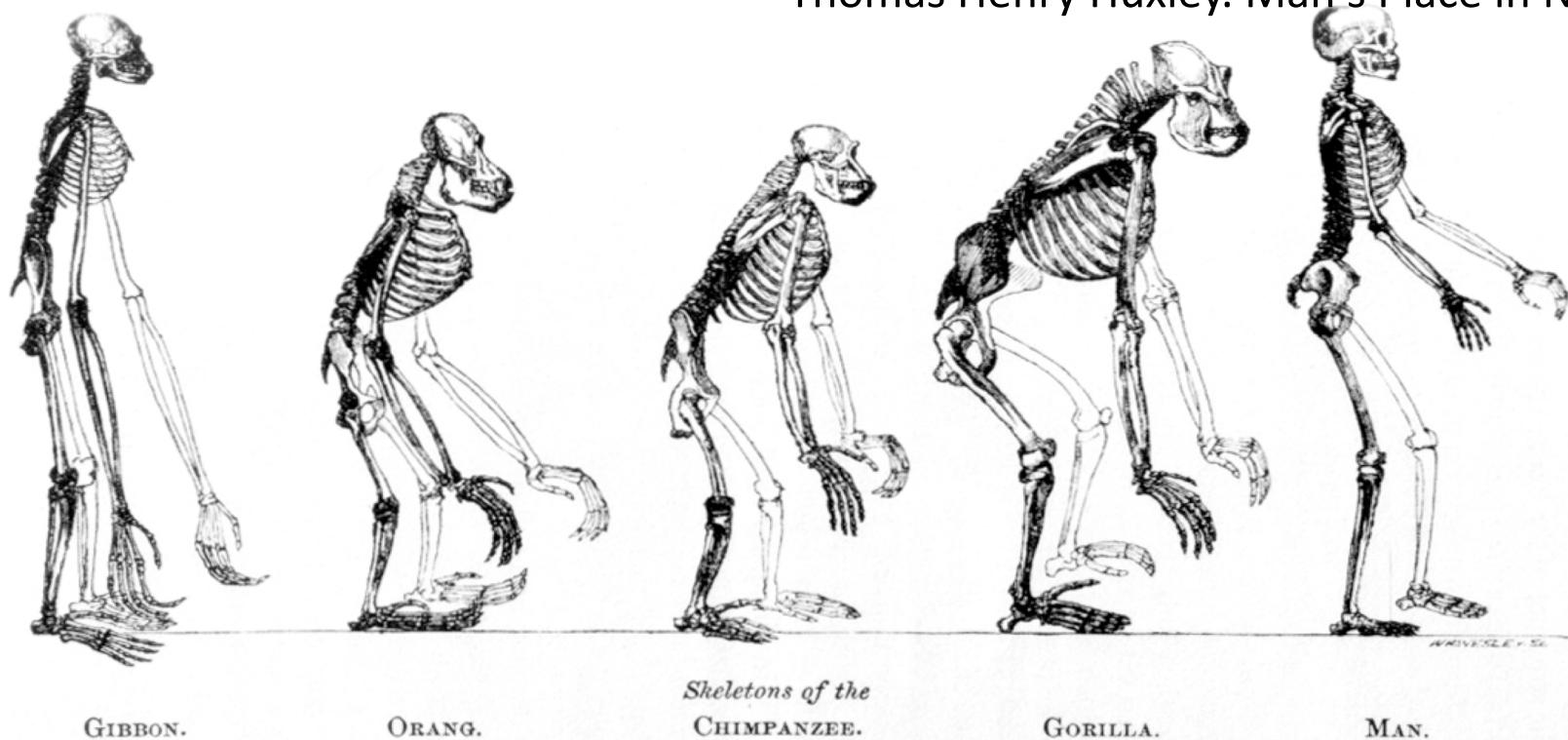


# Population and Quantitative genetics

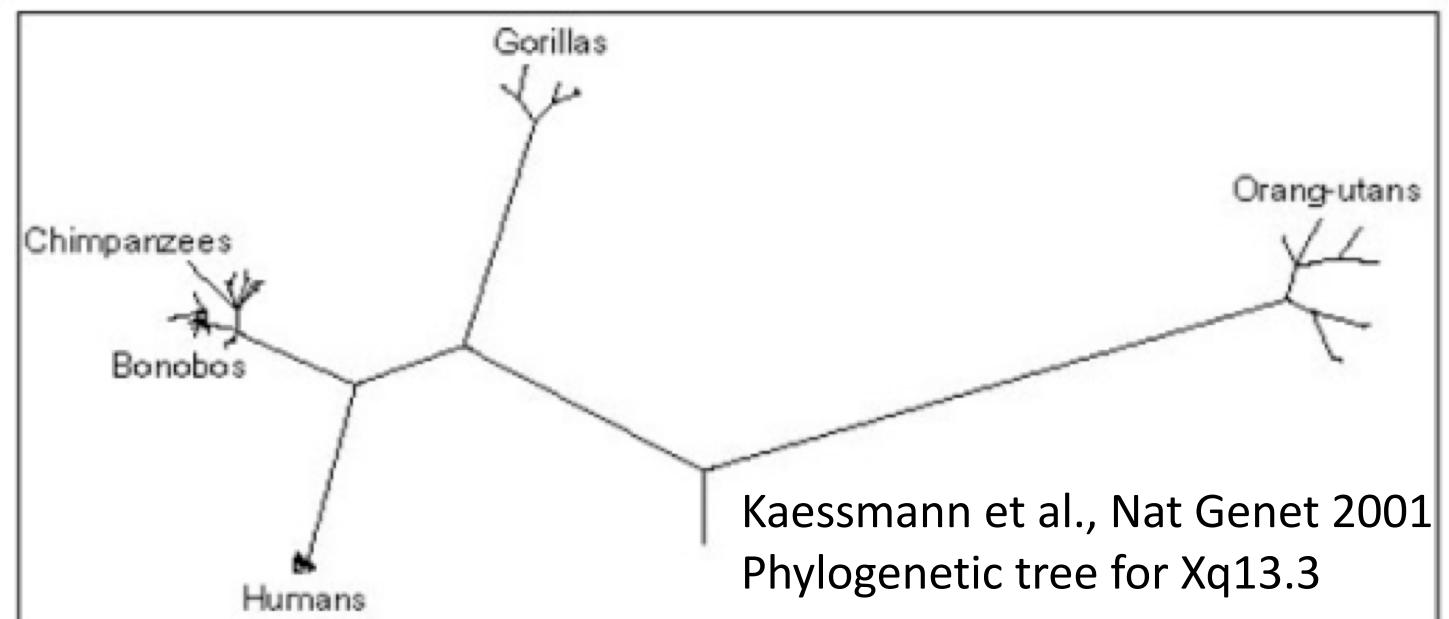
Part 1

@graham\_coop  
[www.gcbias.org](http://www.gcbias.org)

