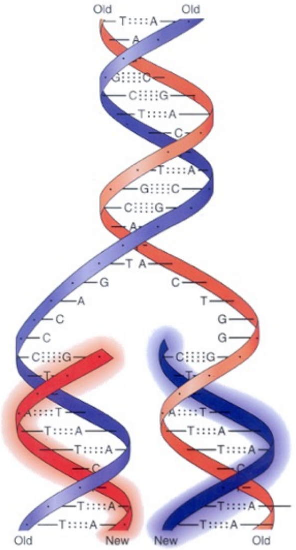
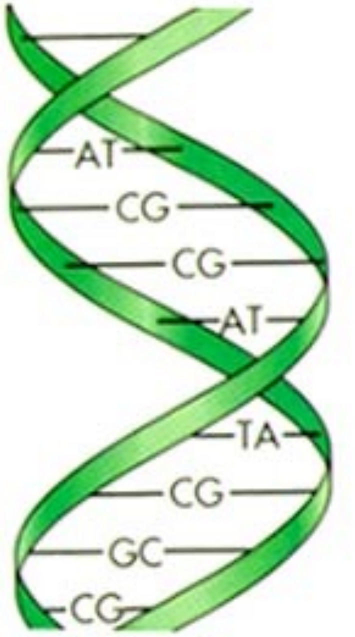


Evolution of germline mutation spectrum in humans: In light of big ‘omics’ datasets

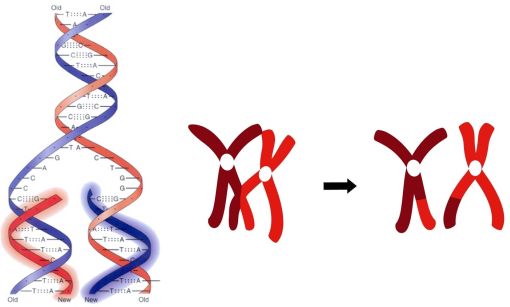
Priya Moorjani
Department of Molecular & Cell Biology
Center for Computational Biology
University of California, Berkeley

Computational Challenges in Very Large-Scale 'Omics’
Simons Institute

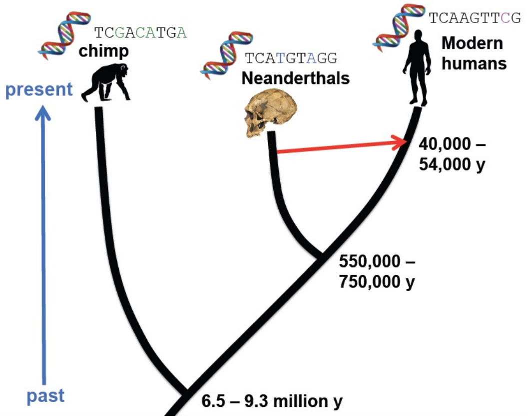
Lab focus: Human Evolutionary Genetics



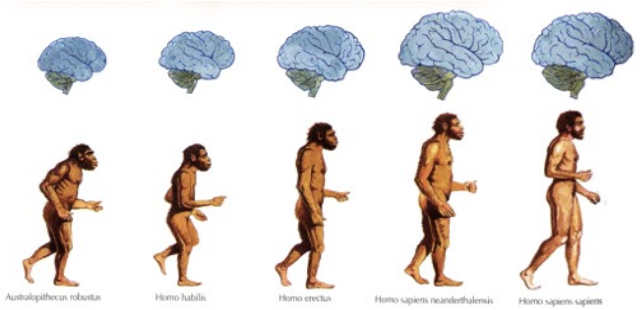
Understand Evolutionary Processes

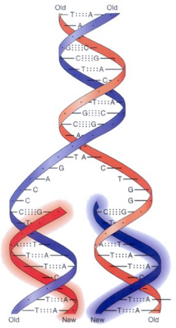


Reconstruct Evolutionary Events



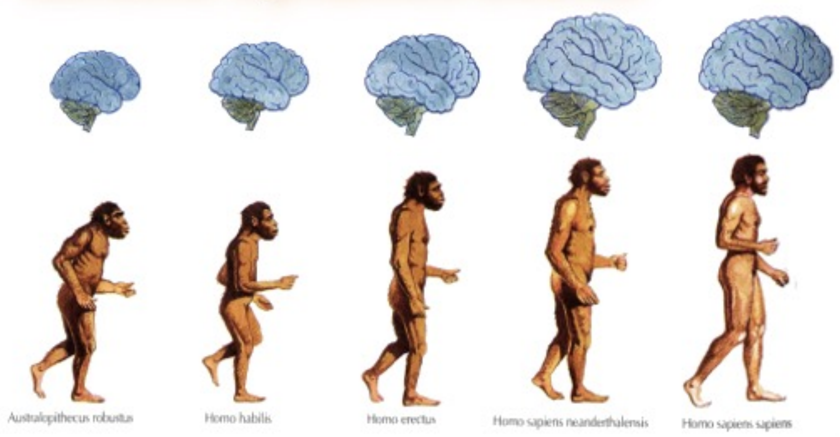
Learn about Adaptation & Disease



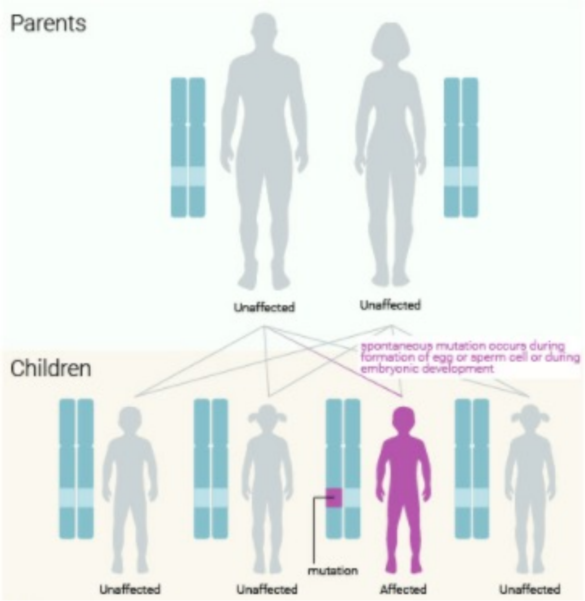


Germline mutations are the ultimate source of genetic variation

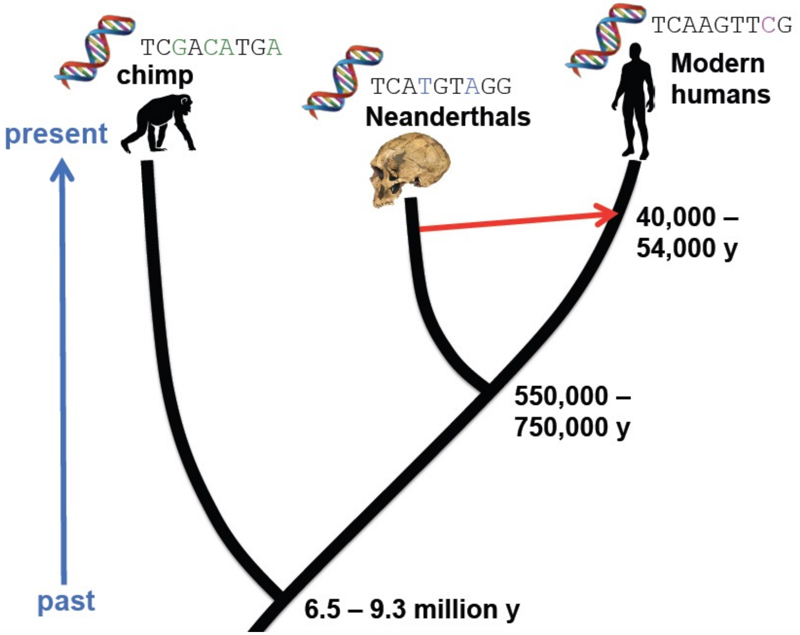
<u>Fuel of evolution</u>

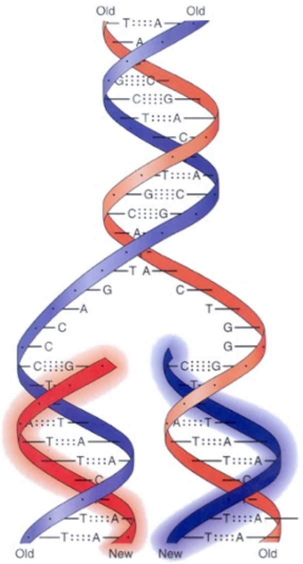


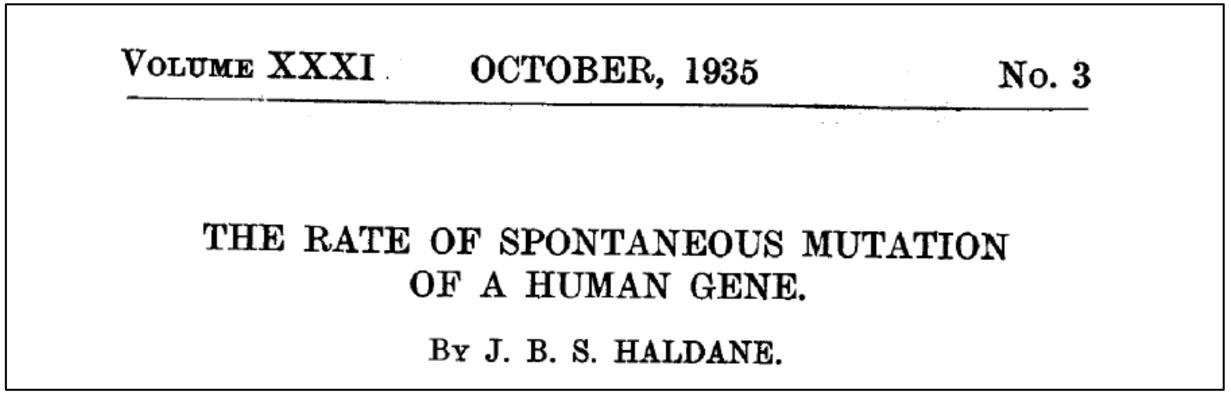
<u>Cause of heritable diseases</u>



<u>“Molecular clock” for dating events</u>







VOLUME XXXI OCTOBER, 1935 No. 3

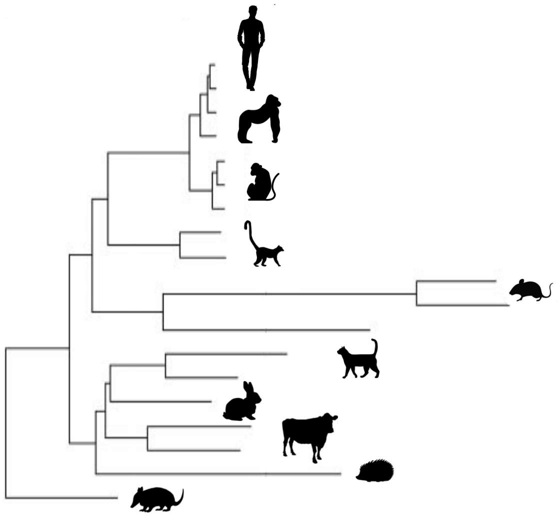
THE RATE OF SPONTANEOUS MUTATION OF A HUMAN GENE.

BY J. B. S. HALDANE.

