

Lecture 20. Human Evolution

Michael Schatz

November 6, 2024

Applied Comparative Genomics



Class Schedule

M	Oct 14	Epigenome	Project Proposal Assigned
W	Oct 16	Single cell	
M	Oct 21	Transformers	Assignment 5 Assigned
W	Oct 23	Enformer	
M	Oct 28	DL in Genomics	Preliminary Report Assigned
W	Oct 30	Midterm Review	
M	Nov 4	Midterm!	
W	Nov 6	Human evolution	Final Report Assigned
M	Nov 11	Metagenomics	
W	Nov 13	No Class (BIODATA24)	
M	Nov 18	Cancer Genomics	
W	Nov 20	Project Presentation 1	
M	Nov 25	Thanksgiving Break	
W	Nov 27	Thanksgiving Break	
M	Dec 2	Project Presentation 2	
W	Dec 4	Project Presentation 3	
M	Dec 16	Project Report Due	

Preliminary Report

Due Monday November 11

Preliminary Project Report

Assignment Date: October 28, 2024

Due Date: Monday, November 11, 2024 @ 11:59pm

Each team should submit a PDF of your preliminary project proposal (2 to 3 pages) to [GradeScope](#) by 11:59pm on Monday November 11

The preliminary report should have at least:

- Title of your project
- List of team members and email addresses
- 1 paragraph abstract summarizing the project
- 1+ paragraph of Introduction
- 1+ paragraph of Methods that you are using
- 1+ paragraph of Results, describing the data evaluated and any preliminary results
- 1+ paragraph of Discussion (what you have seen or expect to see)
- 1+ figure showing a preliminary result (typically a summary of the data you have identified for your project)
- 5+ References to relevant papers and data

The preliminary report must use the Bioinformatics style template. Word and LaTeX templates are available at

https://academic.oup.com/bioinformatics/pages/submission_online. Overleaf is recommended for LaTeX submissions. Google Docs is recommended for non-latex submissions, especially group projects. Paperpile is recommended for citation management.

Later, you will present your project in class starting the week of November 25. You will also submit your final written report (6-8 pages) of your project by Dec 16

Please use Piazza if you have any questions!

Presentations!

Project Presentations

Presentations will be a total of 12 minutes: 10 minutes for the presentation, followed by 2 minutes for questions. We will strictly keep to the schedule to ensure that all groups can present in class!

Schedule of Presentations

Slot	Date	Start	Team Name	Team Members	Project Title
1	11/20	3:00	Two single-cells, one big problem	Kevin Meza Landeros, YunZhou Liu	Cluster-based single-cell RNA-seq variant detection
2	11/20	3:12	Team Yuxiang Li	Yuxiang Li	Contrastive Learning Approach to Integrate Single-Cell scRNA-seq and scATAC-seq for Mechanistic Understanding of Gene Regulation
3	11/20	3:24	Team Roujin An	Roujin An	Cell Type-Specific SNP-to-Splicing Variants Mapping Using Deep Learning Models
4	11/20	3:36	Team Miller	Logan Miller	Population-Specific Evolutionary Hotspots in Human Genomes
5	11/20	3:48	Team1D	Ben Miller	Comparative Genomic Analysis of NOD and (Simulated) NOR Mouse Genomes to Identify Variants Associated with Type 1 Diabetes
1	12/2	3:00	Genomic Visionaries	Iason Mihalopoulos, Siam Mohammed	AR/VR Visualization of Individual Genomes with AI-driven Insights
2	12/2	3:12	Silent Codebreakers	Cecelia Zhang, Jiarui Yang	Benchmarking Non-Coding Mutation Analysis Schemes on Cancer Genomes
3	12/2	3:24	Team Table	Oce Bohra, Zoe Rudnick	The emerging contribution of non-coding mutations in glioblastoma development
4	12/2	3:36	Team Brady	Brady Bock	DNN analysis of gut microbiomes to predict colorectal cancer disease state
5	12/2	3:48	Variant Visionaries	Alexandra Gorham, Christine Park, Natalie Vallejo	Benchmarking Non-coding Variant Scoring Tools for Cancer Pathogenicity Prediction
6	12/2	4:00	Human to Plants	Xiaojun Gao, Yujia, Yushan Zou	Evaluation of applicability of ChromHMM for Plants in Chromatin States and Gene Expression
1	12/4	3:00	SE Palmeiras	Caleb Hallinan, Jamie Moore, Rafael dos Santos Peixoto	Evaluating cell-type clustering algorithm's robustness to technical artifacts via synthetic spatial transcriptomics data
2	12/4	3:12	Nuclencoder	Amanda Xu, Angela Yang, Jiamin Li	DNA Cryptography: Digital Signatures for Encryption to Facilitate Safe Data Storage
3	12/4	3:24	Geoguessr	Alex Ostrovsky, Nicole Lauren Brown	Investigating geographic and environmental effects on soil metagenomes by correlating GIS data
4	12/4	3:36	Quetzalli Tlalli	Arshana Welivita, Atticus Colwell	Benchmarking Methods for Inferring the Ethnicity of an Individual from Their Genotype
5	12/4	3:48	Team Barbour	Alexis Barbour	Benchmarking non-coding mutation analysis schemes for evaluating Type 1 Diabetes
6	12/4	4:00	All of Us Team	Levon Galstyan, Nitish Aswani, Talia Haller	Genomic Insights into Sleep Patterns, Disease Outcomes, and Biomarker Associations using the All of Us Dataset

Presentations!

10 min + 2 min questions

Recommended outline for your talk (~1 minute per slide):

1. Title Slide: Who are you, title, date
2. Intro 1: What's the big idea???
3. Intro 2: More specifically, what are you trying to learn?
4. Methods 1: What did you try?
5. Methods 2: What is the key idea?
6. Data 1: What data are you looking at?
7. Data 2: Anything notable about the data?
8. Results 1: What did you see!
9. Results 2: How does it compare to other methods/data/ideas?
10. Discussion 1: What did you learn from this study?
11. Discussion 2: What does this mean for the future?
12. Acknowledgements: Who helped you along the way?
13. Thank you!

I strongly *discourage* you from trying to give a live demo as they are too unpredictable for a short talk. If you have running software you want to show, use a "cooking show" approach, where you have screen shots of the important steps.

Final Report

Due Monday December 16

Final Project Report

Assignment Date: November 6, 2024

Due Date: Monday, December 16, 2024 @ 11:59pm

Each team should submit a PDF of your final project proposal (6 to 9 pages) to GradeScope by 11:59pm on Monday December 16. No late days can be used as grades must be submitted to the registrar that week.

The report should have at least:

- Title of your project
- List of team members and email addresses
- 1-2 paragraph abstract summarizing the project
- 1-2 pages of Introduction: Background, what is the big problem/question you are addressing, overview of data used, summary of results
- 2-3 pages of Methods that you are using: if you are primarily using existing methods, please describe those methods
- 2-3 pages of Results: be sure to describe the data evaluated along with the results of your analysis. If computational time is measured, please list the machine specifications
- 1 page of Discussion: what you have seen or how that relates to other papers
- Please include 4-6 main figures showing your results. If you have more figures, please include them in a supplemental figures section at the end of the PDF.
- 1 paragraph of acknowledgements
- 1/2 to 1 page of references to relevant papers and data

The report should use the Bioinformatics style template. Word and LaTeX templates are available at

https://academic.oup.com/bioinformatics/pages/submission_online. You can (and should) expand on your preliminary report into the full report.

Please use Piazza if you have any questions!

