

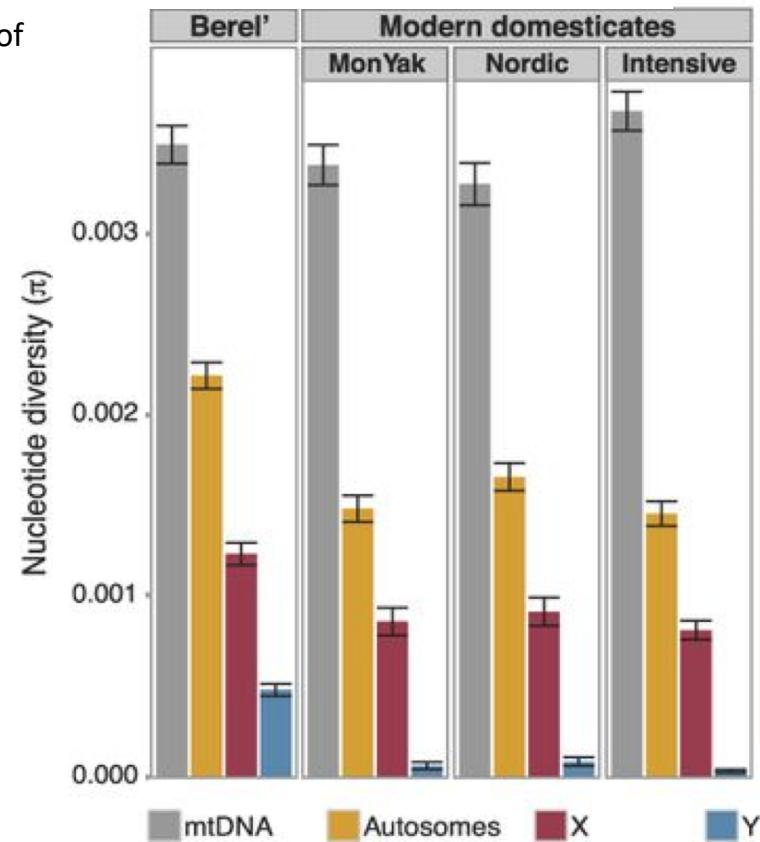
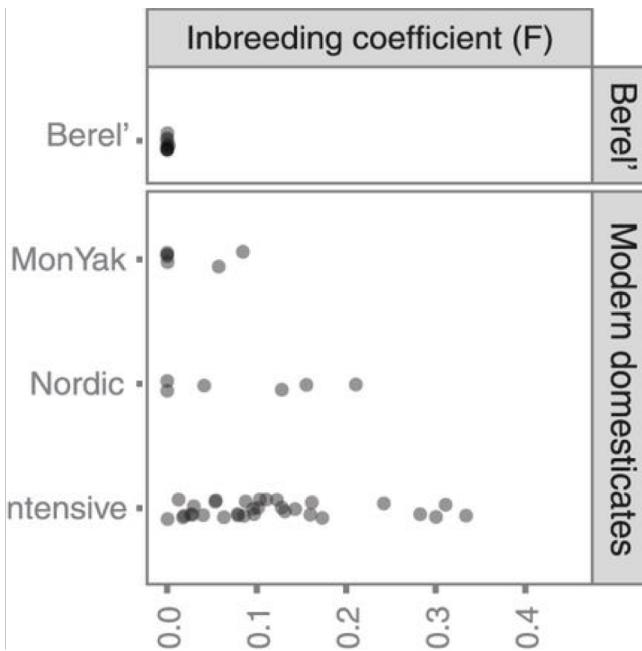


# Ancient genomic changes associated with domestication of the horse

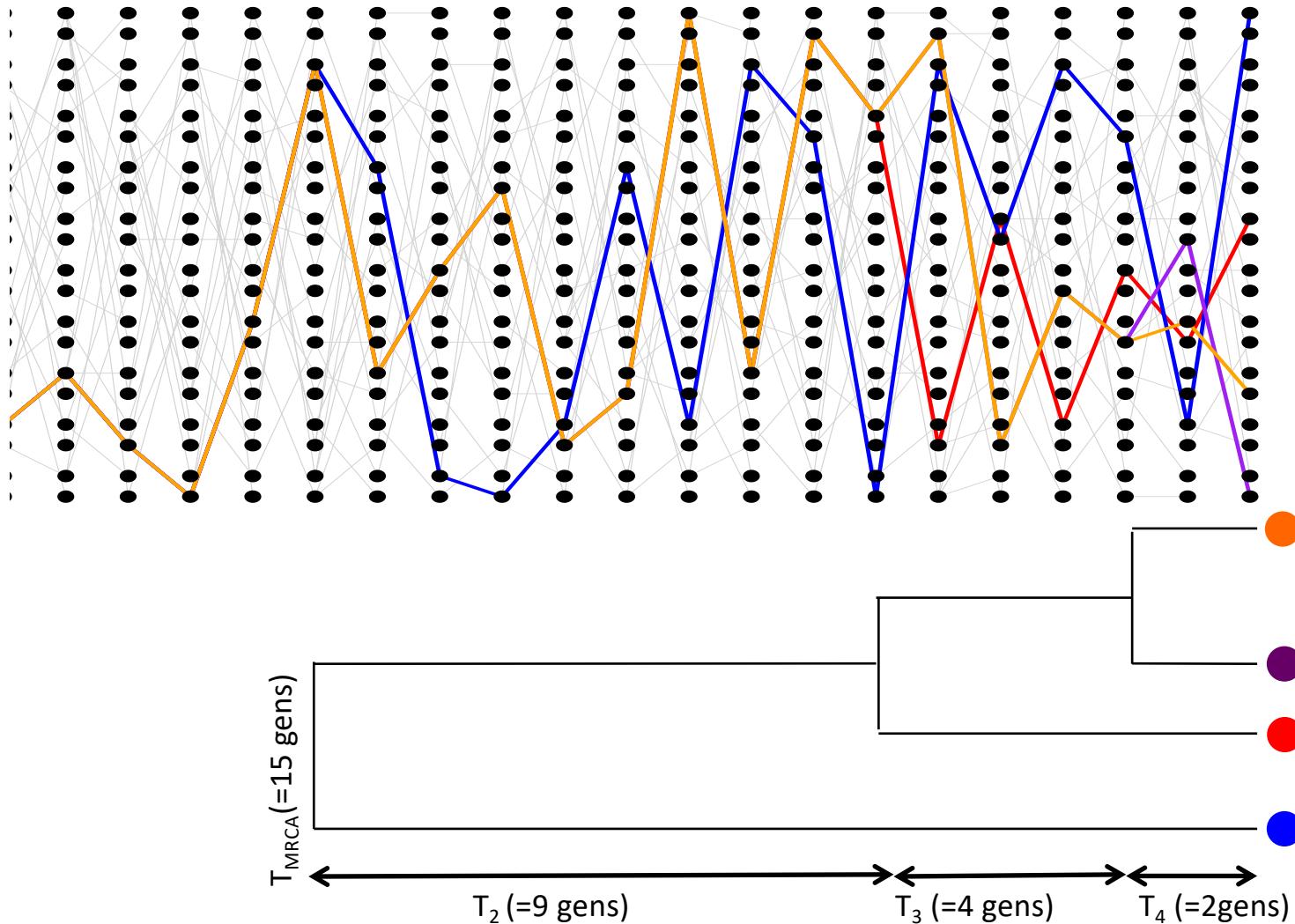
Pablo Librado<sup>1,\*</sup>, Cristina Gamba<sup>1,\*</sup>, Charleen Gaunitz<sup>1</sup>, Clio Der Sarkissian<sup>1</sup>, Mélanie Pruvost<sup>2</sup>, Anders Albrechtsen<sup>3</sup>,

