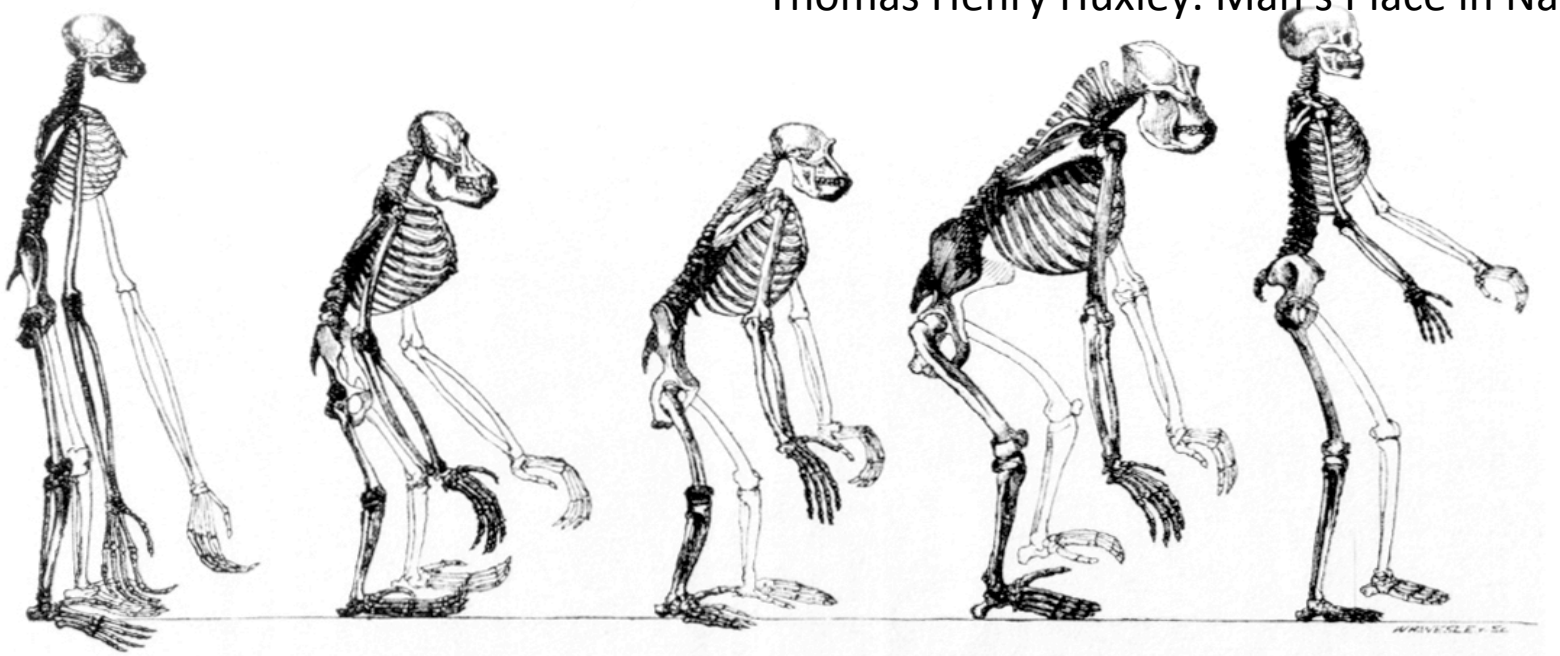
Population genetics

Part 1

@graham_coop
www.gcbias.org

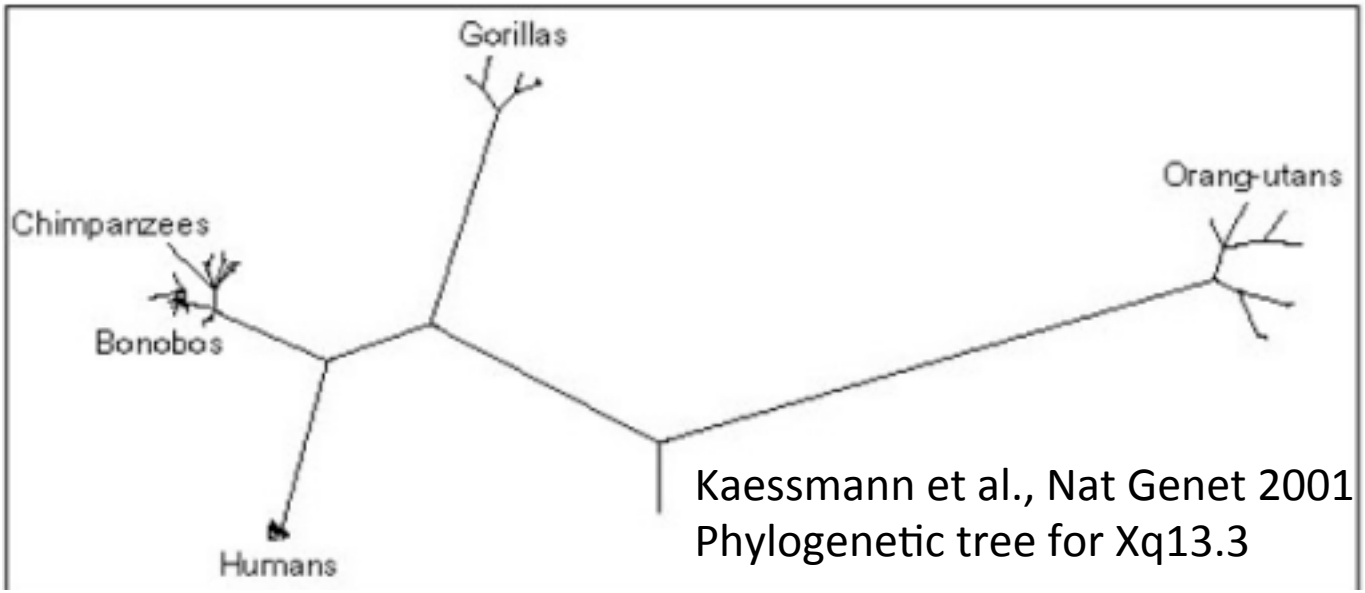
Thomas Henry Huxley. Man's Place in Nature (1863):



Skeletons of the

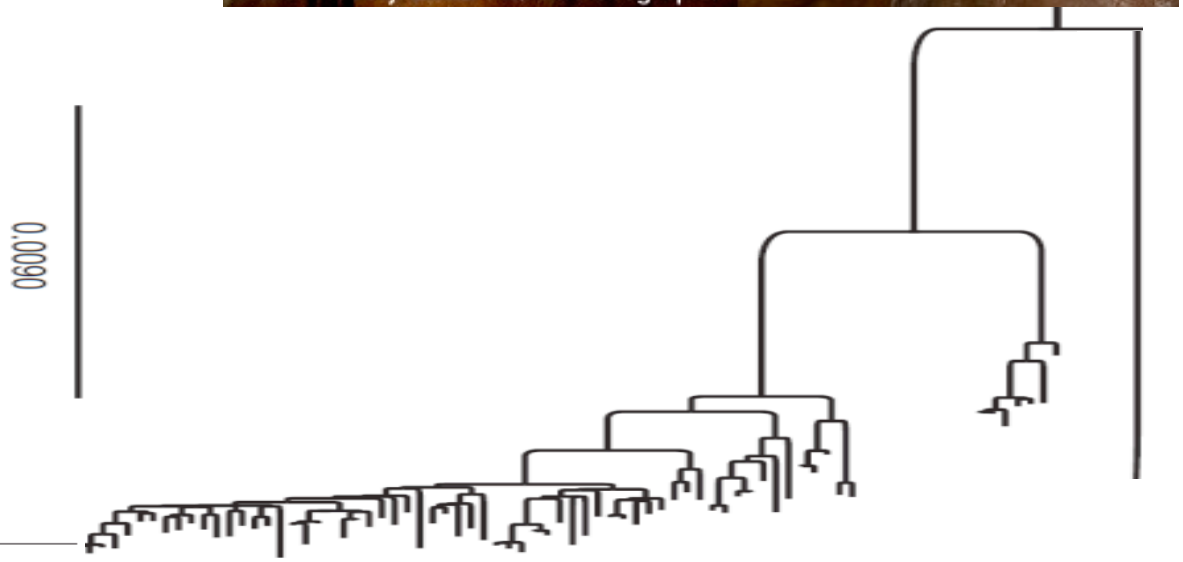
GIBBON. ORANG. CHIMPANZEE. GORILLA. MAN.

Photographically reduced from Diagrams of the natural size (except that of the Gibbon, which was twice as large as nature), drawn by Mr. Waterhouse Hawkins from specimens in the Museum of the Royal College of Surgeons.