Exam Debrief

Midterm Exam

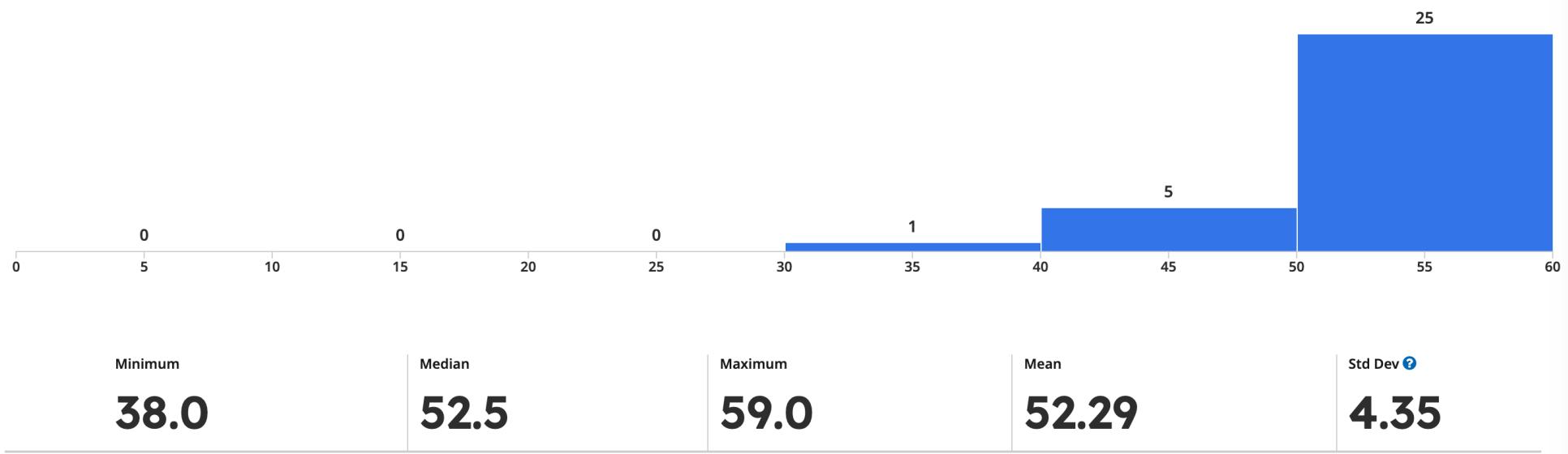
Page 1 of 8

601.449/649 Computational Genomics: Applied Comparative Genomics Midterm Exam

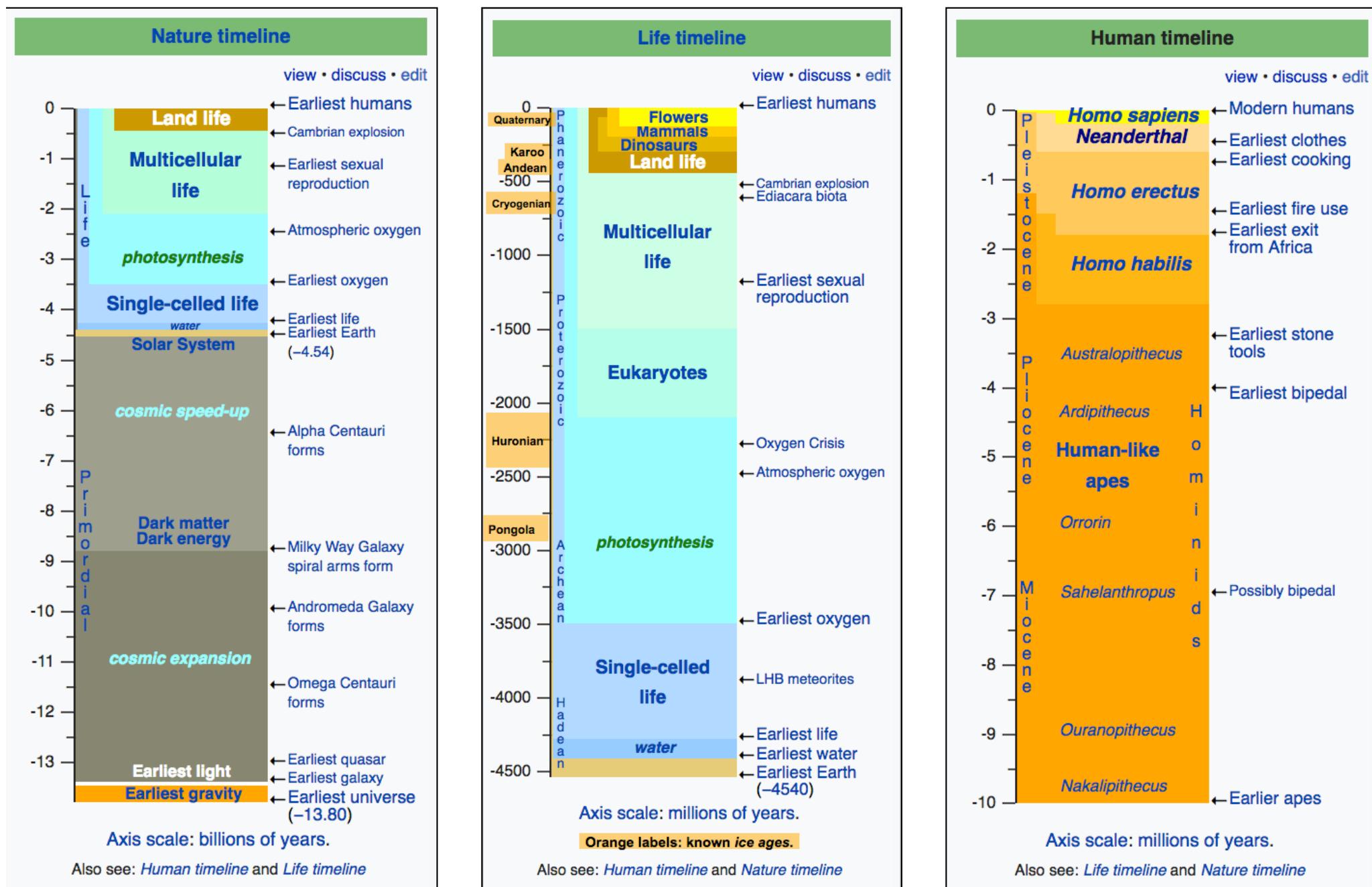
Michael C. Schatz
mschatz@cs.jhu.edu

November 4, 2024
Time: 75 Minutes

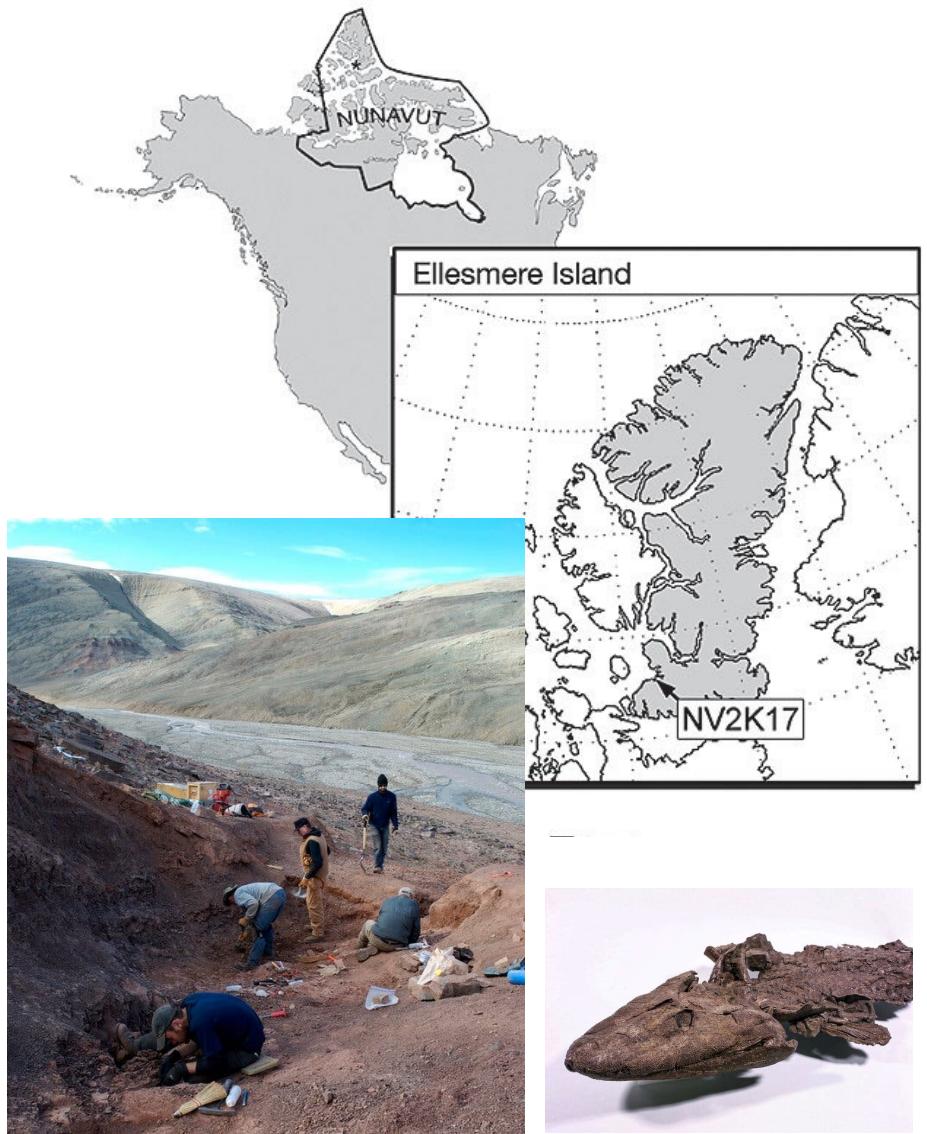
Exam Debrief



Our Origins



Tiktaalik roseae



A Devonian tetrapod-like fish and the evolution of the tetrapod body plan
Daeschler, Shubin, Jenkins (2006) Nature. <https://doi.org/10.1038/nature04639>

Sequencing ancient genomes

Janet Kelso

Max-Planck Institute

A
A C G
T G C G
A T C G A G
G A C T C A C
T G T A G T G
G T G A G G C C A C
T C A C T C G T G
T A T C T A C C G
A C A C G G T C T G
C T A G A T A A G A C T
T C C A T C A T C A C
A T G T G A T C T C
T C G T A C A c G G A
C A G C T G A G C
T A G T T G C T A C
G C T A C A G T



www.ted.com/talks/svante_paabo_dna_clues_to_our_inner_neanderthal

TED Ideas worth spreading

WATCH DISCOVER ATTEND PARTICIPATE ABOUT SIGN IN MEMBERSHIP

Become a TED Member

Want to hear more great ideas like this one? Sign up for TED Membership to get exclusive access to captivating conversations, engaging events, and more!

Join now

Watch next

1,629,829 views | Svante Pääbo • TEDGlobal 2011

Like (48K) Share Add

DNA clues to our inner neanderthal

Sharing the results of a massive, worldwide study, geneticist Svante Pääbo shows the DNA proof that early humans mated with Neanderthals after we moved out of Africa. (Yes, many of us have Neanderthal DNA.) He also shows how a tiny bone from a baby finger was enough to identify a whole new humanoid species.

Read transcript

This talk was presented at an official TED conference. TED's editors chose to feature it for you.

TED is supported by ads and partners

About the speaker

Svante Pääbo
Geneticist

See speaker profile >

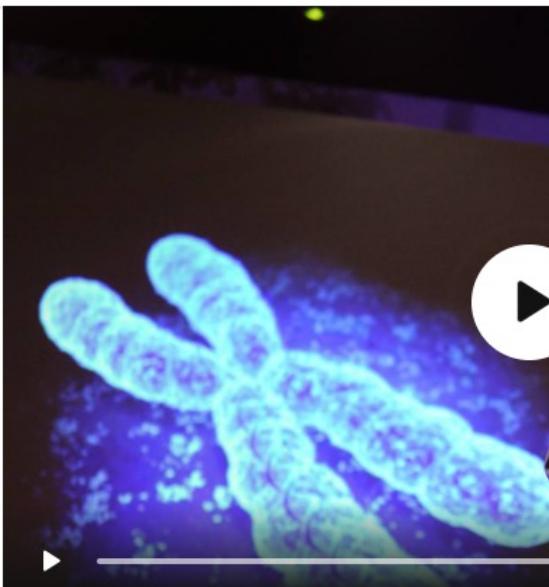
Svante Pääbo explores human genetic evolution by analyzing DNA extracted from ancient sources, including mummies, an Ice Age hunter and the bone fragments of Neanderthals.

www.ted.com/talks/svante_paabo_dna_clues_to_our_inner_neanderthal

TED Ideas worth spreading

WATCH DISCOVER ATTEND PARTICIPATE ABOUT SIGN IN MEMBERSHIP

Become a TED Member



1,629,829 views | Svante Pääbo • TEDGlobal 2011

DNA clues to our inner neanderthal

Sharing the results of a massive, worldwide study, geneticist Svante Pääbo shows DNA proof that early humans mated with Neanderthals after they split from our lineage. (Yes, many of us have Neanderthal DNA.) He also shows how a single finger was enough to identify a whole new humanoid species.

This talk was presented at an official TED conference. TED's general audience talks are available to the public. TEDTalks TEDSummits TEDPartnerships TEDBooks TEDx

About the speaker

 **Svante Pääbo**
Geneticist

Svante Pääbo explores human genetic evolution by analyzing DNA from ancient hominins, including mummies, an Ice Age hunter and the bone fragments of Neanderthals.

www.nobelprize.org/prizes/medicine/2022/paabo/facts/

THE NOBEL PRIZE

Nobel Prizes & Laureates Nomination Alfred Nobel News & insights Events Educational

Medicine The Nobel Prize in Physiology or Medicine 2022 Svante Pääbo - Facts

The Nobel Prize in Physiology or Medicine 2022

Svante Pääbo

Share this



Svante Pääbo Facts



Svante Pääbo
The Nobel Prize in Physiology or Medicine 2022

Born: 20 April 1955, Stockholm, Sweden

Affiliation at the time of the award: Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, Okinawa Institute of Science and Technology, Okinawa, Japan

Prize motivation: "for his discoveries concerning the genomes of extinct hominins and human evolution"

Prize share: 1/1

Work

Where do we humans come from, and how are we related to extinct hominins? In 2010, Svante Pääbo succeeded in sequencing the genome of the Neanderthal. He also discovered a previously unknown hominin, Denisova. He also found that gene transfer had occurred from these now extinct hominins to Homo sapiens following the migration out of Africa around 70,000 years ago. This ancient flow of genes to present-day humans has physiological relevance today, for example affecting how our immune system reacts to infections.



Homo neanderthalensis

- Proto-Neanderthals emerge around 600k years ago
- “True” Neanderthals emerge around 200k years ago
- Died out approximately 40,000 years ago
- Known for their robust physique
- Made advanced tools, probably had a language (the nature of which is debated and likely unknowable) and lived in complex social groups



Homo sapiens sapiens

- Apparently emerged from earlier hominids in Africa around 50k years ago
- Capable of amazing intellectual and social behaviors
- Mostly Harmless ☺

A Draft Sequence of the Neandertal Genome

Richard E. Green, et al.
Science **328**, 710 (2010);
DOI: 10.1126/science.1188021

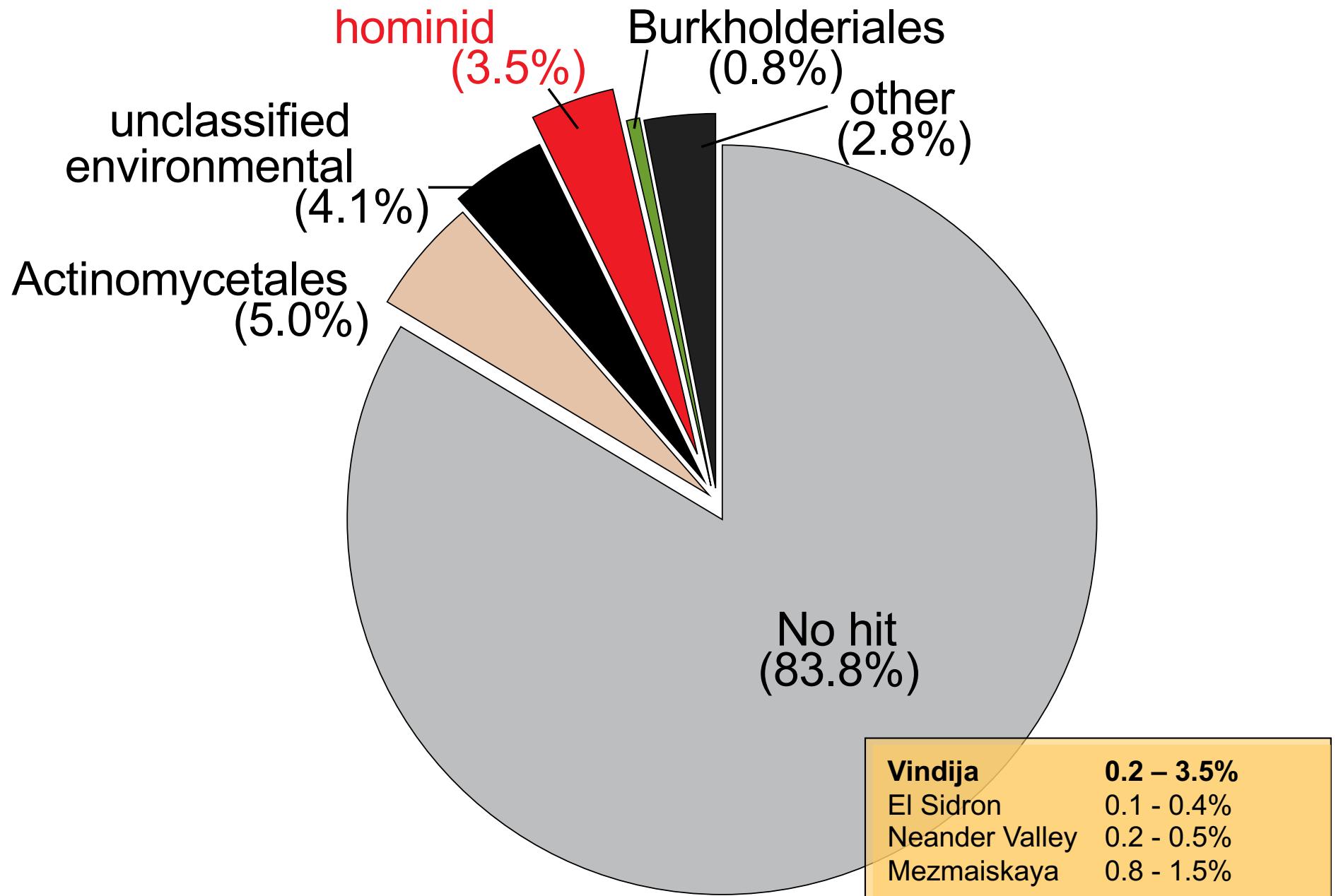
A**B**

Fig. 1. Samples and sites from which DNA was retrieved. **(A)** The three bones from Vindija from which Neandertal DNA was sequenced. **(B)** Map showing the four archaeological sites from which bones were used and their approximate dates (years B.P.).

