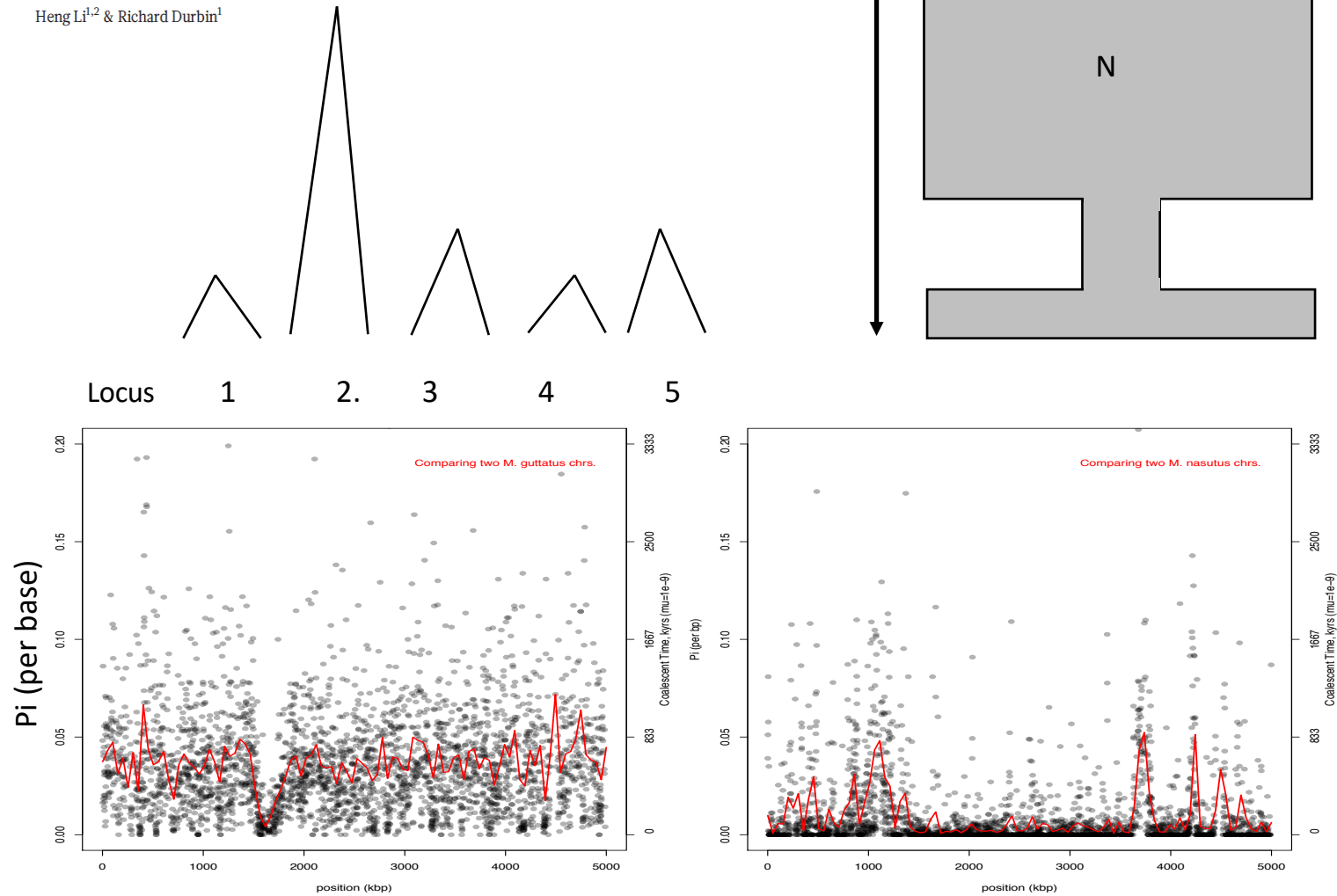


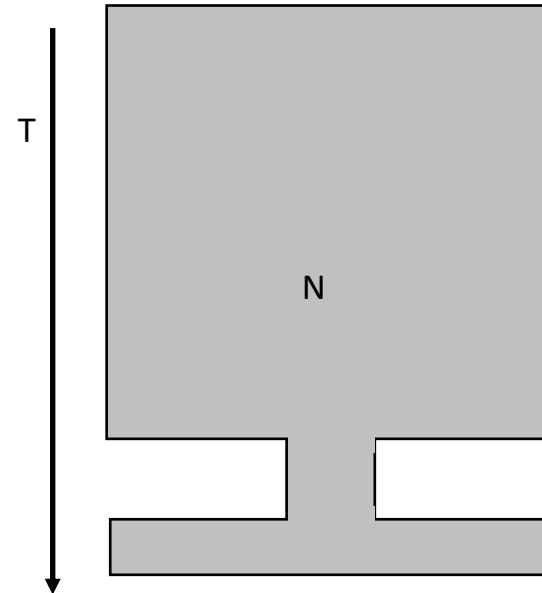
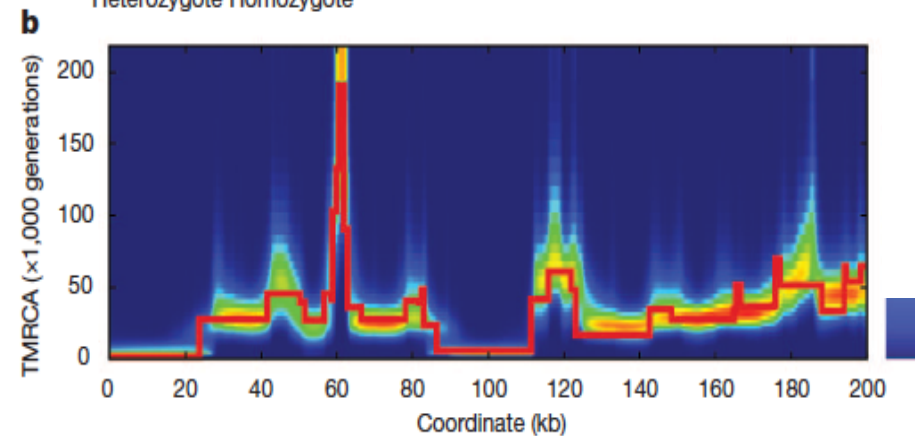
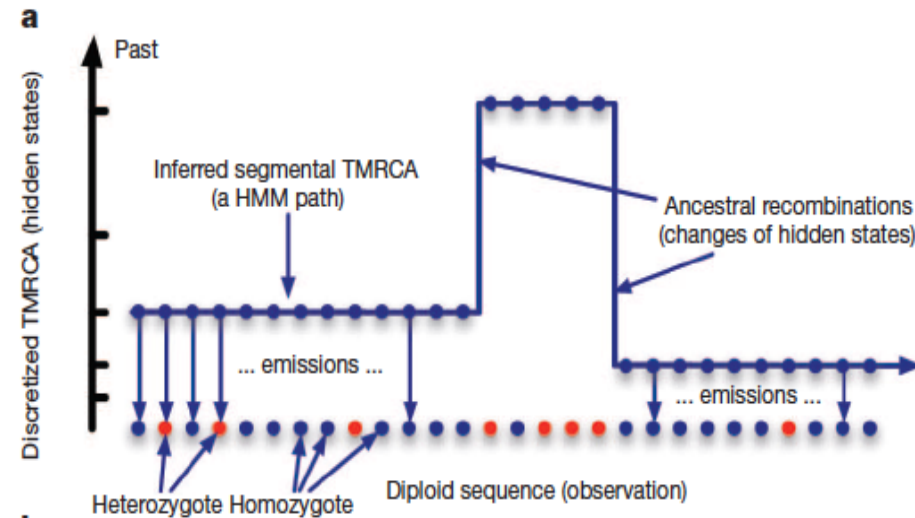
Inference of human population