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



*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.*





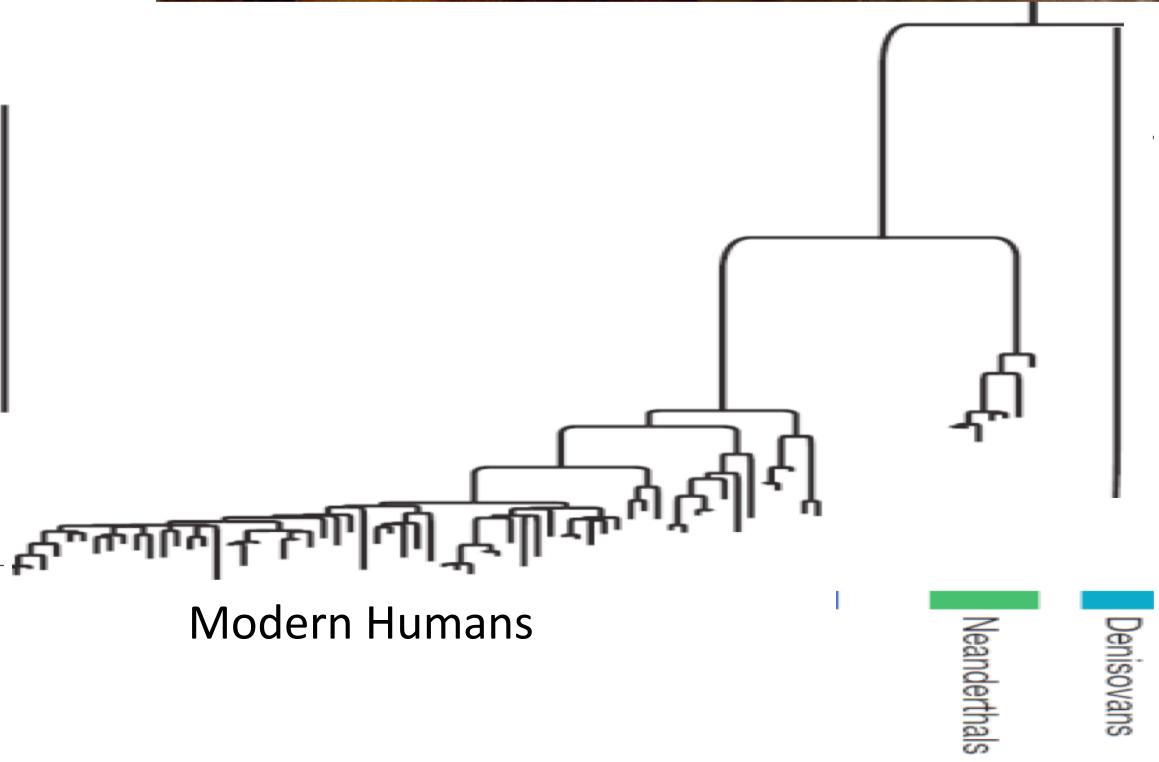
Neanderthal  
by Thomas Henry Huxley

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

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

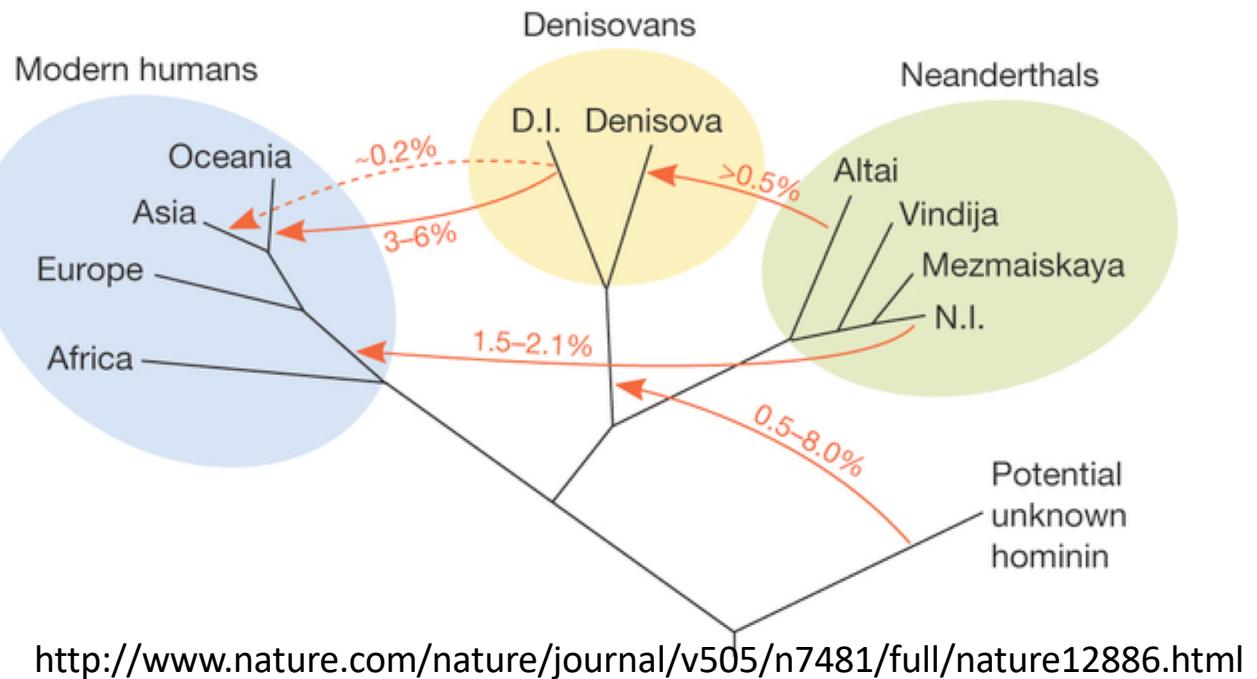


© Joe McNally ©2008 National Geographic





Copyright: Neanderthal Museum/H. Neumann



## 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<sup>1,2,✉,\*</sup>, Simon Aeschbacher<sup>1,2,†</sup>, Graham Coop<sup>1,2,‡</sup>

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

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

OPEN  ACCESS Freely available online

 PLOS | BIOLOGY

## The Geography of Recent Genetic Ancestry across Europe

Peter Ralph<sup>\*†</sup>, Graham Coop<sup>\*</sup>

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

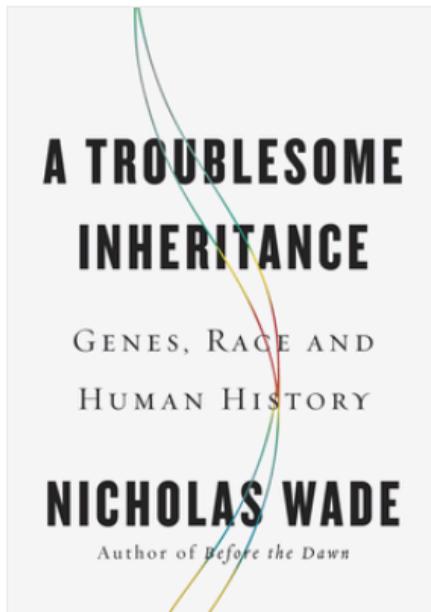
## Geneticists say popular book misrepresents research on human evolution

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/>

ochre stars (*Pisaster ochraceus*)



D. Gordon E. Robertson

sea star wasting disease (SSWD)



Photo credit: Melissa Miner

# Sea Stars Started Dissolving. What Helped Some of Them Survive?

Researchers say they've detected genetic differences that might help explain why some of these creatures on California's coast survived a deadly plague.