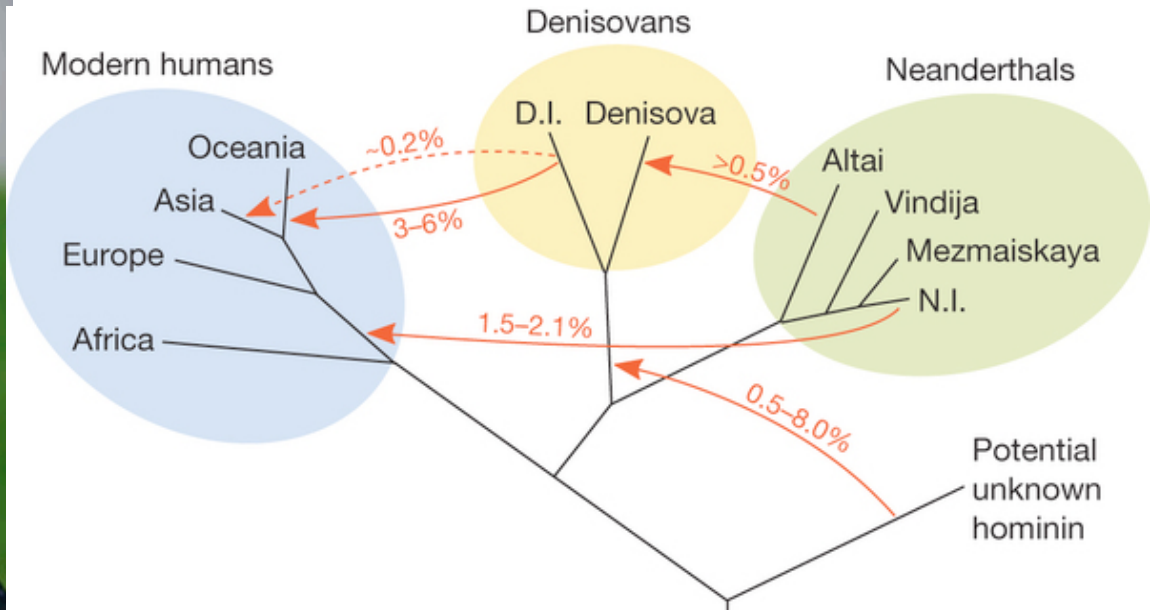
Kaessmann et al., Nat Genet 2001
Phylogenetic tree for Xq13.3

Neanderthal
by Thomas Henry Huxley



The complete mitochondrial DNA genome of an unknown hominin from southern Siberia

Johannes Krause[1], Qiaomei Fu[1], Jeffrey M. Good[2], Bence Viola[1,3], Michael V. Shunkov[4], Anatoli P. Derevianko[4] & Svante Pääbo[1]



http://www.nature.com/nature/journal/v505/n7481/full/nature12886.html

The complete genome sequence of a Neanderthal from the Altai Mountains

Kay Prüfer, Fernando Racimo, Nick Patterson, Flora Jay, Sriram Sankararaman, Susanna Sawyer, Anja Heinze, Gabriel Renaud, Peter H. Sudmant, Cesare de Filippo, Heng Li, Swapan Mallick, Michael Dannemann, Qiaomei Fu, Martin Kircher, Martin Kuhlwilm, Michael Lachmann, Matthias Meyer, Matthias Ongyerth, Michael Siebauer, Christoph Theunert, Arti Tandon, Priya Moorjani, Joseph Pickrell, James C. Mullikin + *et al.*

The Strength of Selection Against Neanderthal Introgression

Ivan Juric[1,2,¤a,*], Simon Aeschbacher[1,2,‡], Graham Coop[1,2,‡]

ARTICLE

OPEN
doi:10.1038/nature19057

Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek[1,2,3,4], Konrad J. Karczewski[1,2*], Eric V. Minikel[1,2,5*], Kaitlin E. Samocha[1,2,5,6*], Eric Banks[2], Timothy Fennell[2], Anne H. O'Donnell-Luria[1,2,7], James S. Ware[2,8,9,10,11], Andrew J. Hill[1,2,12], Beryl B. Cummings[1,2,5], Taru Tukiainen[1,2], Daniel P. Birnbaum[2], Jack A. Kosmicki[1,2,6,13], Laramie E. Duncan[1,2,6], Karol Estrada[1,2], Fengmei Zhao[1,2], James Zou[2], Emma Pierce-Hoffman[1,2], Joanne Berghout[14,15], David N. Cooper[16], Nicole Deflaux[17], Mark DePristo[18], Ron Do[19,20,21,22], Jason Flannick[2,23], Menachem Fromer[1,6,19,20,24], Laura Gauthier[18], Jackie Goldstein[1,2,6], Namrata Gupta[2], Daniel Howrigan[1,2,6], Adam Kiezun[18], Mitja I. Kurki[2,25], Ami Levy Moonshine[18], Pradeep Natarajan[2,26,27,28], Lorena Orozco[29], Gina M. Peloso[2,27,28], Ryan Poplin[18], Manuel A. Rivas[2], Valentin Ruano-Rubio[18], Samuel A. Rose[6], Douglas M. Ruderfer[19,20,24], Khalid Shakir[18], Peter D. Stenson[16], Christine Stevens[2], Brett P. Thomas[1,2], Grace Tiao[18], Maria T. Tusie-Luna[30], Ben Weisburd[2], Hong-Hee Won[31], Dongmei Yu[6,25,27,32], David M. Altshuler[2,33], Diego Ardissino[34], Michael Boehnke[35], John Danesh[36], Stacey Donnelly[2], Roberto Elosua[37], Jose C. Florez[2,26,27], Stacey B. Gabriel[2], Gad Getz[18,26,38], Stephen J. Glatt[39,40,41], Christina M. Hultman[42], Sekar Kathiresan[2,26,27,28], Markku Laakso[43], Steven McCarroll[6,8], Mark I. McCarthy[44,45,46], Dermot McGovern[47], Rutl ish Saleheen[50,51,52], Jere hto[57], Ming T. Tsuang[58], Hug ggregation Consortium†

OPEN ACCESS Freely available online

PLOS | BIOLOGY

The Geography of Recent Genetic Ancestry across Europe

Peter Ralph*¤, Graham Coop*

Department of Evolution and Ecology & Center for Population Biology, University of California, Davis, California, United States of America

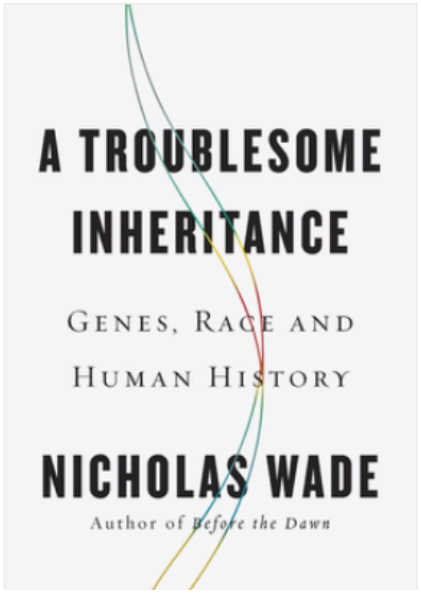
Geneticists say popular book misrepresents research on human evolution

08 Aug 2014 | 20:27 BST | Posted by Ewen Callaway | Category: Anthropology, Evolution

More than 130 leading population geneticists have condemned a book arguing that genetic variation between human populations could underlie global economic, political and social differences.

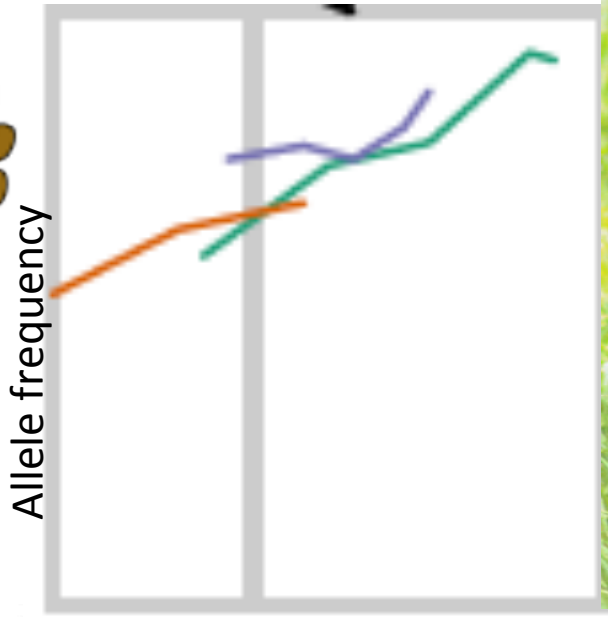
A Troublesome Inheritance, by science journalist Nicholas Wade, was published in June by Penguin Press in New York. The 278-page work garnered widespread criticism, much of it from scientists, for suggesting that genetic differences (rather than culture) explain, for instance, why Western governments are more stable than those in African countries. Wade is former staff reporter and editor at the *New York Times*, *Science* and *Nature*.

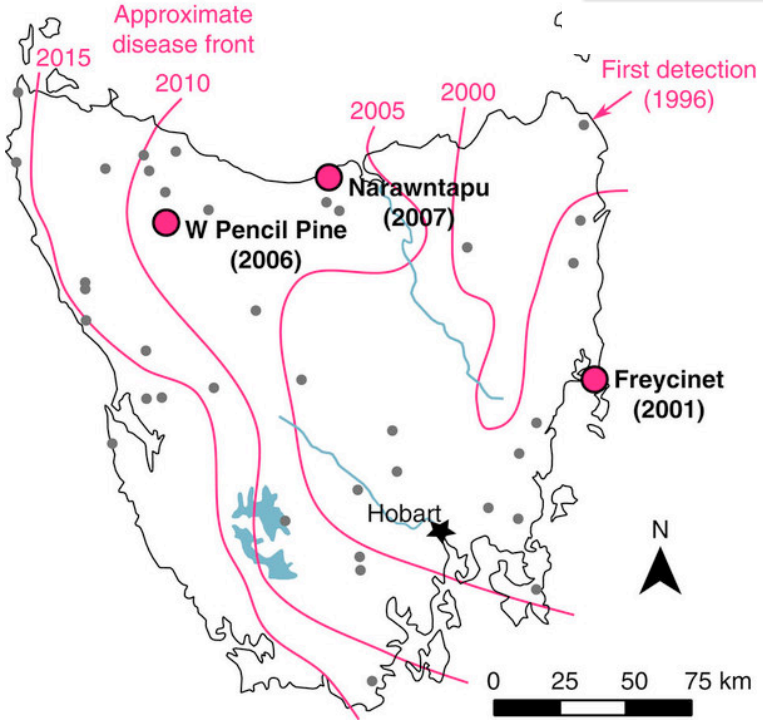
But the letter — signed by a who's-who of researchers in population genetics and human evolution — and published in the 10 August issue of the *New York Times* — represents a rare unified statement from scientists in the field and includes many whose work was cited by Wade. "It's just a measure of



https://blog.23andme.com/23andme-and-you/23andme-how-to/ancestry-new-insights-for-sheridan/

Sympathy for the Devil?





nature COMMUNICATIONS

ARTICLE

Received 31 Mar 2016 | Accepted 20 Jul 2016 | Published 30 Aug 2016

DOI: 10.1038/ncomms12684 OPEN

Rapid evolutionary response to a transmissible cancer in Tasmanian devils ?