history from individual whole-genome sequences

Heng Li^{1,2} & Richard Durbin¹



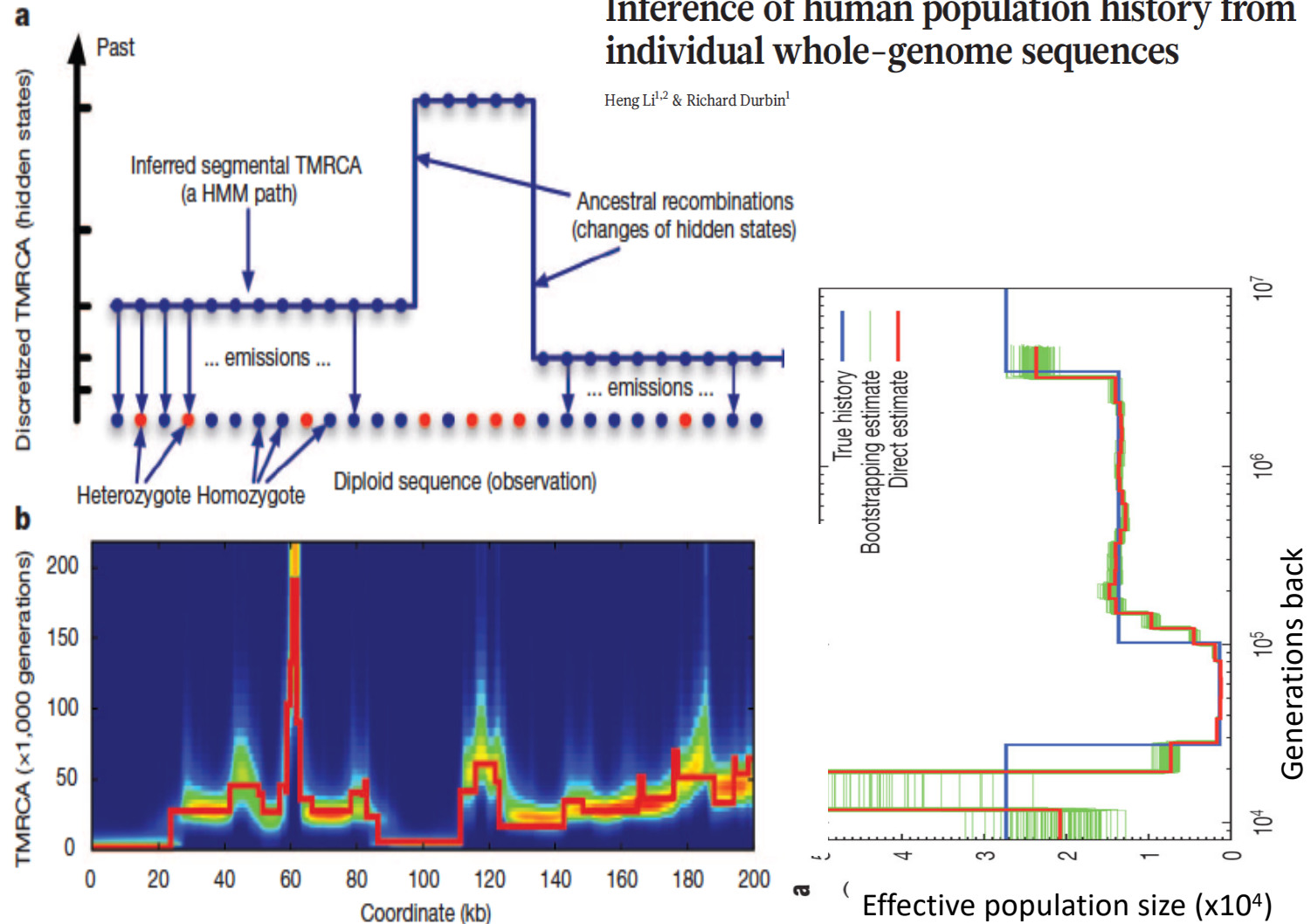
Inference of human population history from individual whole-genome sequences