By Veronique Greenwood

June 18, 2018

## Decimation by sea star wasting disease and rapid genetic change in a keystone species, *Pisaster ochraceus*

Lauren M. Schiebelhut, Jonathan B. Puritz, and Michael N Dawson

PNAS July 3, 2018 115 (27) 7069-7074; published ahead of print June 18, 2018

<https://doi.org/10.1073/pnas.1800285115>

# Panel Endorses ‘Gene Drive’ Technology That Can Alter Entire Species

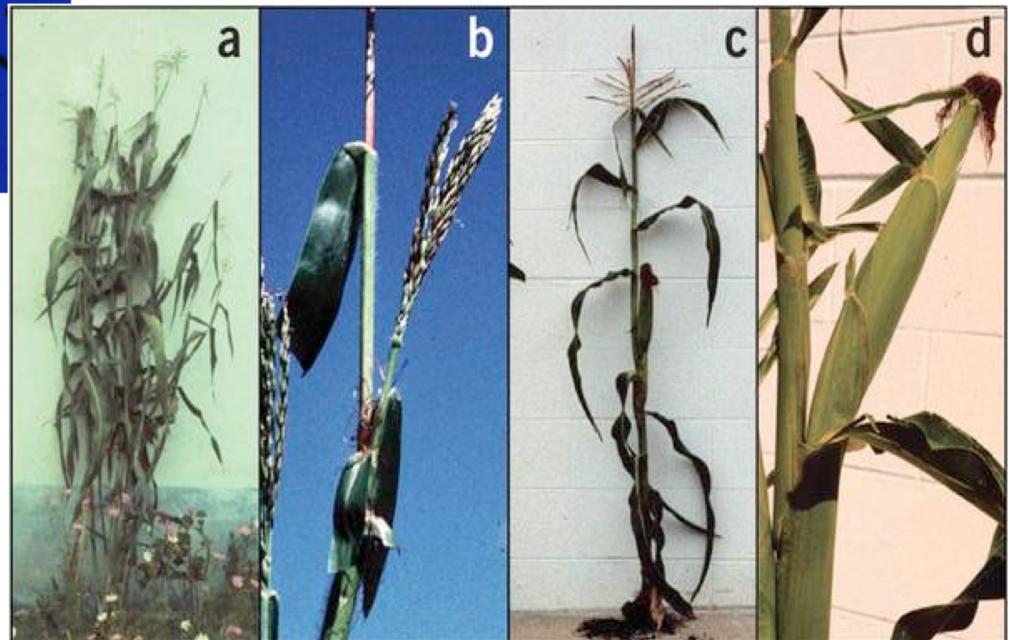
By AMY HARMON

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



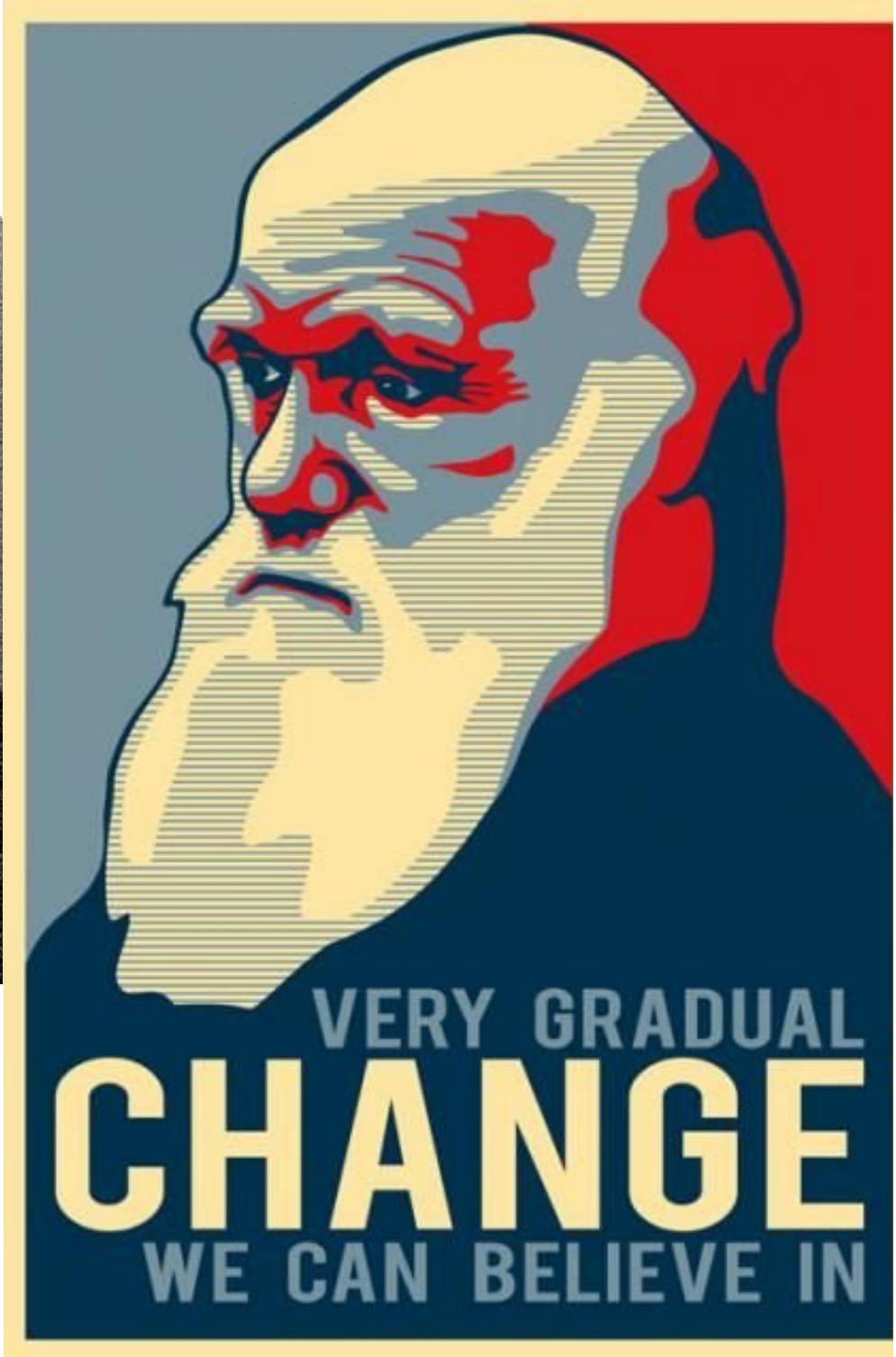
Identification of a functional transposon insertion in the maize domestication gene *tb1*

Anthony Studer<sup>1</sup>, Qiong Zhao<sup>1</sup>, Jeffrey Ross-Ibarra<sup>2,3</sup> & John Doebley<sup>1</sup>



Population genetics: the extension of Mendelian genetics to evolving populations

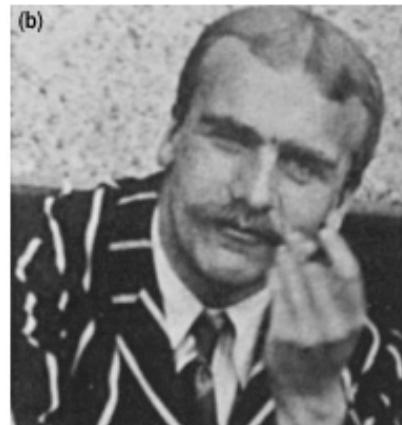
Quantitative Genetics: Extension to phenotype evolution



# 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)

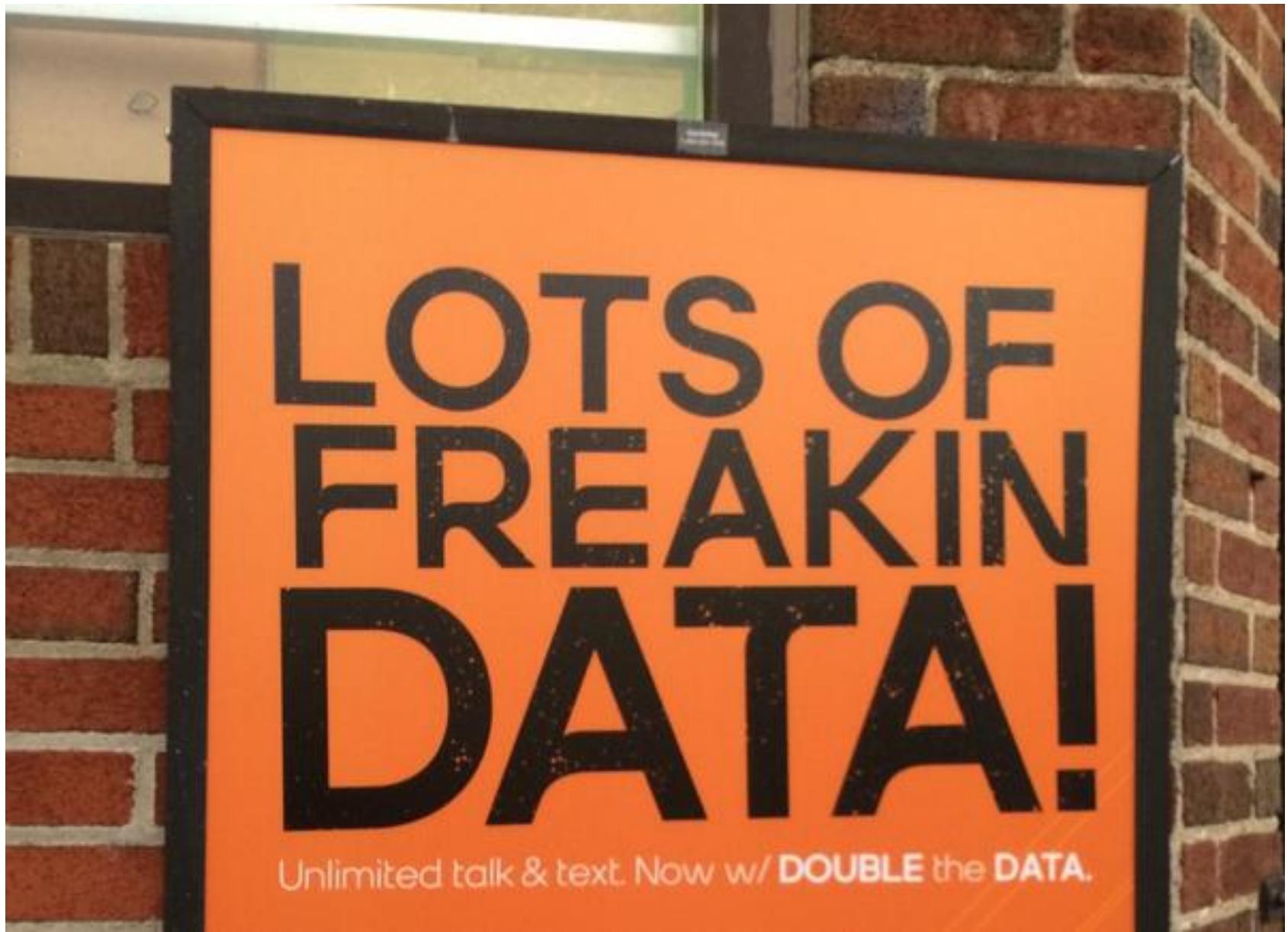
# What is theoretical Population genetics?