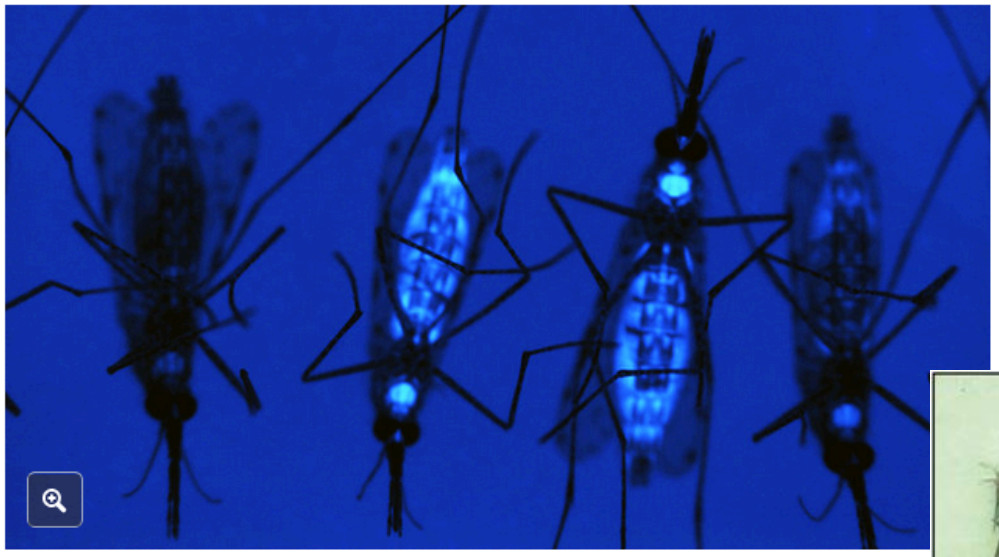
Brendan Epstein[1], Menna Jones[2], Rodrigo Hamede[2], Sarah Hendricks[3], Hamish McCallum[4], Elizabeth P. Murchison[5], Barbara Schönfeld[2], Cody Wiench[3], Paul Hohenlohe[3,*] & Andrew Storfer[1,*]

Panel Endorses 'Gene Drive' Technology That Can Alter Entire Species

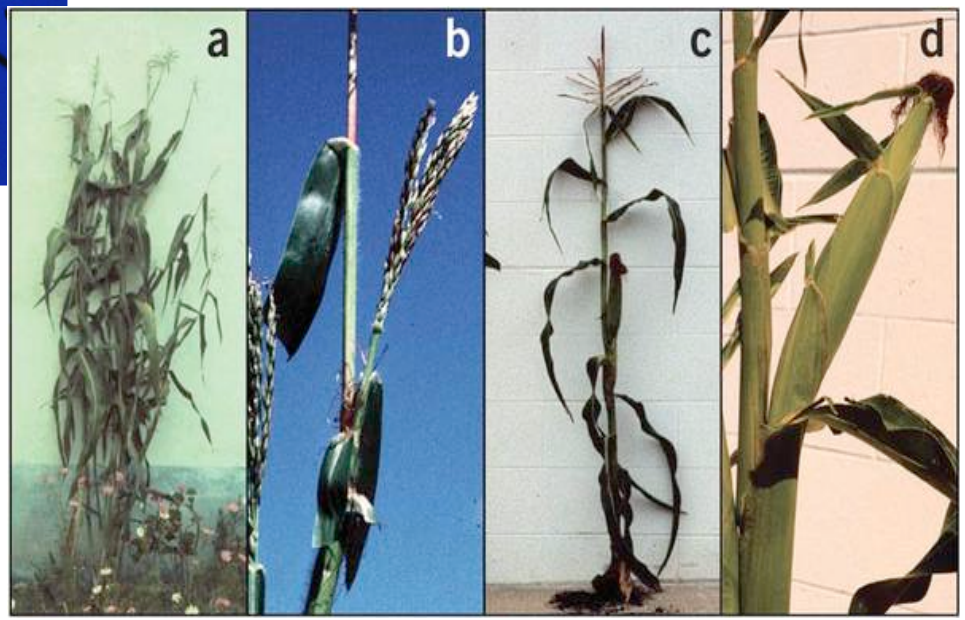
By AMY HARMON

NYT June 2016

http://www.nytimes.com/2016/06/09/science/national-academies-sciences-gene-drive-technology.html

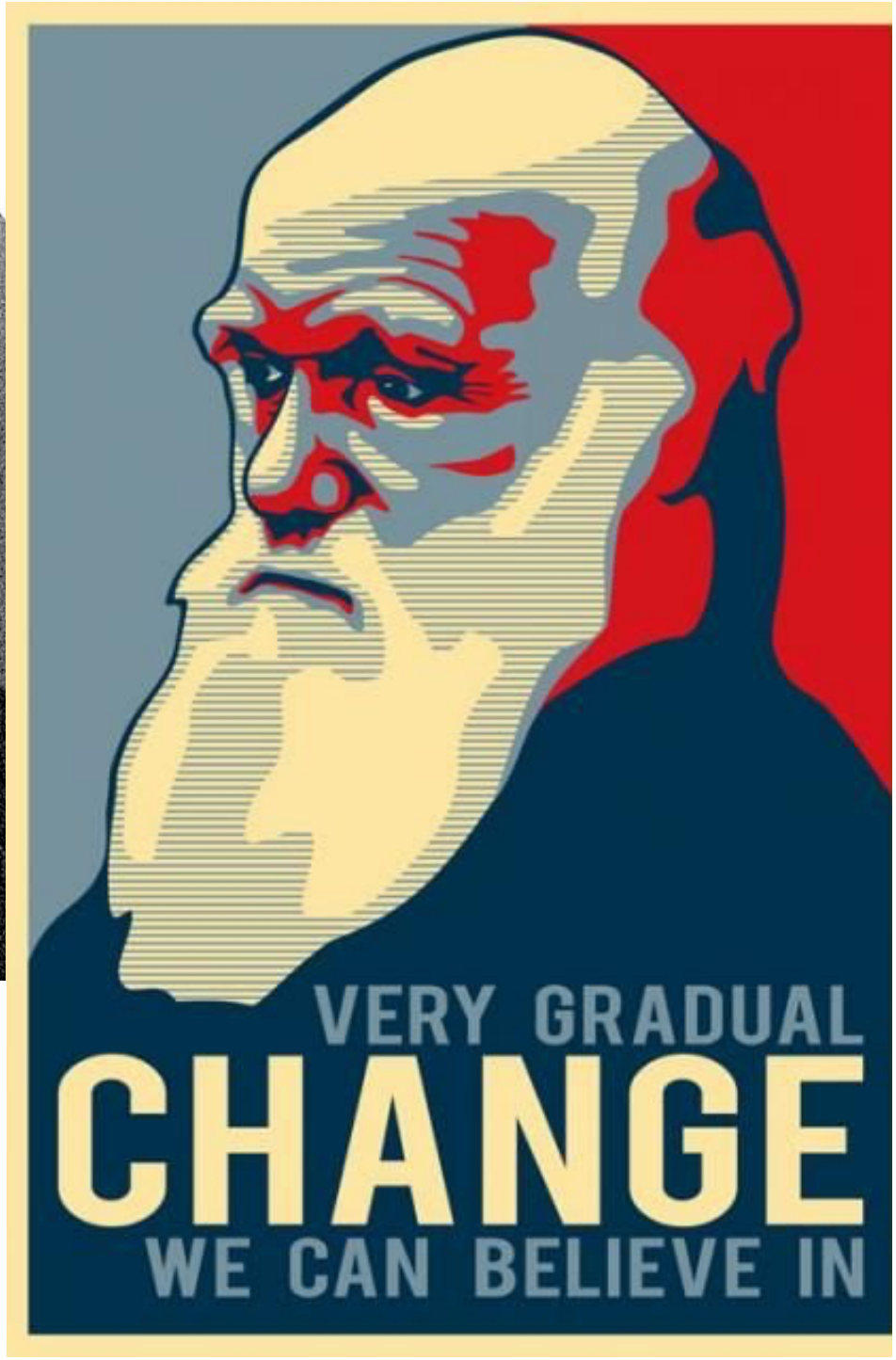


Female mosquitoes that have been altered as part of a gene drive experiment. Anthony James



Identification of a functional transposon insertion in the maize domestication gene *tb1*

Anthony Studer[1], Qiong Zhao[1], Jeffrey Ross-Ibarra[2,3] & John Doebley[1]



Population genetics: the extension of Mendelian genetics to evolving populations

Quantitative Genetics: Extension to phenotype evolution

The Modern Synthesis - 1930s

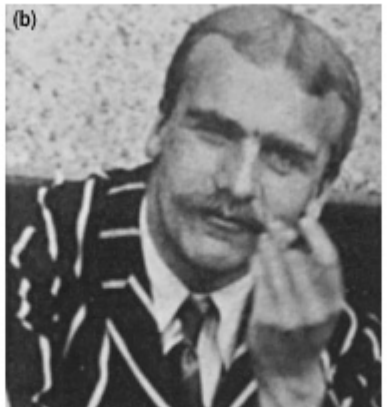
- Quantitative Genetics and Population genetics

  A quantitative theory of genetic and phenotypic variation.

  A quantitative theory of evolutionary change.

from Ridley

R.A. Fisher



J.B.S. Haldane

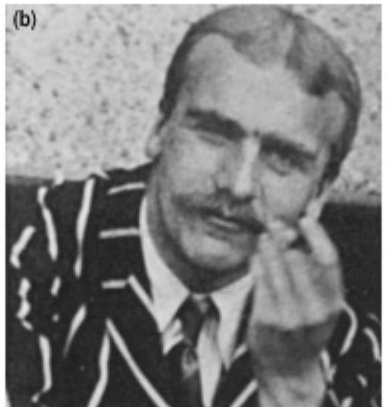
Sewall Wright

The Modern Synthesis - 1930s

- Quantitative Genetics and Population genetics
- Systematics and speciation – biodiversity
- Paleobiology - history of life should be consistent with known evolutionary mechanisms in extant species

from Ridley

R.A. Fisher



J.B.S. Haldane

Sewall Wright

Th. Dobzhansky

Ernst Mayr

G.G. Simpson

G. Ledyard Stebbins

What is population genetics?