Heng Li^{1,2} & Richard Durbin¹



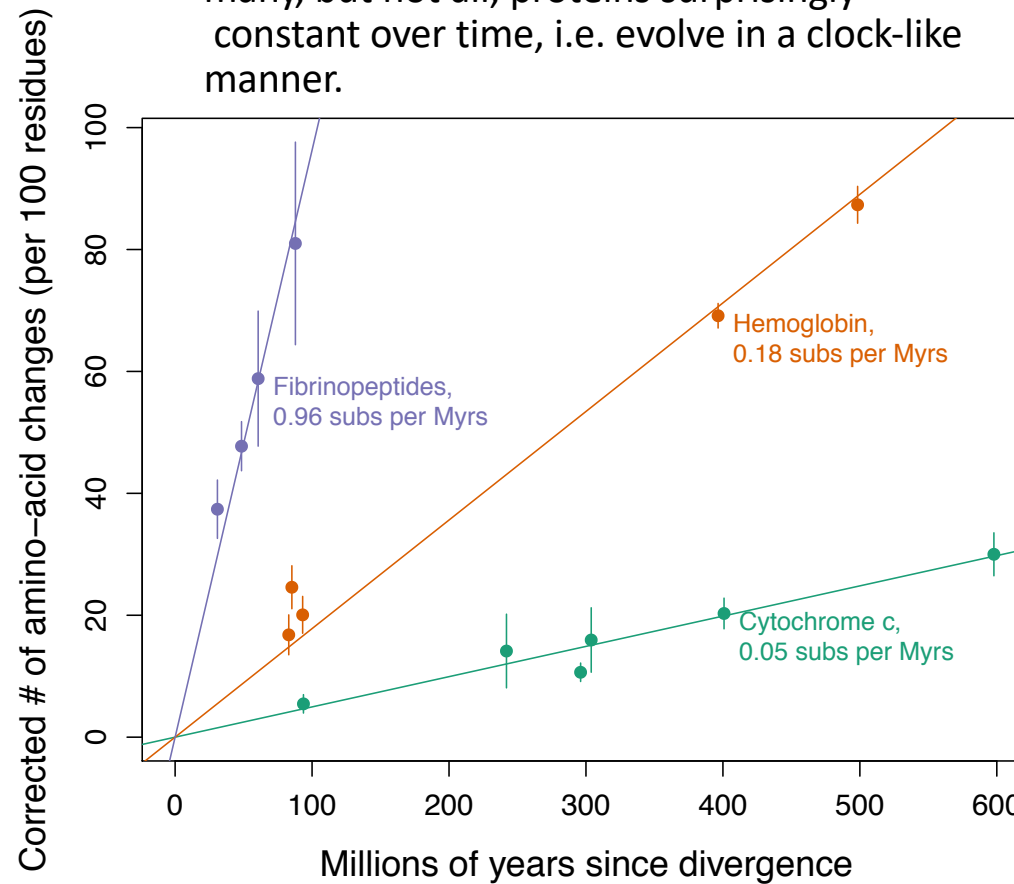
Inference of human population history from individual whole-genome sequences