- 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

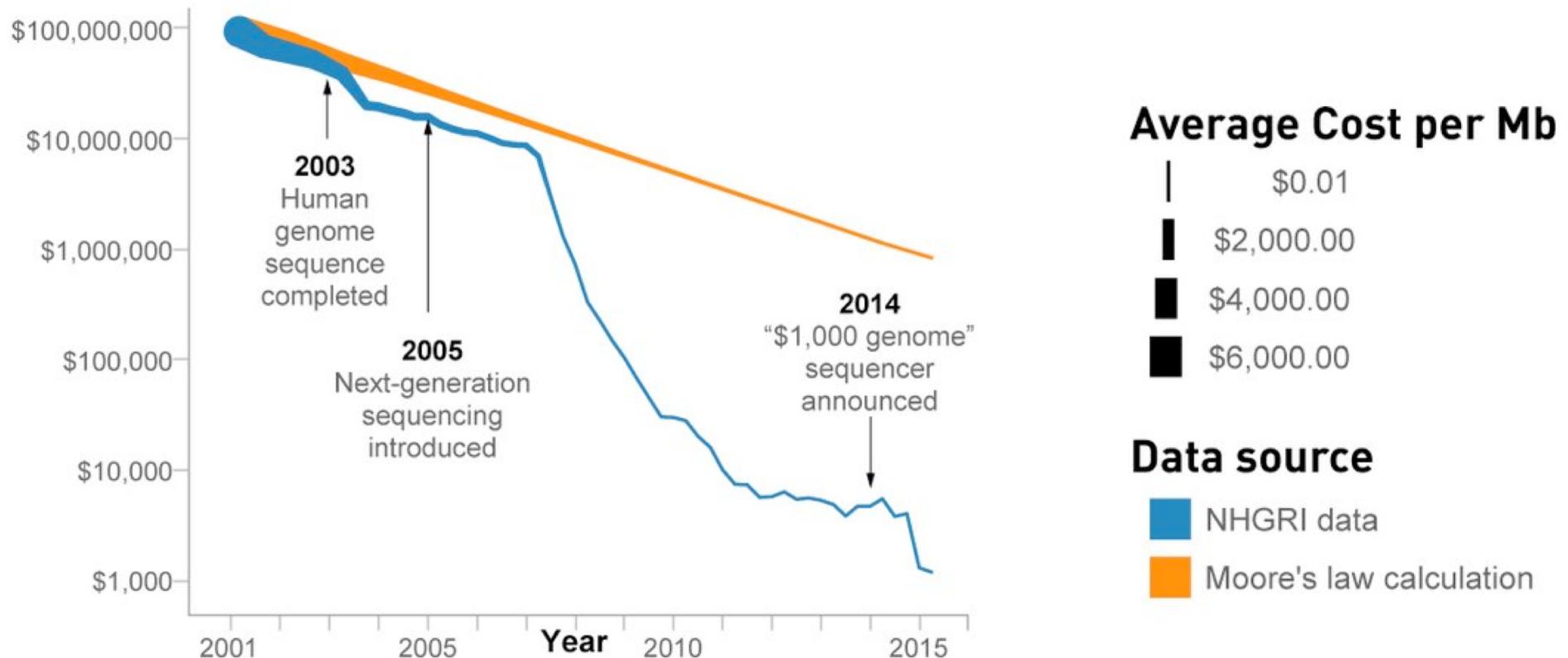
# Why are population genetics models useful?

- Support or discount verbal models
- Build intuition.
- Evolution is fundamentally a statistical process.
- Mendelian inheritance, segregation, & recombination provide a powerful framework.

# What is empirical Population genomics?



# DNA sequencing costs over time



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/>

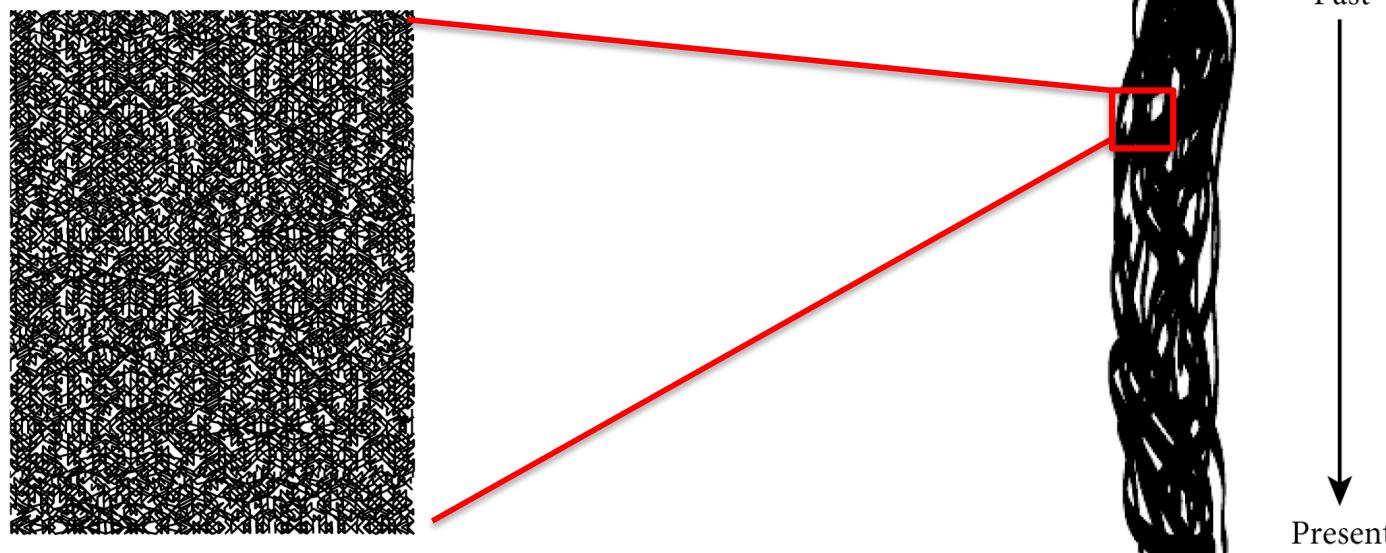
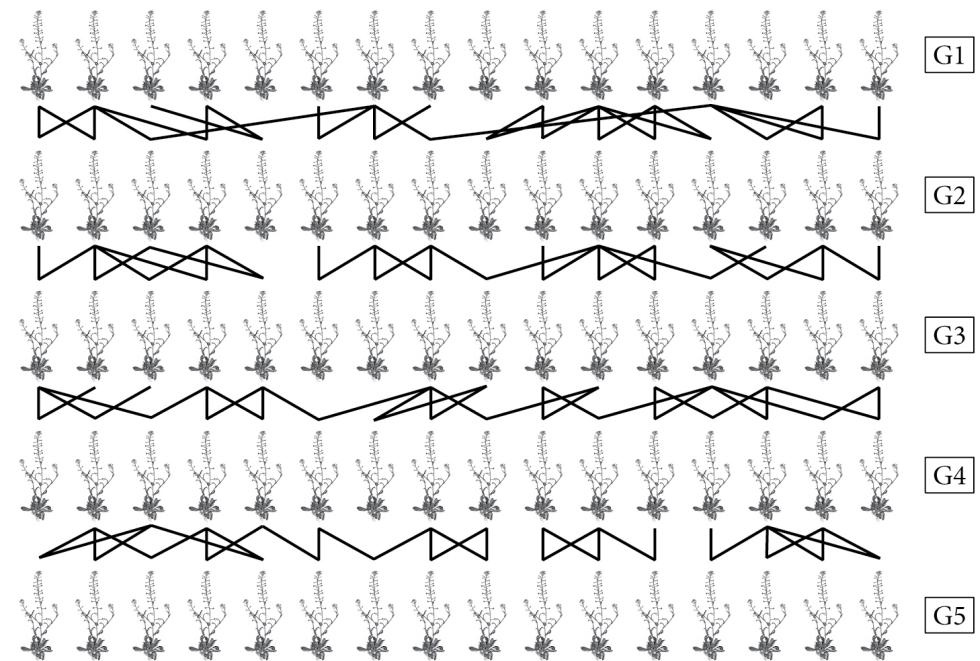
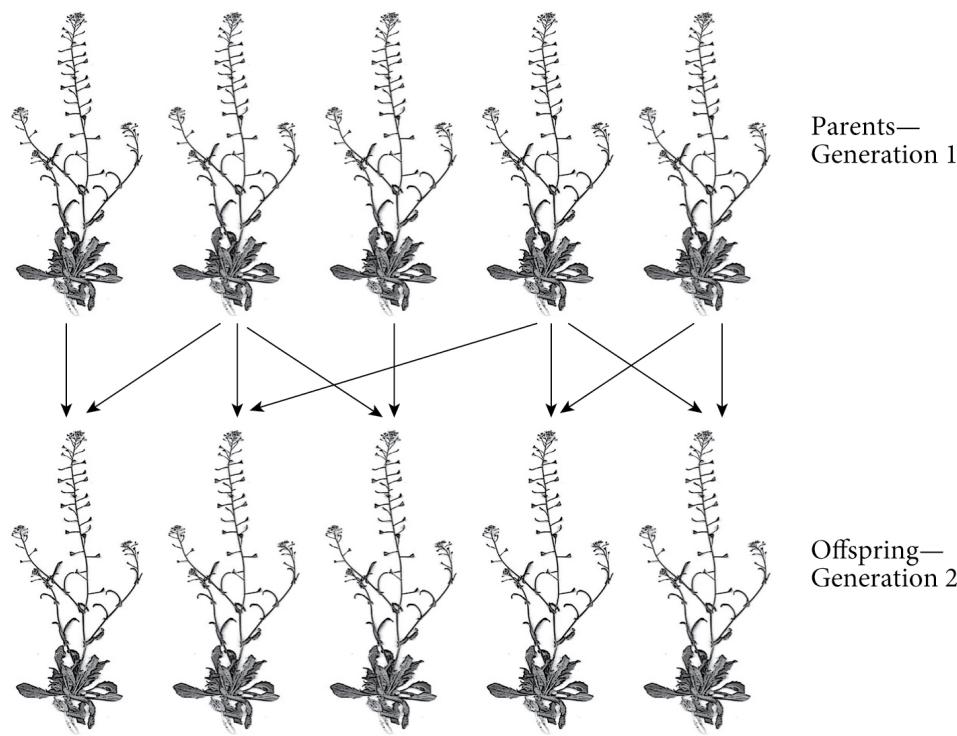