Pedigrees

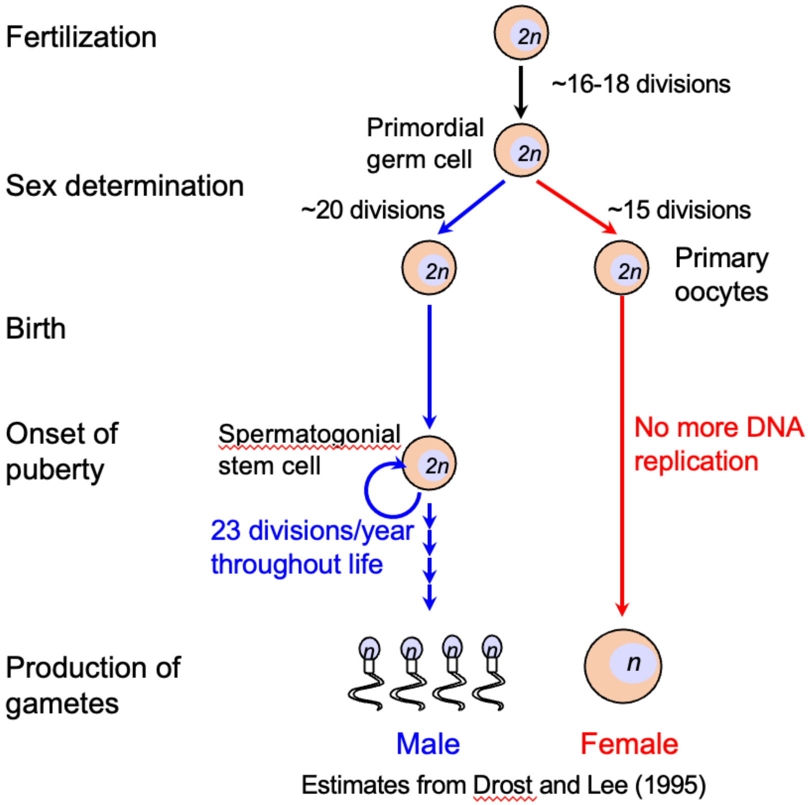
Among populations

Across species



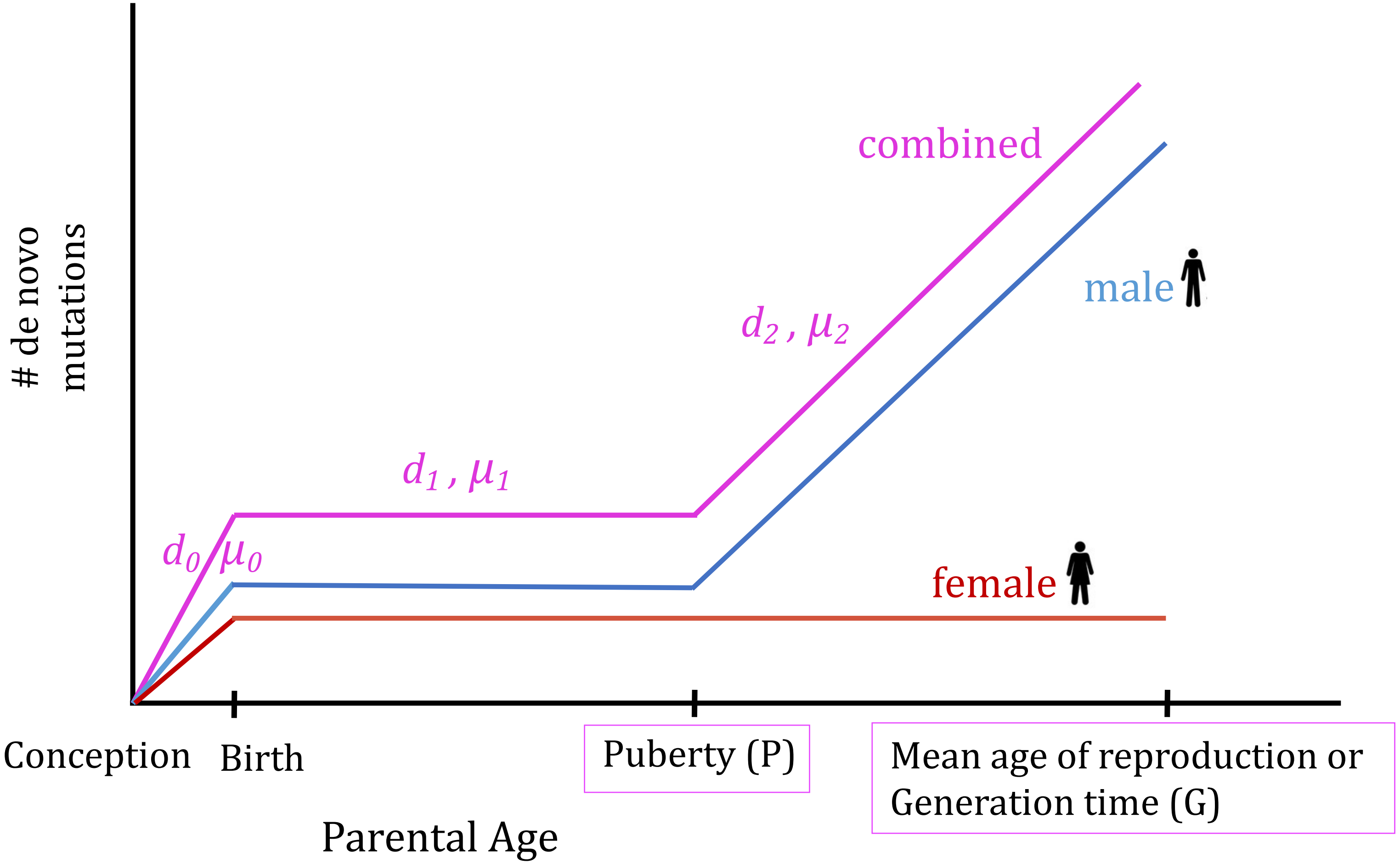
The textbook view of mutation rate

Cell divisions with age



Fertilization
2n
~16-18 divisions
Sex determination
Primordial germ cell
2n
~20 divisions
~15 divisions
Birth
2n
2n
Primary oocytes
Onset of puberty
Spermatogonial stem cell
2n
No more DNA replication
23 divisions/year throughout life
Production of gametes
n
Male
Female

Estimates from Drost and Lee (1995)

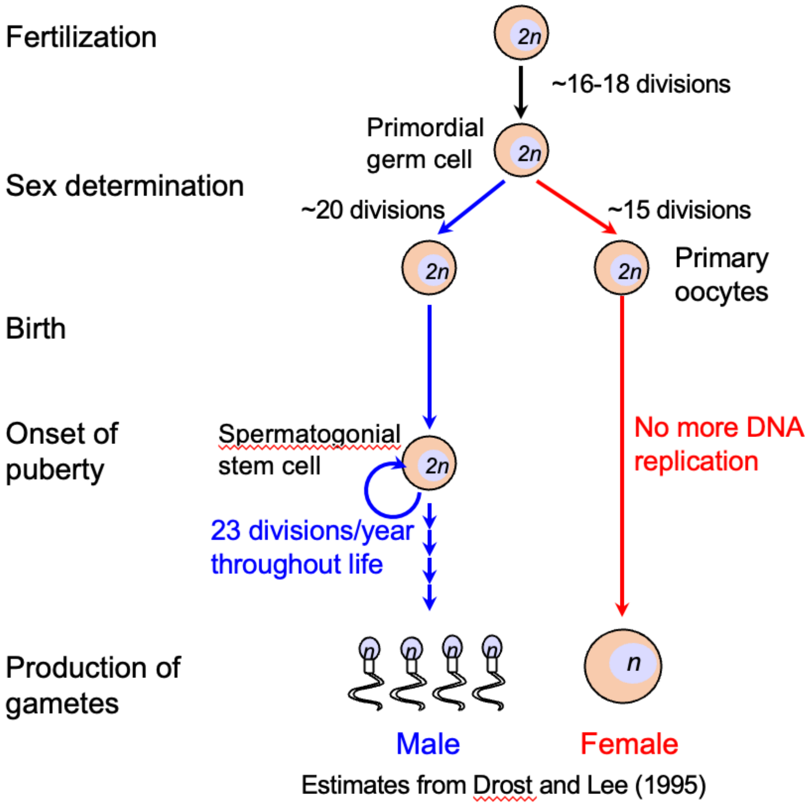


d_i = number of cell divisions in stage i
μ_i = mutation rate per cell division in stage i

Crow 2000; Ségurel et al. 2014; Gao et al. 2016

Whole-genome sequencing of parent-offspring trio enables direct survey of germline mutations in one generation

Cell divisions with age



Fertilization

2n

~16-18 divisions

Sex determination

Primordial germ cell

2n

~20 divisions

~15 divisions

2n

2n

Primary oocytes

Birth

Onset of puberty

Spermatogonial stem cell

2n

23 divisions/year throughout life

No more DNA replication

Production of gametes

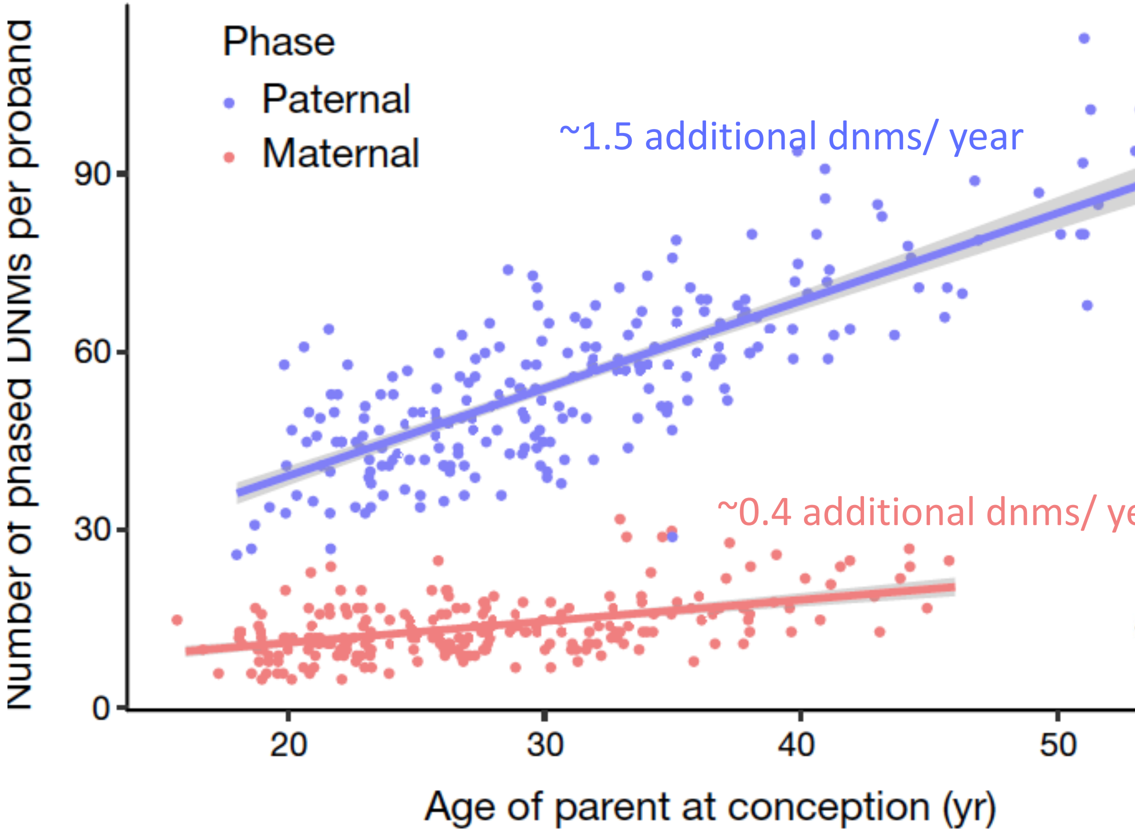
n

Male

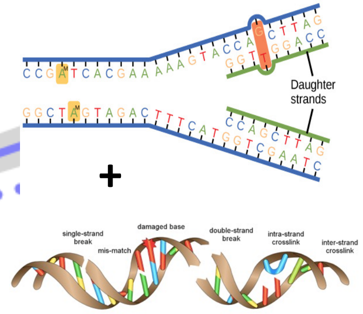
Female

Estimates from Drost and Lee (1995)

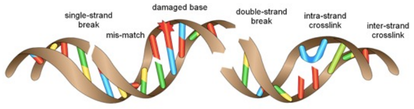
Average human mutation rate = ~70 mutations or 1.2×10^{-8}/per bp/ per generation



Replication errors



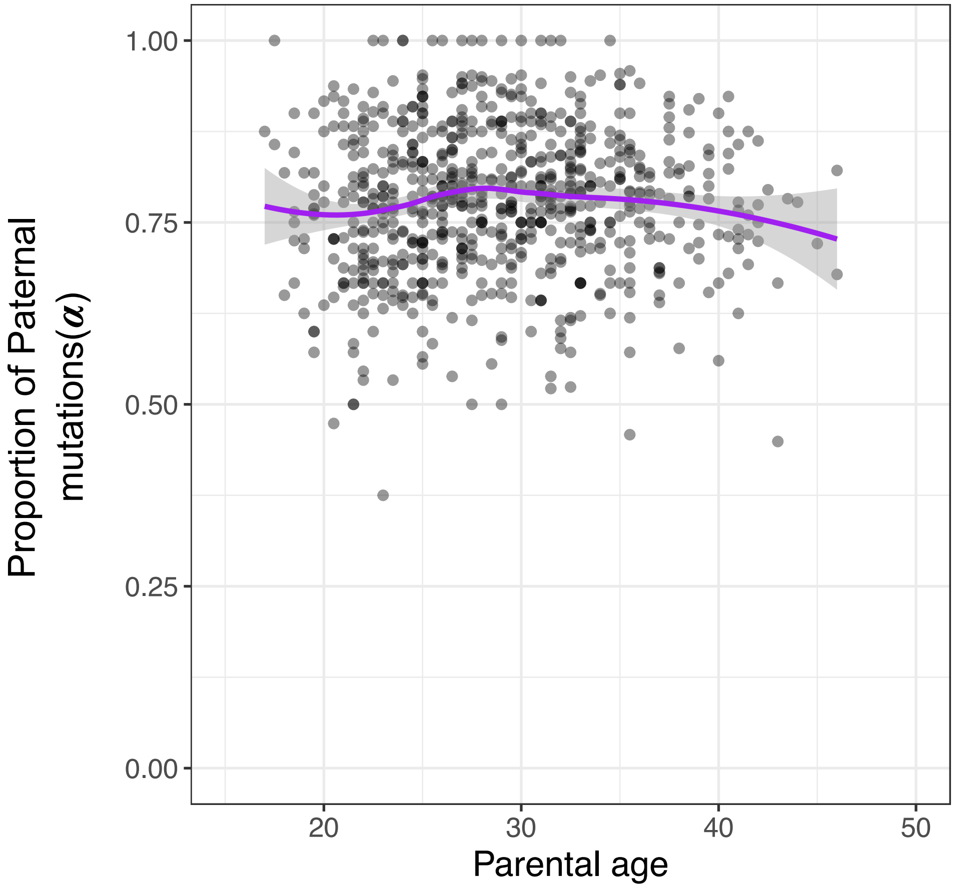
DNA damage



deCODE genetics: Jónsson et al. 2017

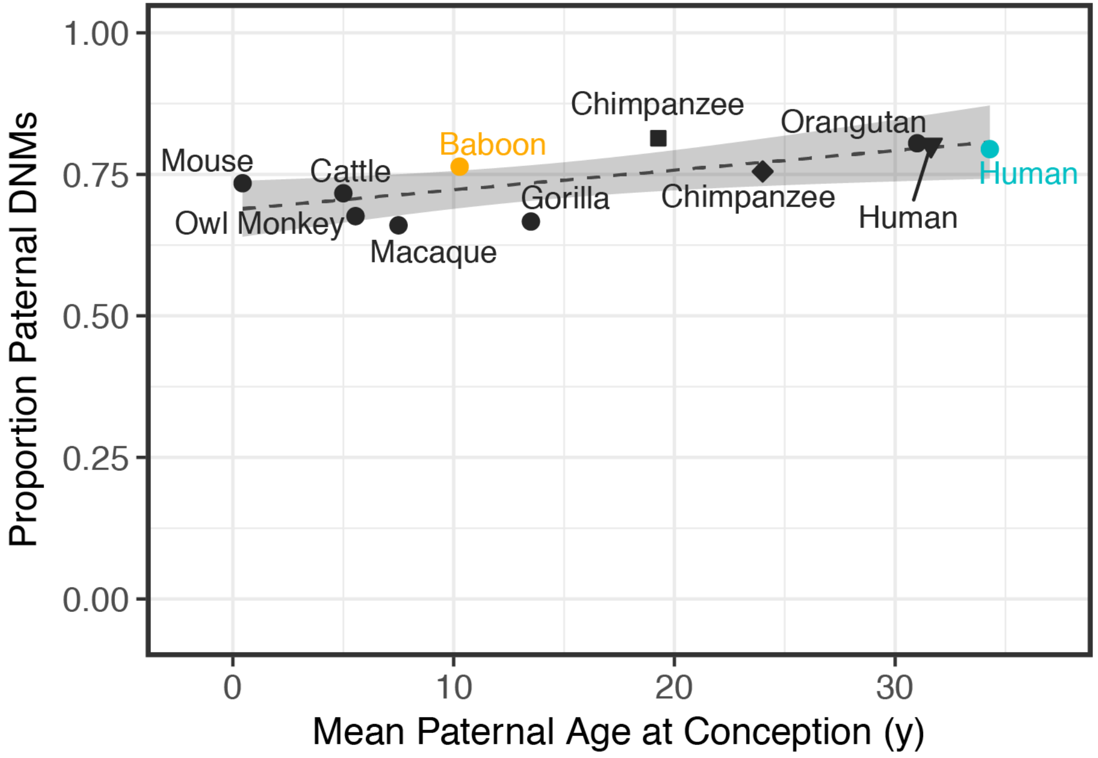
Surprise 1: Stable male-bias with age suggests underappreciated role of non-replicative sources to mutagenesis