Heng Li^{1,2} & Richard Durbin¹



The Molecular Clock

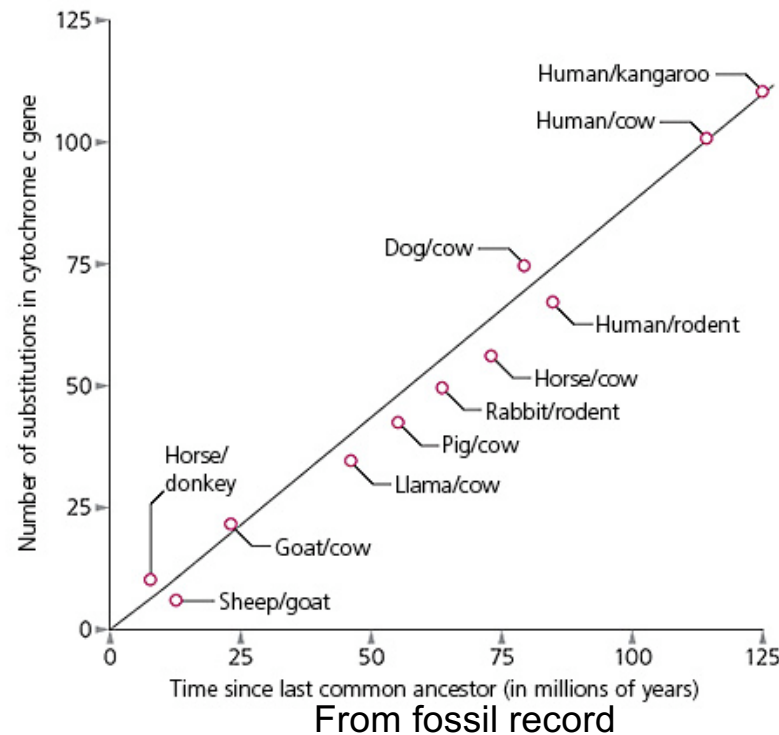
Observation: Rate of **amino acid** substitution in many, but not all, proteins surprisingly constant over time, i.e. evolve in a clock-like manner.



The Molecular Clock

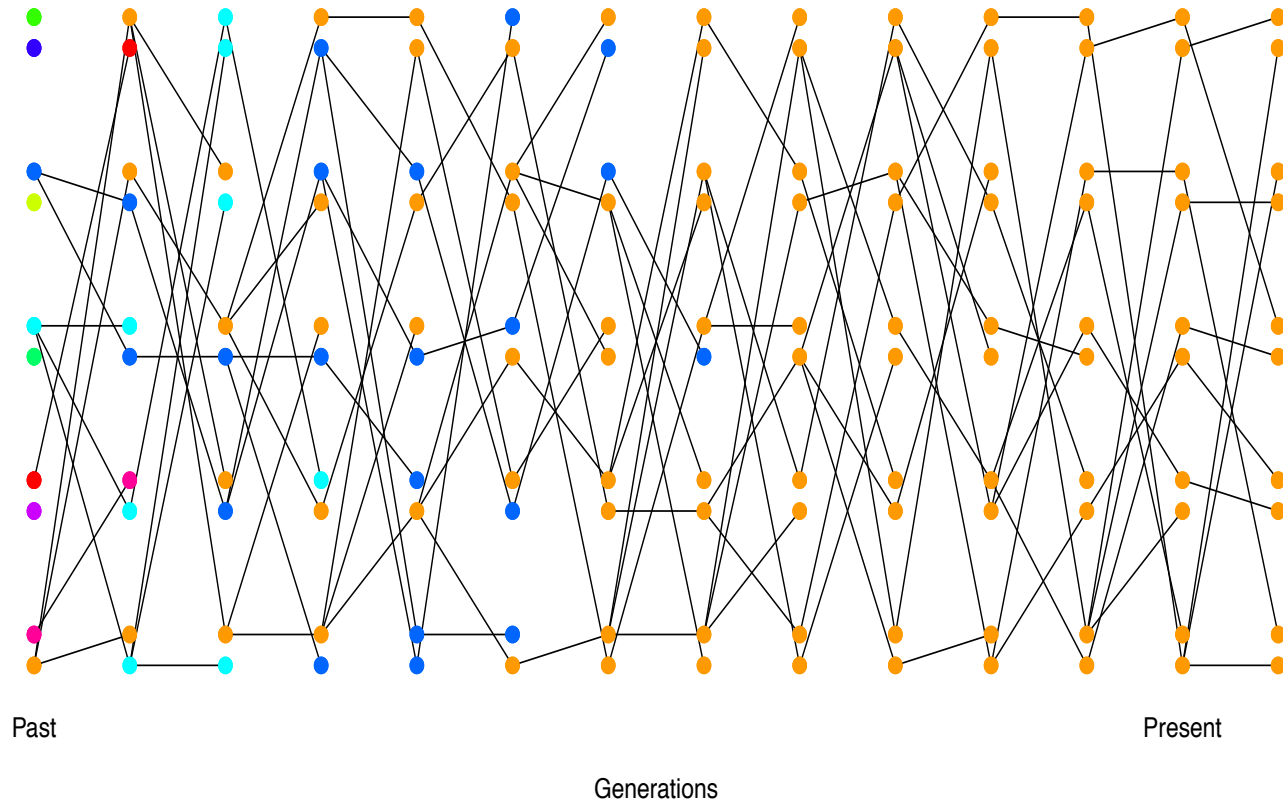
Observation: Rate of **amino acid** substitution in many, but not all, proteins surprisingly constant over time, i.e. evolve in a clock-like manner.