# What is Evolution?

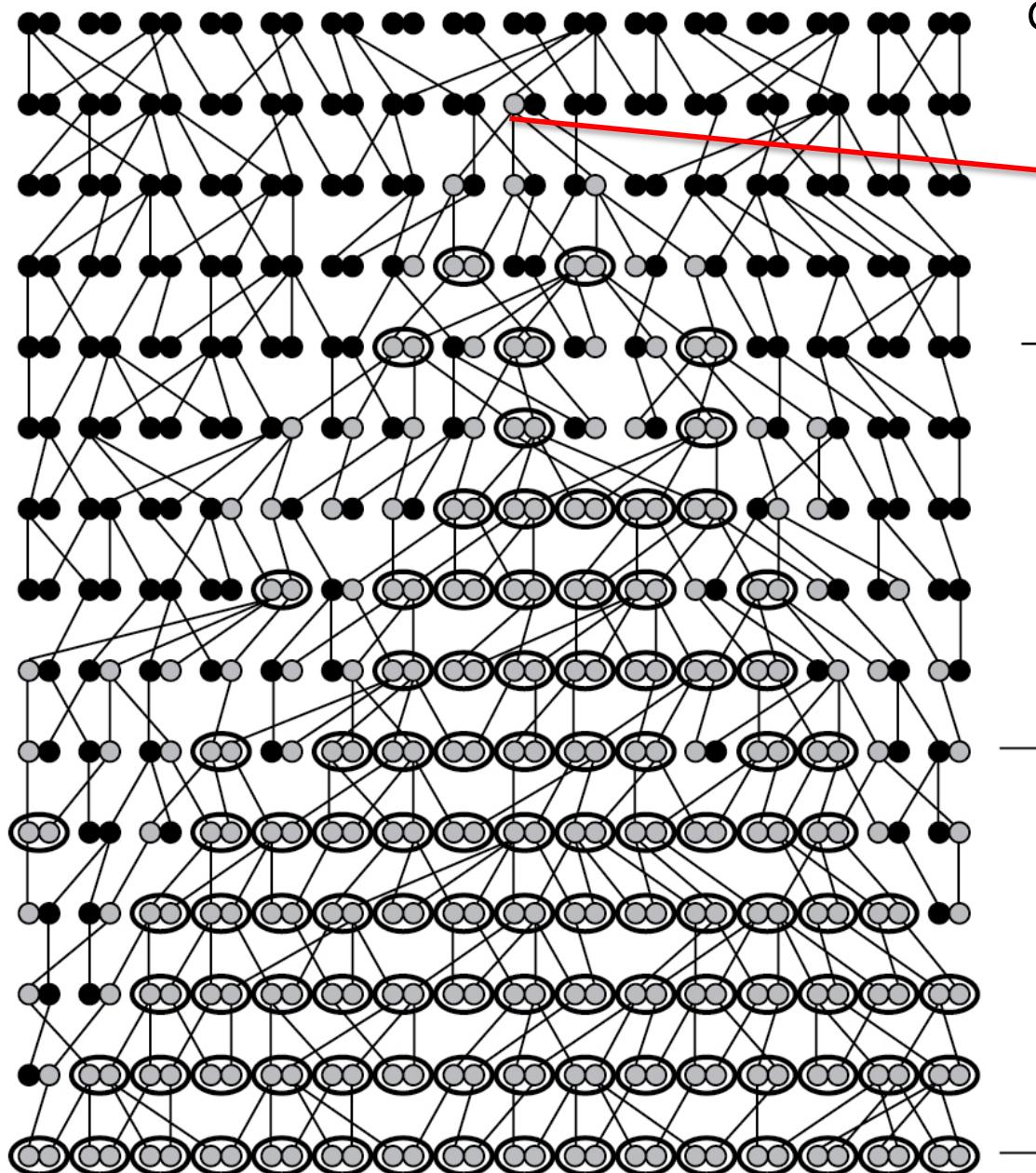
Descent with modification

Genetic changes in populations of organisms over time.