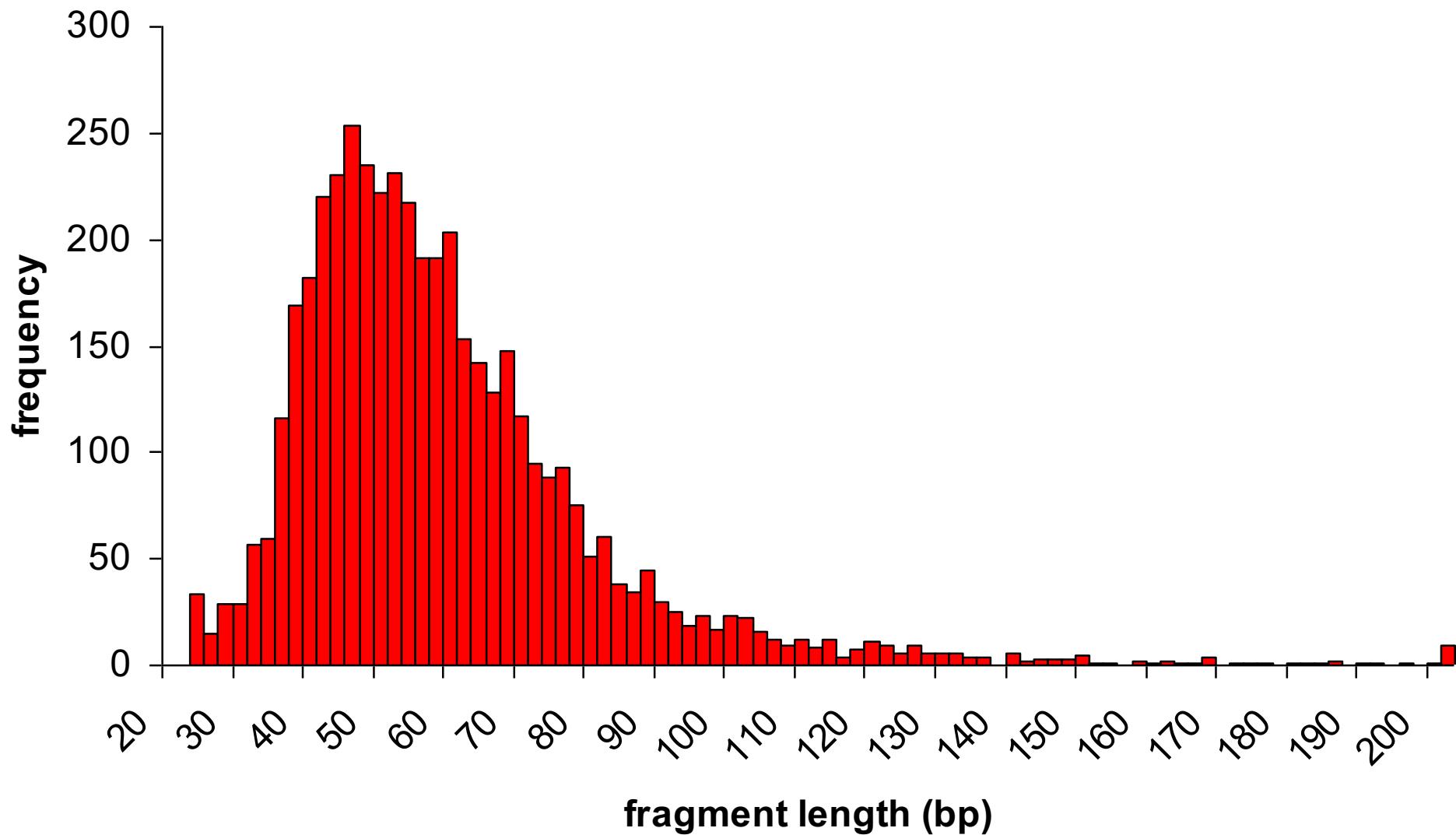
Extracting Ancient DNA



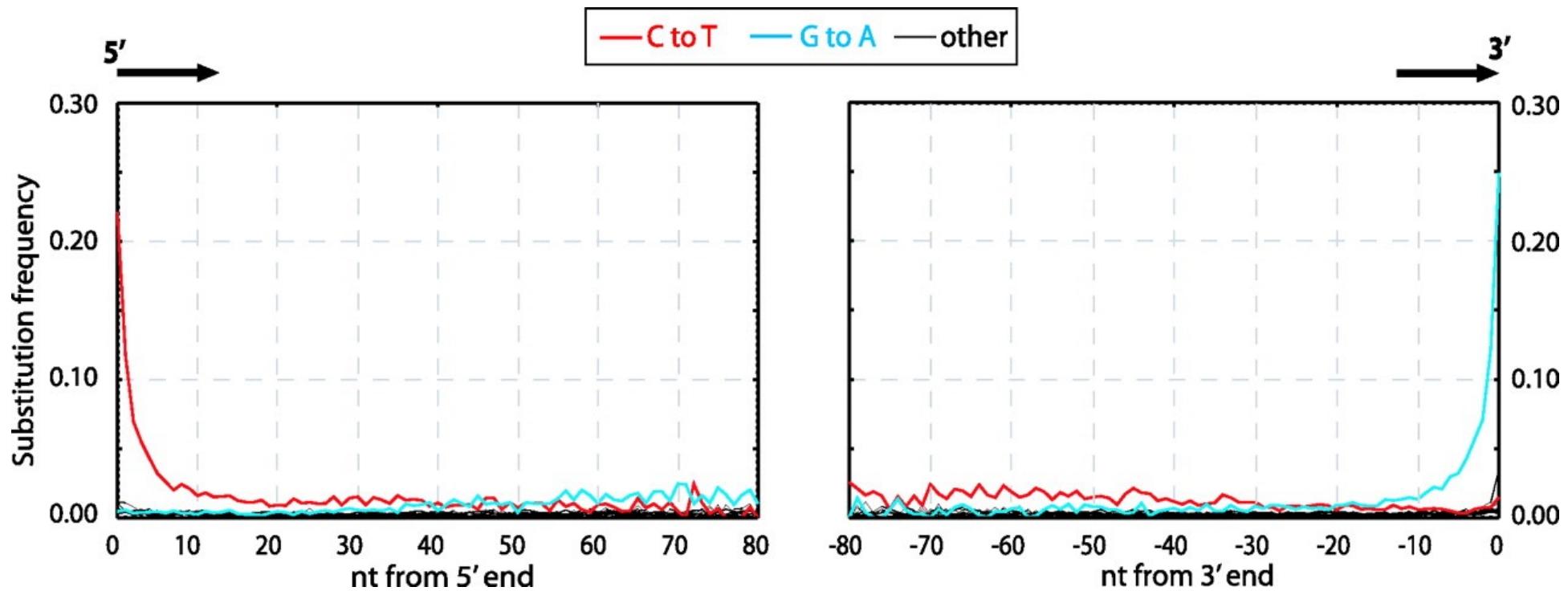
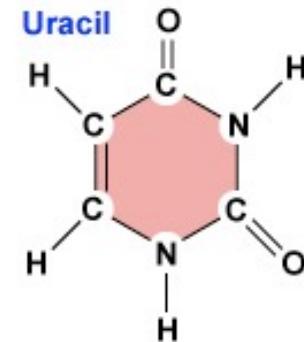
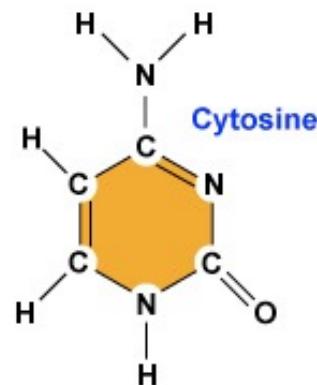
DNA is from mixed sources



DNA is degraded



DNA is chemically damaged





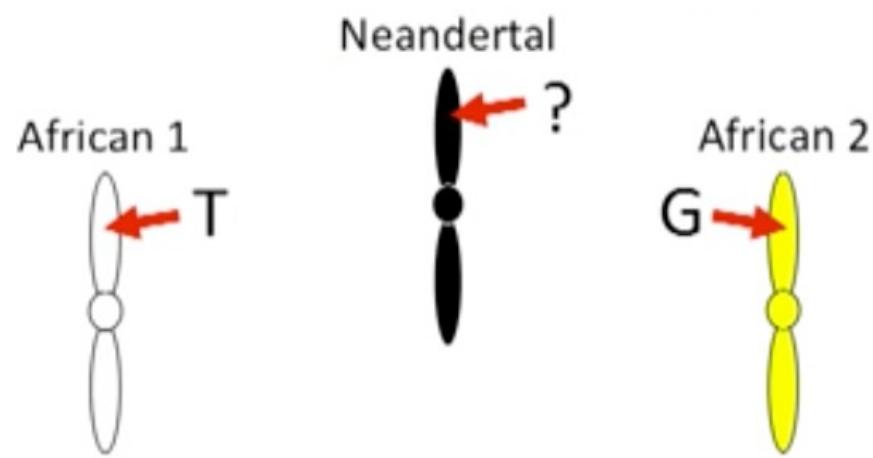
Green et al. 2010

Vindija 33.16	~1.2 Gb
33.25	~1.3 Gb
33.26	~1.5 Gb
El Sidron (1253)	~2.2 Mb
Feldhofer 1	~2.2 Mb
Mezmaiskaya 1	~56.4 Mb

~35 Illumina flow cells

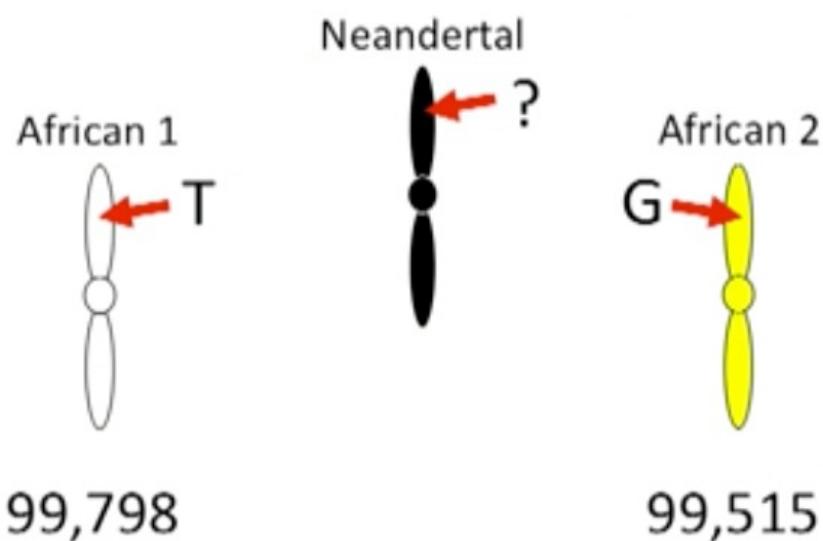
Genome coverage ~1.3 X

Did we mix?



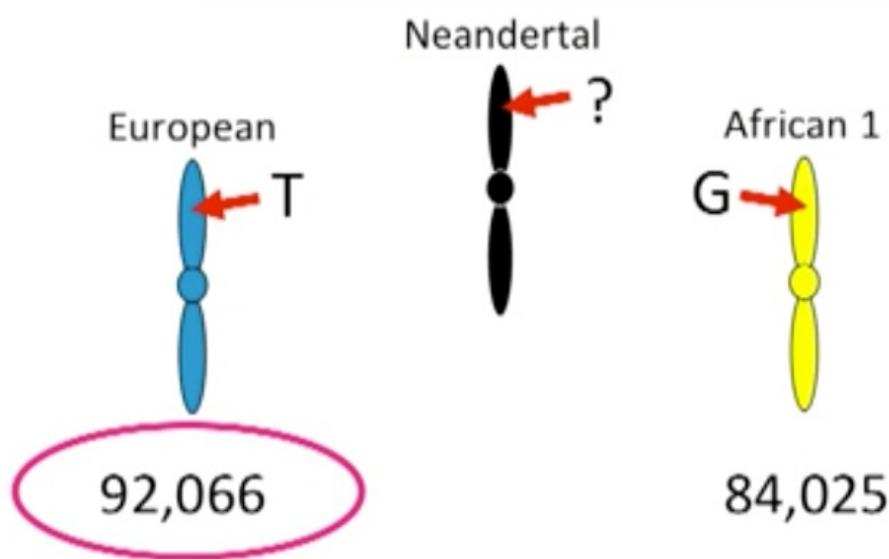
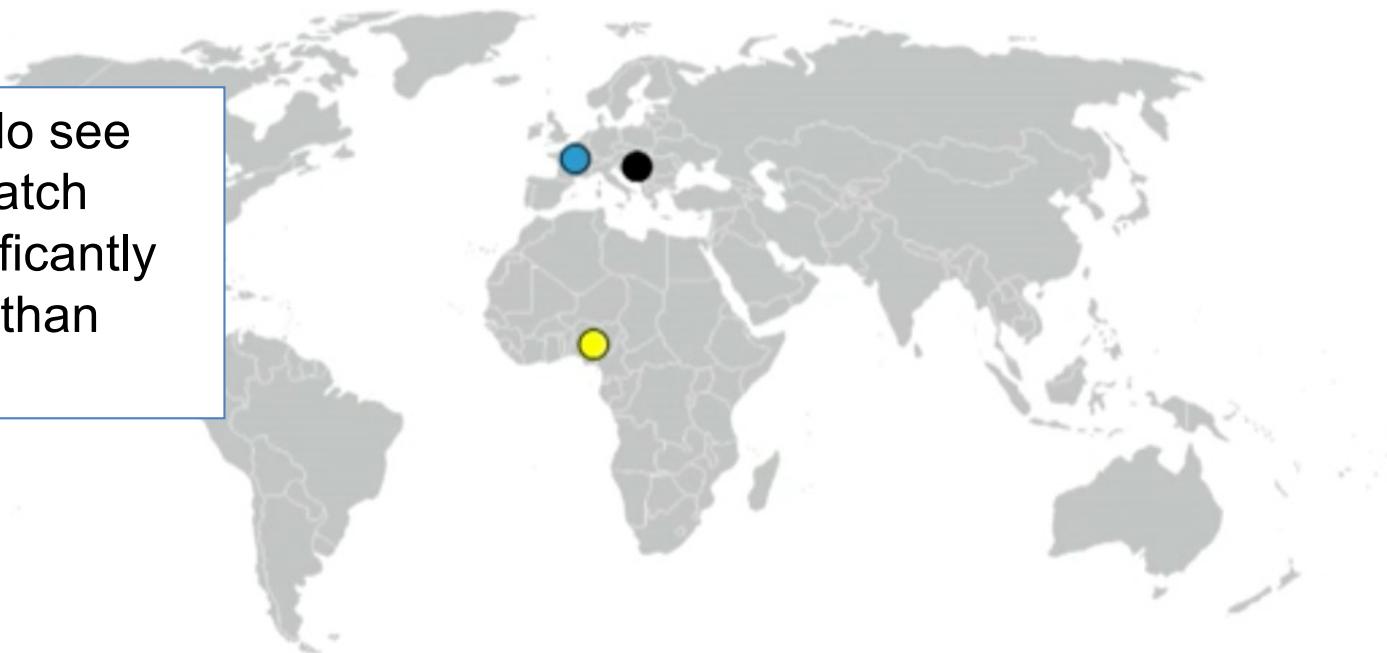
Did we mix?