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



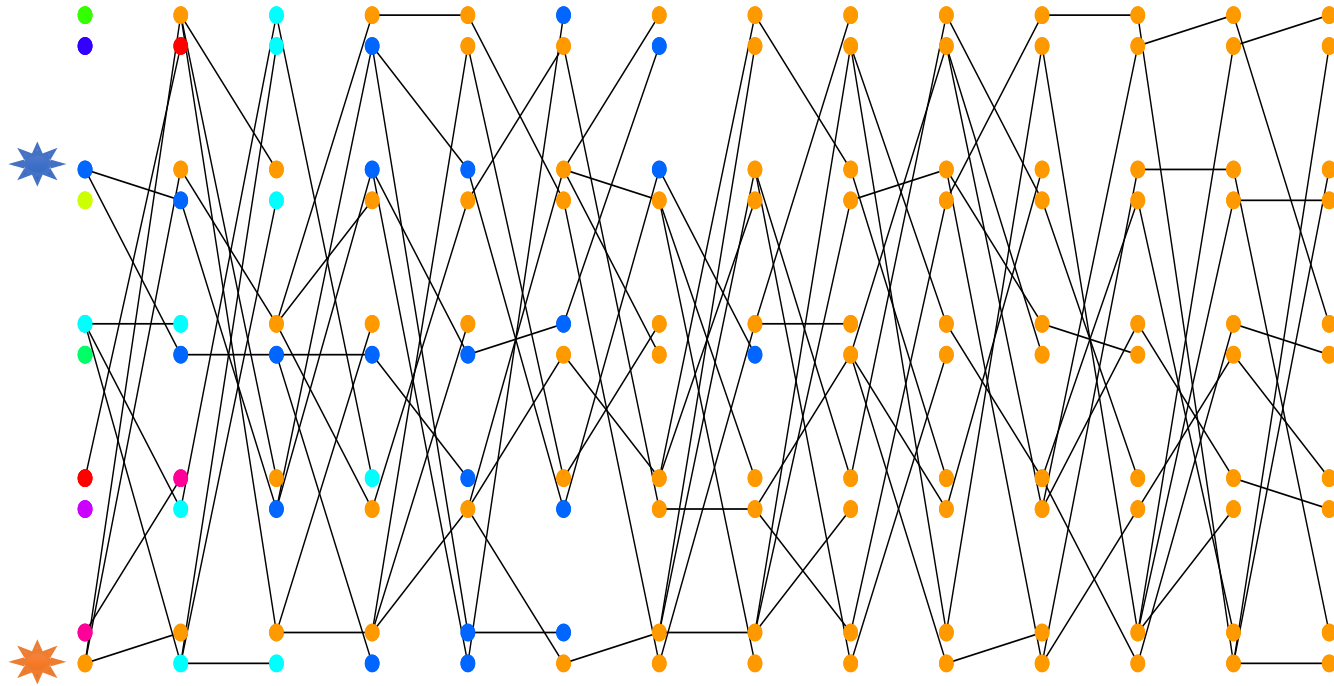
Zimmer book

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 $1/(2N)$

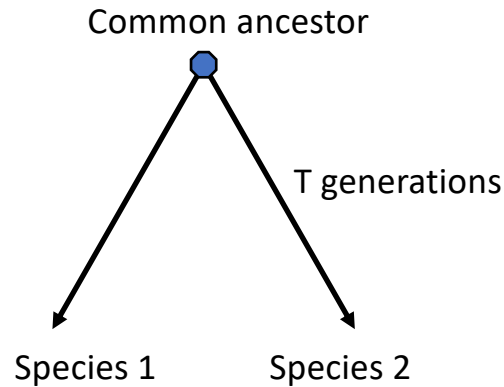
Each generation there are $2N\mu$ new mutations per site

- 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

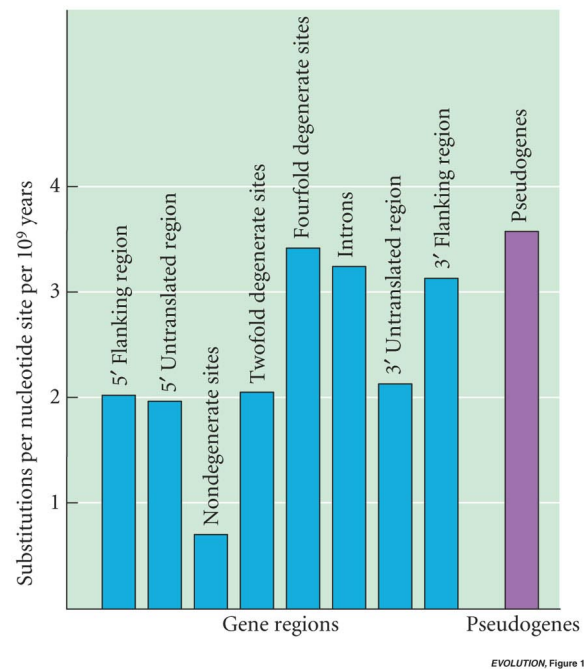


The expected number of neutral substitutions = $2T \mu$

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.

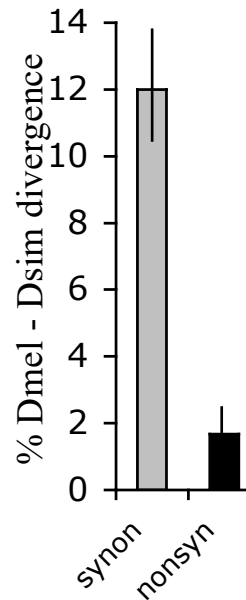


Consistent with neutral theory
slower rate of substitution at
more constrained sites.