# The process of descent



# Descent with modification



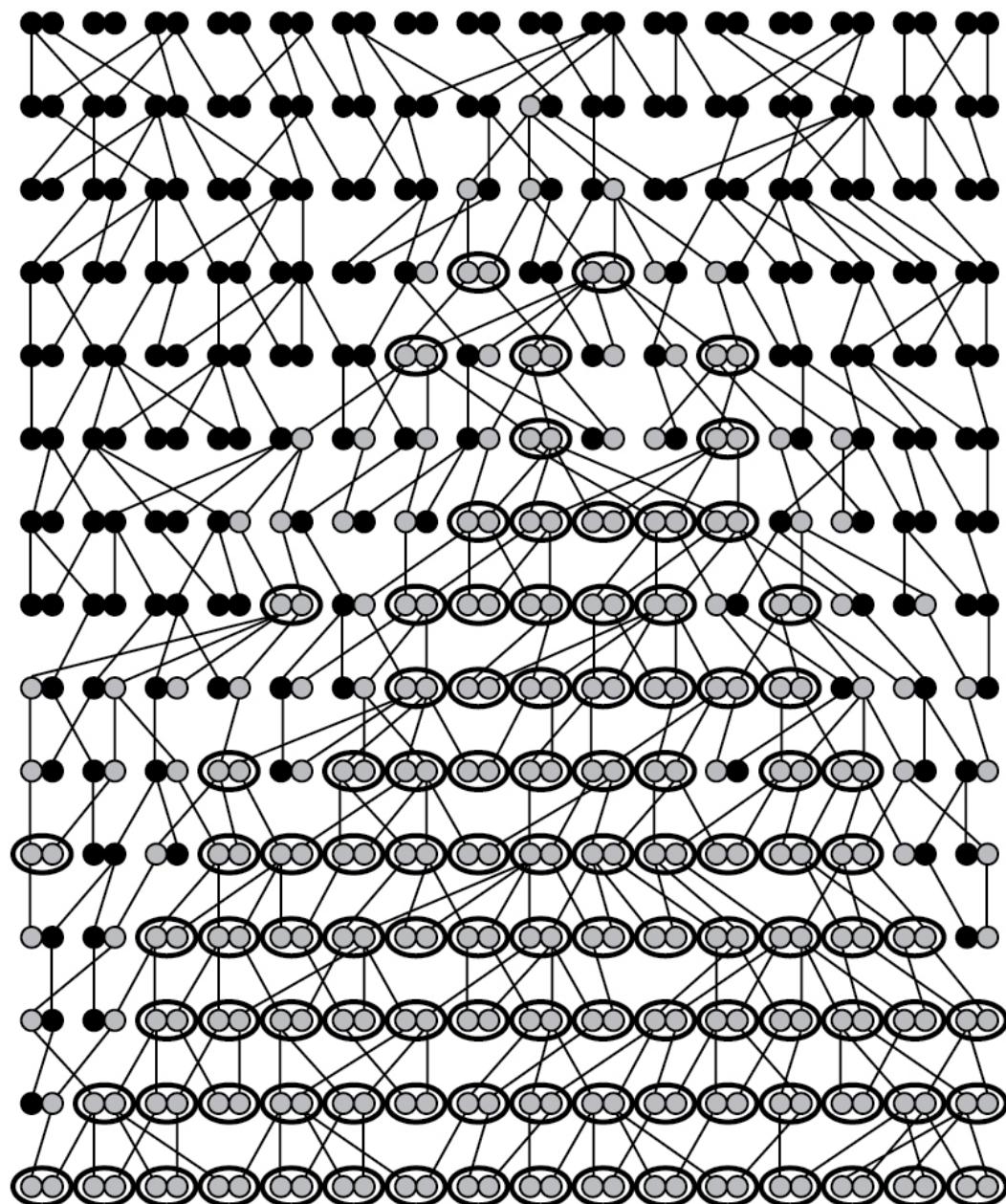
Generation 1 ATCCGGAAA

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

ATCCG~~A~~AAA

Modified from  
Baum and Smith  
Tree-Thinking  
book

# Descent with modification



ATCCGGAAA

Generation 1

Gen. 2

Gen. 10

Generation 15  
ATCCGAAAA

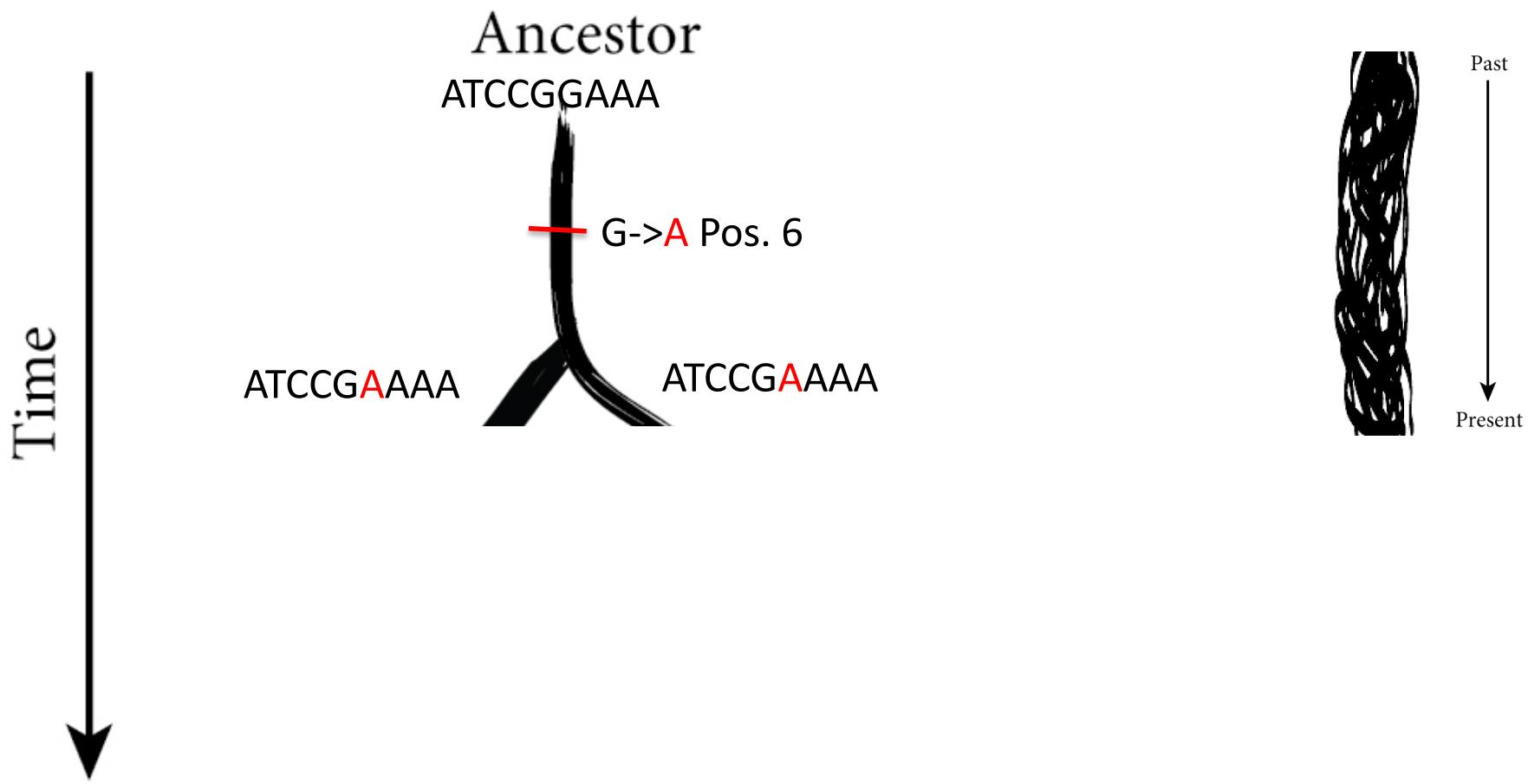
Past

ATCCGGAAA

ATCCG**G**AAA

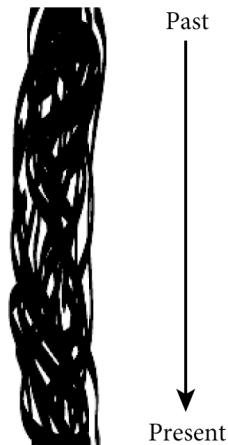
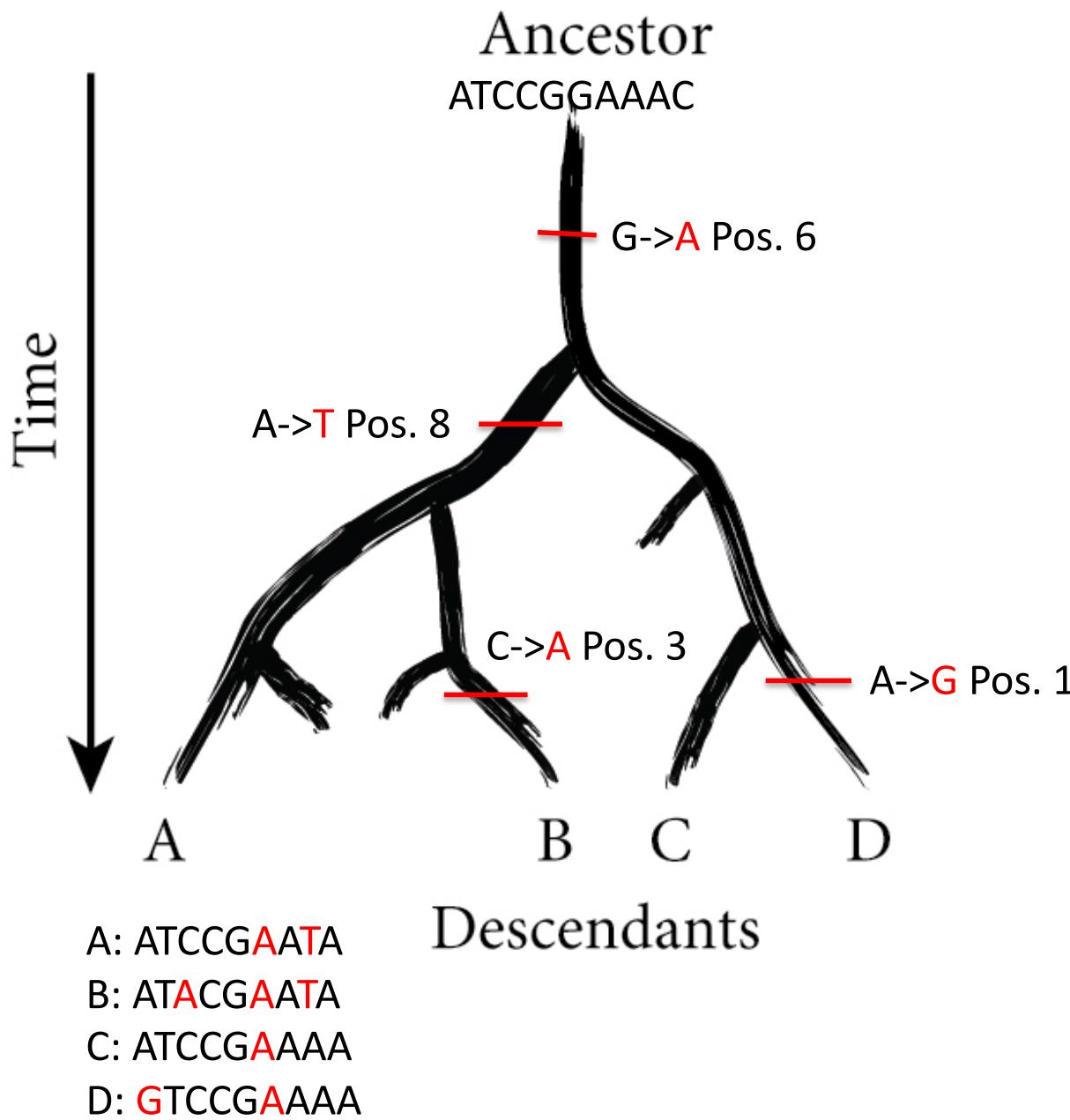
Present