In humans



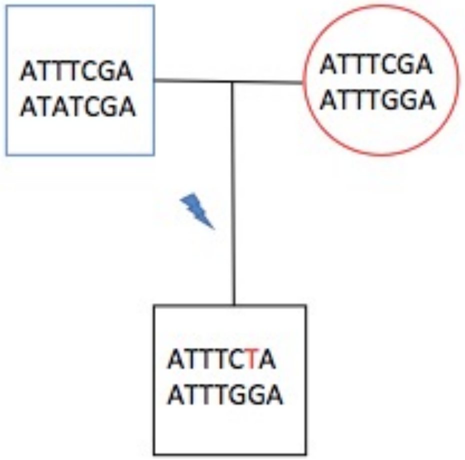
Gao, Moorjani et al. PNAS 2019

Across mammals



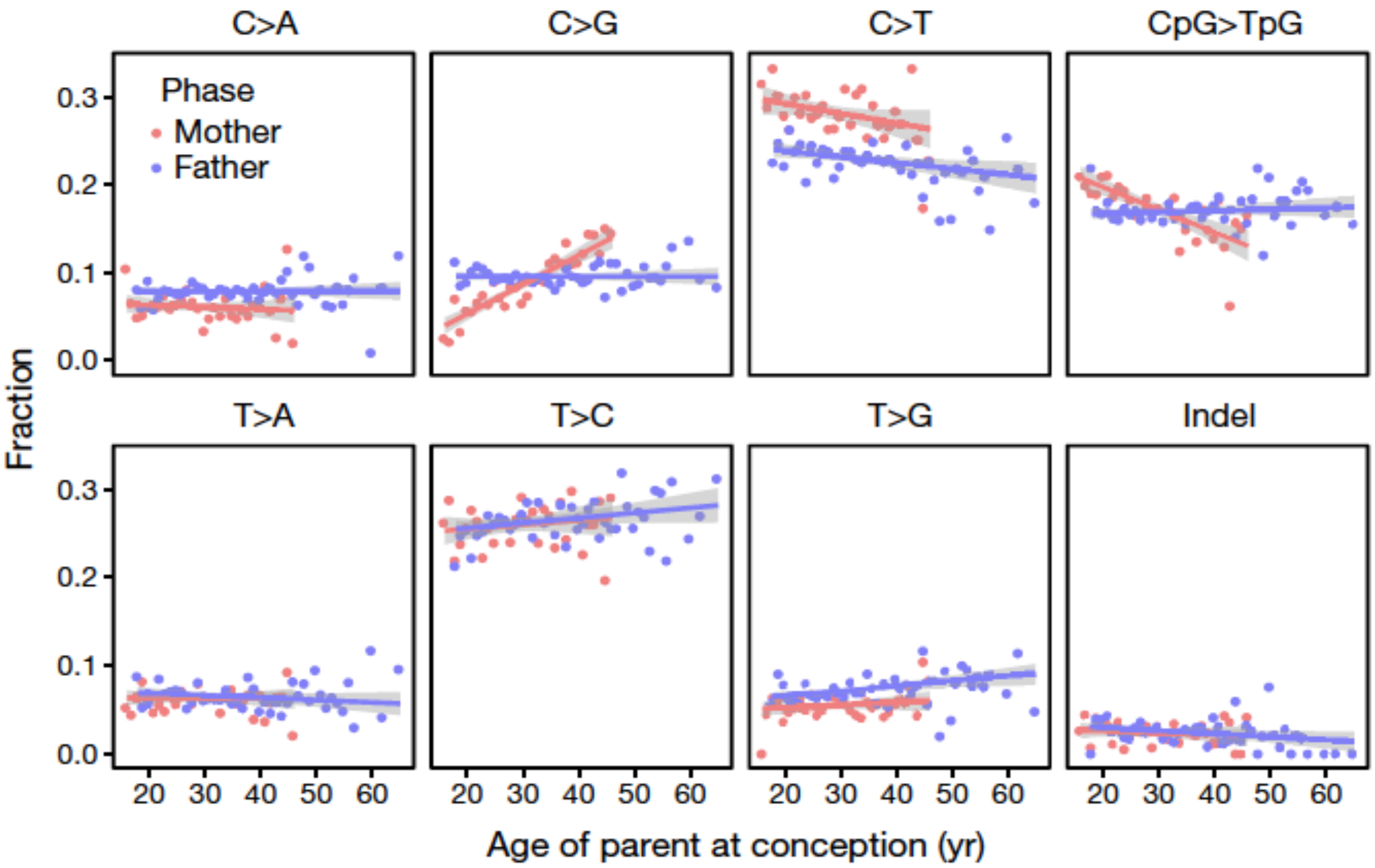
Wu, Ober, Wall, Moorjani* & Przeworski*, PLoS Biology 2020

Whole-genome sequencing of parent-offspring trio enables direct survey of germline mutations in one generation


ATTTCGA
ATATCGA

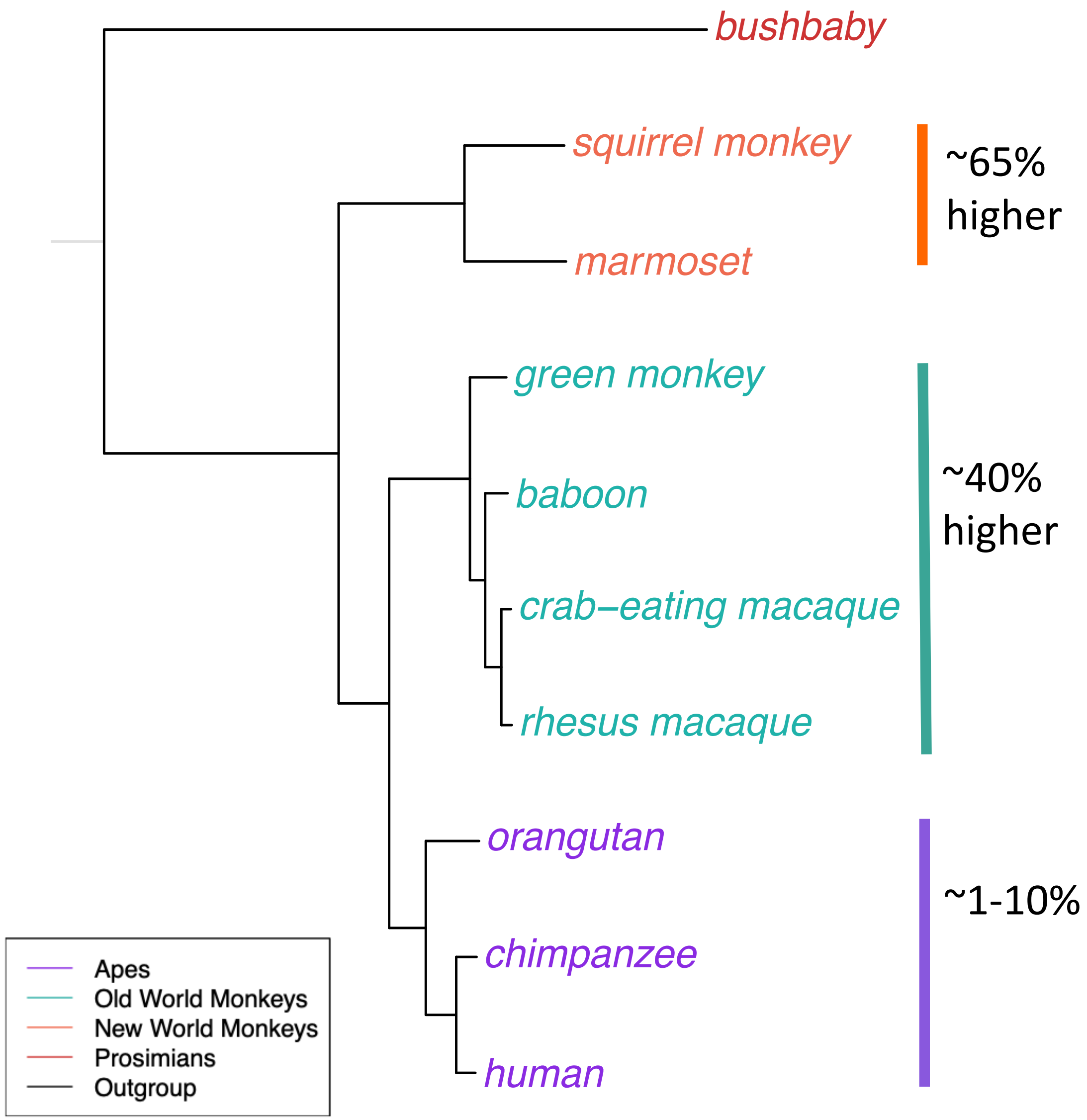
ATTTCGA
ATTTGGA

ATTTCTA
ATTTGGA



deCODE genetics: Jónsson et al. 2017; Gao et al. 2019

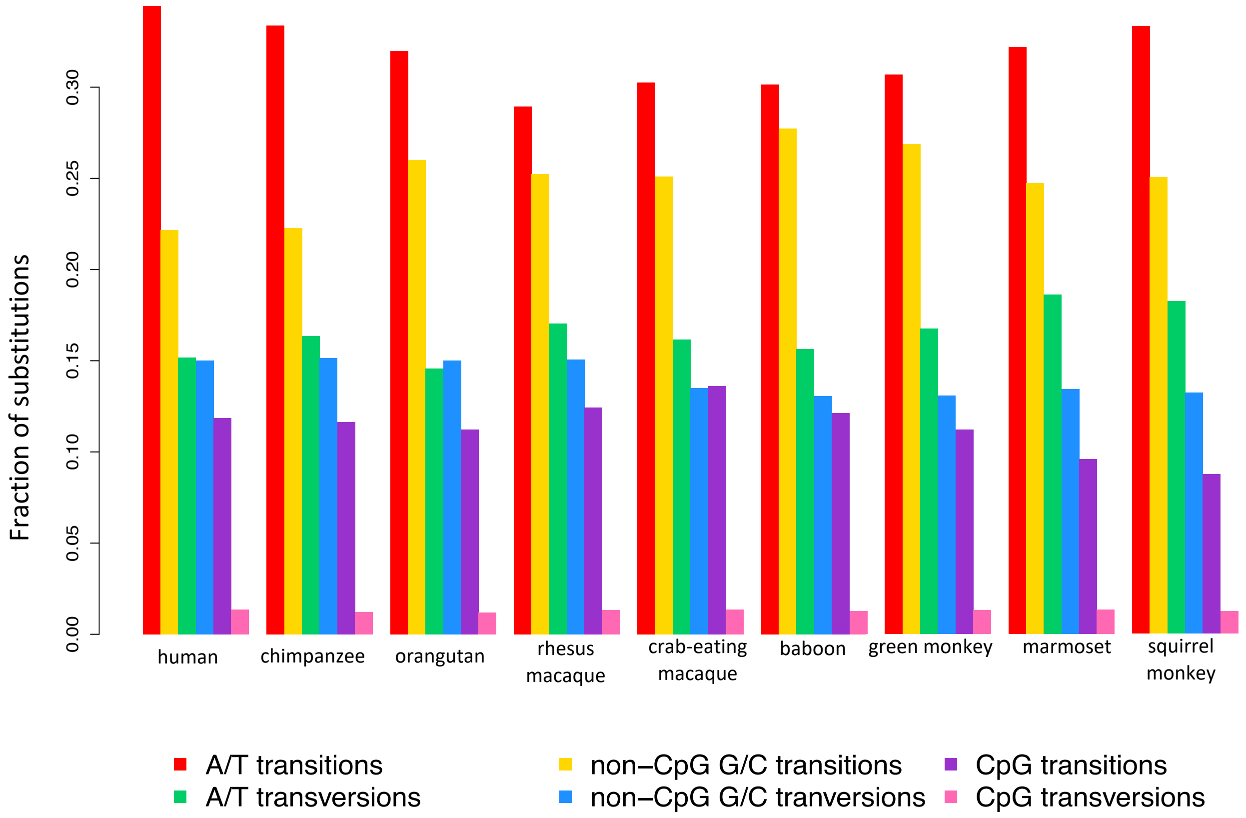
Surprise 2: The unstable molecular clock: Large variation in substitution rates and spectrum across primates