\* See all authors and affiliations

13 sacrificed Scythian stallions from the kurgan 11 of Berel', Kazakhstan, dated to ~2.3 ky



## The coalescent for sample sizes greater than two



## The coalescent for sample sizes greater than two

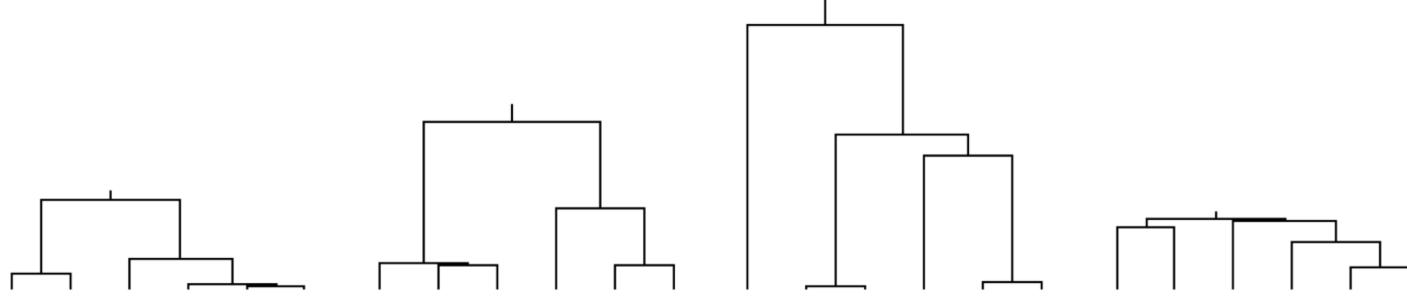


Figure 4: Four realizations of the coalescent for  $n = 6$ , drawn on the same scale (the labels 1–6 should be assigned randomly to the tips).

There is little information about  
underlying processes in a single genealogy

From Nordborg 2000

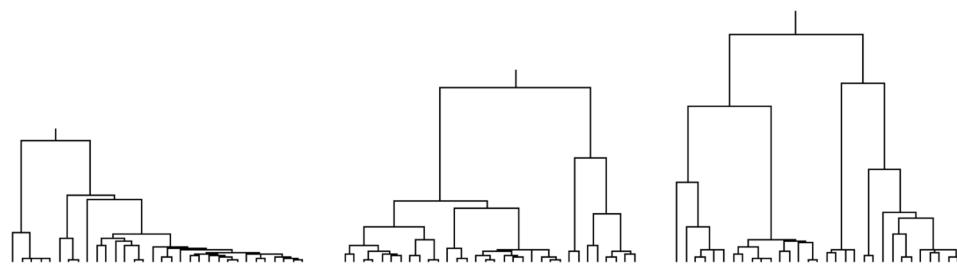
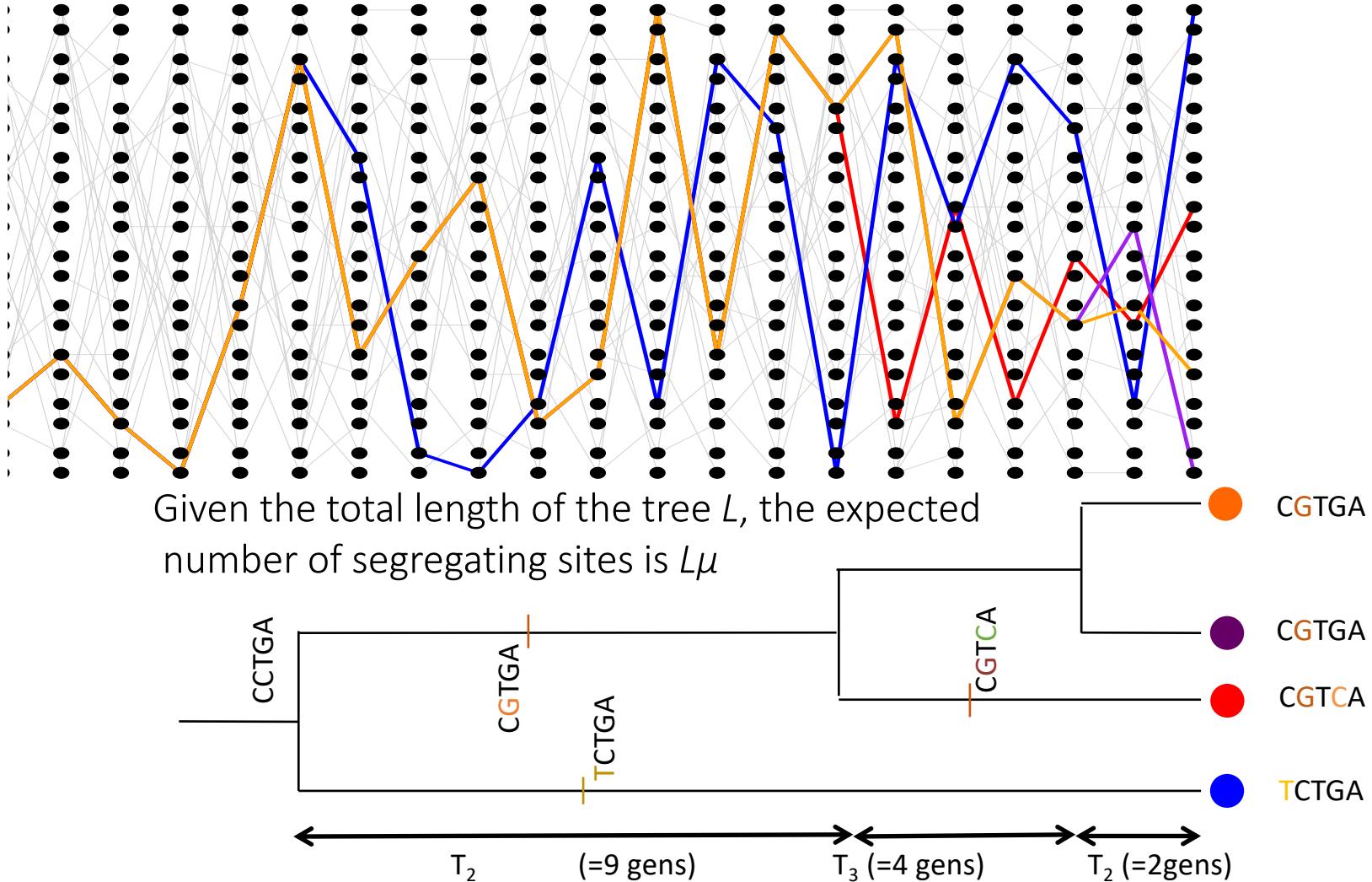


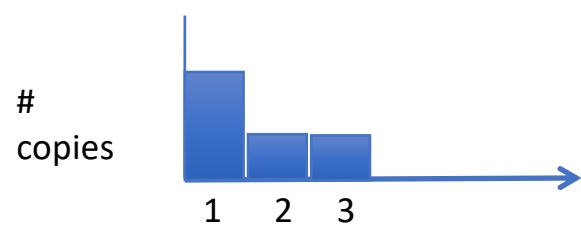
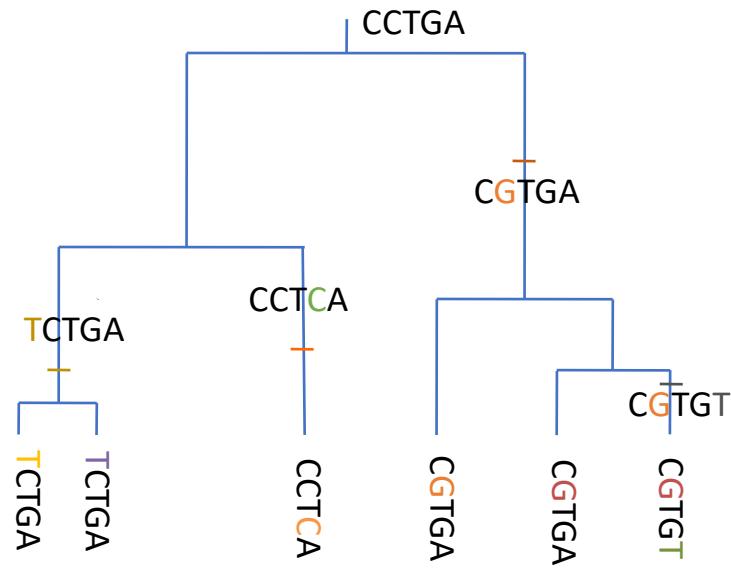
Figure 5: Three realizations of the coalescent for  $n = 32$ , drawn on the same scale (the labels 1–32 should be assigned randomly to the tips). From Nordborg review