Modified from  
Baum and Smith  
Tree-Thinking  
book



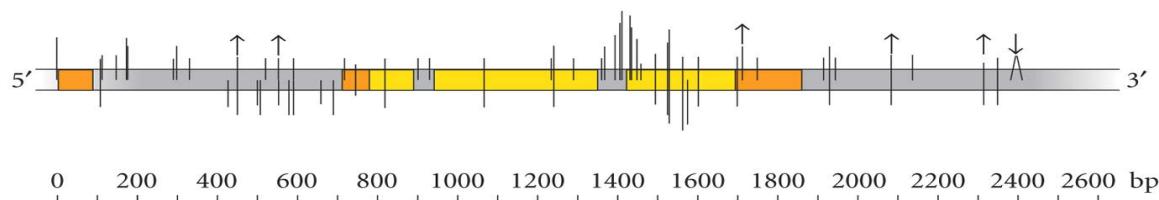
- A: ATCCG~~A~~AAA
- B: ATCCG~~A~~AAA
- C: ATCCG~~A~~AAA
- D: ATCCG~~A~~AAA

Modified from  
Baum and Smith  
Tree-Thinking  
book



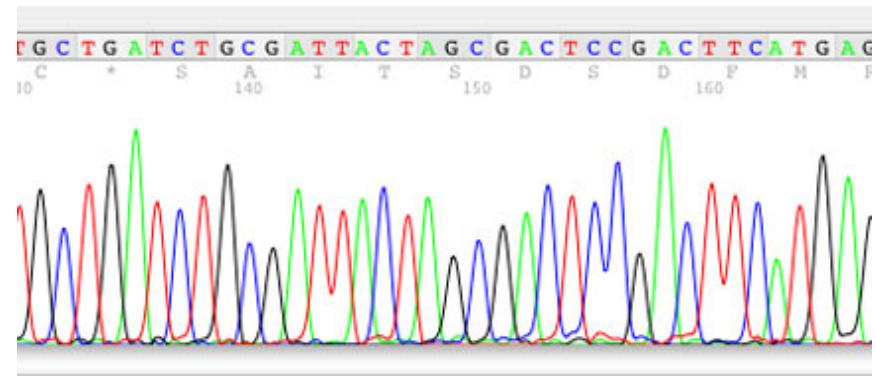
Modified from  
Baum and Smith  
Tree-Thinking  
book

# •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	CAT	GTA	TTT	ACG	GCT	GAT
ATG	CAG	CGC	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	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

A sample from a population

Species 1      Nonsyn

Syn

Syn

Syn

Syn

ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	GAT
ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	TAT
ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	TAT
ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	TAT
ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	C(T)	GTA	TTT	ACG	GCT	GAT
ATG	CAG	CGC	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	GAT
ATG	CAG	CGC	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCT	GAT
ATG	CAG	CGT	ATT	TCA	A	CAT	TTG	GGA	CAT	GTA	TTT	ACG	GCC	TAT
ATG	CGG	CGT	ATT	TCG	G	CAT	TTA	GGA	CAT	GTA	TTT	ACG	GCT	TAT