- The genetic basis of evolutionary change.
- The study of genetic variation within and between populations and species
- The basis of much of "micro"-evolutionary thought

Nothing in biology makes sense except in light of evolution. T. Dobzhansky (1973)

Nothing in evolution makes sense except in light of population genetics. M. Lynch (2005)

What is theoretical Population genetics?

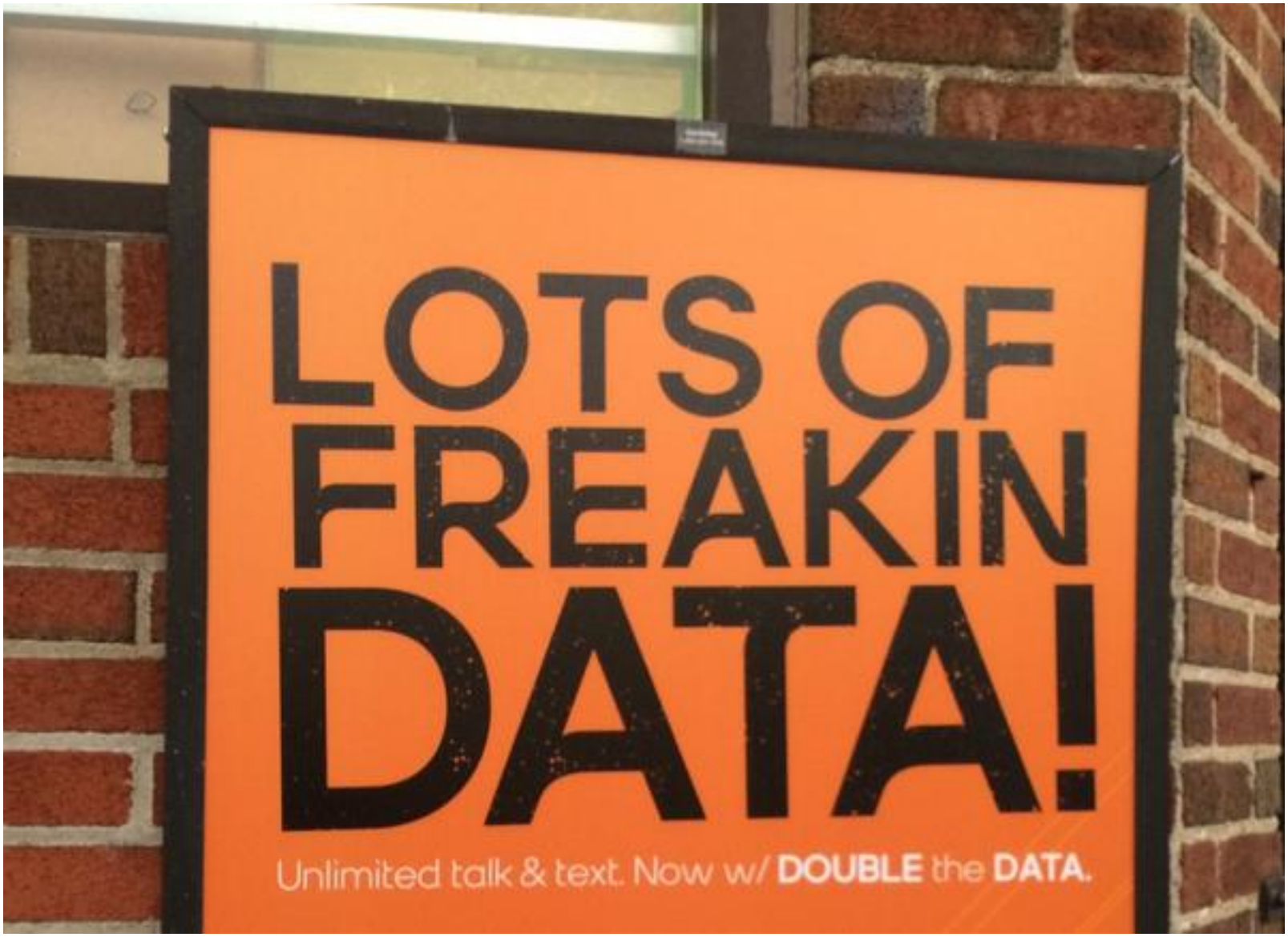
- The interplay between:
  - Mutation, assortative mating, migration, drift, recombination and selection
  - And how these 'forces' shape polymorphism and divergence

  - "All models are wrong, but some are useful" –Box

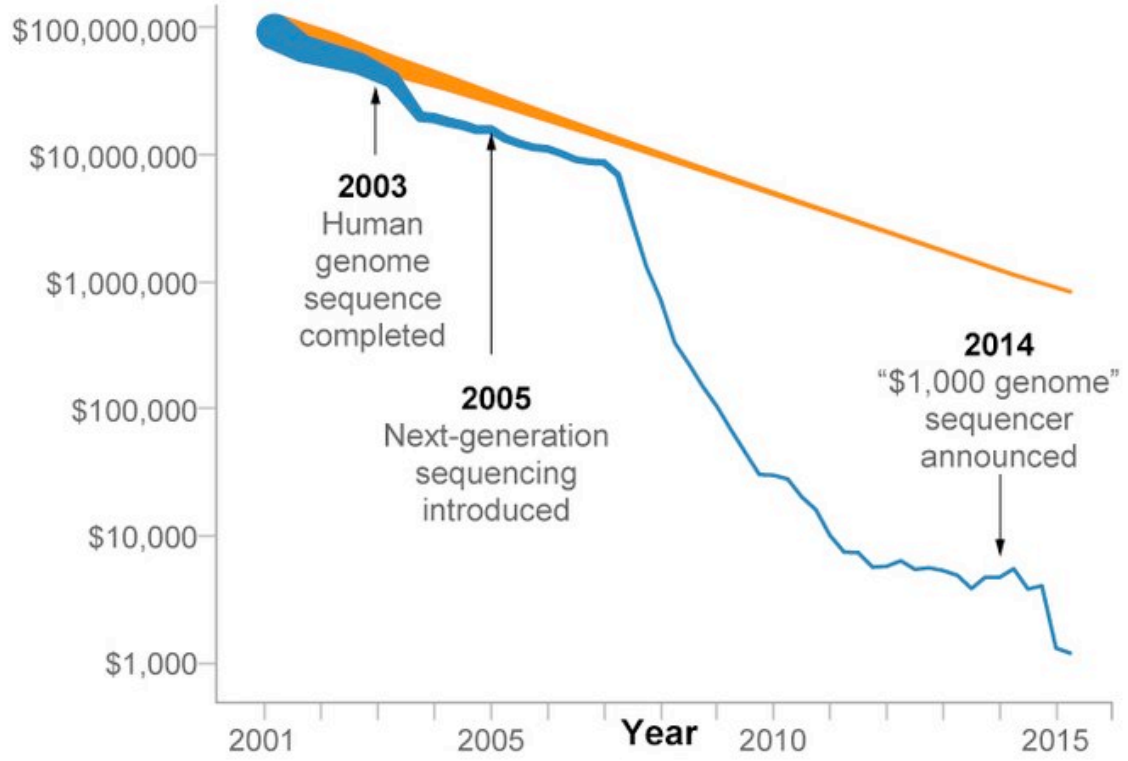
Why are population genetics models useful?

- Support or discount verbal models
- Build intuition.
- Evolution is fundamentally a statistical process.
- Mendelian inheritance, segregation, & recombination provide a powerful framework.
  - "theoretical population genetics is a formal reaction to the existence of sex"

What is empirical Population genomics?



DNA sequencing costs over time



Average Cost per Mb

$0.01

$2,000.00

$4,000.00

$6,000.00

Data source

NHGRI data

Moore's law calculation

Decline in real costs compared to expected declines based on Moore's Law.
Trend line: Cost per human genome. Line width: Cost per megabase (Mb)
Data: NHGRI https://www.genome.gov/27541954/dna-sequencing-costs-data/