→ You don't increase the information about the tree much from increasing your  
sample size

## The coalescent for sample sizes greater than two



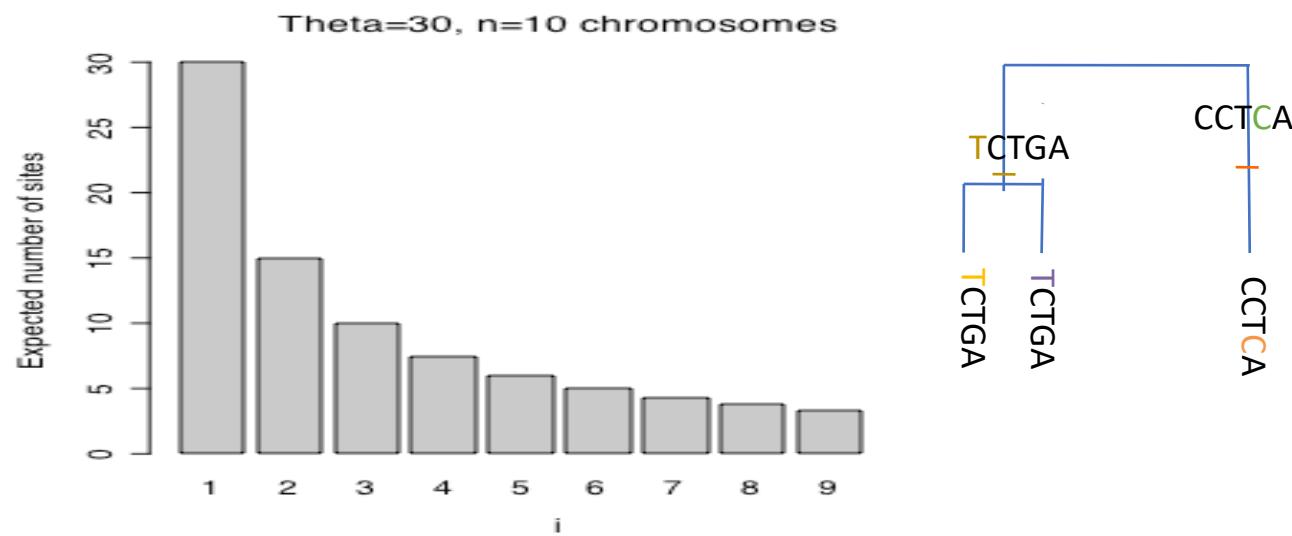
The frequency of an allele in the sample reflects the branch in the genealogy on which the branch occurred.



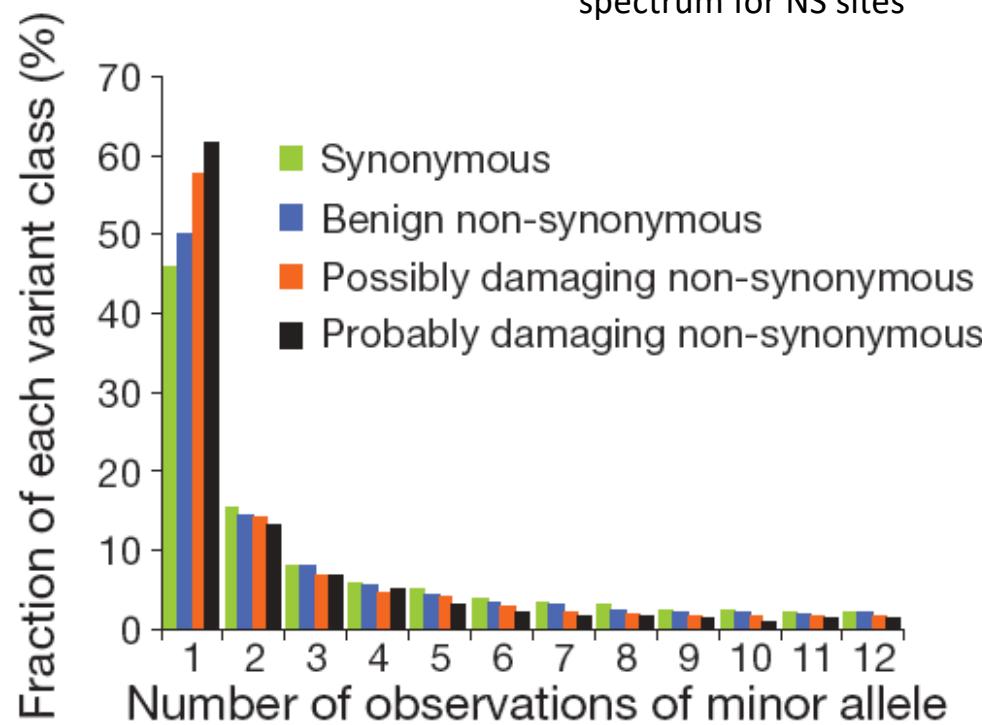
## Frequency spectrum for derived neutral alleles

Sample size	Theoretical spectrum (singletons, doubletons, ...)
2	$\theta$
3	$\theta, \theta/2$
4	$\theta, \theta/2, \theta/3$
5	$\theta, \theta/2, \theta/3, \theta/4$

- General solution: Expected count of sites at frequency  $i$  is equal to  $\frac{\theta}{i}$ .



A balance between mutation, drift and Selection could explain our frequency spectrum for NS sites