As far as we know,
Neanderthals were never
in Africa, and do not see
Neanderthal alleles to be
more common in one
African population over
another



Did we mix?

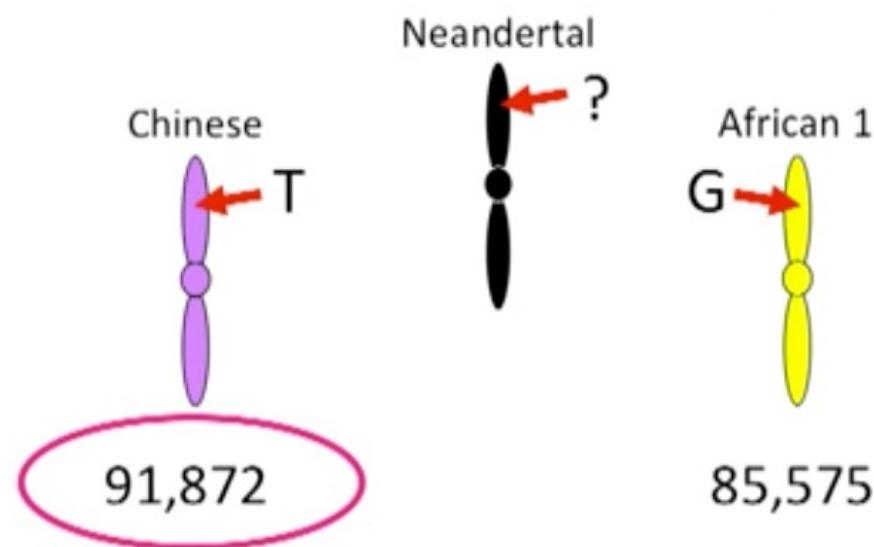
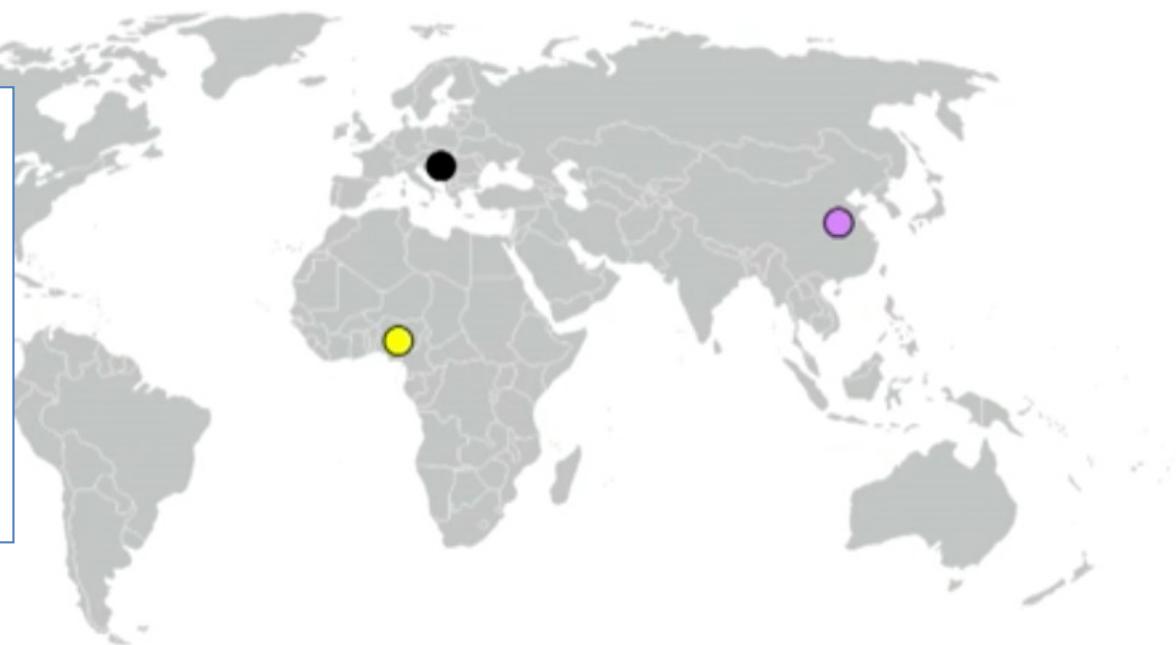
In contrast, we do see Neanderthals match Europeans significantly more frequently than Africans



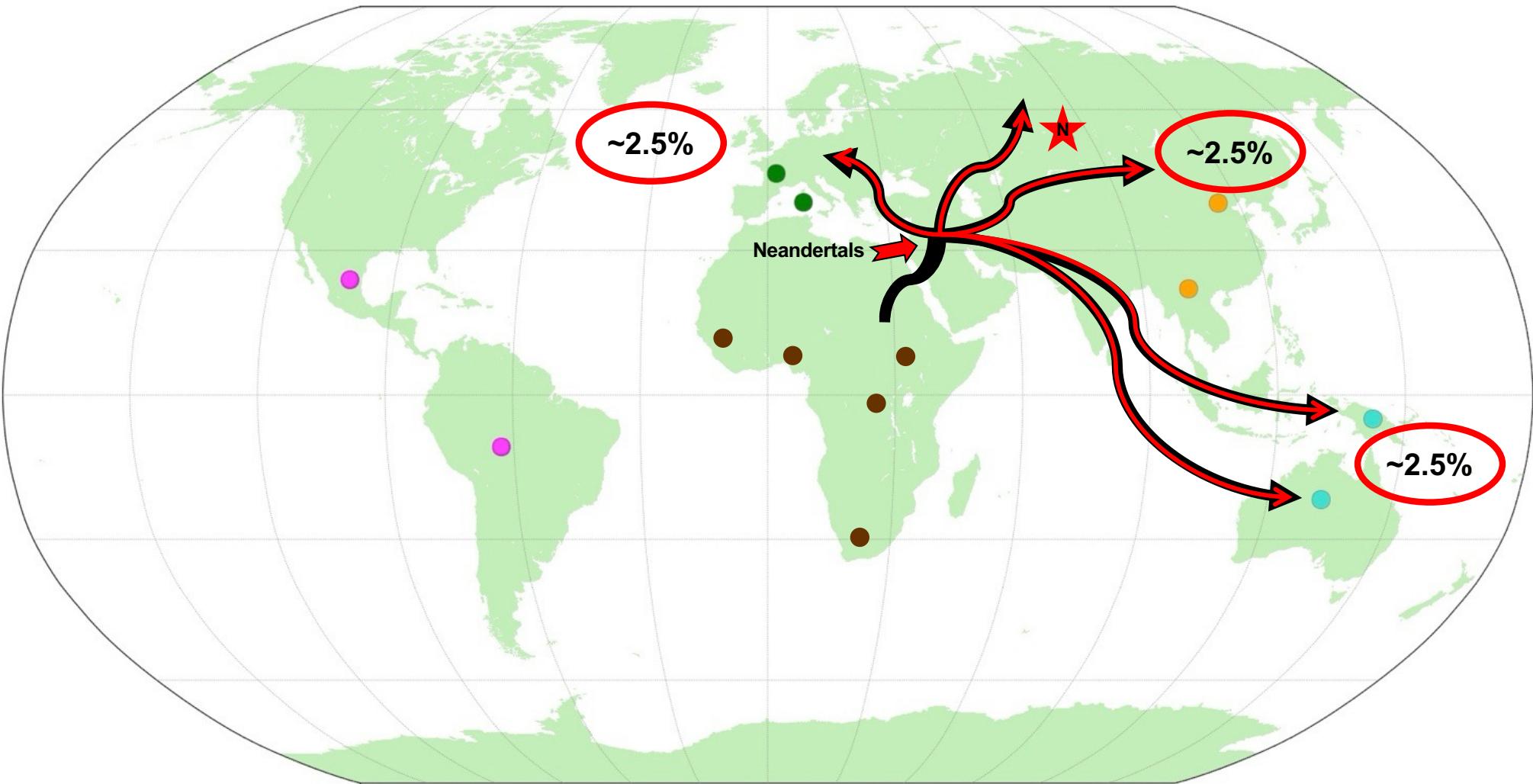
Did we mix?

Also see Neanderthals
match Chinese
significantly more
often...

... but Neanderthals
never lived in China!



Neanderthal Interbreeding



As modern humans migrated out of Africa, they apparently interbred with Neanderthal's so we see their alleles across the rest of the world and carry about 2.5% of their genome with us!

What about other ancient hominids?



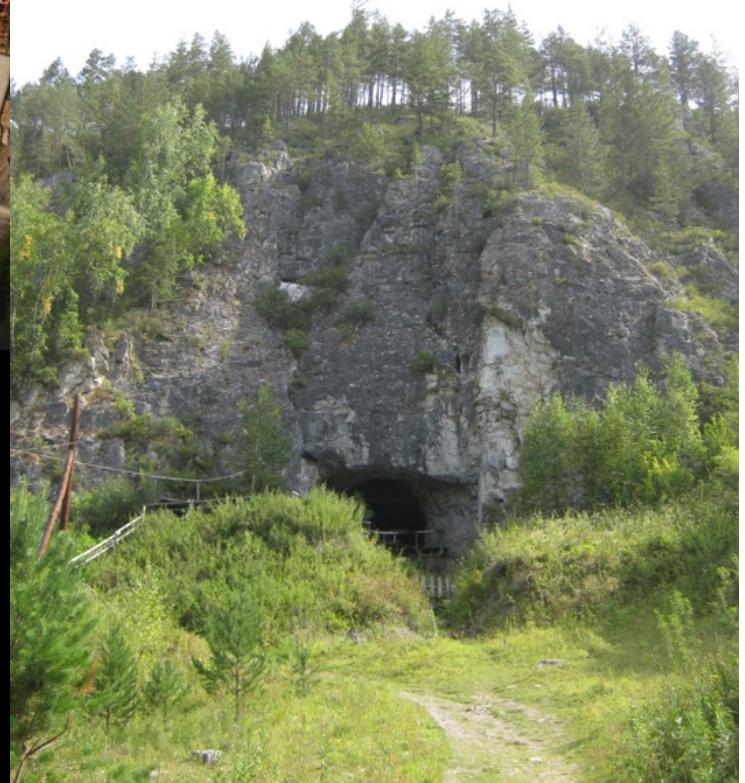
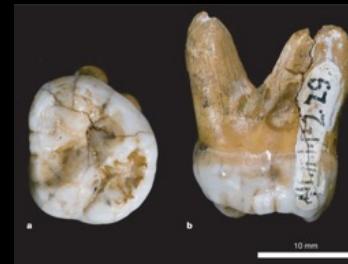
Denisova cave