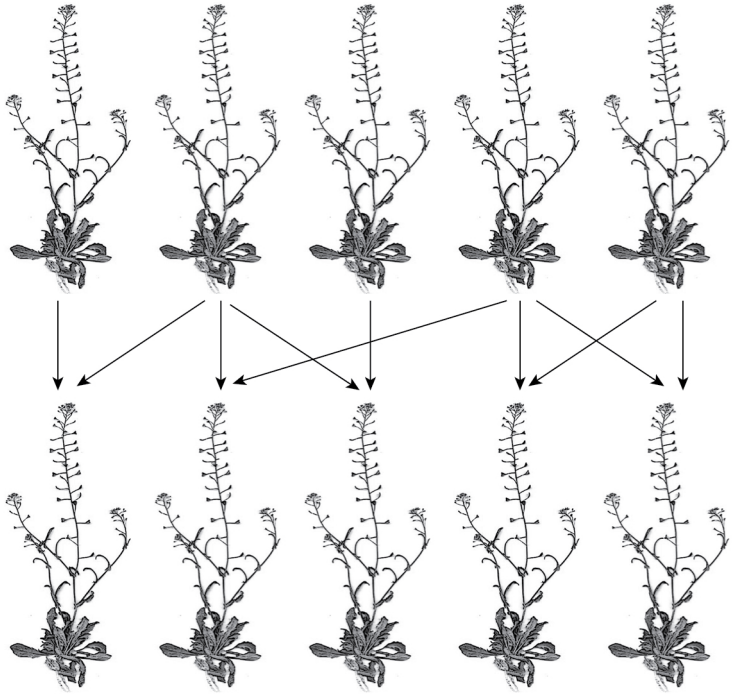
BROAD INSTITUTE

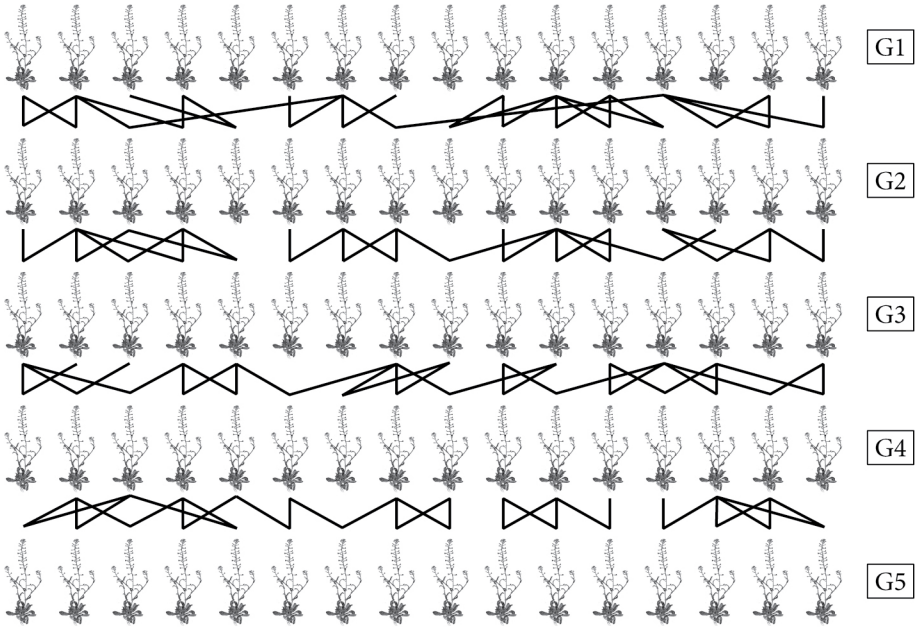
What is Evolution?

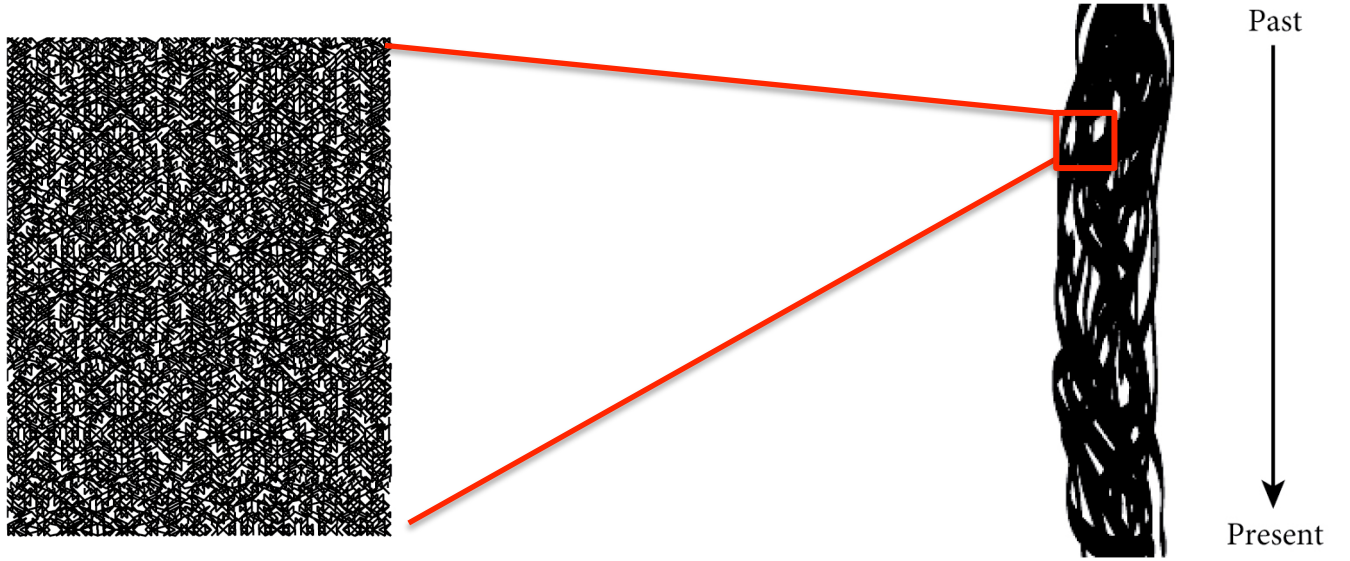
Descent with modification

Genetic changes in populations of organisms over time.