## **Targeted capture and massively parallel sequencing of 12 human exomes**

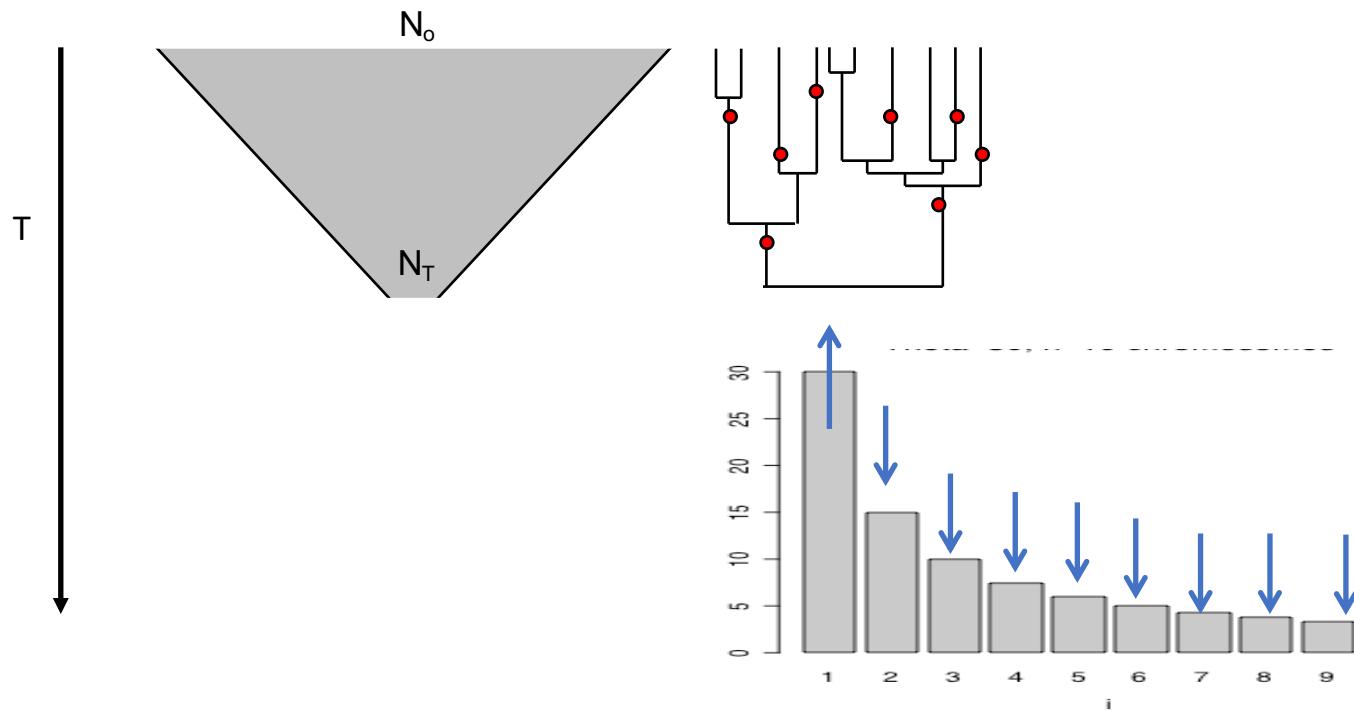
Sarah B. Ng<sup>1</sup>, Emily H. Turner<sup>1</sup>, Peggy D. Robertson<sup>1</sup>, Steven D. Flygare<sup>1</sup>, Abigail W. Bigham<sup>2</sup>, Choli Lee<sup>1</sup>, Tristan Shaffer<sup>1</sup>, Michelle Wong<sup>1</sup>, Arindam Bhattacharjee<sup>4</sup>, Evan E. Eichler<sup>1,3</sup>, Michael Bamshad<sup>2</sup>, Deborah A. Nickerson<sup>1</sup> & Jay Shendure<sup>1</sup>

What does the genealogy look like in a recently expanded population?

Slower coal. rate towards the tips, compared with a stationary population.

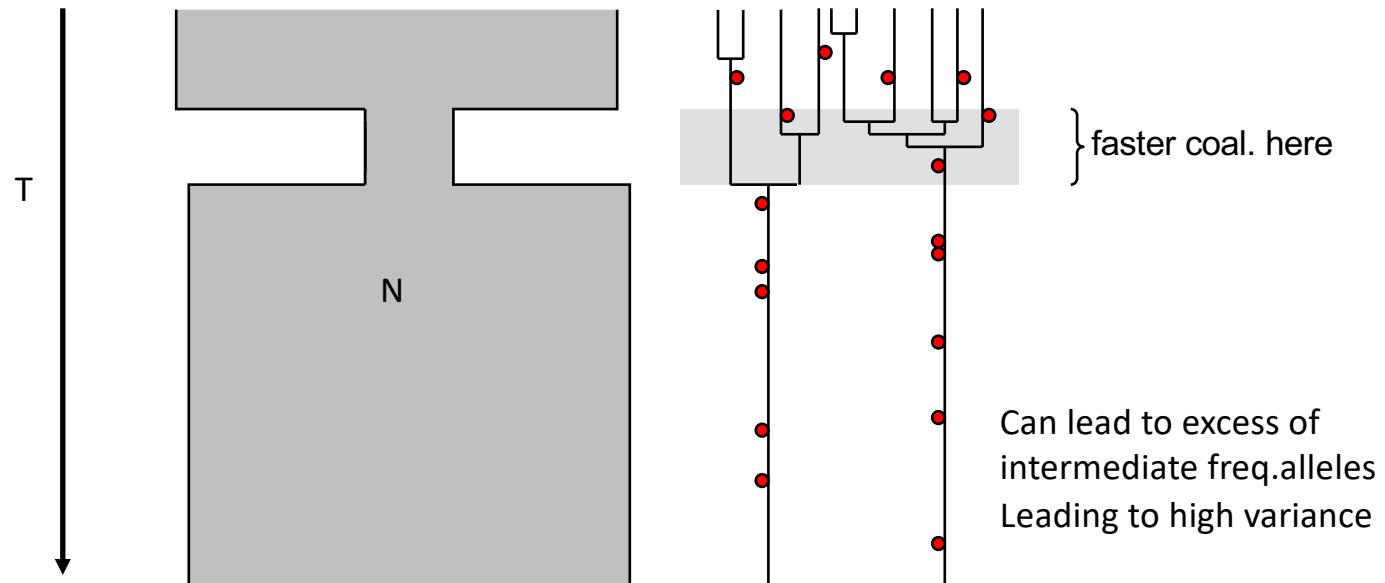
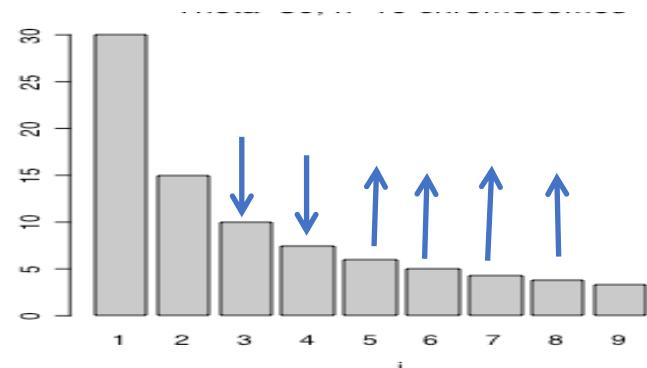
Excess of singletons

A growing population



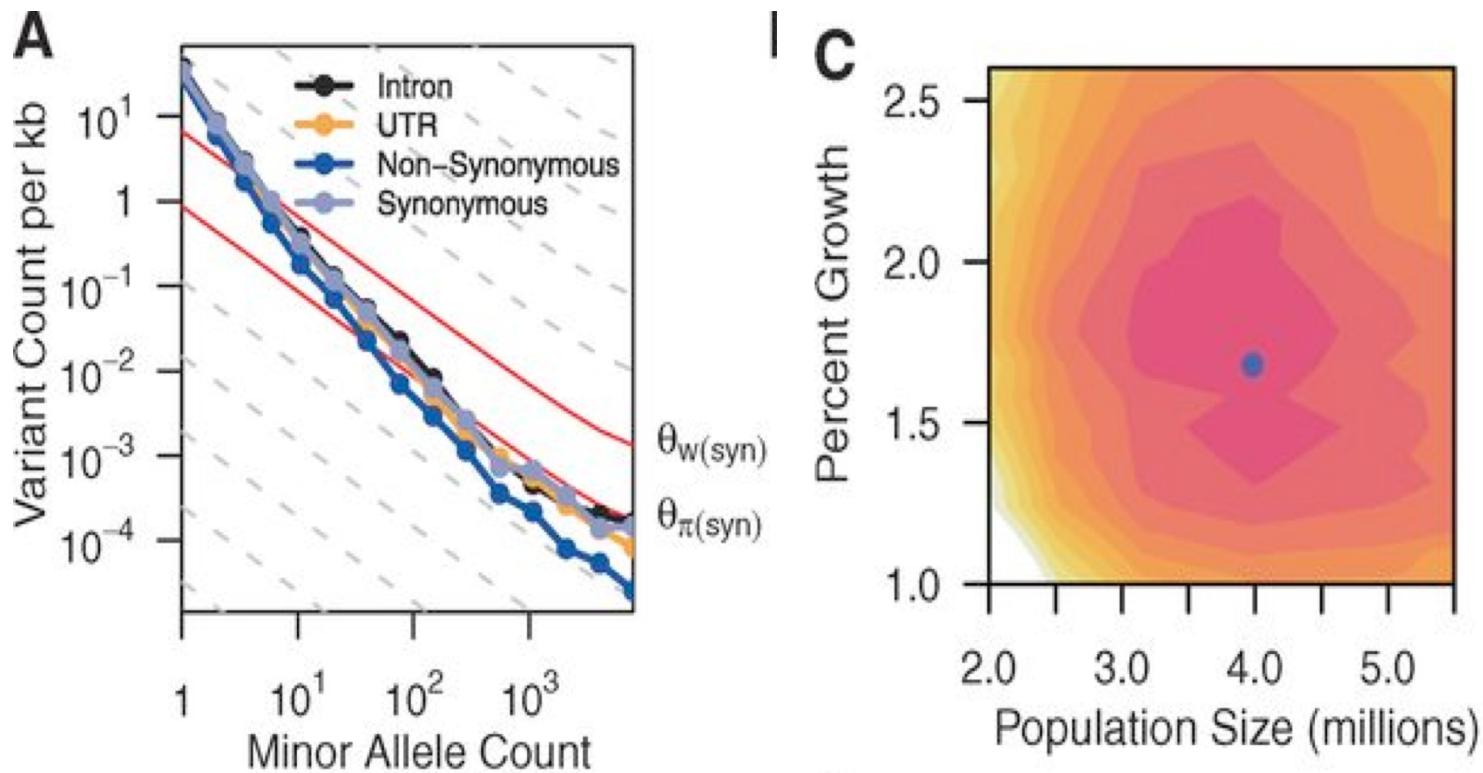
What does the genealogy look like if there was a recent bottleneck?