~65% higher

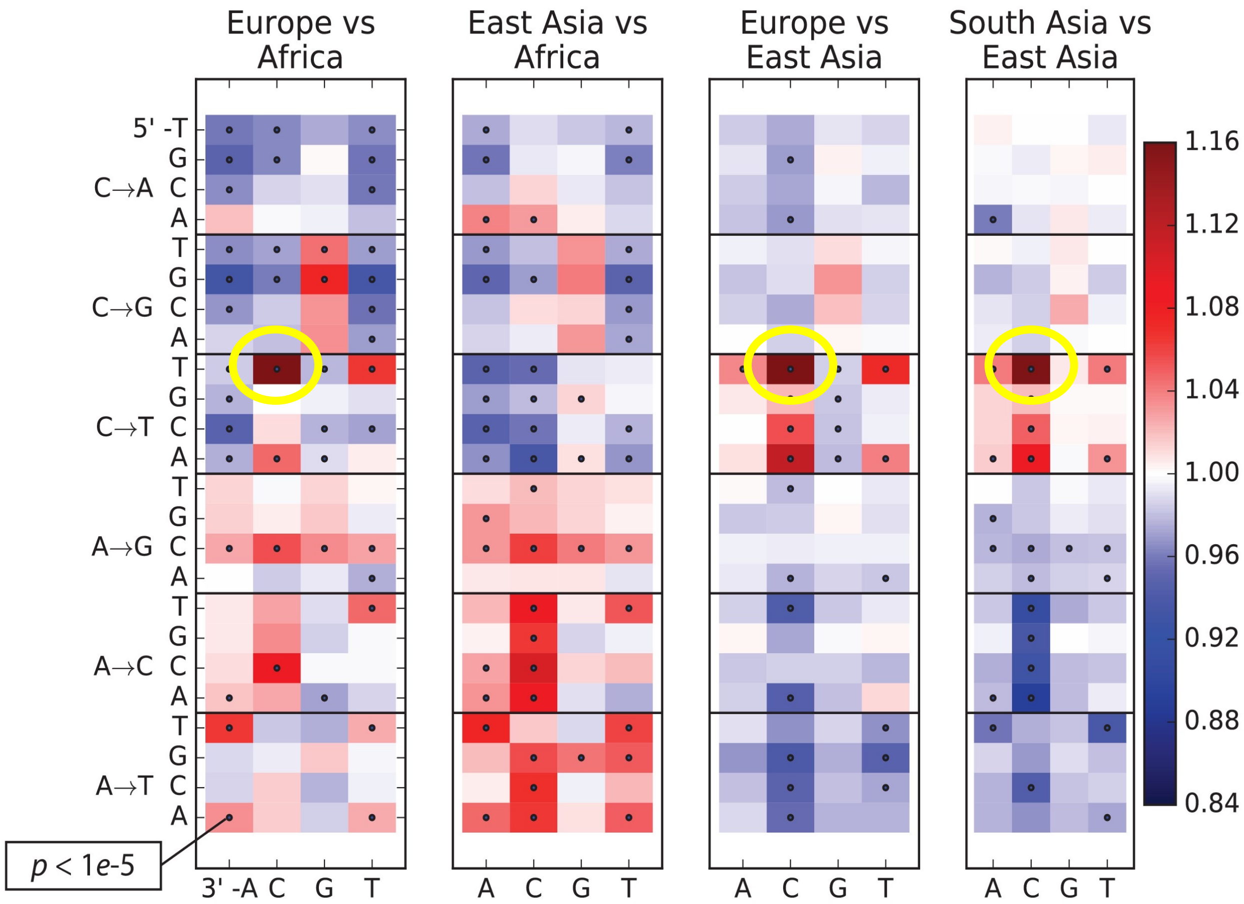
~40% higher

~1-10%

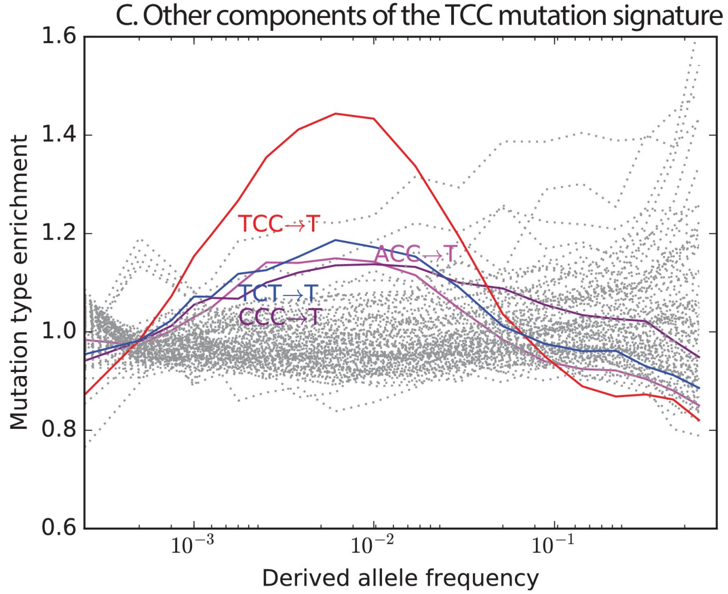


Moorjani et al. 2016

Surprise 3: The mutation spectrum of polymorphisms differ across human populations



Harris and Pritchard (2017)

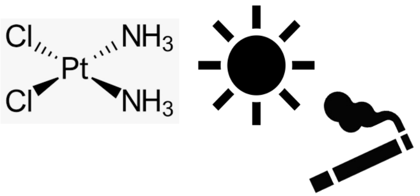


The transient elevation in TCC>TTC mutation rate in Europeans vs. non-Europeans

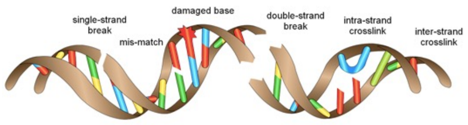
also Speidel et al. 2020; Mathieson and Reich 2017

The spectrum of polymorphisms is shaped by multiple evolutionary forces

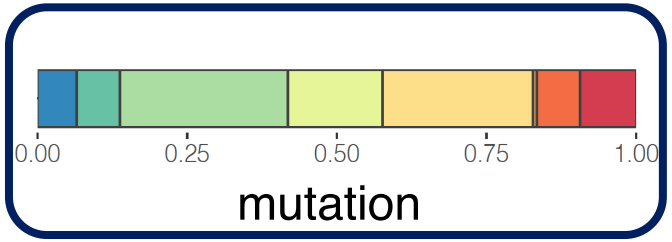
Pedigrees



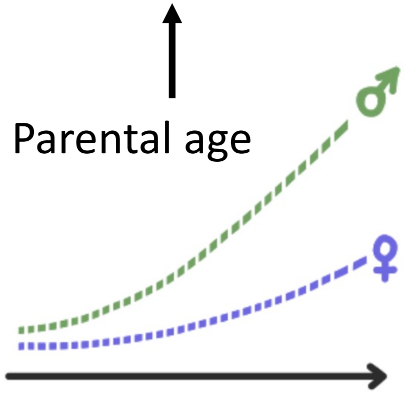
Environmental exposure



DNA damage, repair & modifiers

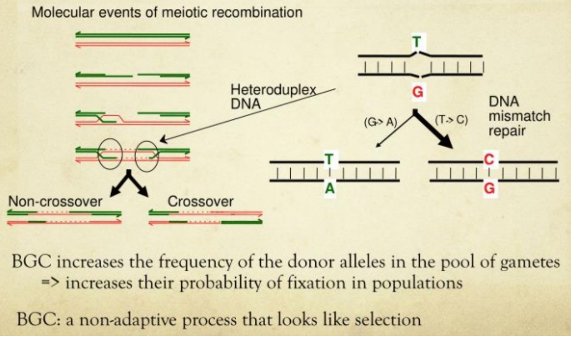


mutation



Parental age

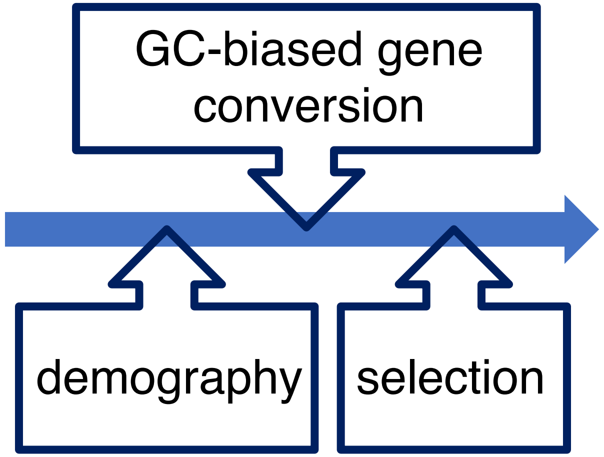
Mutation type
- 🟥 C>G
- 🟧 nonCpG C>A
- 🟧 CpG C>A
- 🟨 nonCpG C>T
- 🟨 CpG C>T
- 🟩 T>C
- 🟩 T>G
- 🟦 T>A



Molecular events of meiotic recombination

BGC increases the frequency of the donor alleles in the pool of gametes
=> increases their probability of fixation in populations

BGC: a non-adaptive process that looks like selection

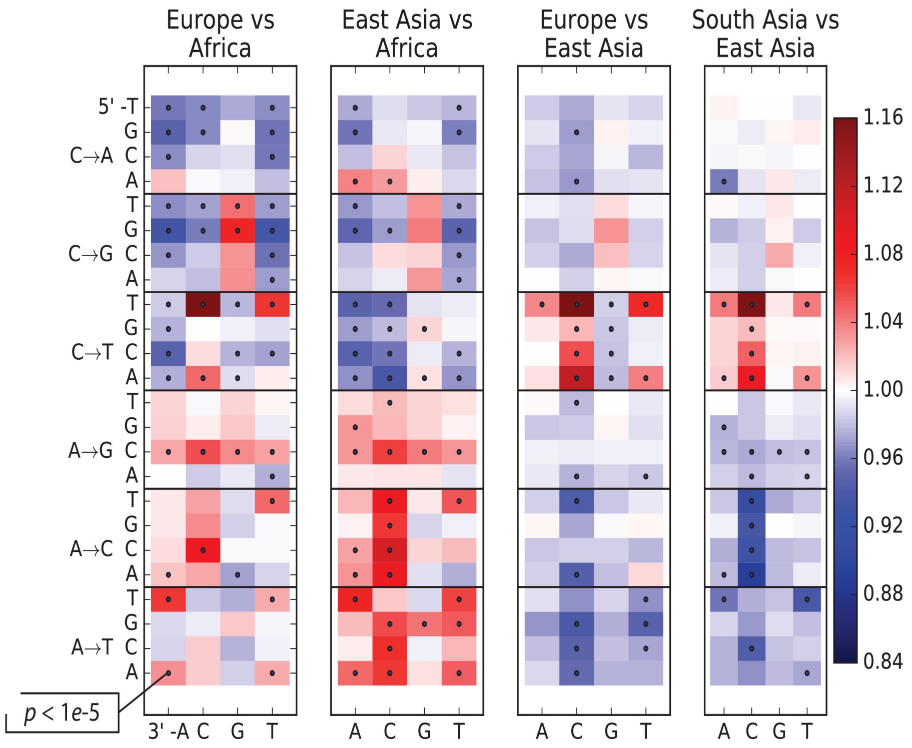


GC-biased gene conversion

demography

selection

Polymorphisms



Timing and causes of evolution of mutation rates in humans

bioRxiv. DOI: https://doi.org/10.1101/2022.06.17.496622

Ziyue Gao, Univ. of Pennsylvania

Yulin Zhang, UC Berkeley

Key questions:

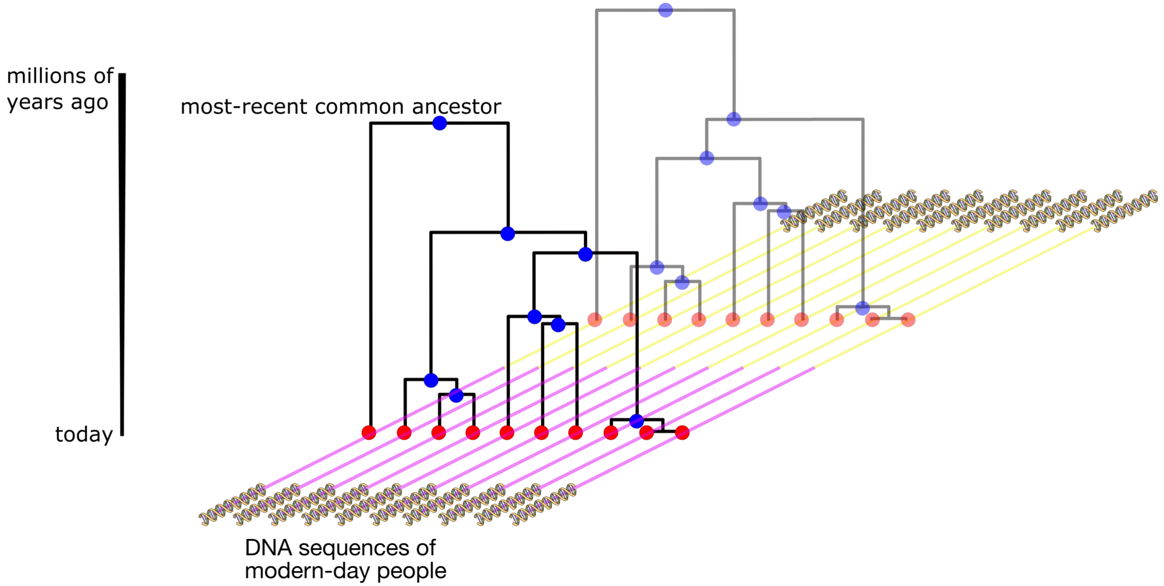
- How many “independent” changes in the mutation spectrum happened during human evolution?
- When and in which population(s) did they occur?
- What are the causes? Genetic modifiers, environmental exposure, or changes in generation time?

How can we better characterize variation in mutation spectrum?

We developed a new framework with following features:

(1) Has a time dimension to allow reliable inter-population comparisons;
(2) Controls for effects of GC-biased gene conversion and selection.