The process of descent

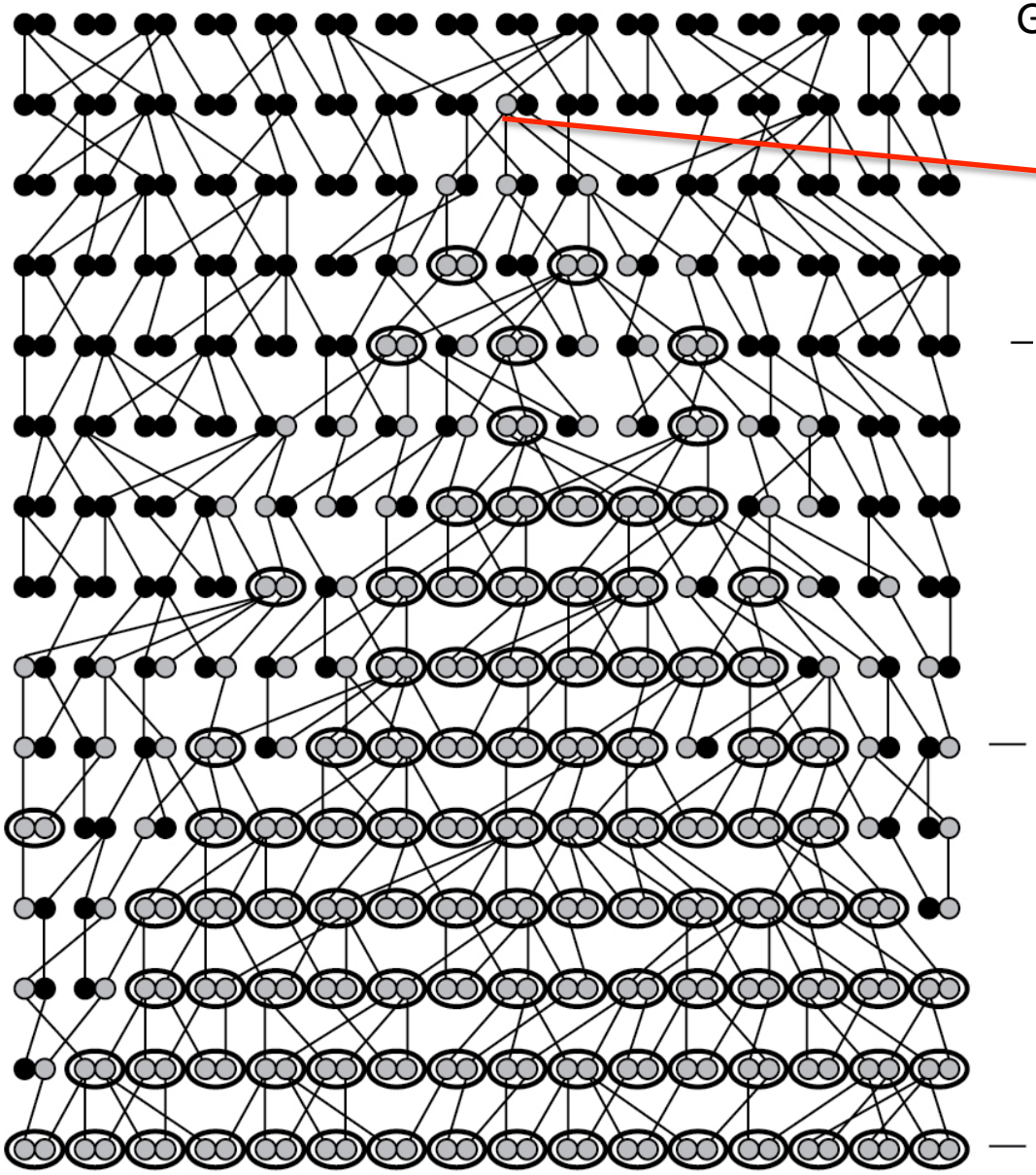






Modified from Baum and Smith Tree-Thinking book

Descent with modification



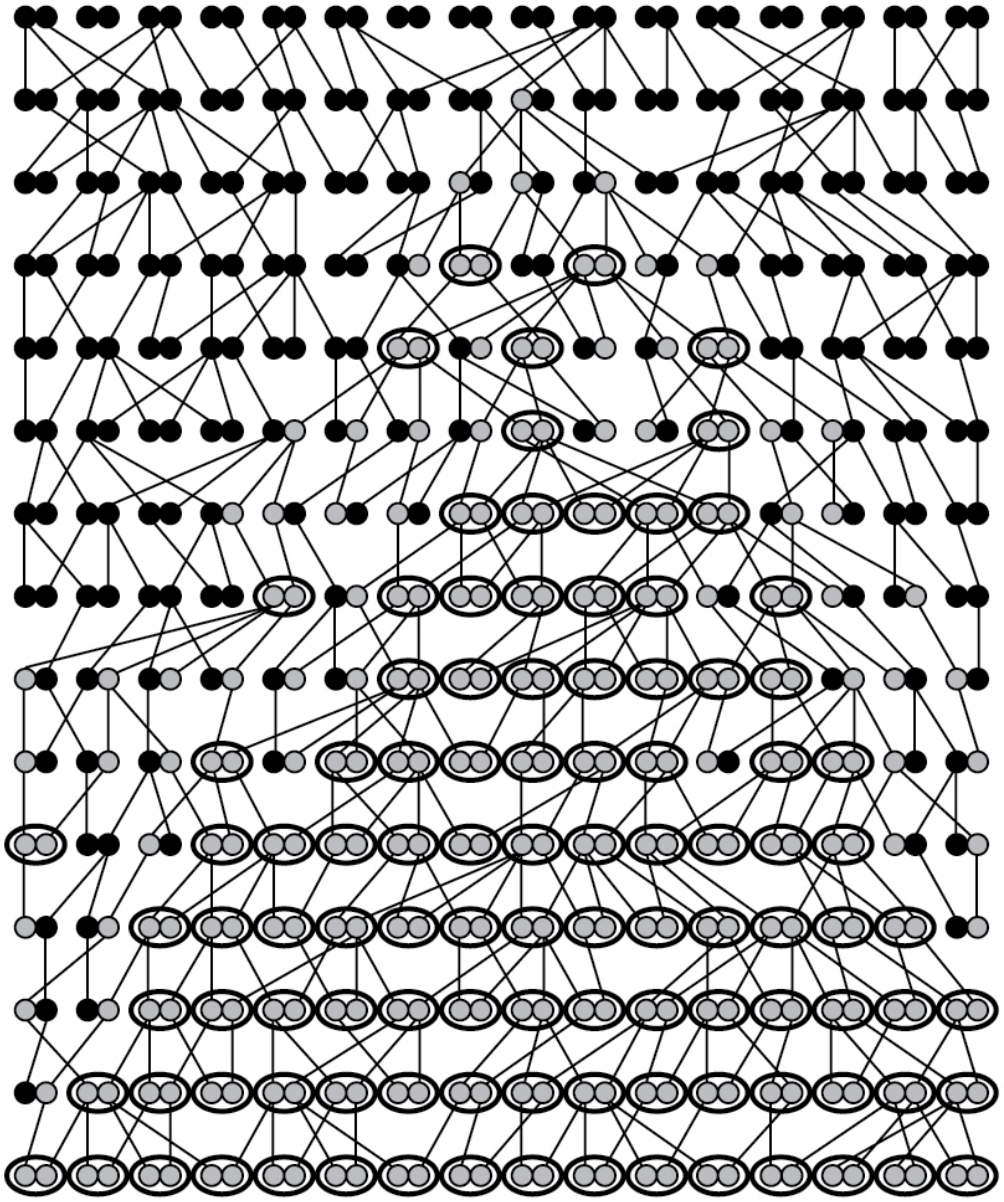
Generation 1 ATCCGGAAA

Mutation from G->A position 6.
Creates a polymorphism
G/A in population

ATCCGAAAA

Modified from
Baum and Smith
Tree-Thinking
book

Descent with modification



ATCCGGAAA

Generation 1

Gen. 2

Gen. 10

Generation 15

ATCCGAAAA

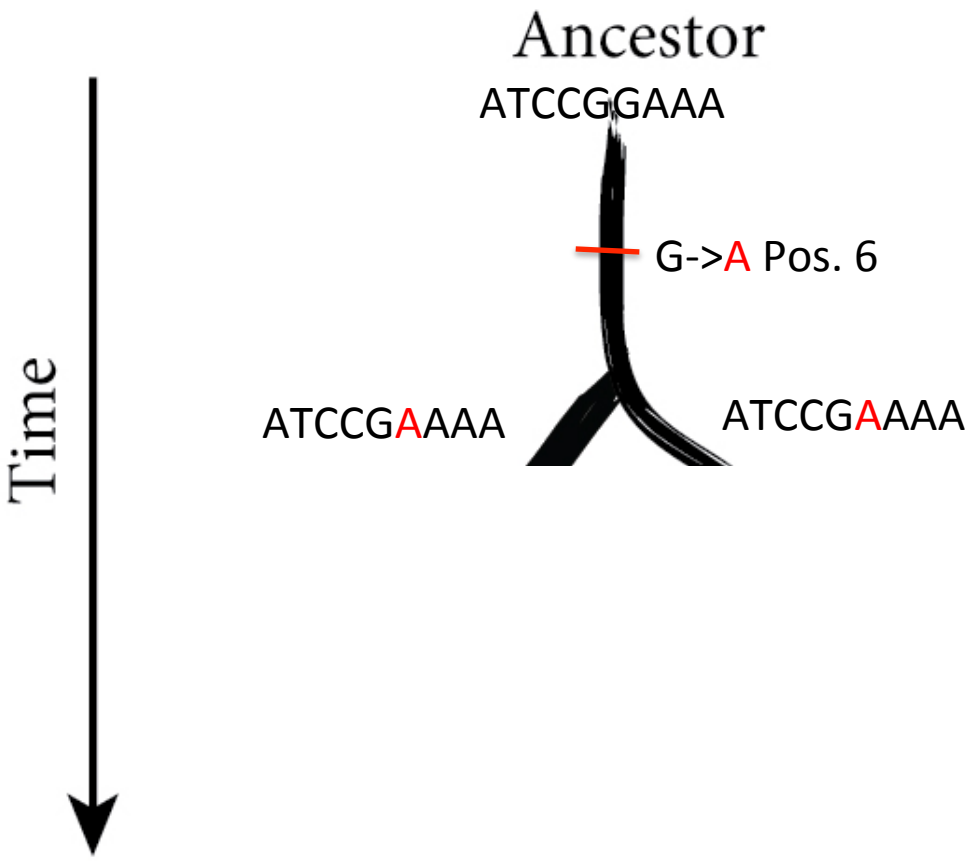
Past

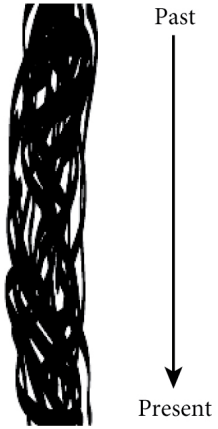
Present

ATCCGGAAA

ATCCGGAAA

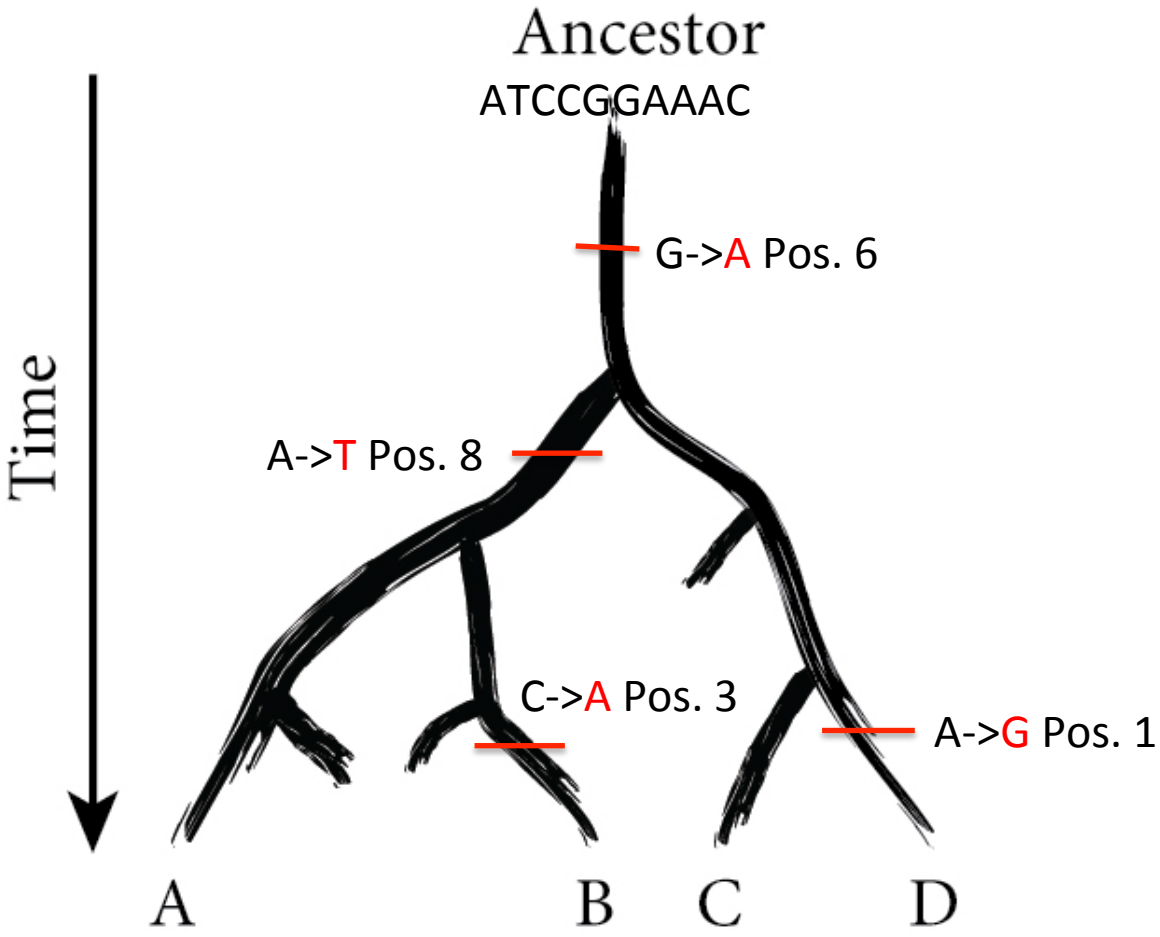
Modified from Baum and Smith Tree-Thinking book