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

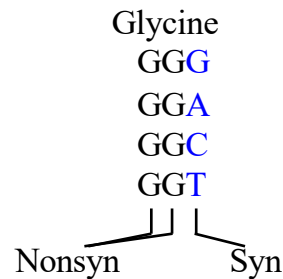
Fut. 10.14

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

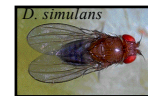


Andolfatto Nature
2005

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

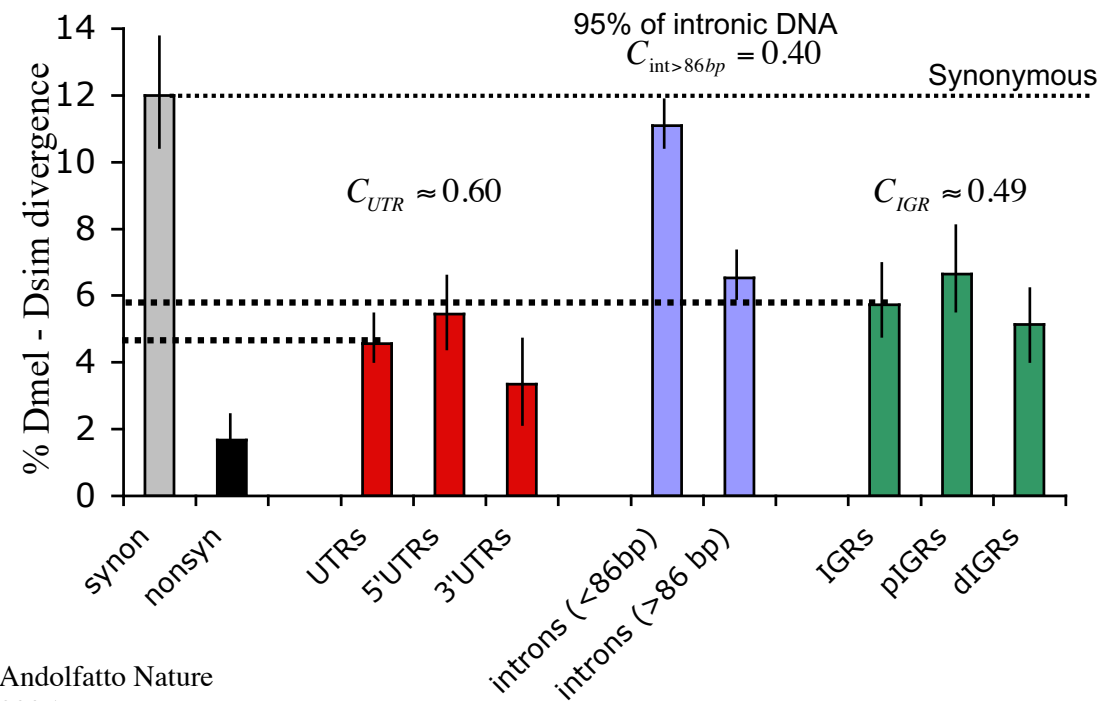


$$C_{NonSyn} = 1 - \frac{D_{obs}}{D_{exp}} = 1 - \frac{D_{nonsyn}}{D_{syn}} = 0.865$$