A bottlenecked population



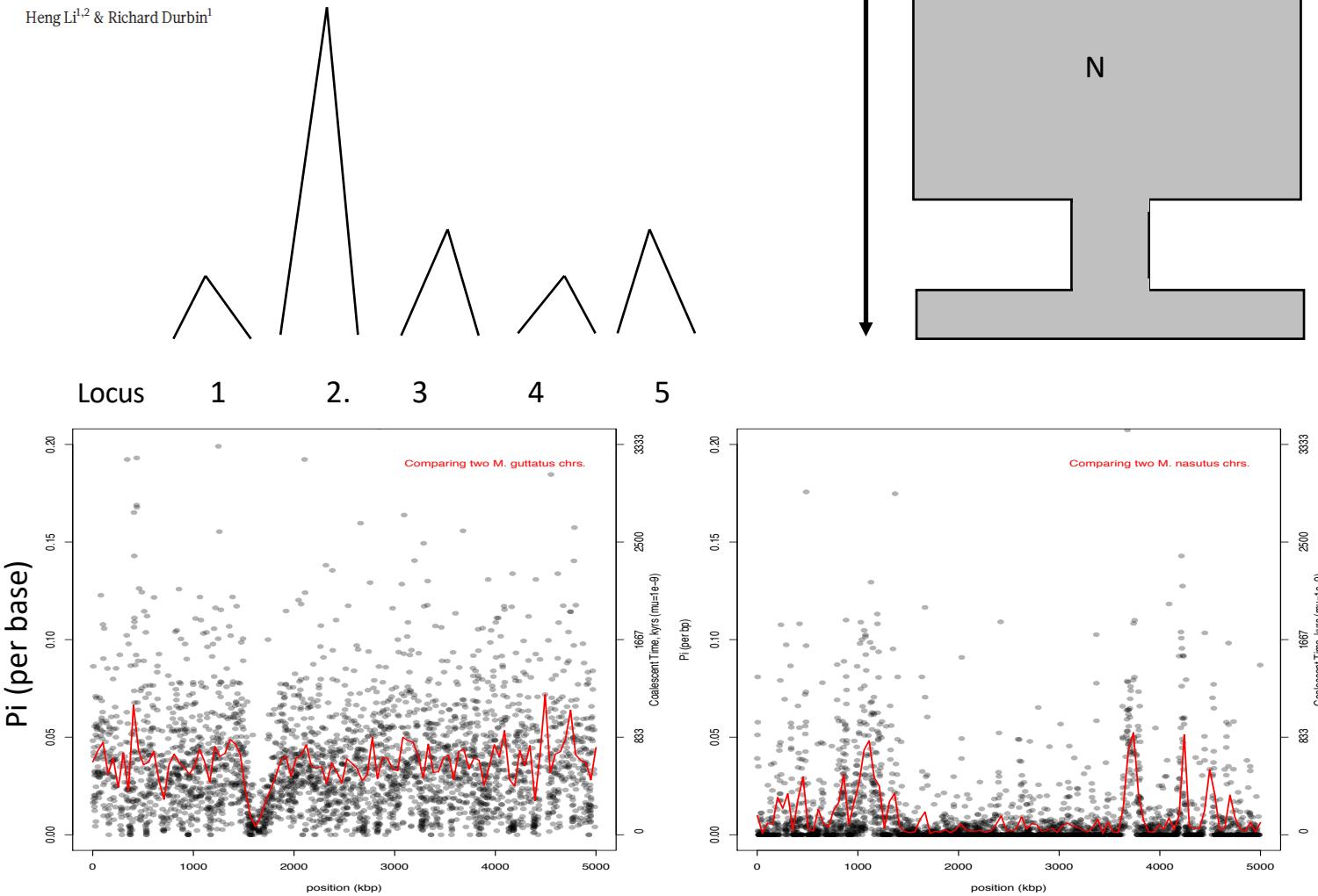
# An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People

Matthew R. Nelson,<sup>1,\*†</sup> Daniel Wegmann,<sup>2,\*</sup> Margaret G. Ehm,<sup>1</sup> Darren Kessner,<sup>2</sup> Pamela St. Jean,<sup>1</sup> Claudio Verzilli,<sup>3</sup> Judong Shen,<sup>1</sup> Zhengzheng Tang,<sup>4</sup> Silviu-Alin Bacanu,<sup>1</sup> Dana Fraser,<sup>1</sup> Liting Warren,<sup>1</sup> Jennifer Aponte,<sup>1</sup> Matthew Zawistowski,<sup>5</sup> Xiao Liu,<sup>6</sup> Hao Zhang,<sup>6</sup> Yong Zhang,<sup>6</sup> Jun Li,<sup>7</sup> Yun Li,<sup>4</sup> Li Li,<sup>1</sup> Peter Woollard,<sup>3</sup> Simon Topp,<sup>3</sup> Matthew D. Hall,<sup>3</sup> Keith Nangle,<sup>1</sup> Jun Wang,<sup>6,8</sup> Gonçalo Abecasis,<sup>5</sup> Lon R. Cardon,<sup>9</sup> Sebastian Zöllner,<sup>5,10</sup> John C. Whittaker,<sup>3</sup> Stephanie L. Chissoe,<sup>1</sup> John Novembre,<sup>2,†‡</sup> Vincent Mooser<sup>9,‡</sup>