Mutation age inferred based on genealogy reconstruction

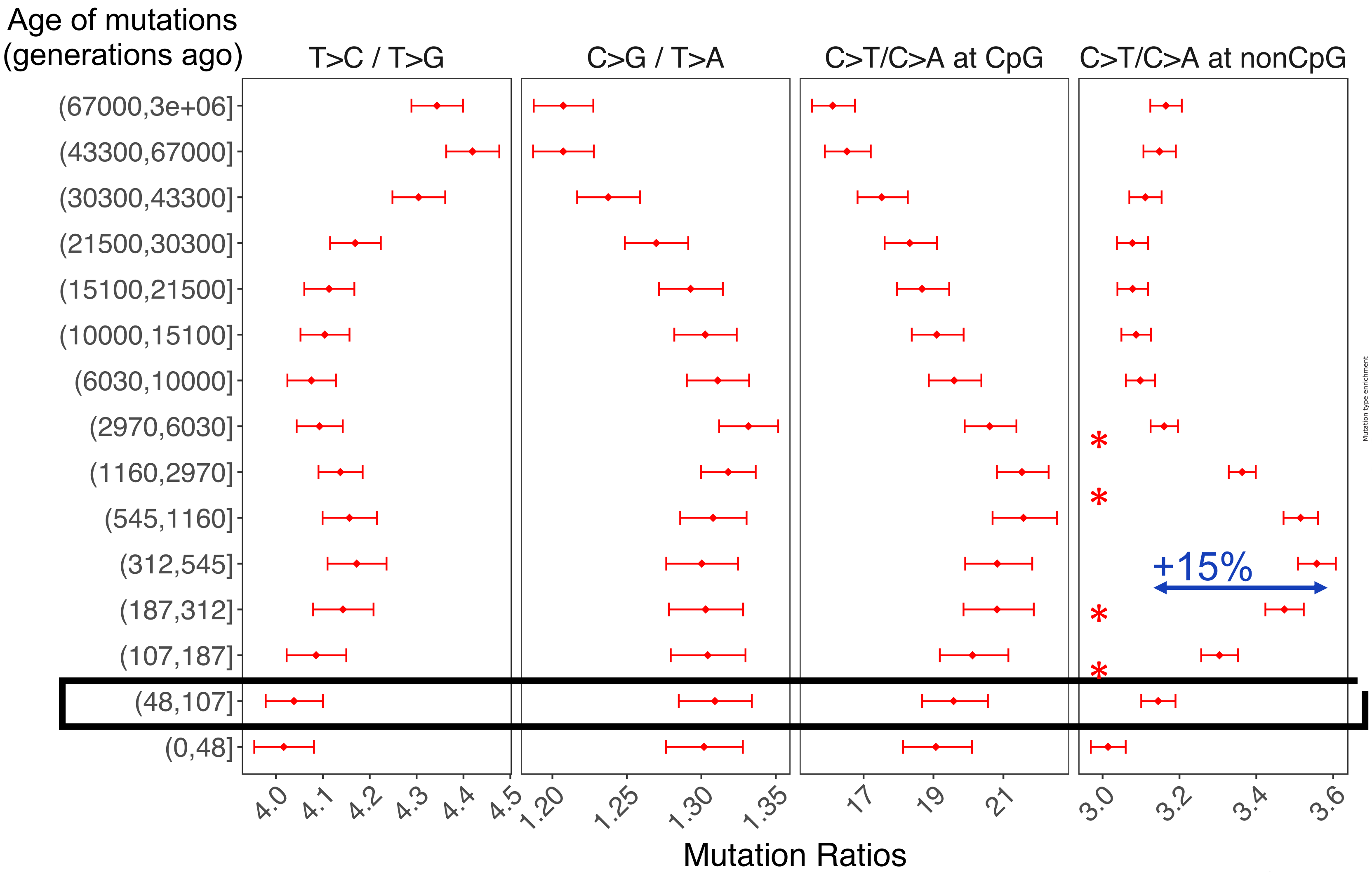


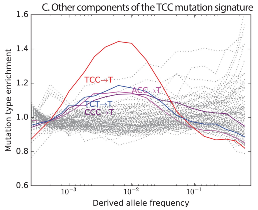
Speidel et al., (2019)

Pairwise comparison matched for gBGC effects

Mutation type 1	**Mutation type 2**	**gBGC effect**	**Mutation opportunity**
T>C	T>G	Both favored	same
C>G	T>A	No effect	GC vs. AT
C>T at CpG	C>A at CpG	Both disfavored	same
C>T at nonCpG	C> A at nonCpG	Both disfavored	same

Pairwise ratios of derived polymorphisms over time in northern Europeans





European-specific acceleration at TCC>TTC

(Harris & Pritchard 2017)

Relate

1000 Genomes Project

Defining Genetic Variation in People

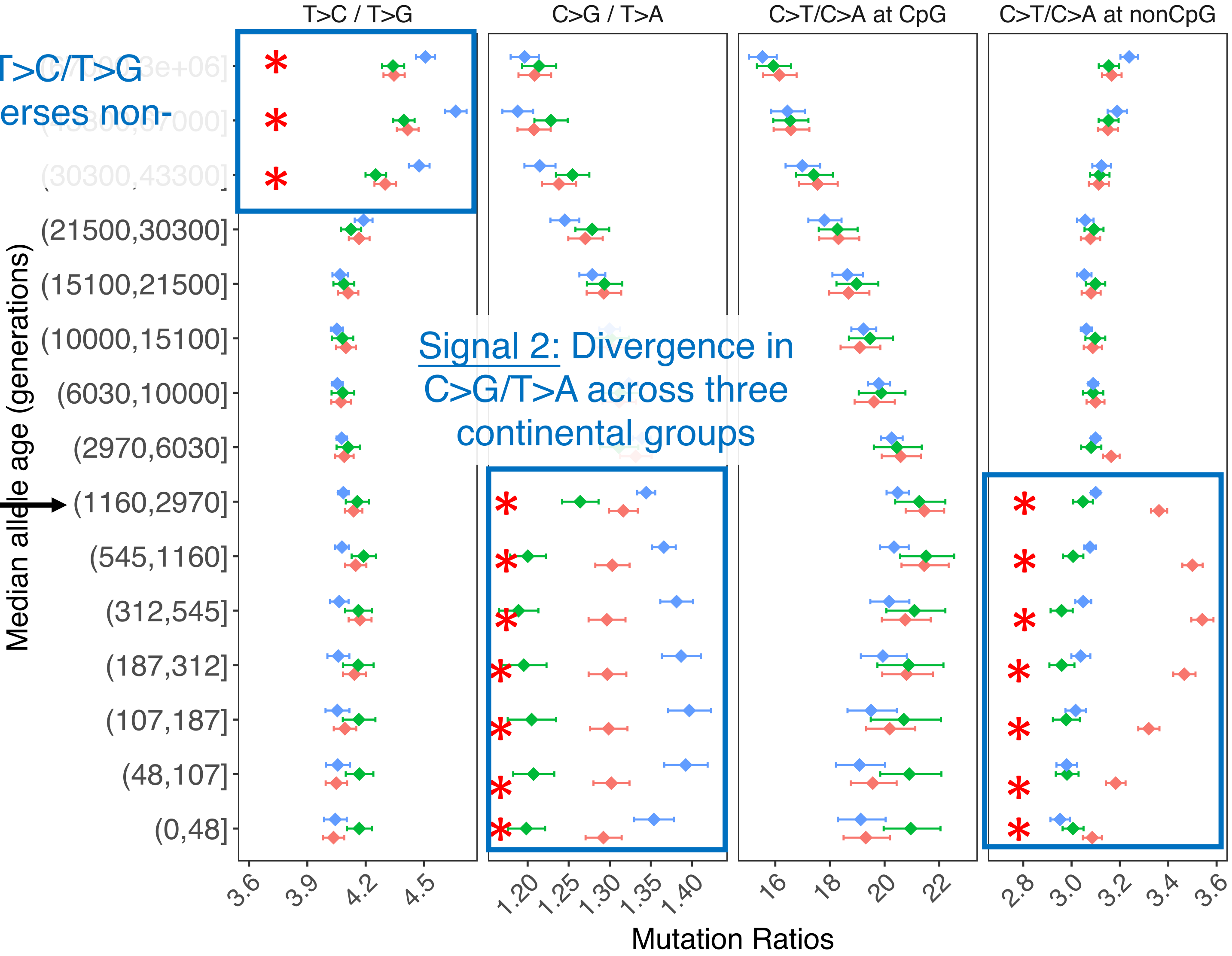
www.1000genomes.org

Spiedel et al. 2020

* *significant*, after multiple hypothesis testing

Differences in pairwise ratios of polymorphisms across human populations

Signal 3: Higher T>C/T>G ratio in Africans verses non-Africans



Population

- 🔶 CEU
- 🟩 CHB
- 🔷 YRI

OOA migration →

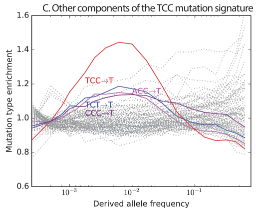
Signal 2: Divergence in C>G/T>A across three continental groups

1000 Genomes

CEU = Europeans

CHB = East Asians

YRI = Africans

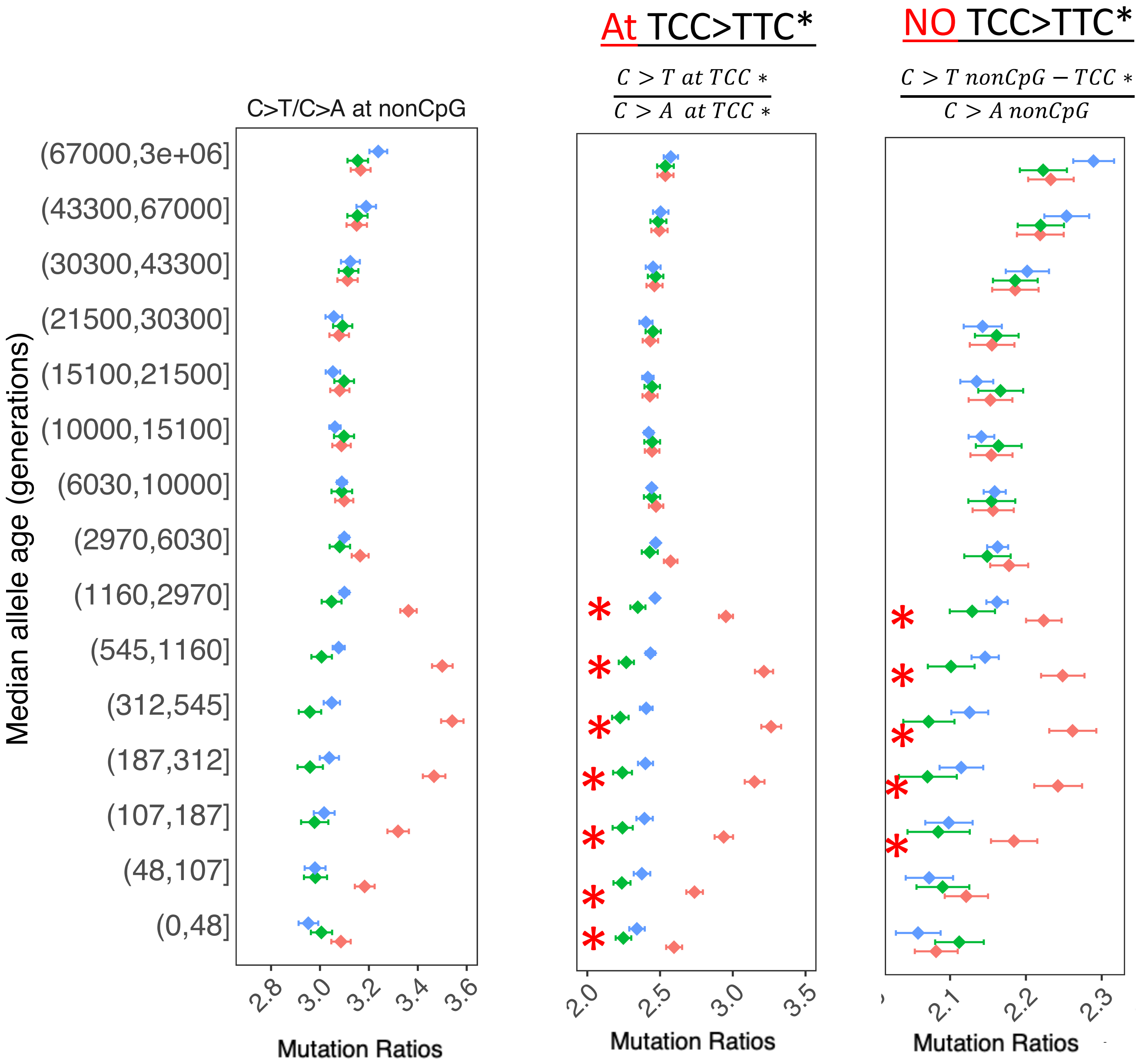


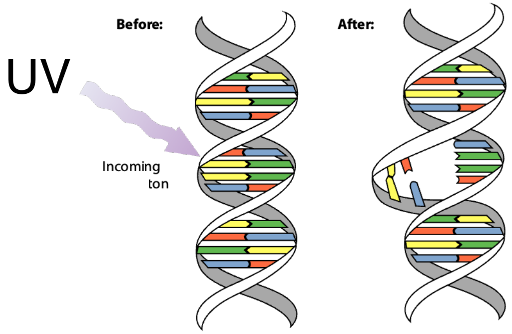
Signal 1: Transient elevation in C>T mutation rate in Europeans (Harris & Pritchard 2017)

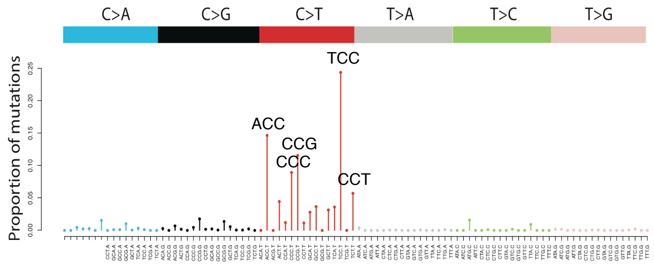
* *significant*, after multiple hypothesis testing