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

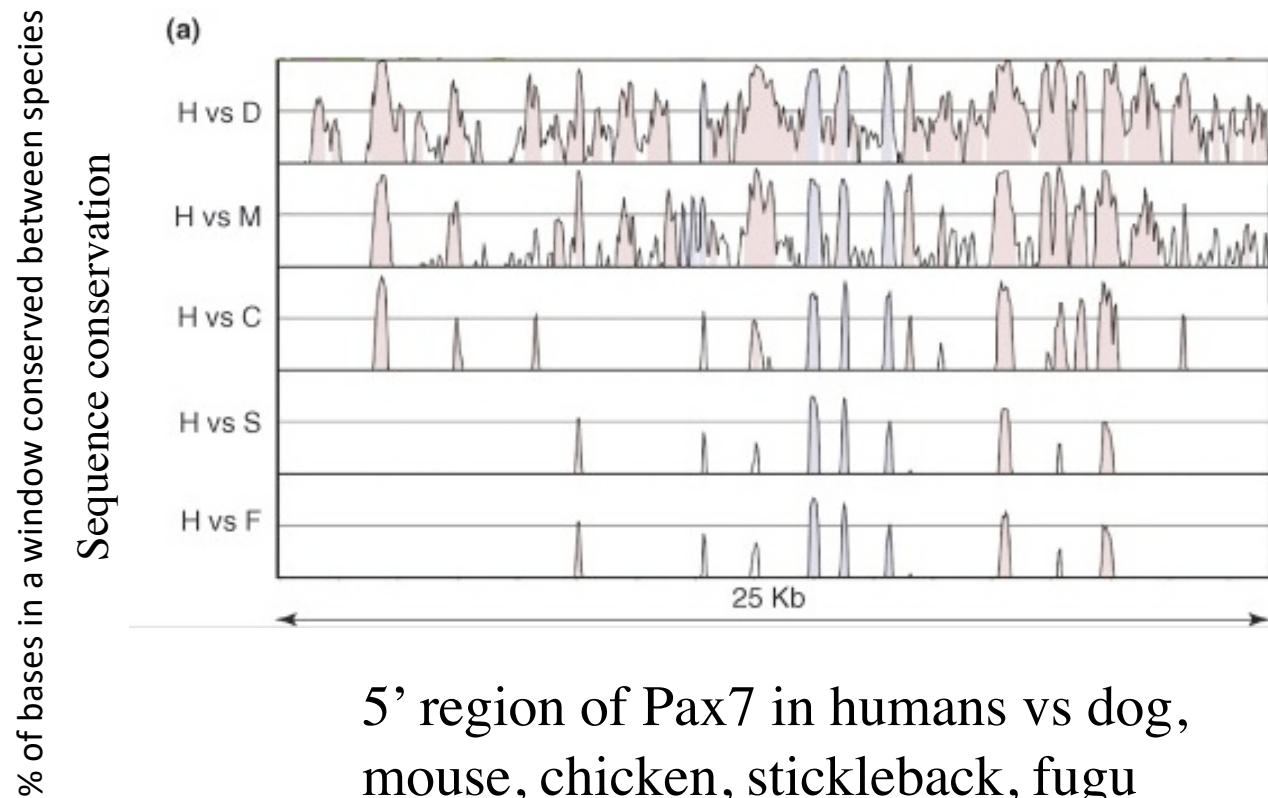


Andolfatto Nature
2005

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

Levels of constraint

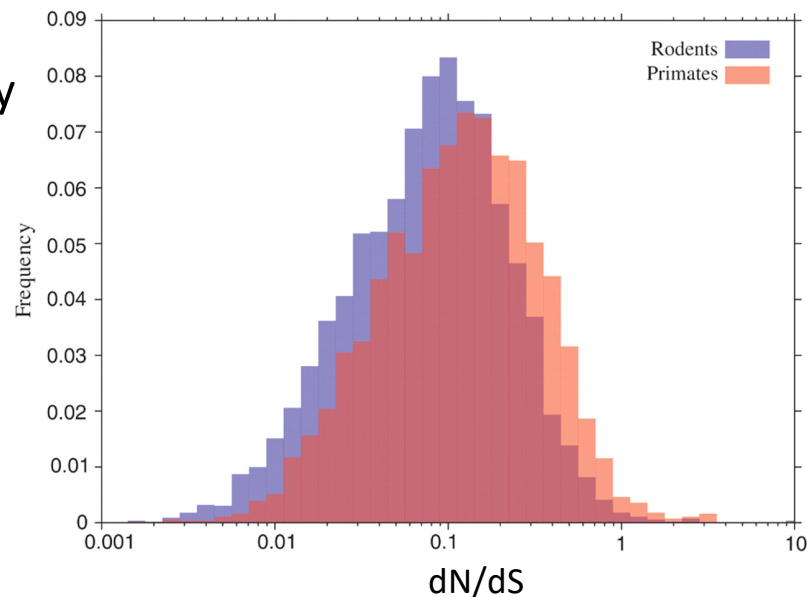
Functional elements in genomes are often identified by their conservation over species.



Levels of constraint

- Neutral theory claims most new mutations are deleterious and are lost immediately. Only neutral mutations contribute to substitution.
- dN = nonsynonymous subst./site
- dS = synonymous subst./site
- Expected dN/dS for a “typical” gene. < 1

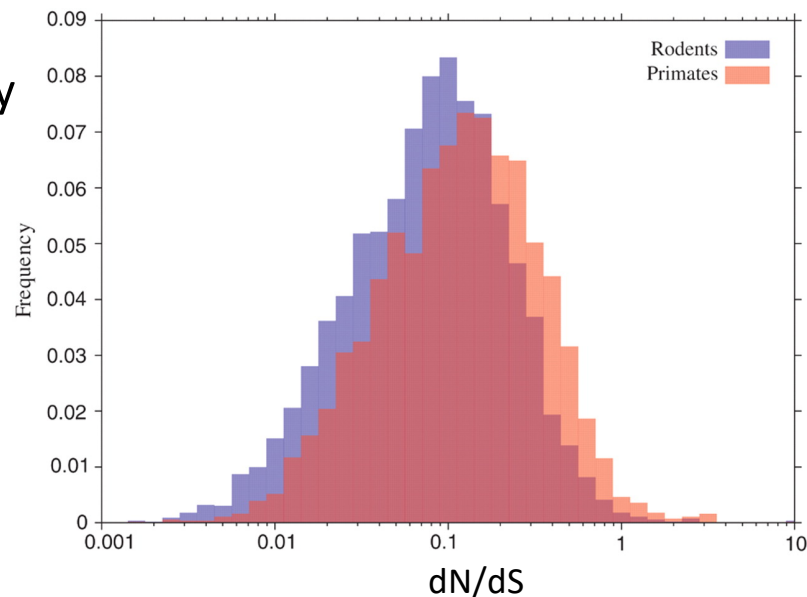
Consistent with neutral theory
slower rate of substitution at
more constrained sites.