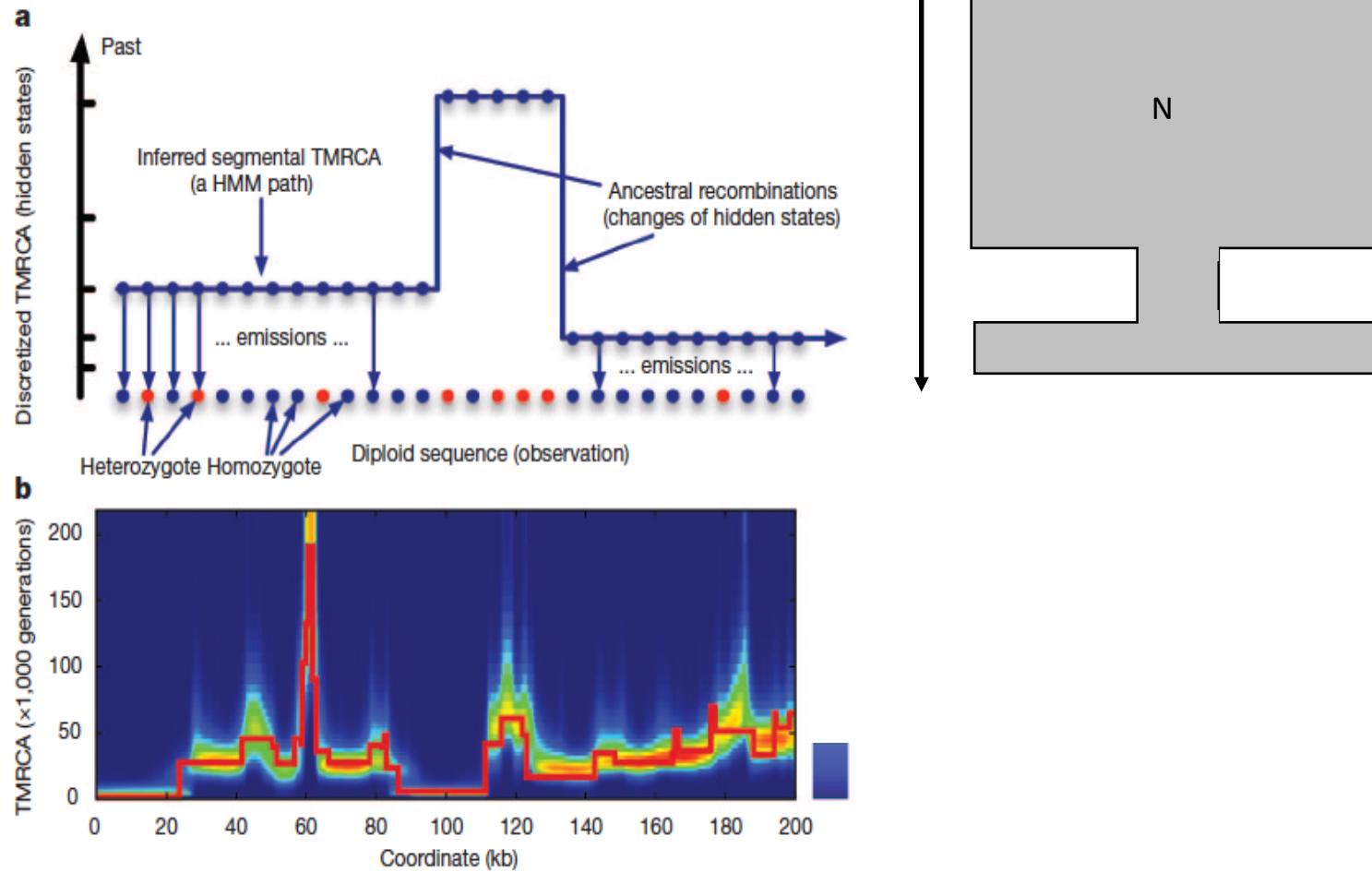
# Inference of human population history from individual whole-genome sequences

Heng Li<sup>1,2</sup> & Richard Durbin<sup>1</sup>



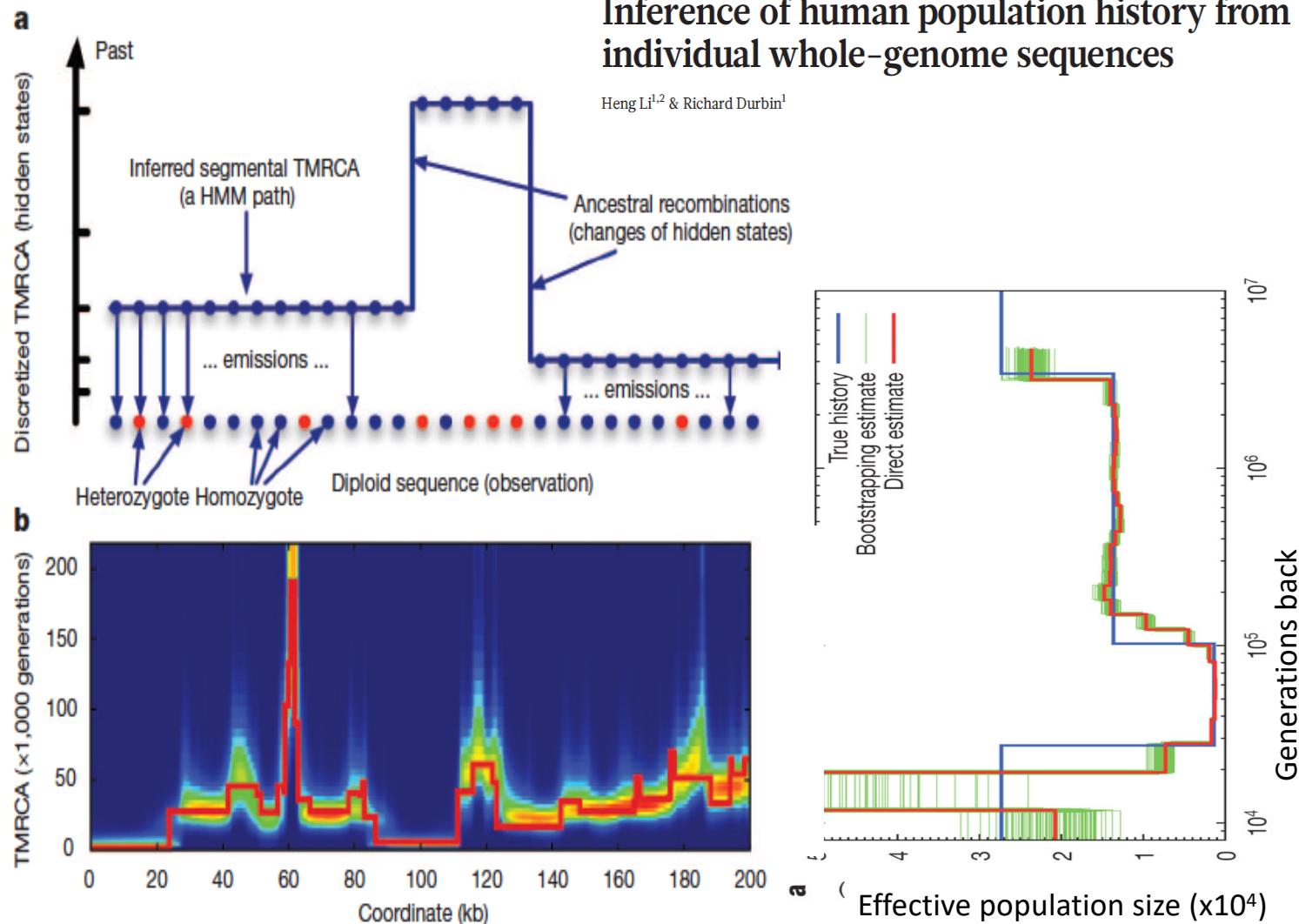
# Inference of human population history from individual whole-genome sequences

