes, Nature Genetics 35, 311 - 313 (2003).

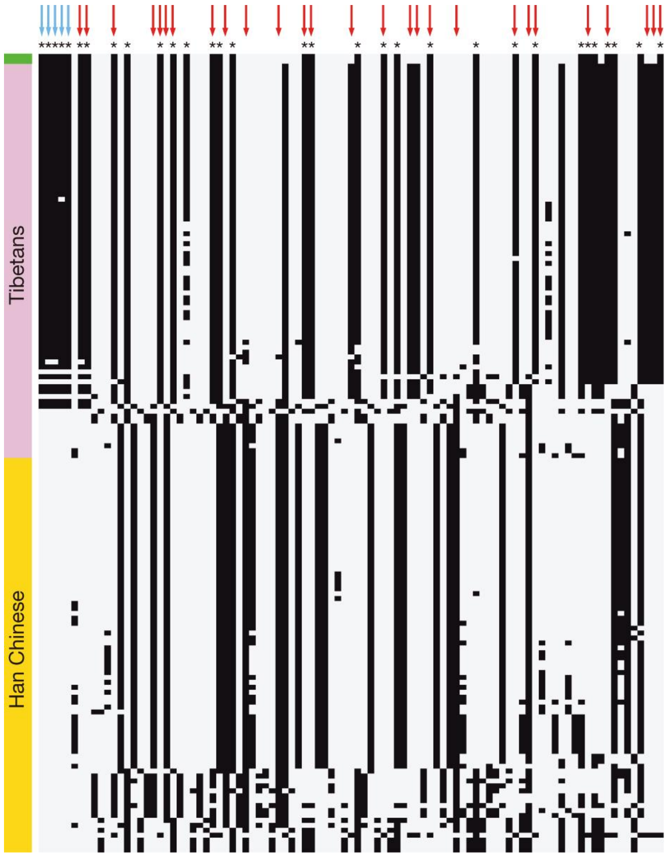
Human lactase persistence



Ingram et al, Hum Genet 124:579–591, 2009

High altitude adaptation

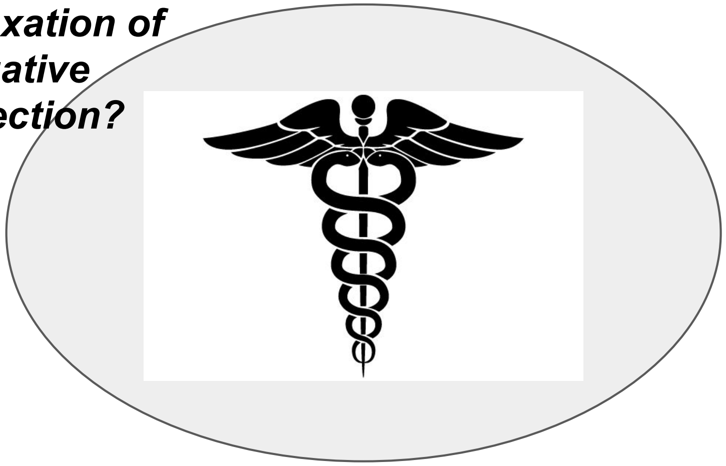
Denisovans ->



So… are we still evolving?

So… are we still evolving?

relaxation of negative selection?



Urban social /economic pressures?

limited resources and pollution?

The answer is an unequivocal ***yes***, but how will environmental pressures alter our evolutionary trajectory?