Gao et al. 2022 *biorxiv*

Signal 1: Northern European-specific acceleration in non-CpG C>T mutations







Signature 1

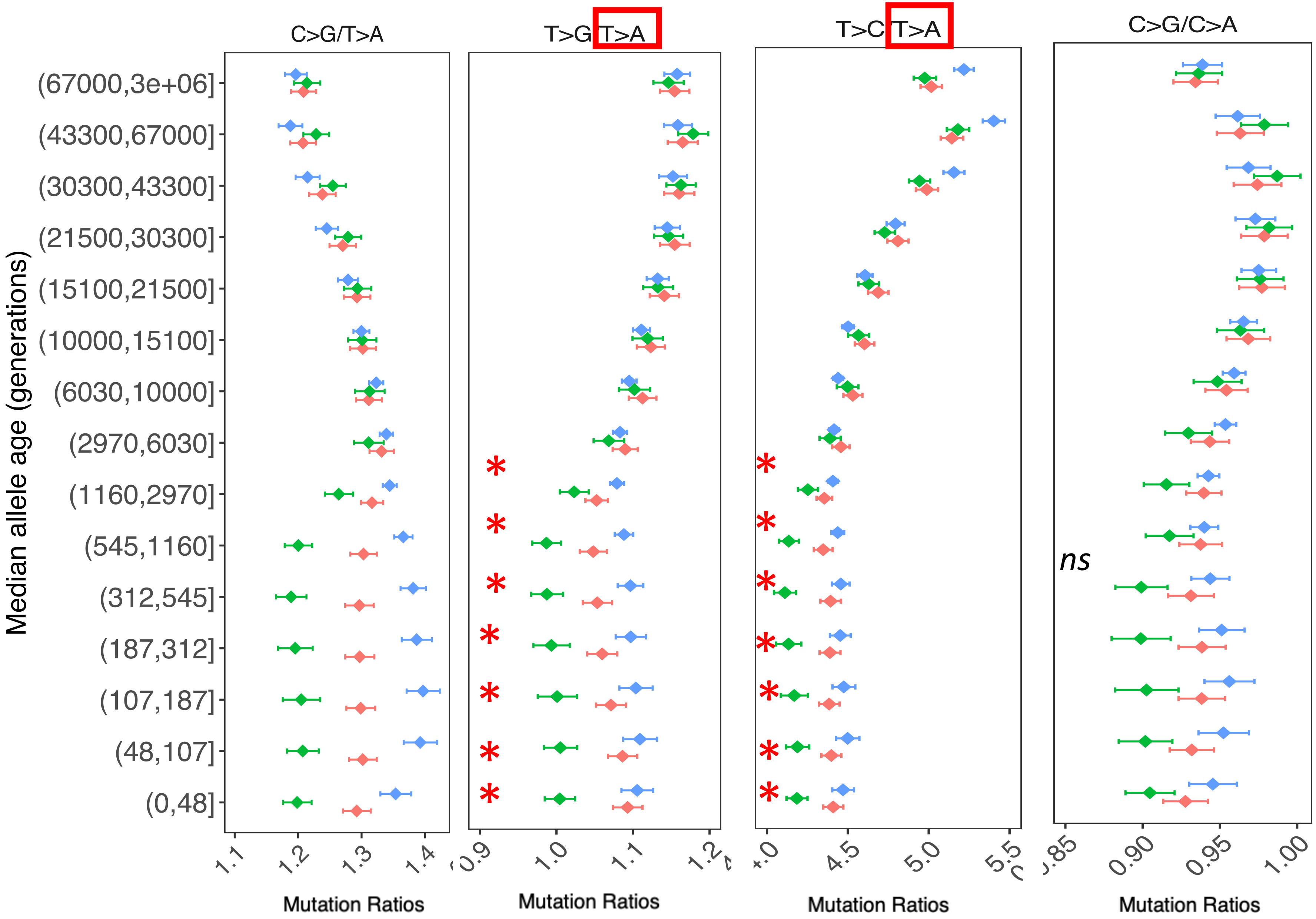
Mathieson and Reich (2017)

Population
- CEU
- CHB
- YRI

TCC* == TCC>T, TCT>T,
ACC>T and CCC>T

Genetic modifiers: Harris and Pritchard 2017
Generation time: Macià, Skov et al. 2021

Signal 2: Variation in C>G / T>A ratios among human populations



T>A mutation rate:

East Asian > European ≥ African

C>G mutation rate:

East Asian < European ≤ African

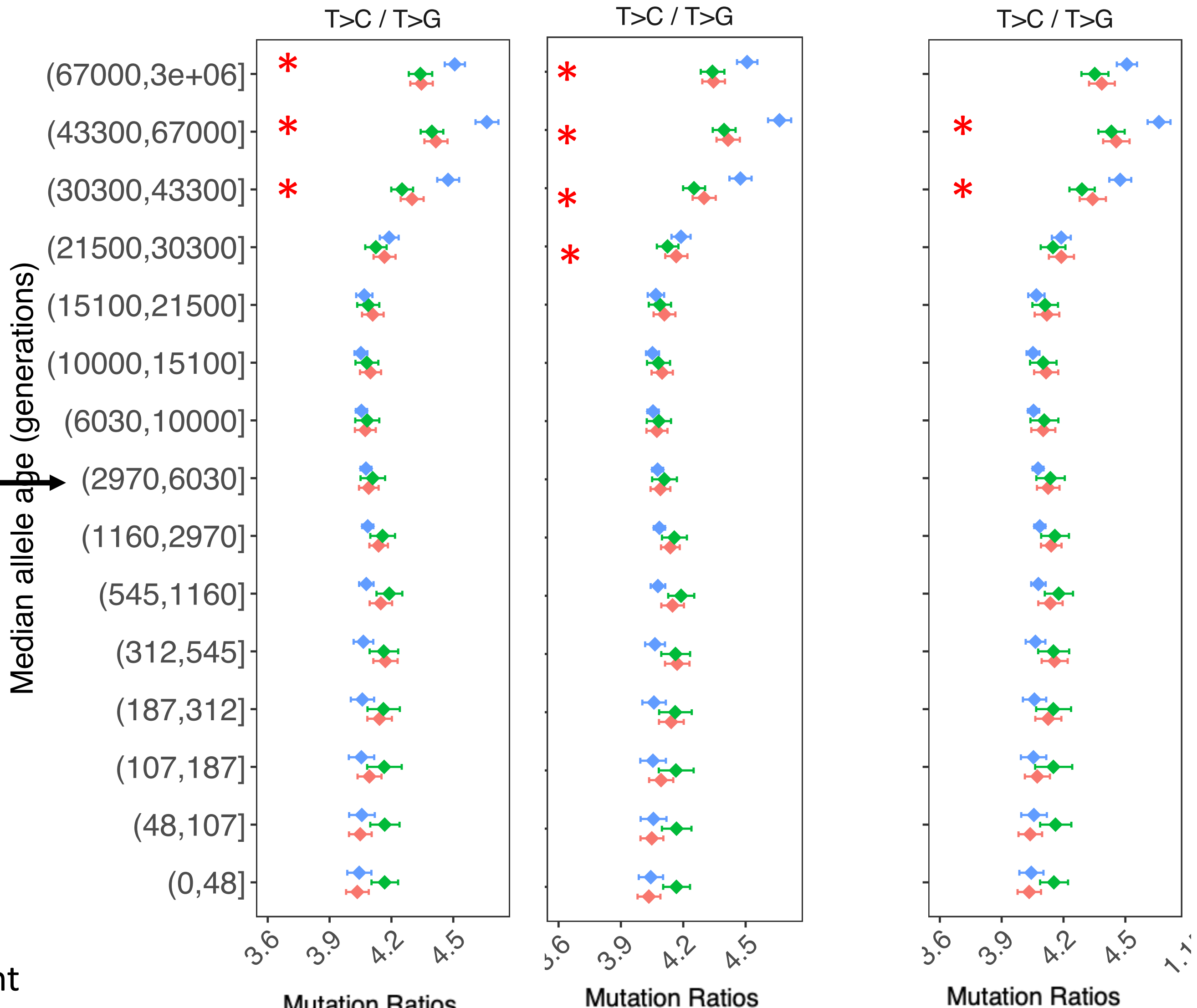
ns = not significant

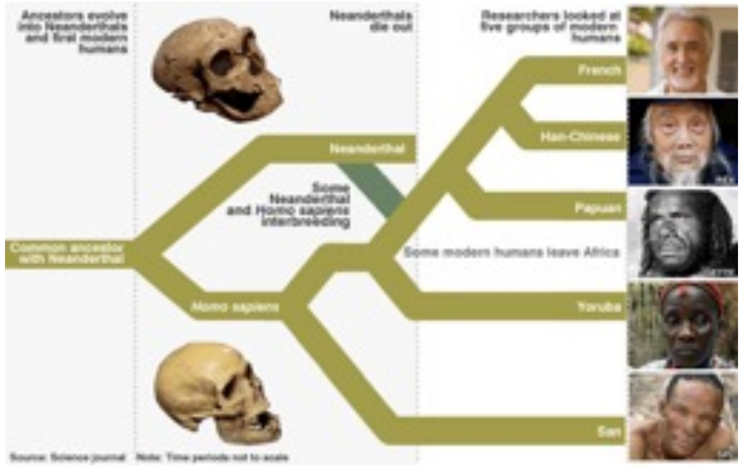
Signal 3: Variation in T>C / T>G ratios among human populations

Misspecification of ancestral allele (use Chimp)

NO Neanderthal in non-Africans

OOA migration



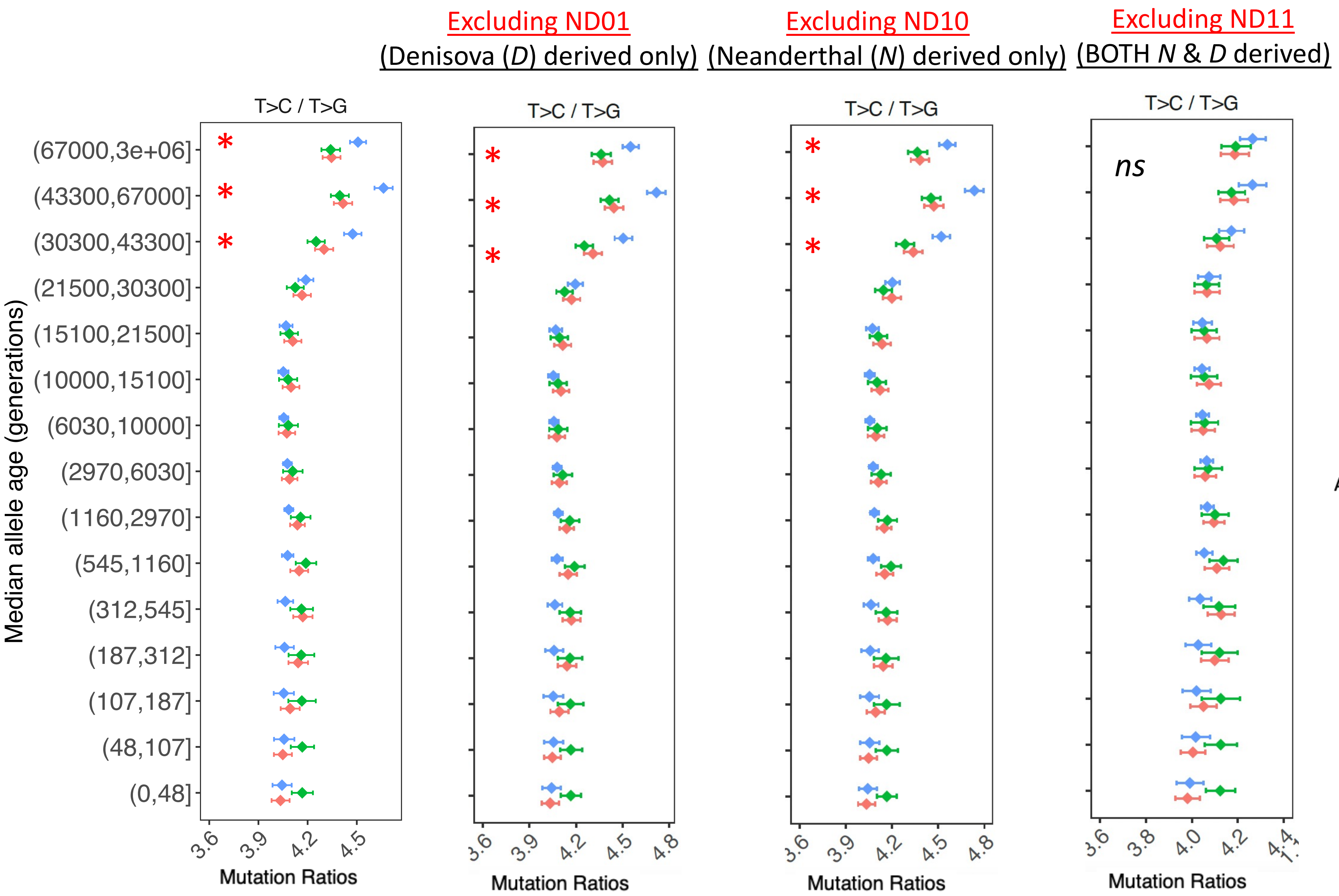


Green et al. 2010

ns = not significant

Gao et al. 2022 *biorxiv*

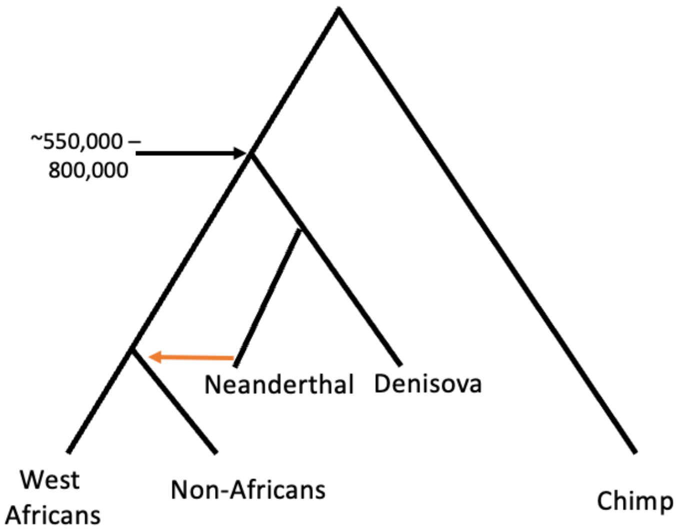
Signal 3: Variation in T>C/ T>G polarized by sharing with archaic populations