Heng Li<sup>1,2</sup> & Richard Durbin<sup>1</sup>



# Inference of human population history from individual whole-genome sequences

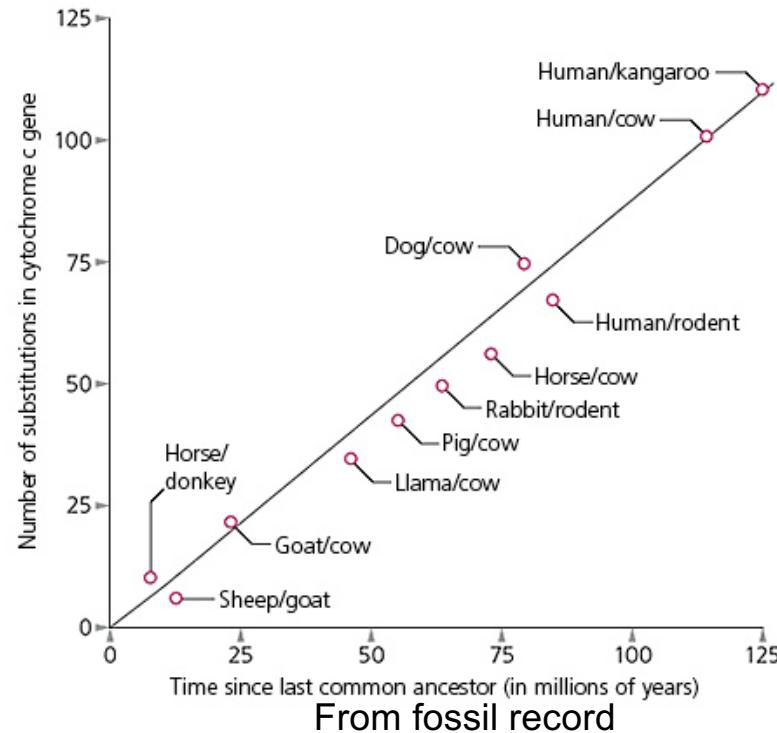
Heng Li<sup>1,2</sup> & Richard Durbin<sup>1</sup>



# The Molecular Clock

**Observation:** Rate of **amino acid** substitution in many, but not all, proteins surprisingly constant over time, i.e. evolve in a clock-like manner.

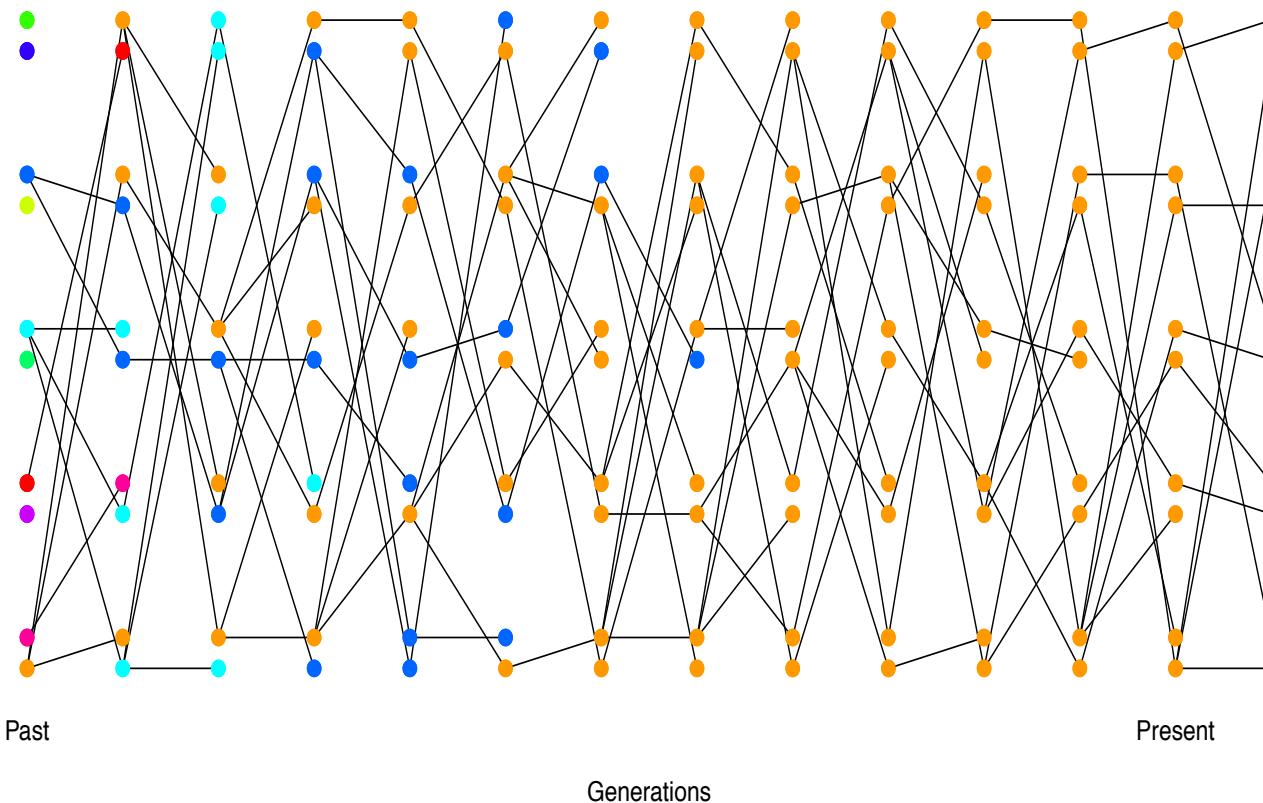
- Neutral theory Claim:** Due to most amino acid replacement substitutions between species being neutral



From fossil record

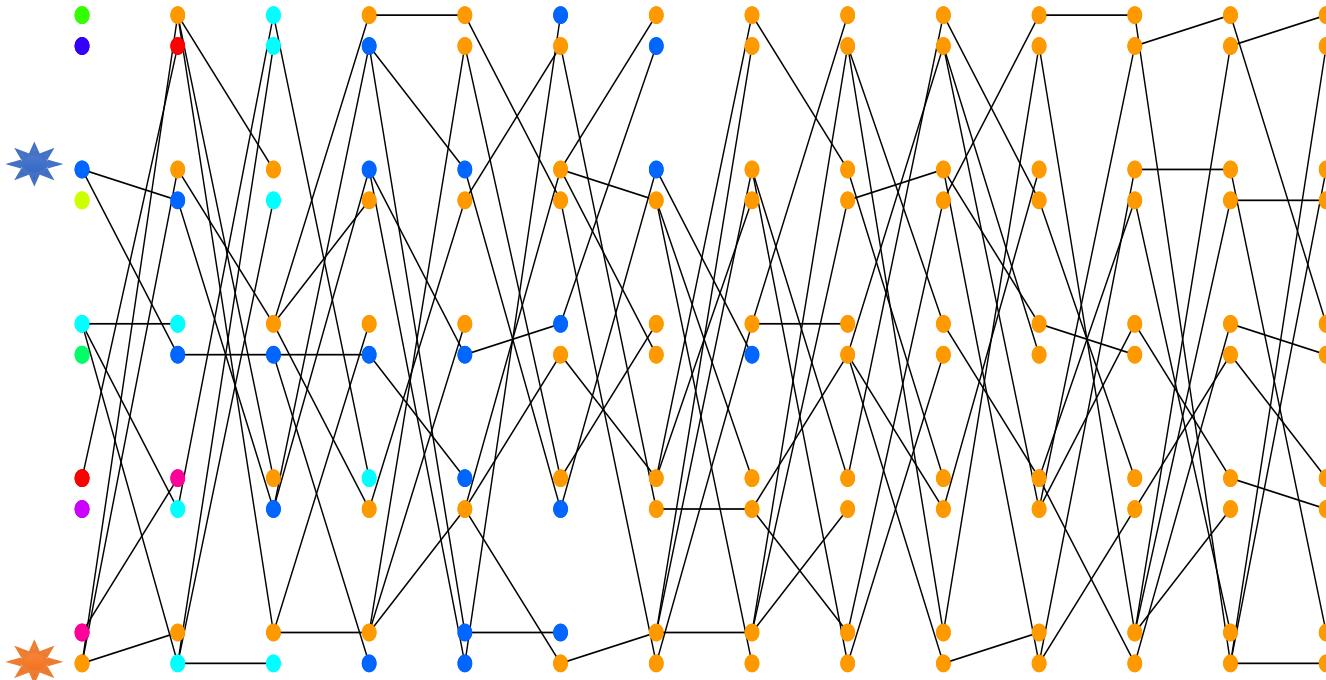
Zimmer book

# Neutral evolution and a molecular clock



there are  $2N$  alleles in our population

# Neutral evolution and a molecular clock



The probability that all individuals are descended from a particular neutral allele at a locus is  $1/(2N)$