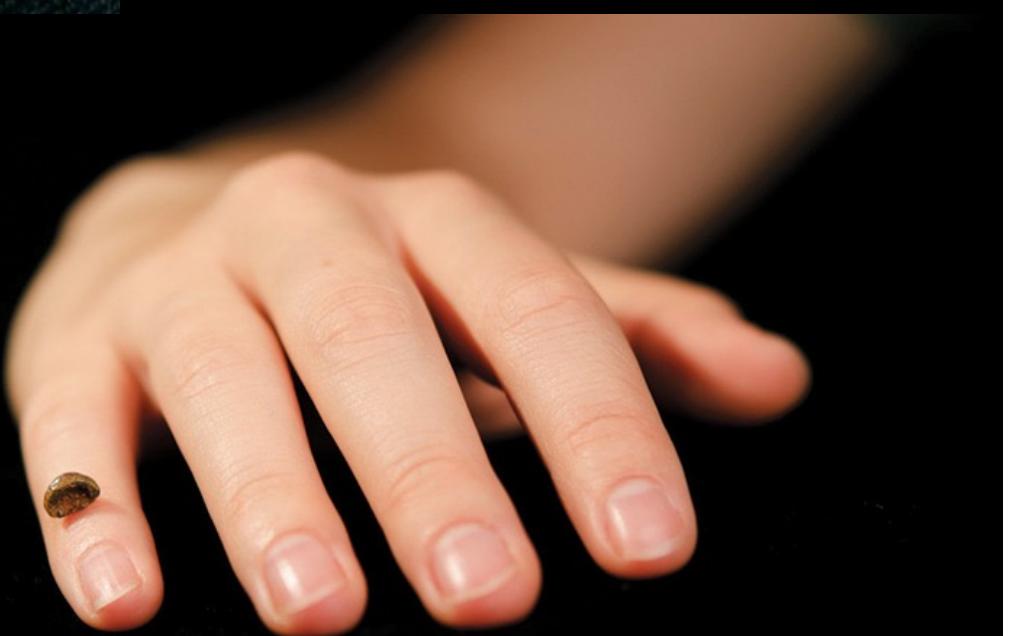
Altai mountains

Russia

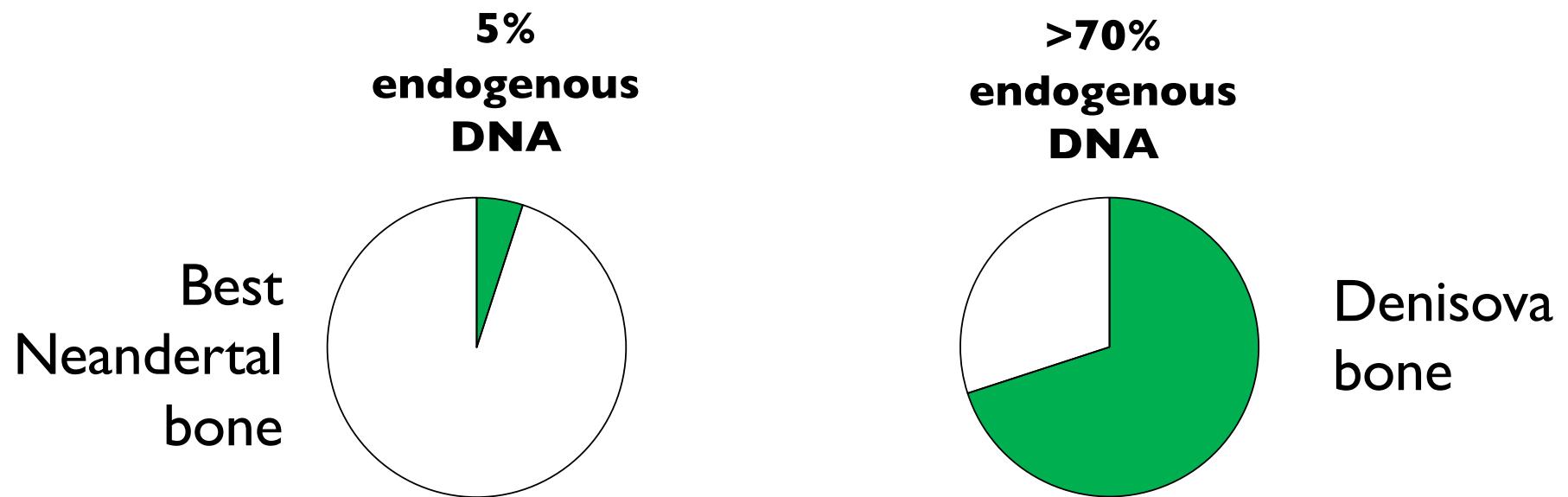
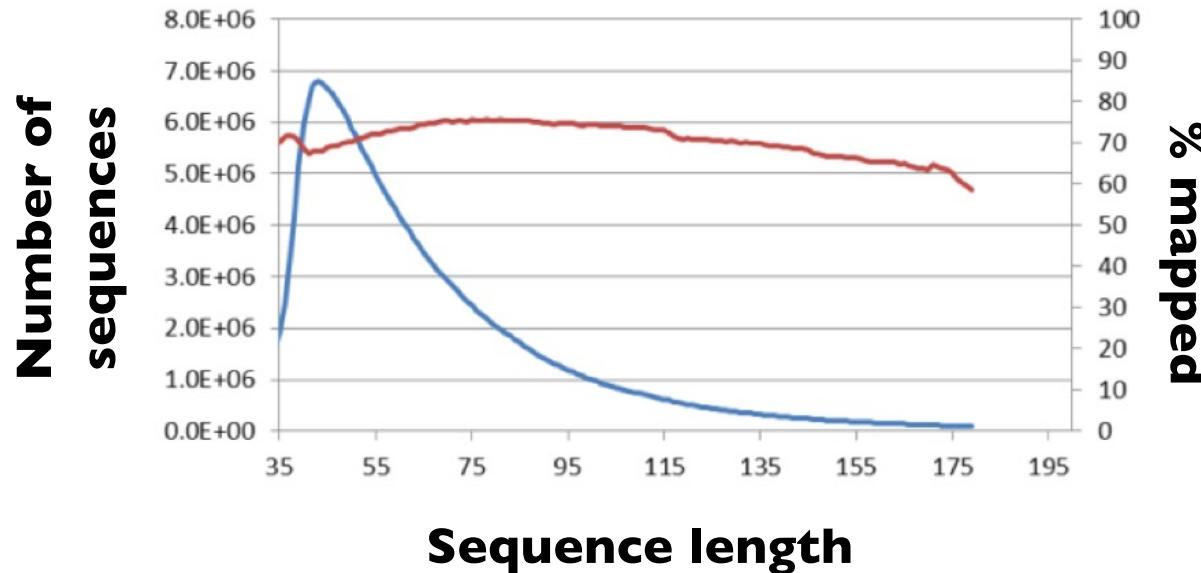


Academician A.P. Derevianko

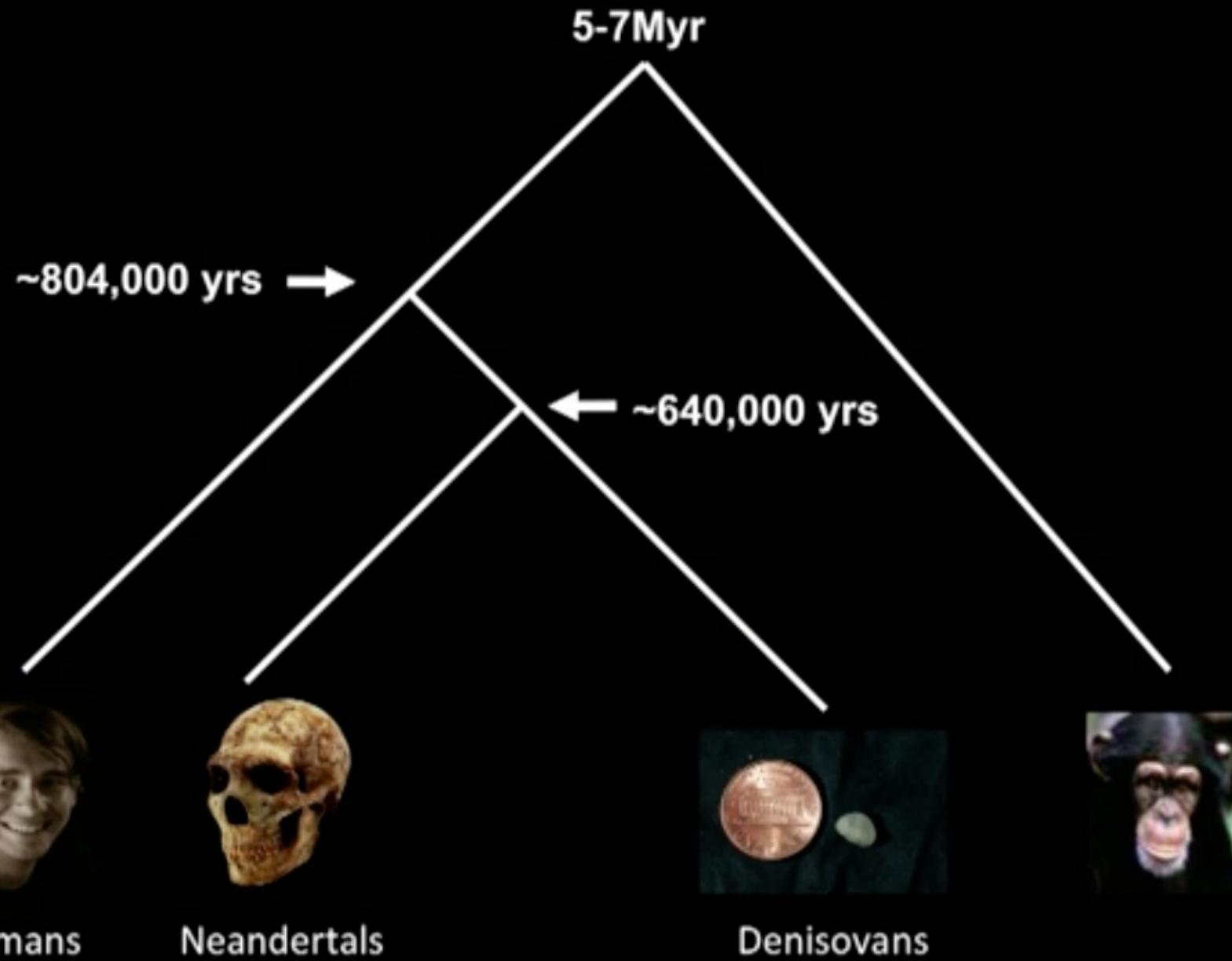




Extraordinary preservation



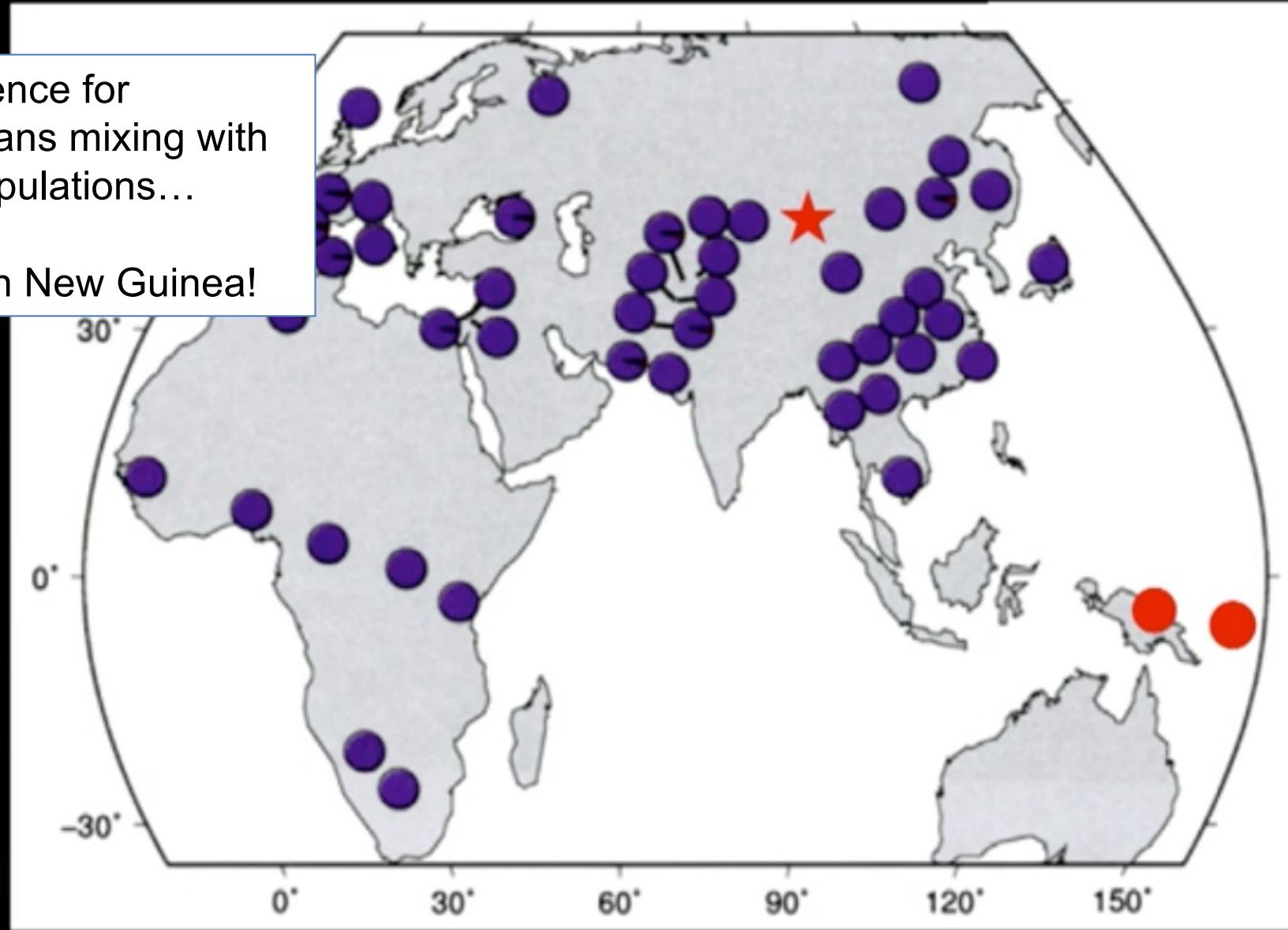
Denisovans & Neandertals



Did we mix?