Levels of constraint

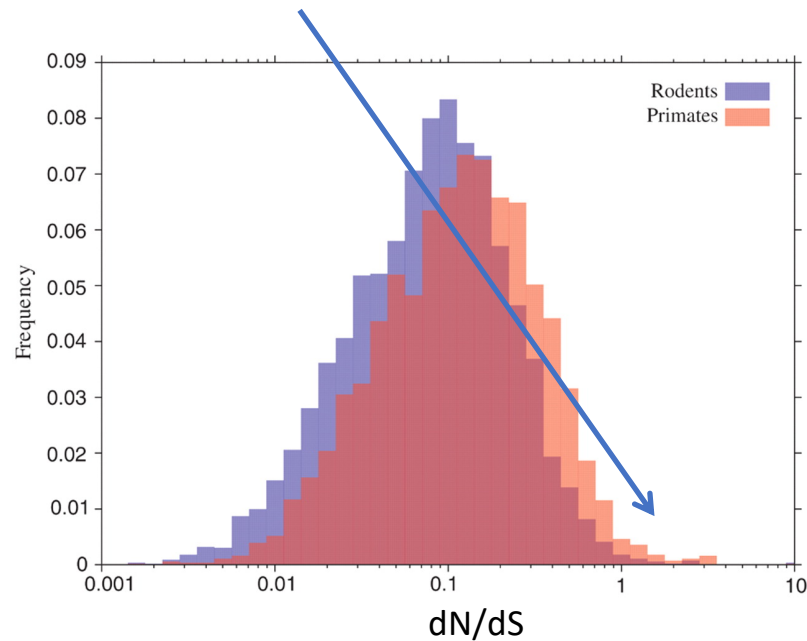
- Neutral theory claims most new mutations are deleterious and are lost immediately. Only neutral mutations contribute to substitution.
- dN = nonsynonymous subst./site
- dS = synonymous subst./site
- Expected dN/dS for a “typical” gene. < 1

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



Adaptive protein divergence and the dN/dS ratio

- dN = nonsynonymous subst./site
- dS = synonymous subst./site
- Expected dN/dS for a “typical” gene. < 1
- Expected dN/dS for a gene coding for an unconstrained protein = 1
- Recurrent directional selection, $dN/dS > 1$ *



* This is a very conservative test.

An example

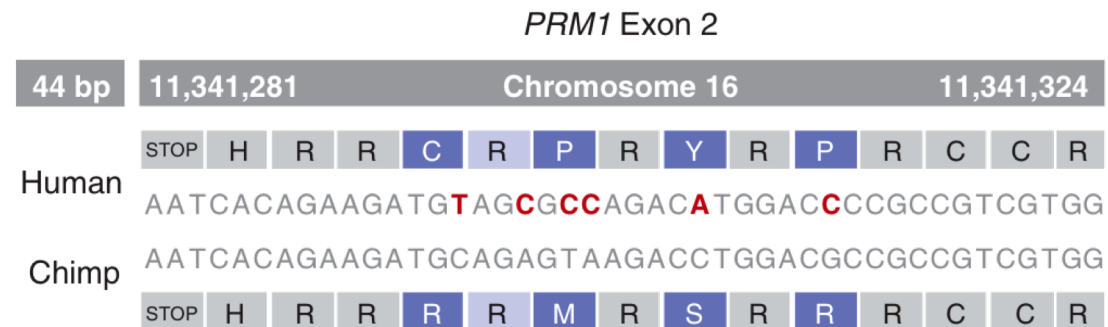
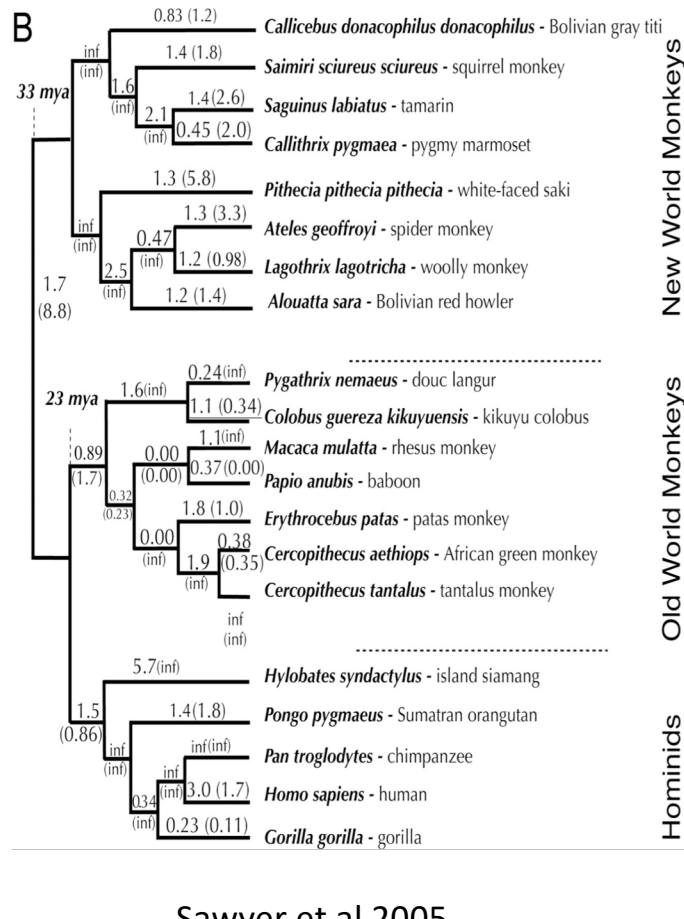


Fig. 2. Excess of function-altering mutations in *PRM1* exon 2. The *PRM1* gene exon 2 contains six differences between humans and chimpanzees, five of which alter amino acids (7, 8).

TRIM5 α gene involved in primate retrovirus defense.
dN/dS > 1 for the majority of branches in phylogeny



Sawyer et al 2005

Examples of genes showing $dN/dS > 1$

--Vertebrate immune system

--Viral coat proteins in viruses

--Reproduction related genes

Invertebrate gamete

recognition proteins

Vertebrate gametogenesis