Species 2

M    Q/R    R    I    S    H    L    G    R/L    V    P    S    A    Y/D

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.**

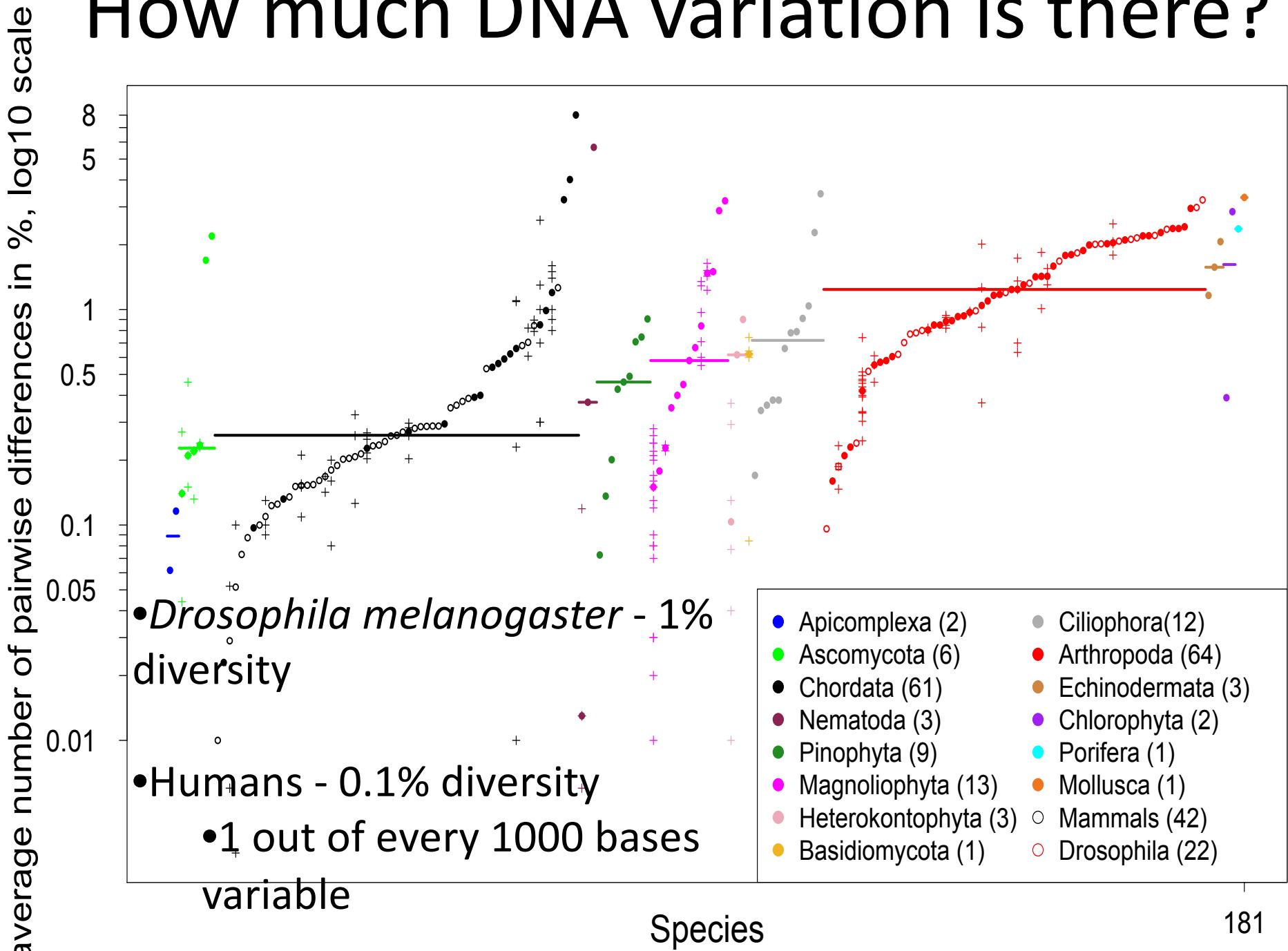
**Number of segregating sites.**

**Heterozygosity:** Fraction of all sites where an individual is heterozygous

**Pairwise Diversity:**

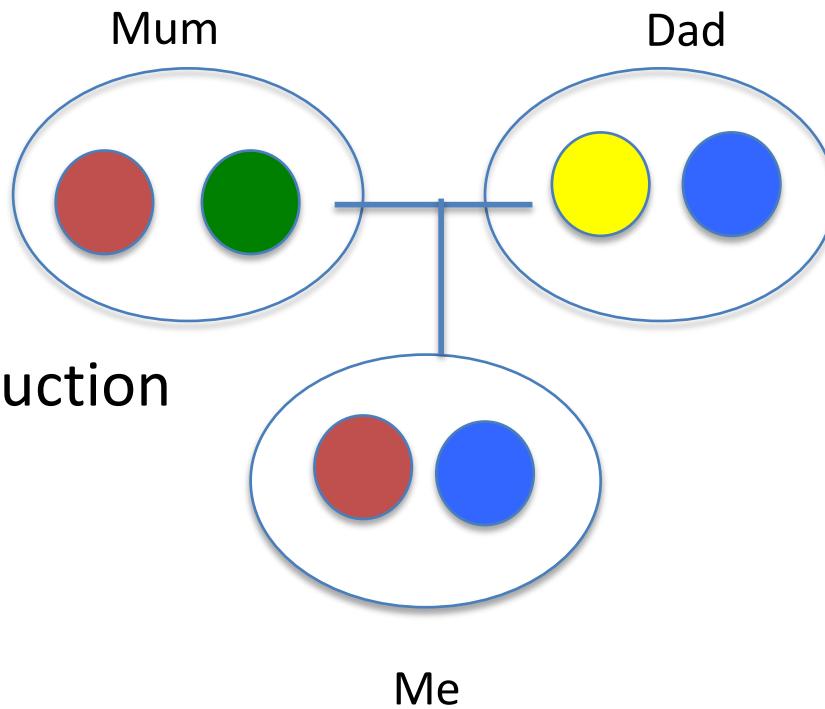
Fraction of sites that differ between two sequences chosen at random

# How much DNA variation is there?



- 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?

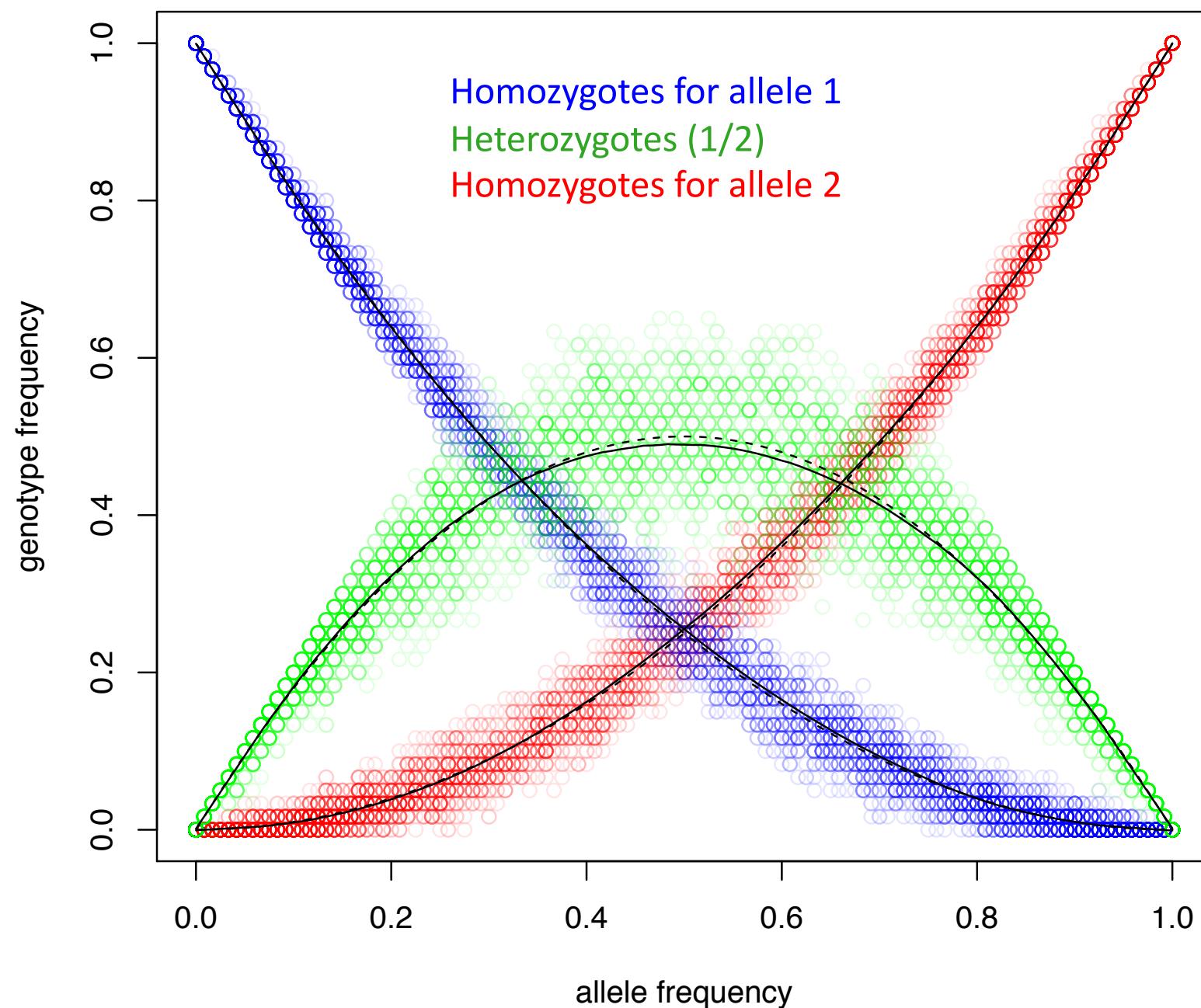
What do I have to assume?

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!

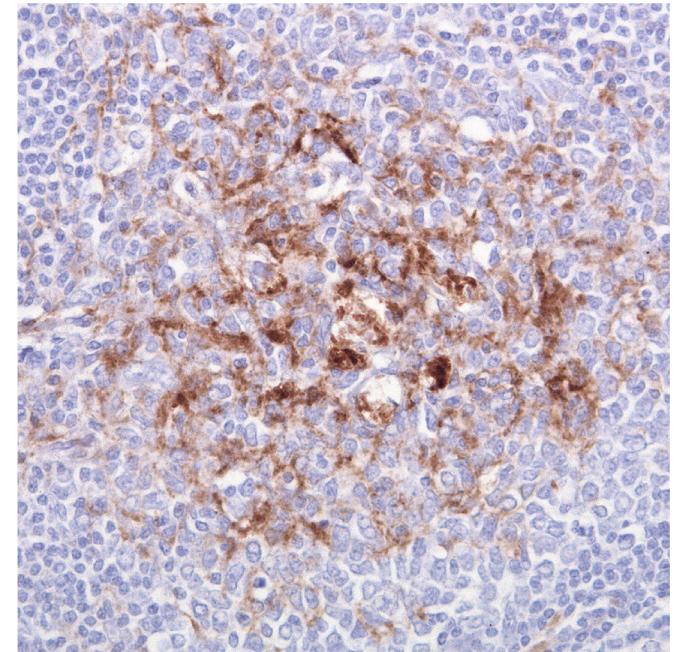
Therefor the Hardy Weinberg expectations/proportions are an equilibrium if nothing interesting happens in popgen.

# 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.

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



[https://en.wikipedia.org/wiki/Creutzfeldt%E2%80%93Jakob\\_disease#/media/File:VCJD\\_Tonsil.jpg](https://en.wikipedia.org/wiki/Creutzfeldt%E2%80%93Jakob_disease#/media/File:VCJD_Tonsil.jpg)

Chi-squared statistic significance =0.0034