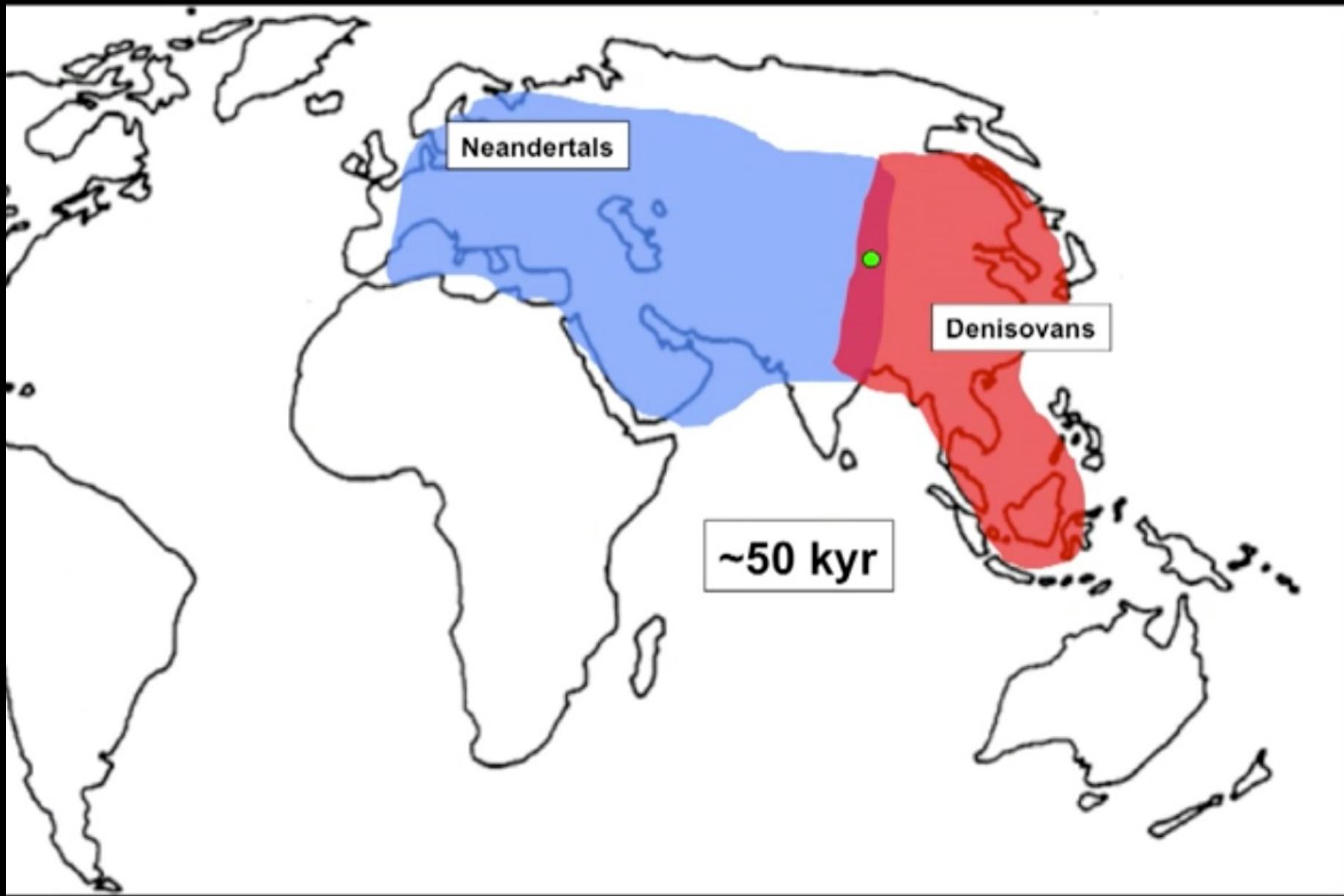
No evidence for
Denisovans mixing with
other populations...

Except in New Guinea!