Excluding ND01 (Denisova (*D*) derived only)

Excluding ND10 (Neanderthal (*N*) derived only)

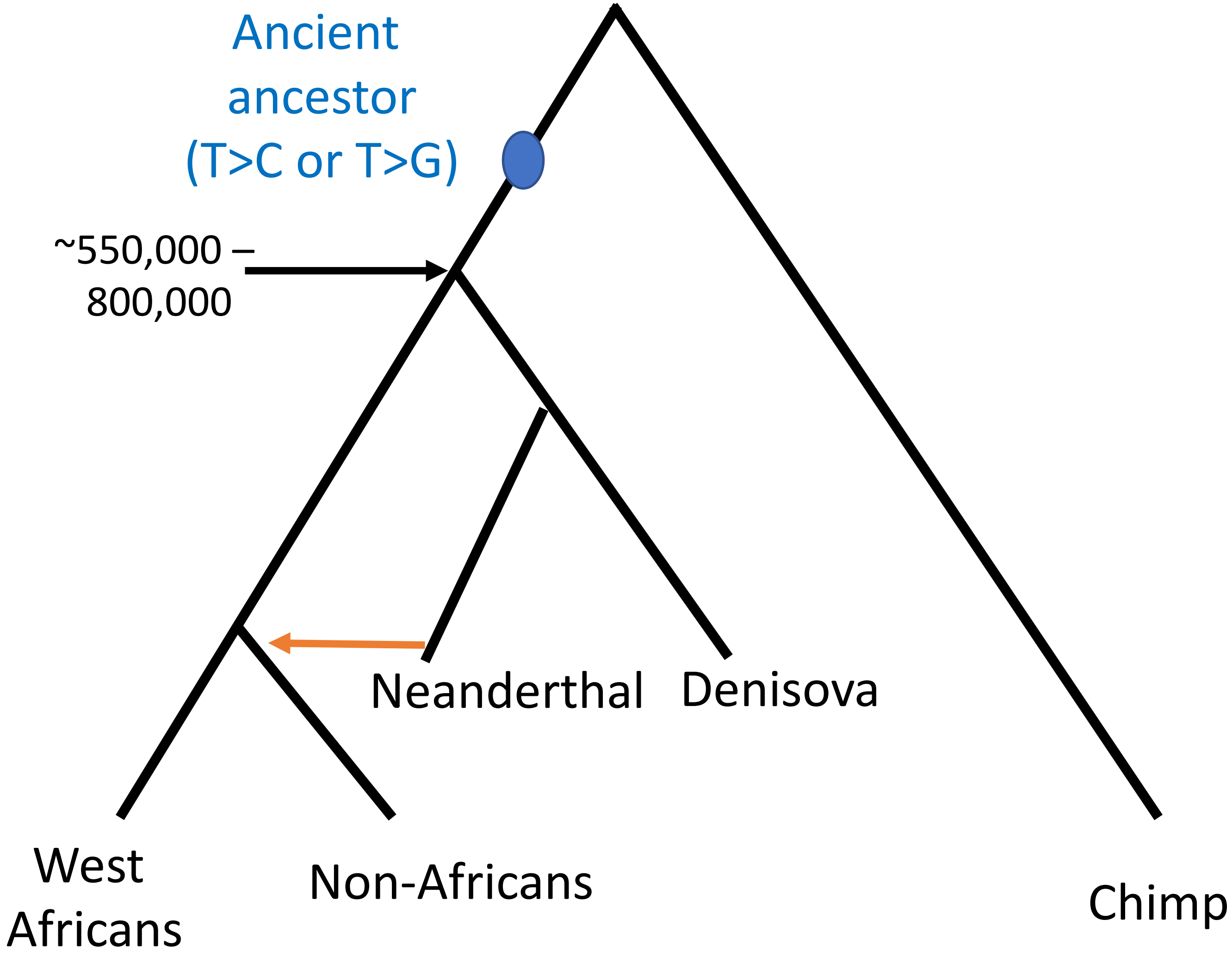
Excluding ND11 (BOTH *N* & *D* derived)



0 = ancestral allele
(same as chimp)
1 = derived allele

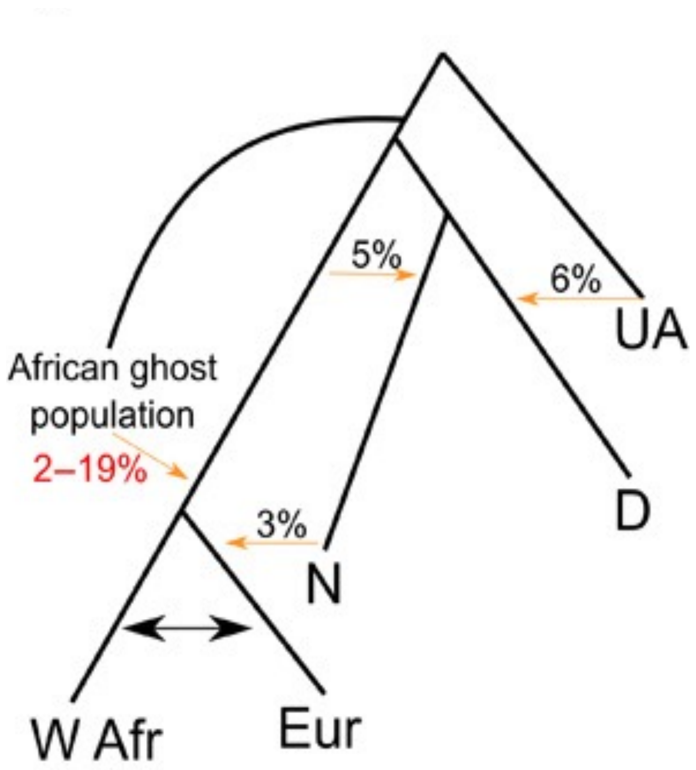
Gao et al. 2022 *biorxiv*

Model


Ancient ancestor (T>C or T>G)
~550,000 – 800,000
Neanderthal
Denisova
West Africans
Non-Africans
Chimp

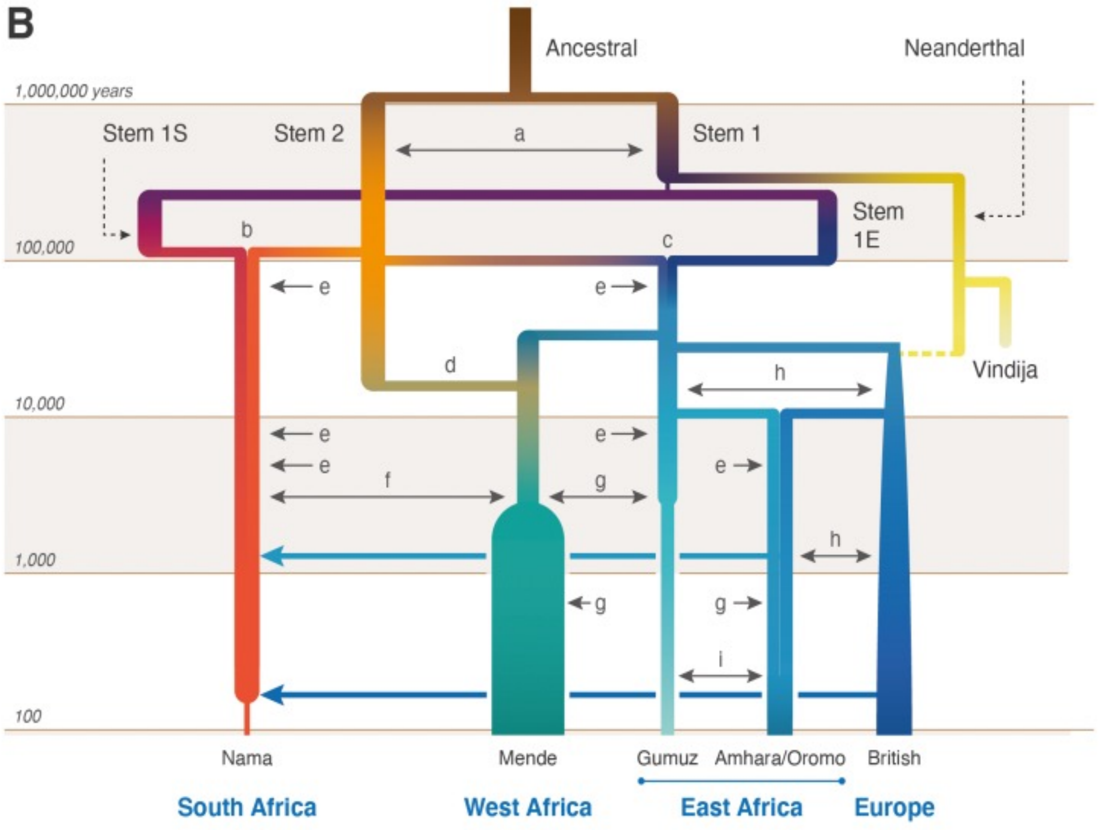
Model: Possible sources

A. Gene flow from unknown archaic hominin



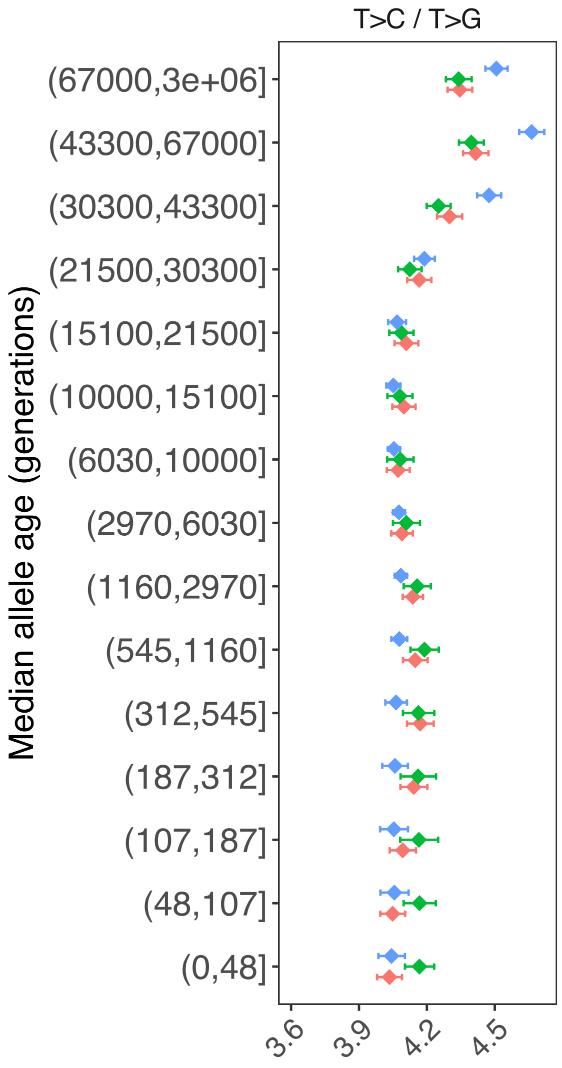
Hammer et al. 2011; Durvasula and Sankararaman 2020; Ragsdale et al. 2020; Speidel et al. 2021