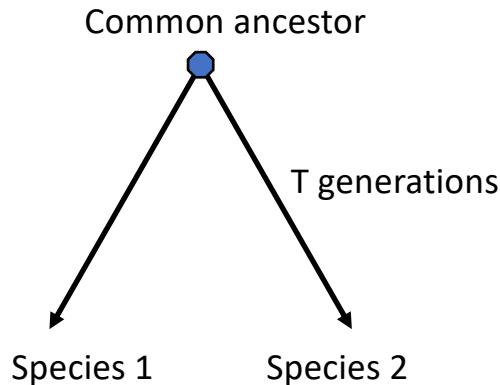
Each generation there are  $2N\mu$  new mutations per site

- Substitution rate per generation =  $2N\mu \times 1/[2N] = \mu$  per generation

Independent of population size!

# Neutral evolution and a molecular clock

- Substitution rate per generation =  $2N\mu \times 1/[2N] = \mu$  per generation

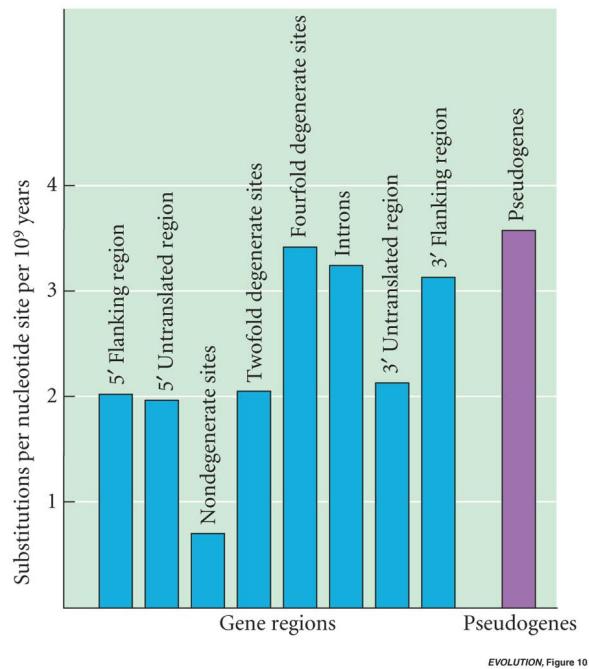


The expected number of neutral substitutions =  $2T \mu$

i.e. substitutions occur at a linear rate, a molecular clock.  
Also gives a way to indirectly estimate mutations from divergence if  $T$  is known.

# Levels of constraint

- Neutral theory claims most new mutations are deleterious and are lost immediately. Only neutral mutations contribute to substitution.

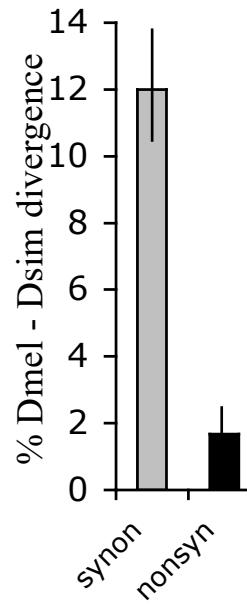


Fut. 10.14

Consistent with neutral theory  
slower rate of substitution at  
more constrained sites.

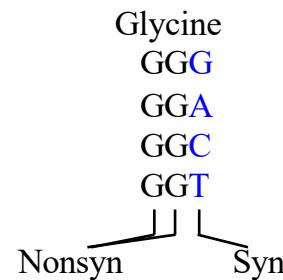
Variation in divergence (substitution rate) across  
classes of nucleotide sites (human vs. rodent)

Selectively constrained (functional) sites evolve more slowly than less functional sites. (C=constraint)



Andolfatto Nature  
2005

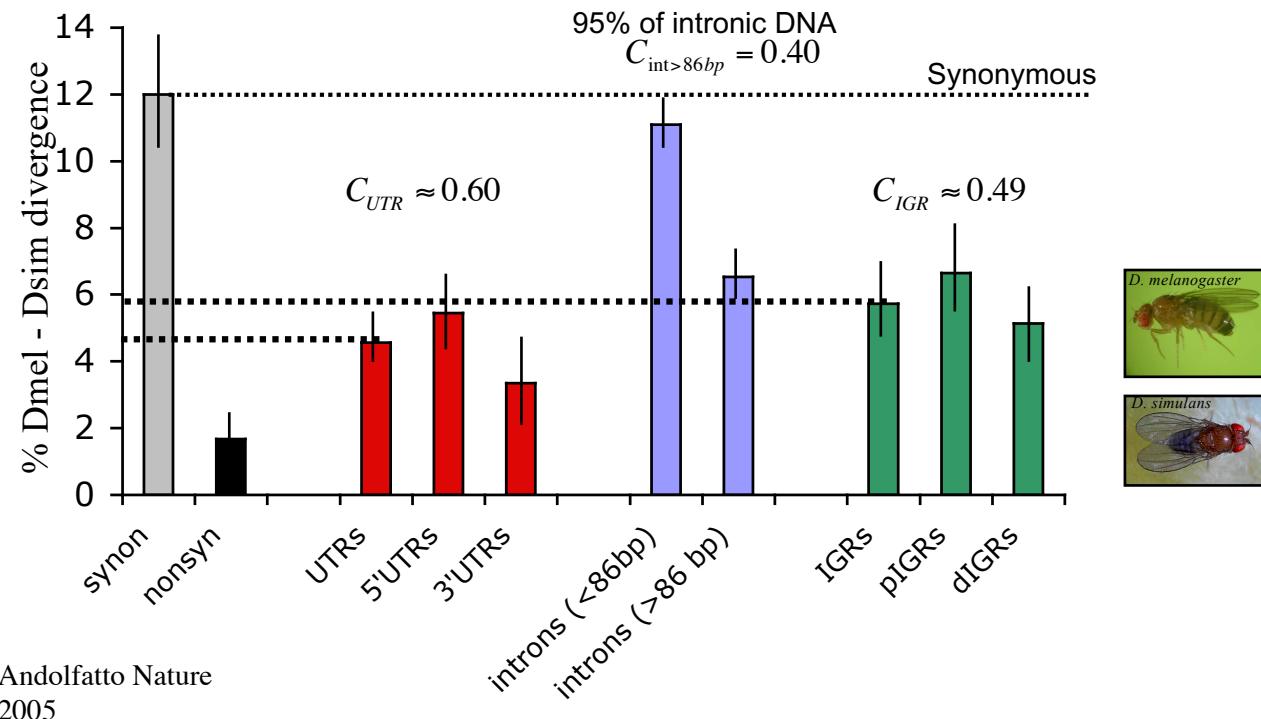
e.g. amino-acid coding sites versus synonymous sites



$$C_{NonSyn} = 1 - \frac{D_{obs}}{D_{exp}} = 1 - \frac{D_{nonsyn}}{D_{syn}} = 0.865$$

Assuming all synonymous substitutions are neutral we estimate that 86.5% of mutations at nonsynonymous sites are deleterious enough to have been removed by selection

Most non-coding DNA evolves slower than synonymous sites in the *D. melanogaster* group



Implication: 40-70% of non-coding sites are constrained by selection