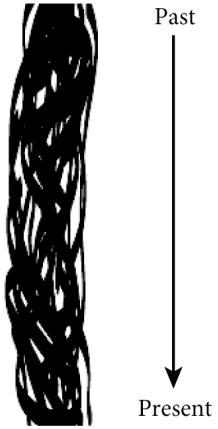


A: ATCCGAAAA
B: ATCCGAAAA
C: ATCCGAAAA
D: ATCCGAAAA

Modified from Baum and Smith Tree-Thinking book

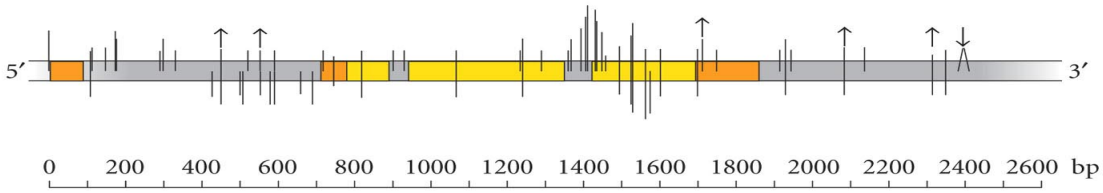


A: ATCCGAATA
B: ATACGAATA
C: ATCCGAAAA
D: GTCCGAAAA



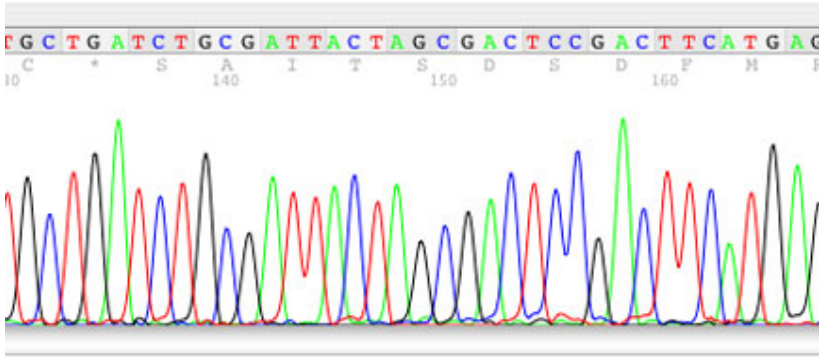
Modified from Baum and Smith Tree-Thinking book

•DNA sequencing



Variants in the *Adh* gene of *D. Melanogaster* *Kreitman 1983*

•Provides unbiased description of genetic variation
- •Coding, noncoding
- •Indels



412 ARTICLES NATURE VOL. 304 4 AUGUST 1983

Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*

Martin Kreitman
Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

The sequencing of eleven cloned Drosophila melanogaster *alcohol dehydrogenase* (Adh) *genes from five natural populations has revealed a large number of previously hidden polymorphisms. Only one of the 43 polymorphisms results in an amino acid change, the one responsible for the two electrophoretic variants (fast, Adh-f, and slow, Adh-s) found in nearly all natural populations. The implication is that most amino acid changes in Adh would be selectively deleterious.*

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

LOCUS (plural loci)
An allele

```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
```

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

Each diploid individual's genotype consists of 2 haplotypes (orthologous sequences)

```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
```

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

Each diploid individual's genotype consists of 2 haplotypes
At each locus/position individuals are homozygous or heterozygous

```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
```

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

Each diploid individual's genotype consists of 2 haplotypes
At each position homozygous/heterozygous

```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CTT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCC TAT
```

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

A sample from a population

Species 1

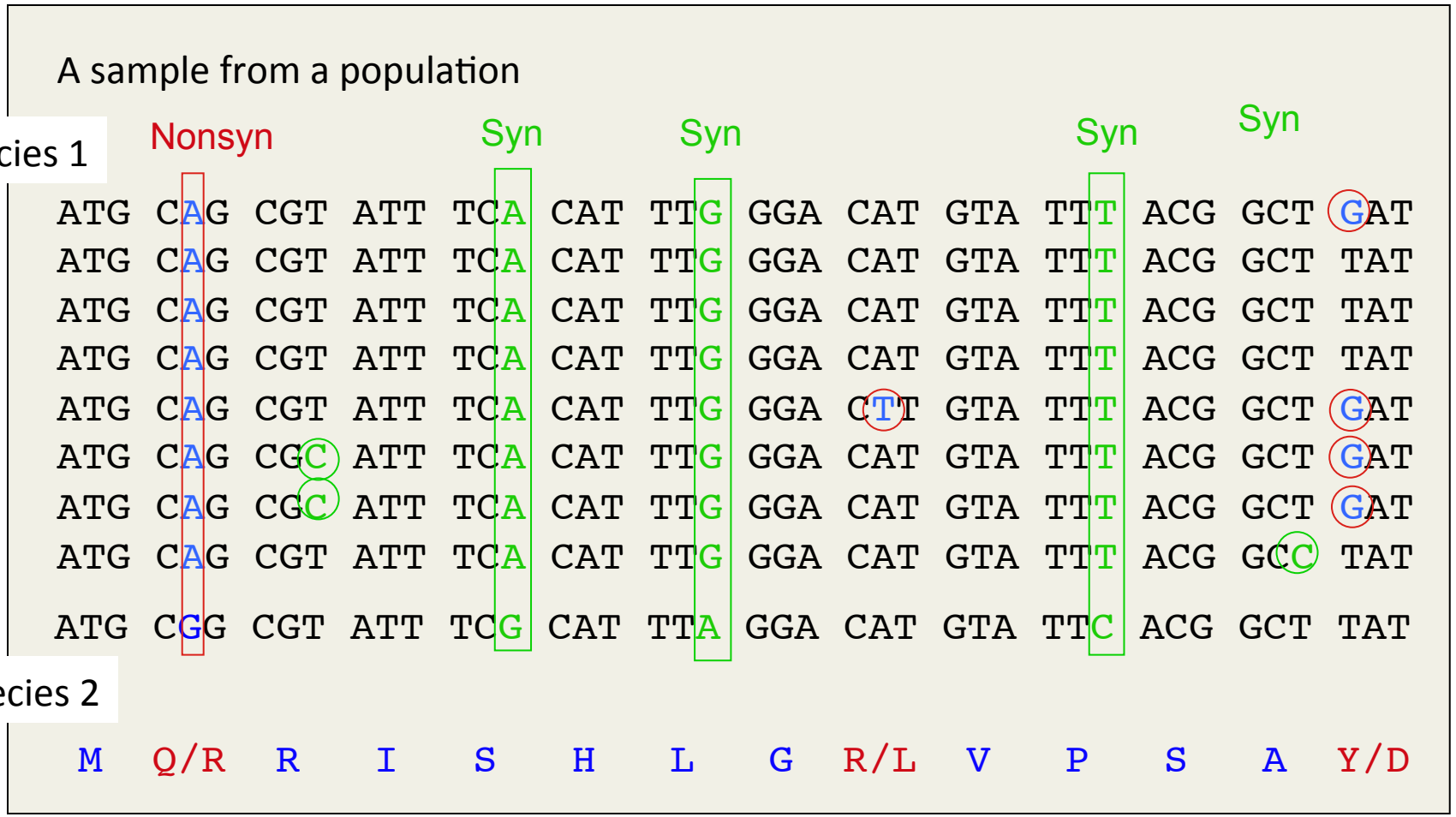
```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CTT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCC TAT
```

Species 2

```
ATG CGG CGT ATT TCG CAT TTA GGA CAT GTA TTC ACG GCT TAT
```

Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data



Basic currency of modern population genetics
Aligned orthologous sequence across individuals

Sequence data

A sample from a population

```
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT TAT
ATG CAG CGT ATT TCA CAT TTG GGA CTT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGC ATT TCA CAT TTG GGA CAT GTA TTT ACG GCT GAT
ATG CAG CGT ATT TCA CAT TTG GGA CAT GTA TTT ACG GCC TAT
```

Four simple summaries of polymorphism:

The frequency of each site.

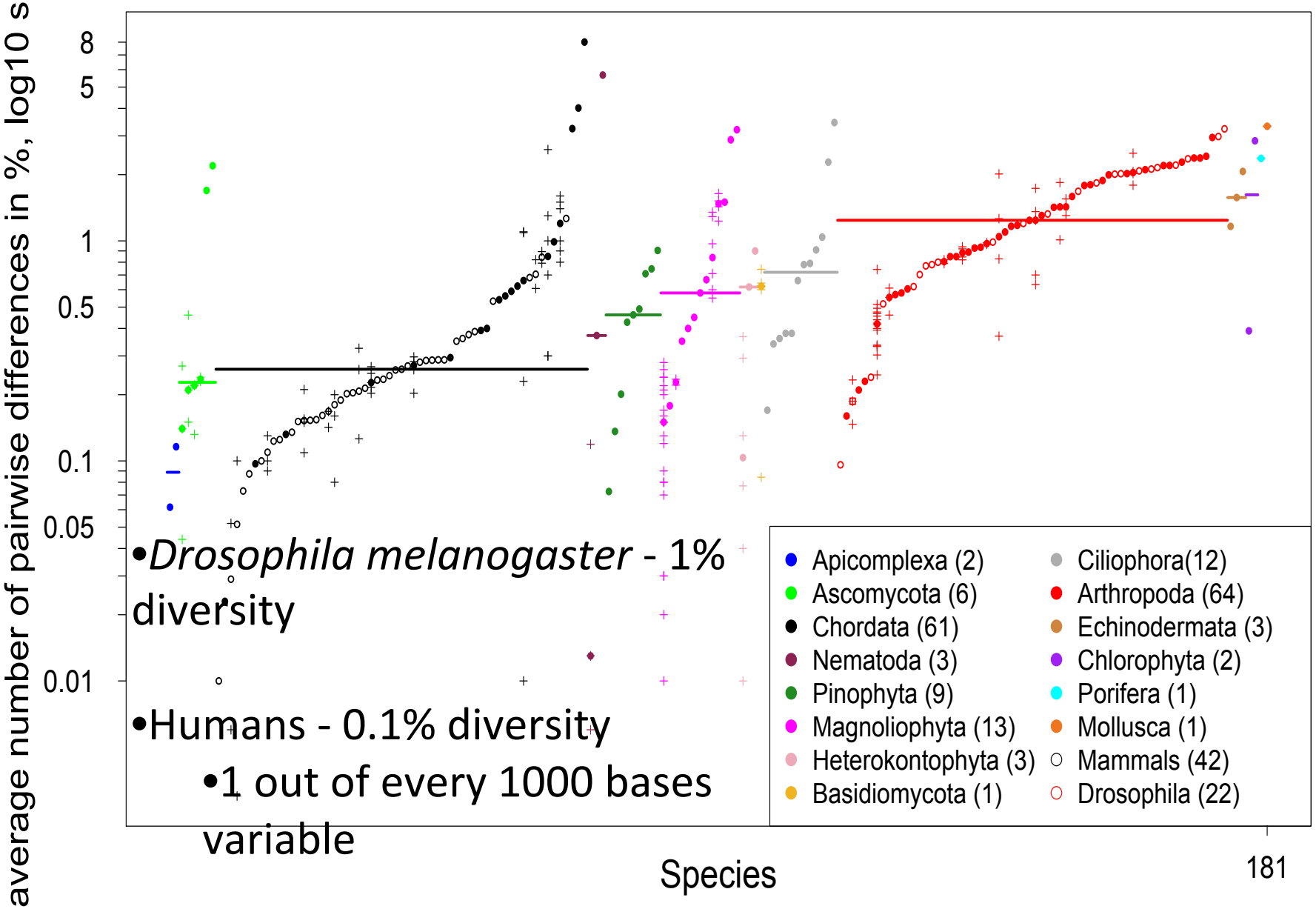
Number of segregating sites.