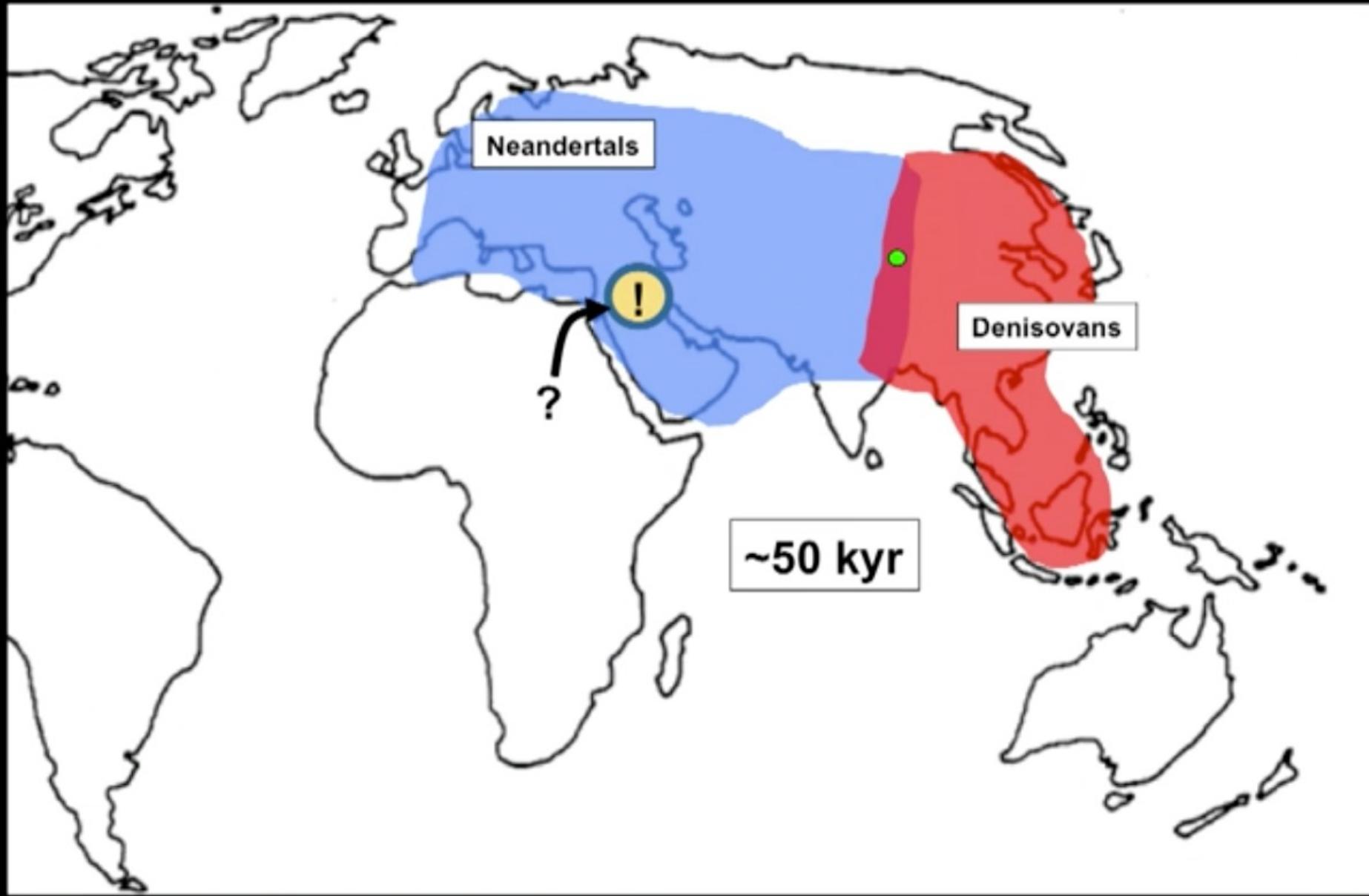


Map after Pickrell et al., 2009

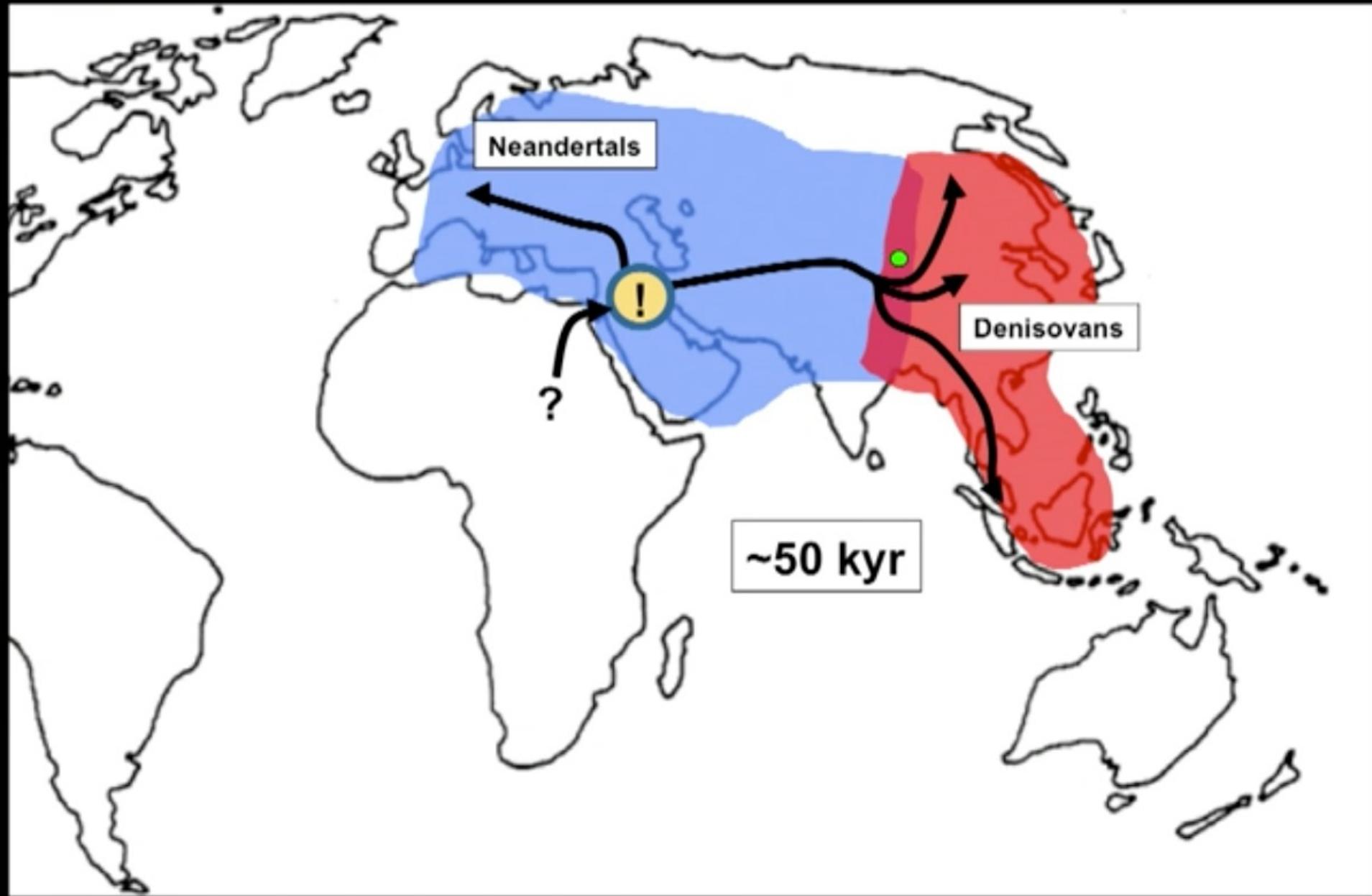
Timeline of ancient hominids



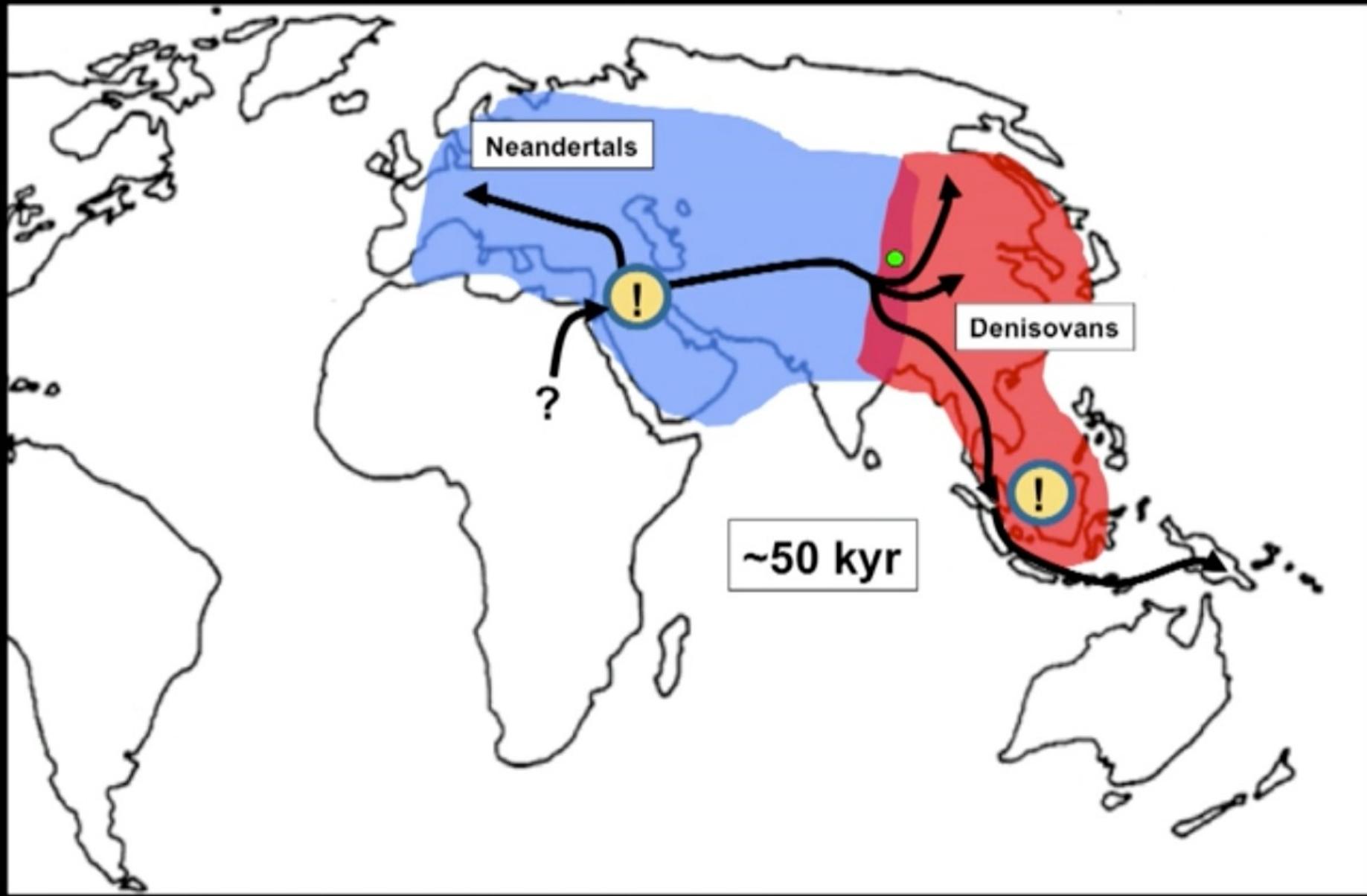
Timeline of ancient hominids



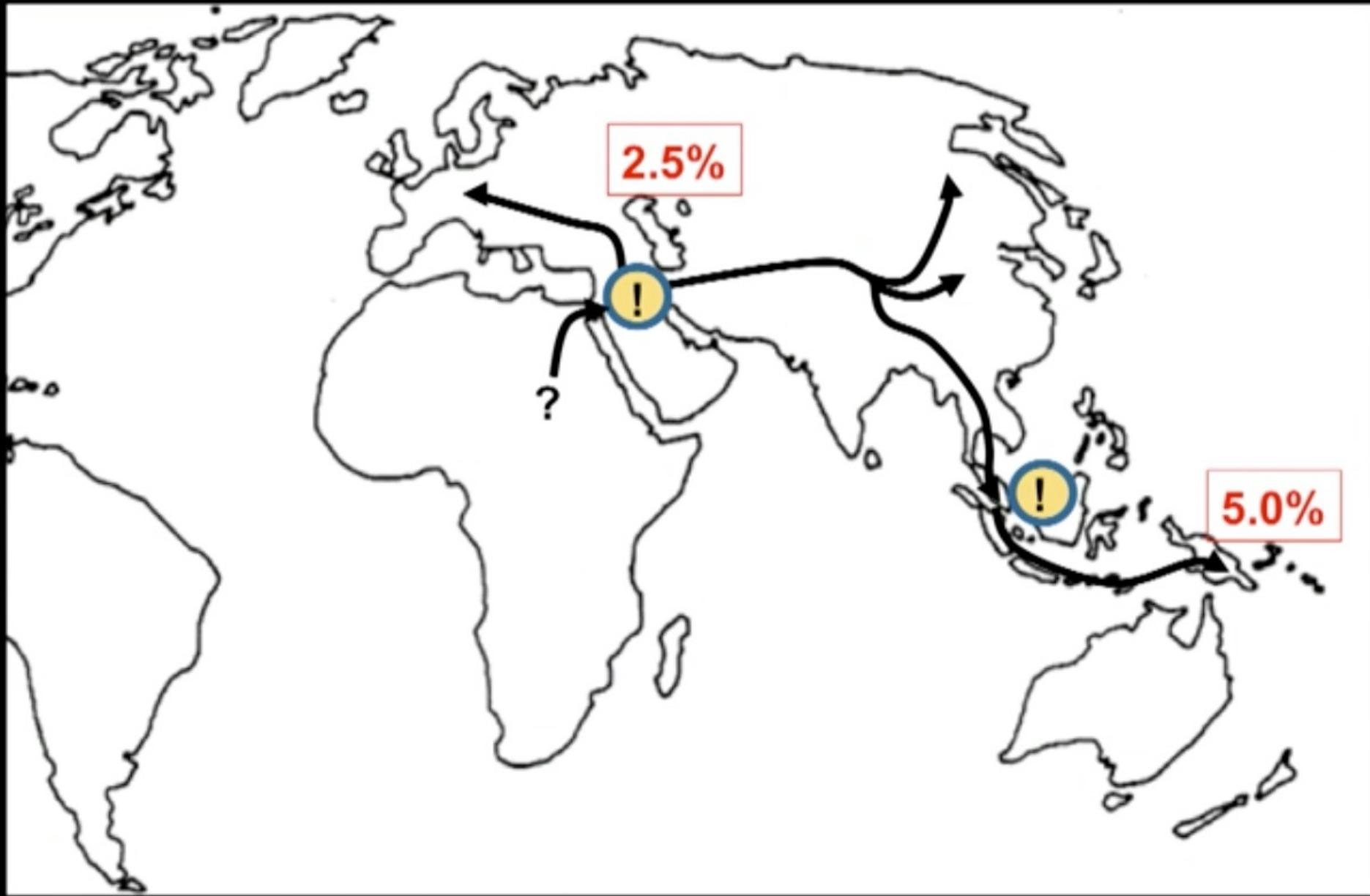
Timeline of ancient hominids



Timeline of ancient hominids



Timeline of ancient hominids



We have always mixed!

Cite as: B. Vernot *et al.*, *Science* 10.1126/science.aad9416 (2016).

Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals

Benjamin Vernot,¹ Serena Tucci,^{1,2} Janet Kelso,³ Joshua G. Schraiber,¹ Aaron B. Wolf,¹ Rachel M. Gittelman,¹ Michael Dannemann,³ Steffi Grote,³ Rajiv C. McCoy,¹ Heather Norton,⁴ Laura B. Scheinfeldt,⁵ David A. Merriwether,⁶ George Koki,⁷ Jonathan S. Friedlaender,⁸ Jon Wakefield,⁹ Svante Pääbo,^{2*} Joshua M. Akey^{1*}

¹Department of Genome Sciences, University of Washington, Seattle, Washington, USA. ²Department of Life Sciences and Biotechnology, University of Ferrara, Italy.

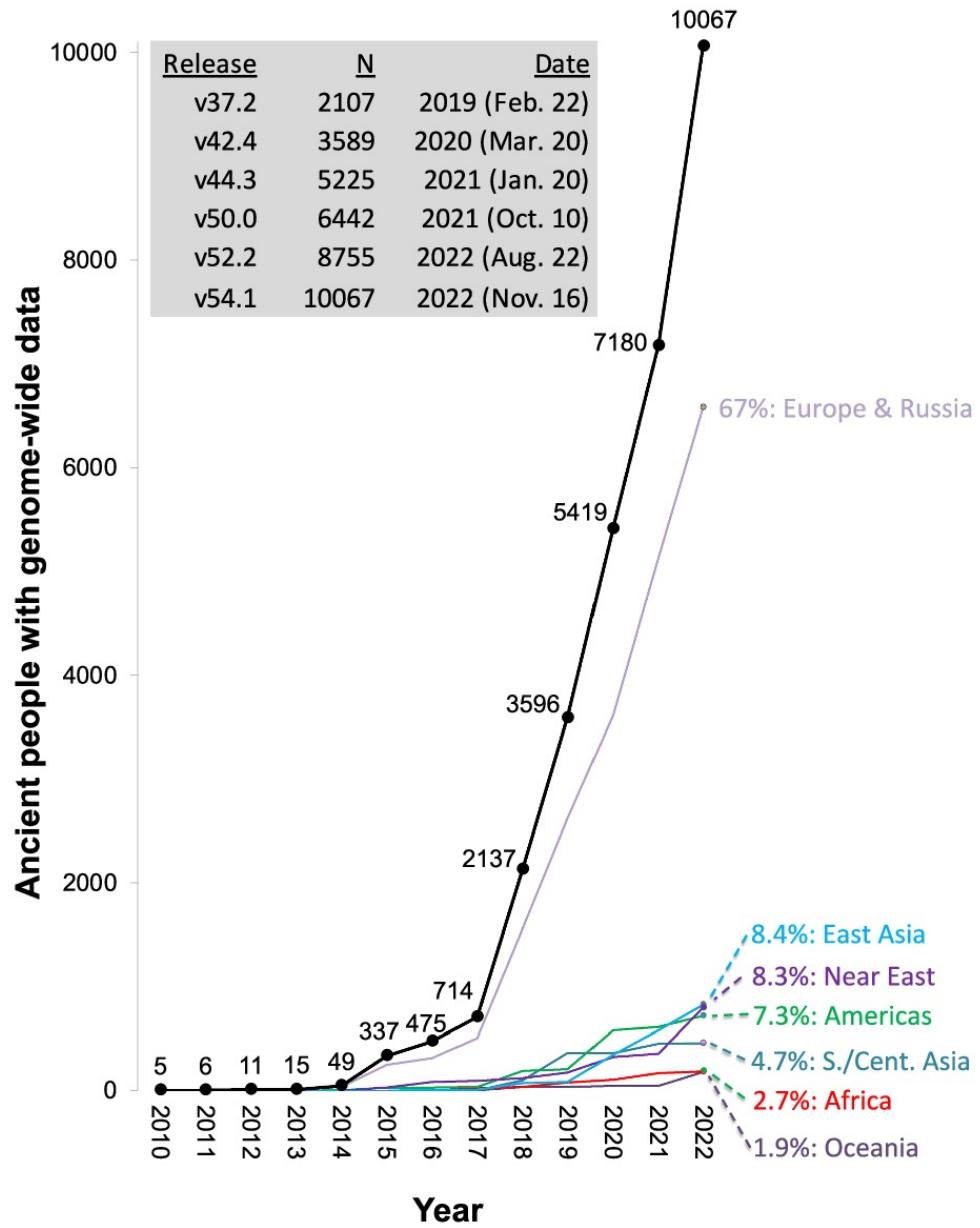
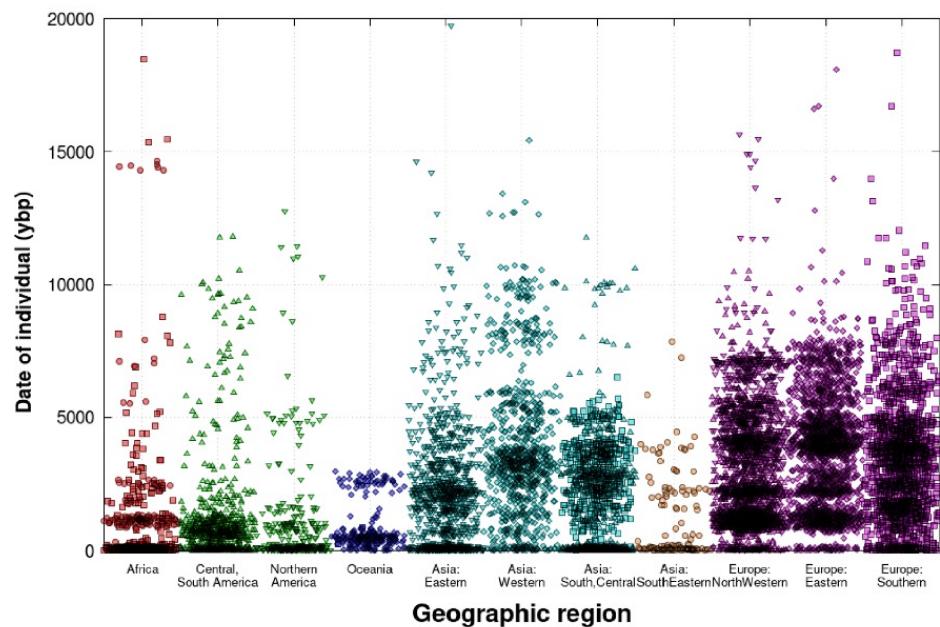
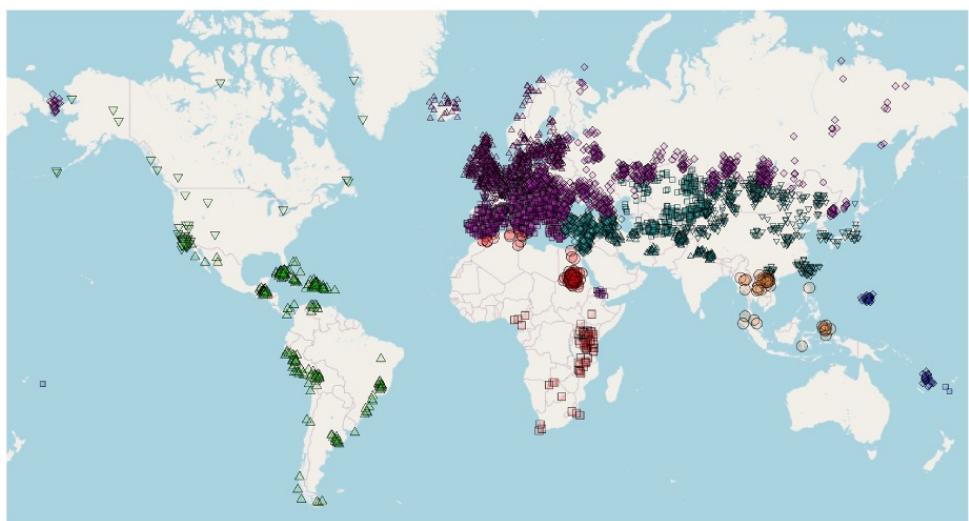
³Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany. ⁴Department of Anthropology, University of Cincinnati,

Cincinnati, OH, USA. ⁵Coriell Institute for Medical Research, Camden, NJ, USA. ⁶Department of Anthropology, Binghamton University, Binghamton, NY, USA. ⁷Institute for

Medical Research, Goroka, Eastern Highlands Province, Papua New Guinea. ⁸Department of Anthropology, Temple University, Philadelphia PA, USA. ⁹Department of Statistics, University of Washington, Seattle, Washington, USA.

*Corresponding author. E-mail: paabo@eva.mpg.de (S.P.); akeyj@uw.edu (J.M.A.)

Although Neandertal sequences that persist in the genomes of modern humans have been identified in Eurasians, comparable studies in people whose ancestors hybridized with both Neandertals and Denisovans are lacking. We developed an approach to identify DNA inherited from multiple archaic hominin ancestors and applied it to whole-genome sequences from 1523 geographically diverse individuals, including 35 new Island Melanesian genomes. In aggregate, we recovered 1.34 Gb and 303 Mb of the Neandertal and Denisovan genome, respectively. We leverage these maps of archaic sequence to show that Neandertal admixture occurred multiple times in different non-African populations, characterize genomic regions that are significantly depleted of archaic sequence, and identify signatures of adaptive introgression.