B. Structure in the stem population of Modern Humans

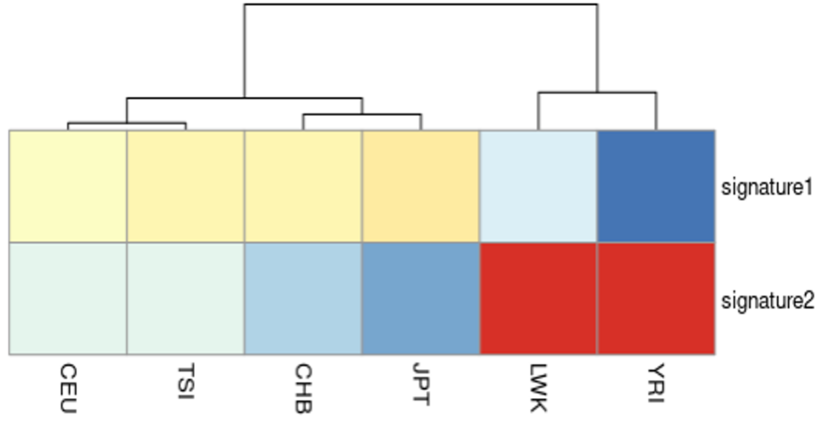


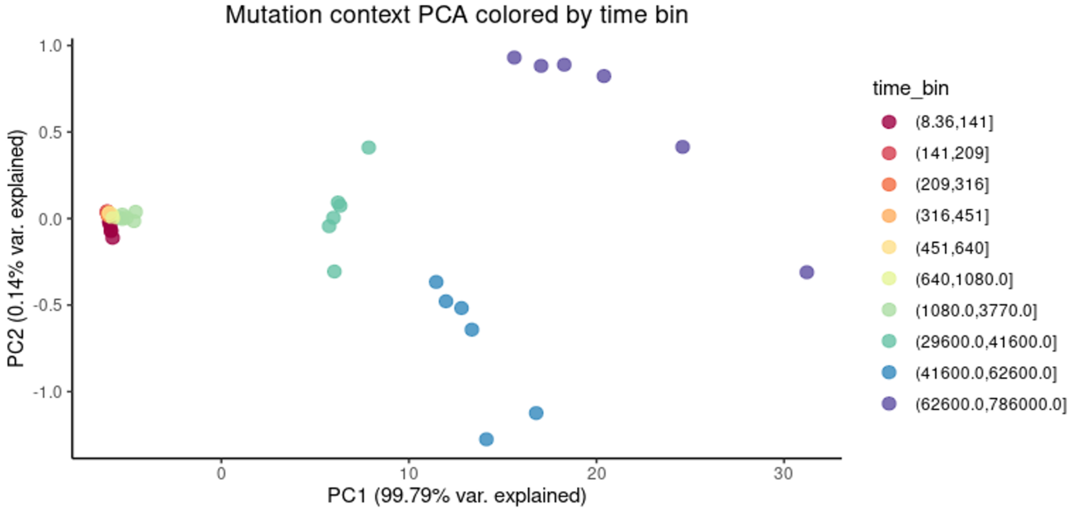
Ragsdale et al. 2022 *biorxiv*

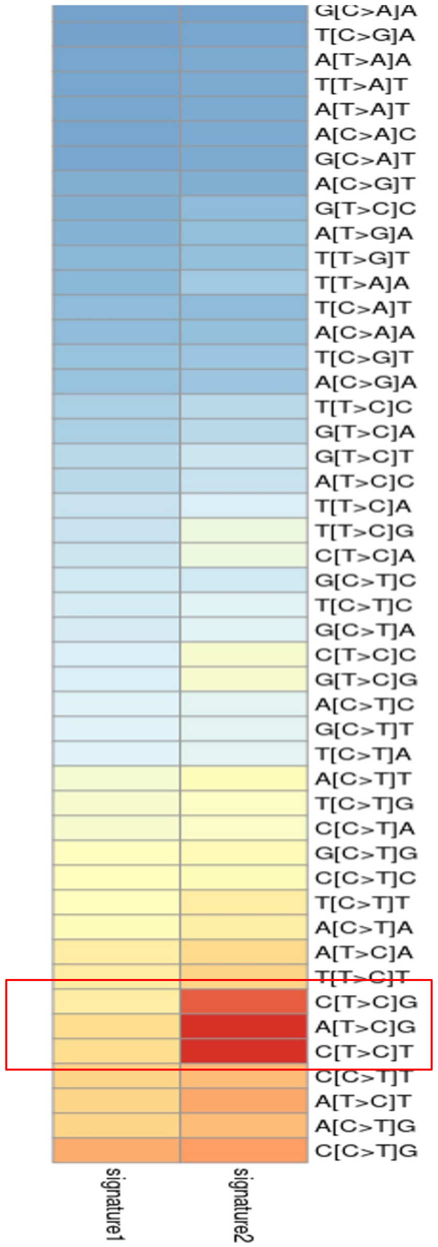
Signal 3: Mutation signature related T>C change in archaic ancestry regions in modern humans



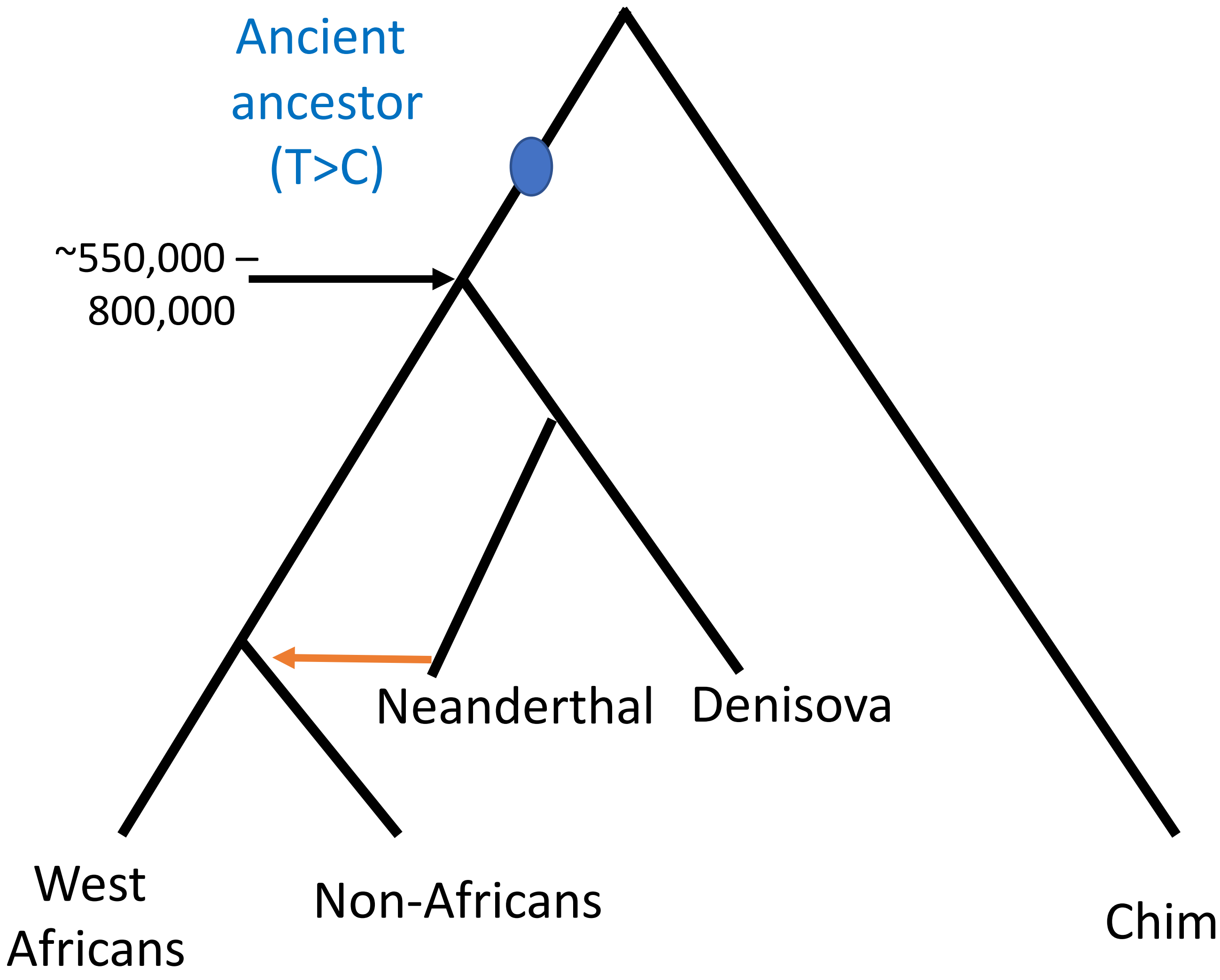
Non-negative matrix factorization





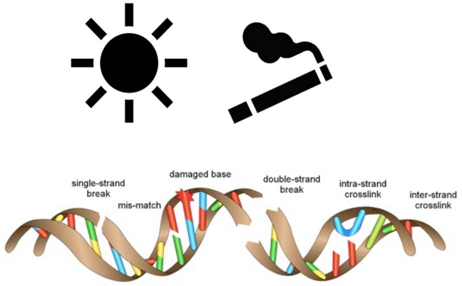


Model

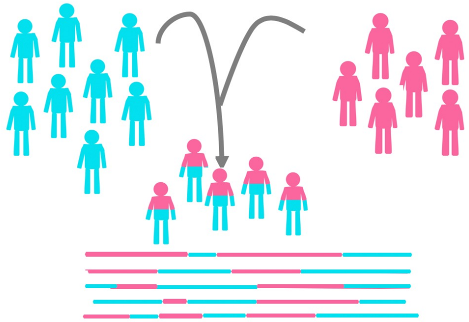


Ancient ancestor (T>C)
~550,000 – 800,000
Neanderthal
Denisova
West Africans
Non-Africans
Chimp

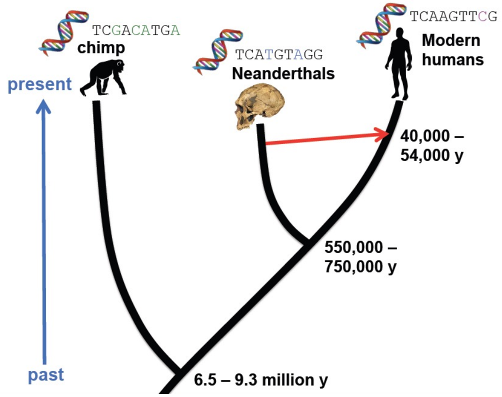
Implications



Non-replicative sources play a non-negligible role in shaping the mutation landscape.



Demography and admixture can have pervasive impacts on shaping genetic variation, including on fundamental parameters such as mutation rate

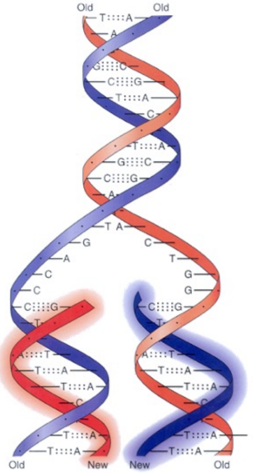


Implications for Molecular clock:

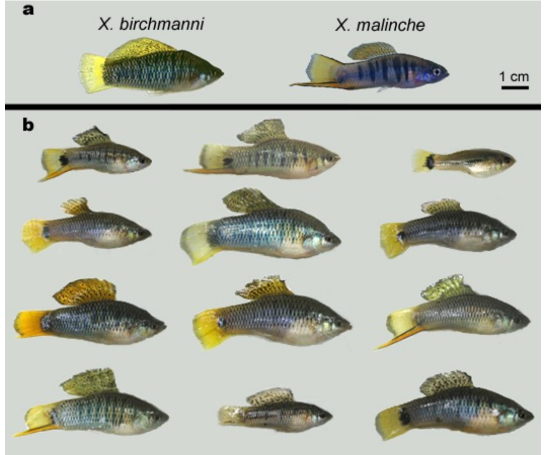
- Unsteady at short timescales within humans, e.g., 10-15% across human populations
- Puzzlingly, little variation across humans and chimpanzees, e.g., on average ~1-2% across species (Moorjani et al. 2016)

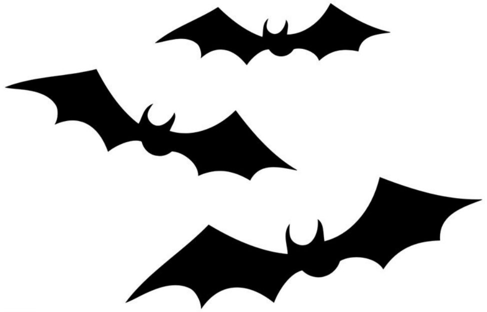
Future Directions: Mutation rates in non-human species

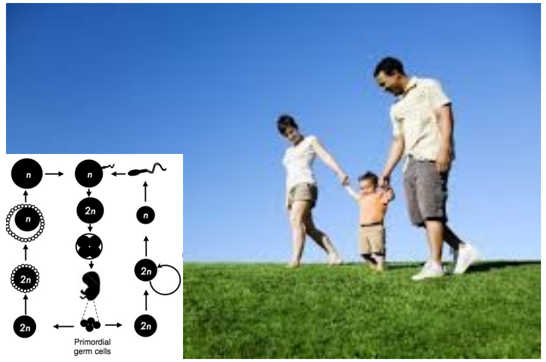
Larger sample sizes



Leveraging hybrid genomes







Studying outlier species

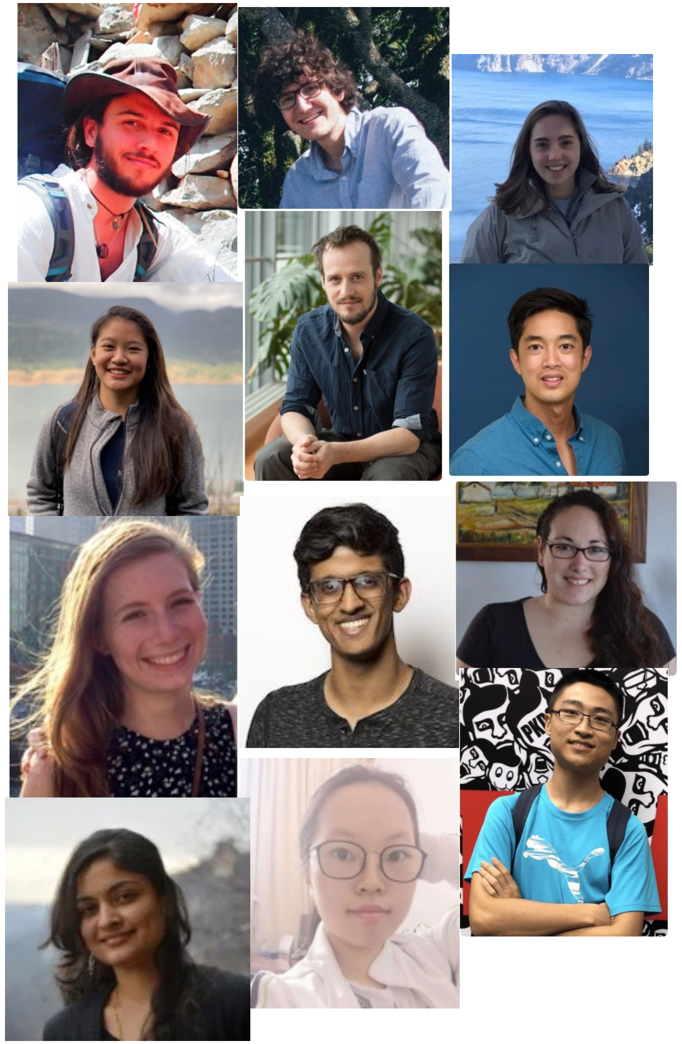
Single cell sequencing of germ cells

This will provide a direct look at the evolution of mutation rates over long evolutionary timescales.

Acknowledgements

Ziyue Gao, University of Pennsylvania
Yulin Zhang, UC Berkeley
Molly Przeworski, Columbia University
Eduardo Amorim, California State University
Felix Wu, Columbia University
Manjusha Chintalapati, UC Berkeley
Nathan Cramer, UC Berkeley

Thanks for helpful discussions:
Roded Sharan, David Pilbeam, Nick Patterson, Monty Slatkin, and members of Moorjani lab



MOORJANI LAB

Koret Berkeley- Tel Aviv Initiative

Alfred P. Sloan FOUNDATION

Hellman Fellows Fund

BURROUGHS WELLCOME FUND

NIH National Institutes of Health