Heterozygosity: Fraction of all sites where an individual is heterozygous

Pairwise Diversity:

Fraction of sites that differ between two sequences chosen at random

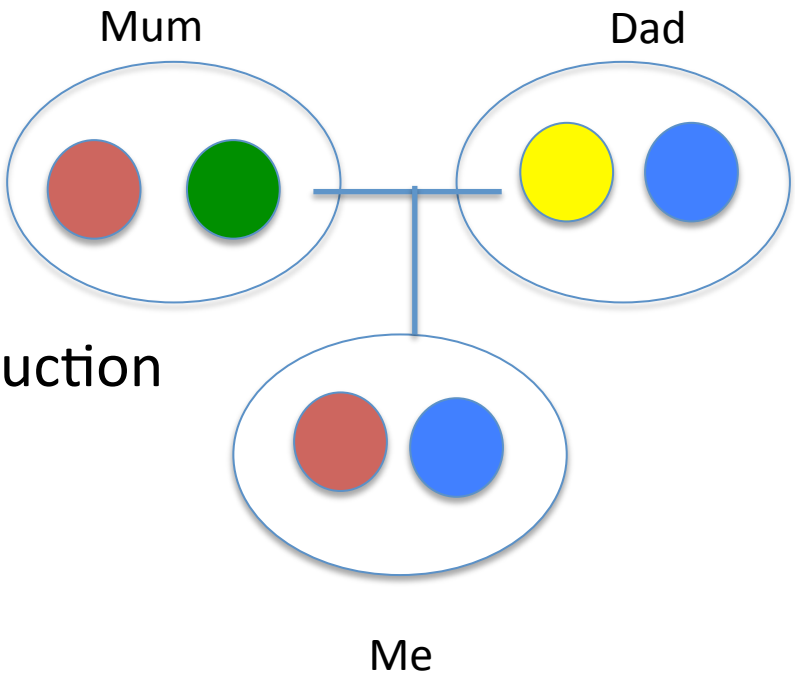
How much DNA variation is there?



Leffer et al PLOS Biology

- Actual genotype frequencies.
- SNP A-4213906 0.07 0.40 0.53

Hardy Weinberg Expectations



In adults just before reproduction
Frequency of A =p
Frequency of a = q

What's the frequency of AA homozygotes?

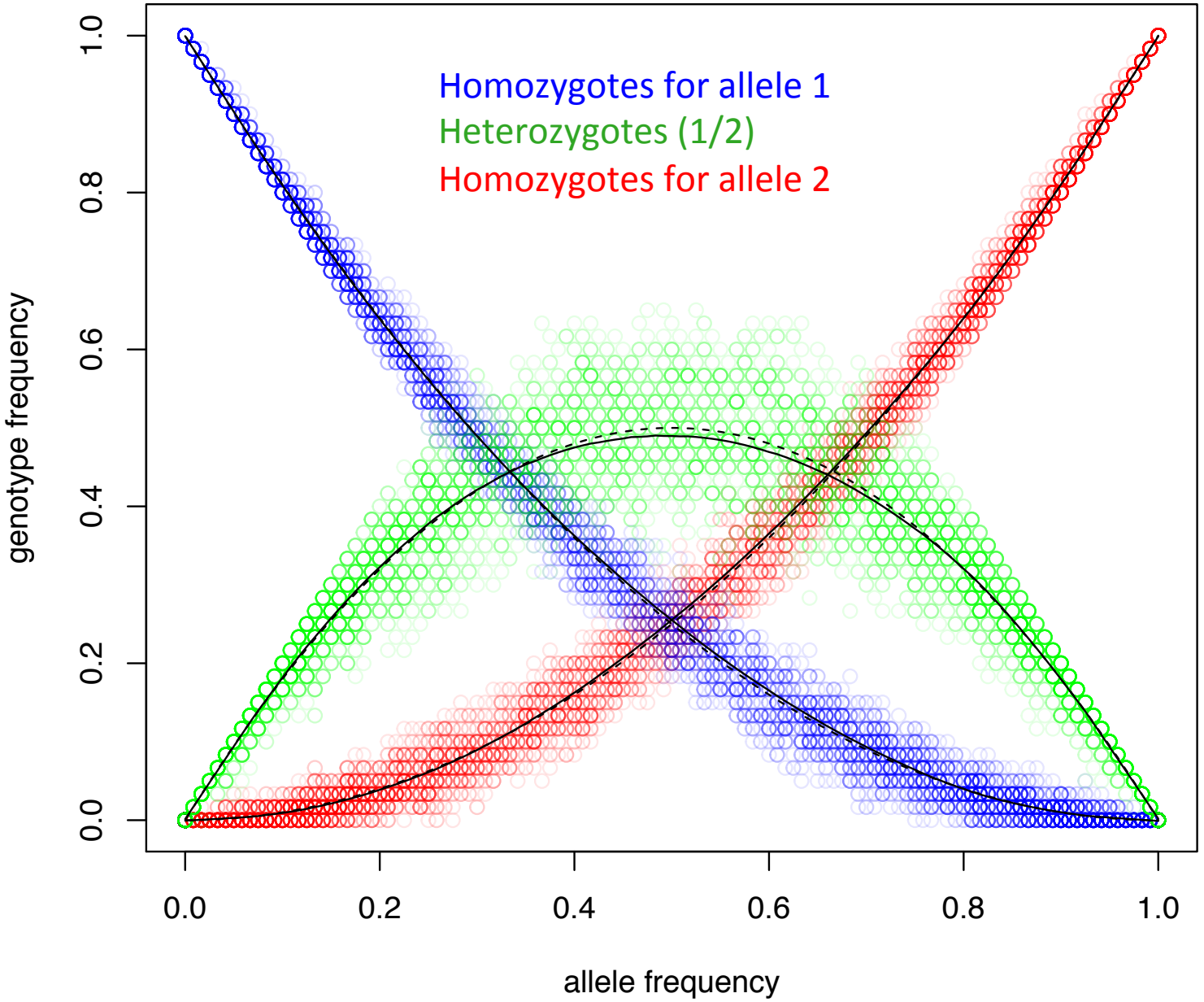
What do I have to assume?

	A	a
• SNP A-4213906	p=0.27	q=0.73

- Freq. of AA
- Freq. of Aa
- Freq. of aa
- Actual genotype frequencies.

• SNP A-4213906	0.07	0.40	0.53

The Empirical Relationship between Genotype and allele Frequencies in a European population



What if genotype frequencies were not in Hardy Weinberg at their HW proportions?

AA	Aa	aa
0.2	0.14	.66

- $p=f_{AA}+ \frac{1}{2} f_{Aa}=0.2+.14/2= 0.27$ $q=f_{aa}+ \frac{1}{2} f_{Aa}=0.73$

- Freq. of AA $=p^2 = 0.27^2 =0.079$
- Freq. of Aa $=2pq = 2 \times 0.27 \times 0.73=0.394$
- Freq. of aa $= q^2 = 0.73^2 = 0.533$

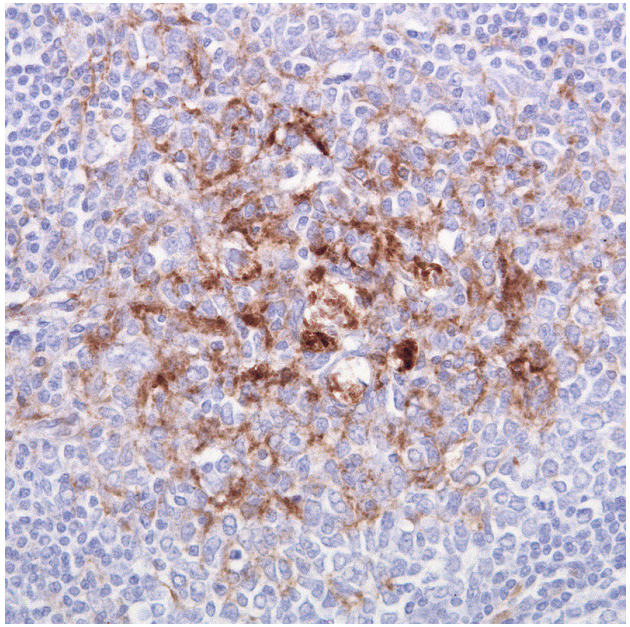
Genotype frequencies returned to Hardy Weinberg Proportions within one generation Of random mating!

Therefor the Hardy Weinberg expectations/proportions are an equilibrium if nothing interesting happens in popgen.

Kuru outbreak in the Fore people

The Fore people of Papua New Guinea practiced ritual funereal cannibalism (till 1950s)

coding polymorphism (Methionine/Valine) at codon 129 of *PRNP.* Homozygotes for either allele develop prion disease at a higher rate.



https://en.wikipedia.org/wiki/Creutzfeldt%E2%80%93Jakob_disease#/media/File:VCJD_Tonsil.jpg

Met/Met	Met/Val	Val/Val
4	23	3

Chi-squared statistic significance =0.0034

Frequency of red hair = 0.01