A**B****C**

The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes
 Mallick et al (2024) Scientific Data. <https://doi.org/10.1038/s41597-024-03031-7>

Recipe for a modern human

109,295 single nucleotide changes (SNCs)
7,944 insertions and deletions

Changes in protein coding genes

277 cause fixed amino acid substitutions
87 affect splice sites

Changes in Non-coding & regulatory sequences

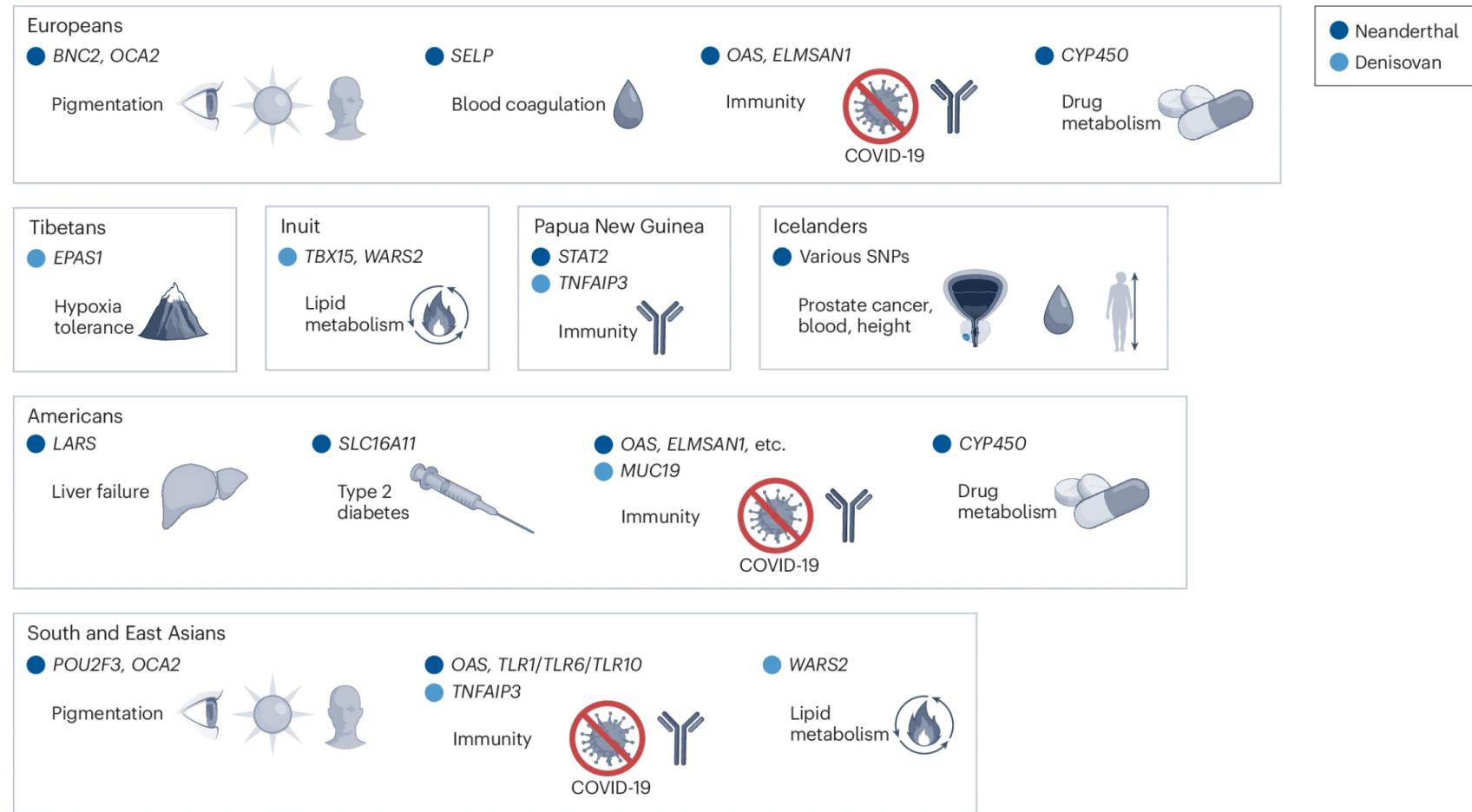
26 affect well-defined motifs inside
regulatory regions

Enrichment analysis

Nonsynonymous	None	- Giant melanosomes in melanocytes (p=6.77e-6; FWER=0.091;
Splice sites	skin pigmentation	
3' UTR	None	<ul style="list-style-type: none"> - 1-3 toe syndactyly (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - 1-5 toe syndactyly (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Aplasia/Hypoplasia of the distal phalanx of the thumb (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Bifid or hypoplastic epiglottis (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Central polydactyly (feet) (p=1.34288e-05; FWER=0.538; FDR=0.0887928)
skeletal morphologies (limb length, digit development)		
<ul style="list-style-type: none"> - Distal urethral duplication (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Dysplastic distal thumb phalanges with a central hole (p=1.34288e-05; FWER=0.538; FDR=0.0887928) 		
morphologies of the larynx and the epiglottis		
<ul style="list-style-type: none"> - Laryngeal cleft (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Midline facial capillary hemangioma (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Preductal coarctation of the aorta (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Radial head subluxation (p=1.34288e-05; FWER=0.538; FDR=0.0887928) - Short distal phalanx of the thumb (p=1.34288e-05; FWER=0.538; FDR=0.0887928) 		

Fig. 4: Archaic traits adaptively introgressed in modern populations.

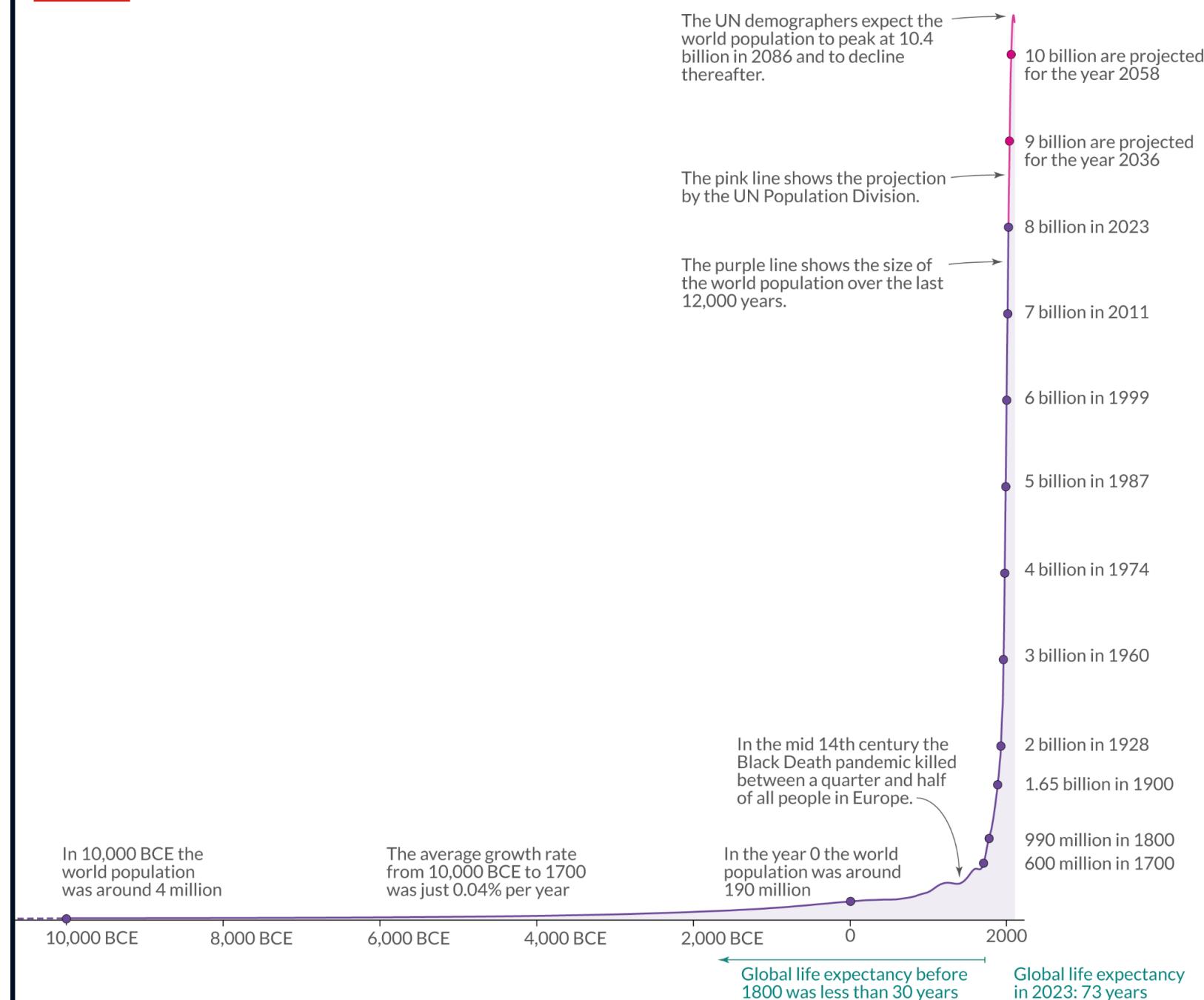
From: [A history of multiple Denisovan introgression events in modern humans](#)



A selection of the most important traits that have been identified as adaptively introgressed from archaic humans is subdivided by the main population group in which they have been described. This figure shows traits and genes adaptively introgressed from Denisovans and Neanderthals to provide a more comprehensive summary, while in the text, we focus only on the Denisovan ones. Specific citations can be found in Supplementary Table 4.

A history of multiple Denisovan introgression events in modern humans
Ongaro & Huerta-Sánchez (2024) Nature Genetics
<https://doi.org/10.1038/s41588-024-01960-y>

The size of the world population over the long-run



Based on estimates by the History Database of the Global Environment (HYDE) and the United Nations.

This is a visualization from OurWorldinData.org.

Licensed under CC-BY-SA by the author Max Roser.

Science

Current Issue First release papers Archive About Submit manu

HOME > SCIENCE > VOL. 336, NO. 6082 > RECENT EXPLOSIVE HUMAN POPULATION GROWTH HAS RESULTED IN AN EXCESS OF RARE GENETIC VARIANTS

[REPORT](#)

Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants

ALON KEINAN AND ANDREW G. CLARK [Authors Info & Affiliations](#)

SCIENCE • 11 May 2012 • Vol 336, Issue 6082 • pp. 740-743 • DOI: 10.1126/science.1217283

801

CHECK ACCESS



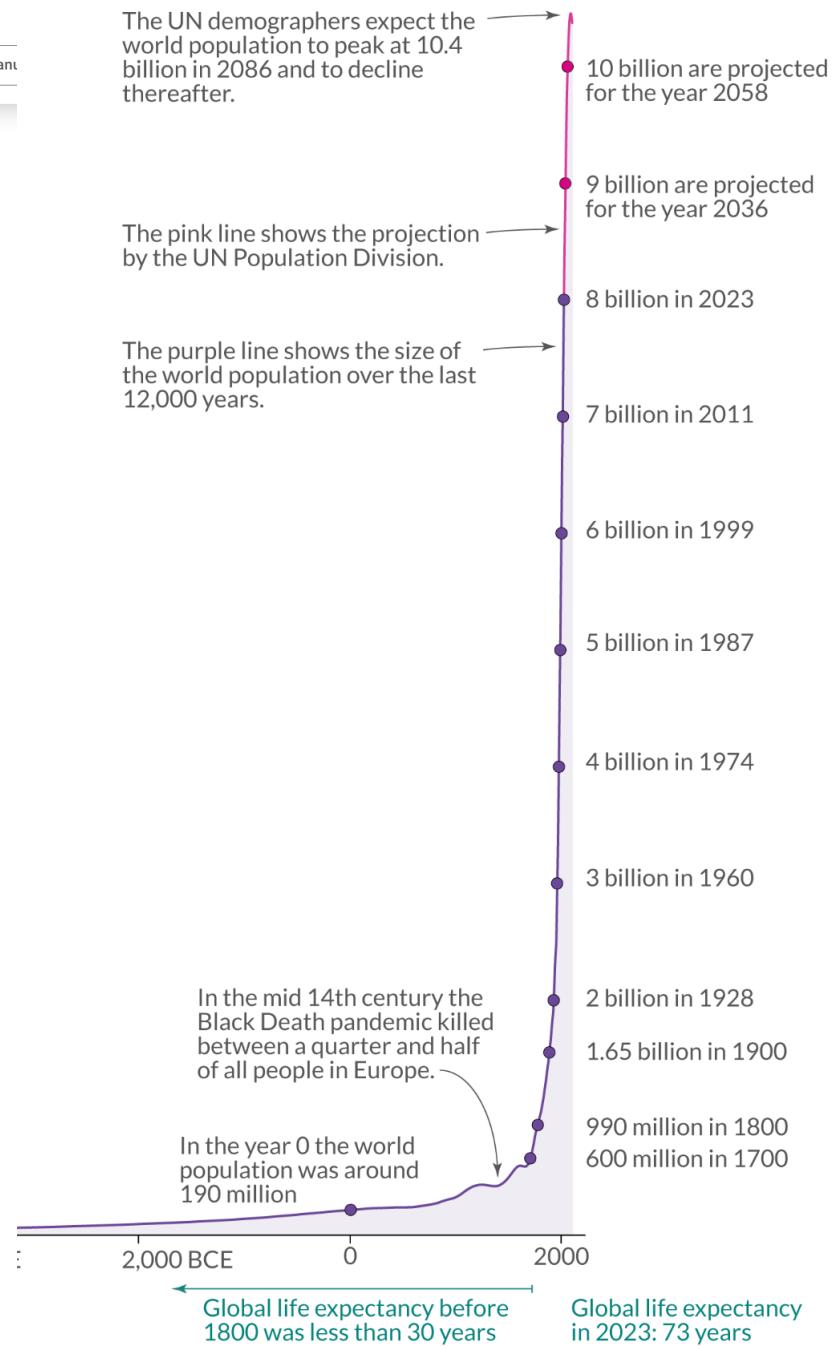
Exponential Growth Effects

Humans are an extraordinarily successful species, as measured by our large population size—approximately 7 billion—much of which can be put down to recent explosive growth. Leveraging human genomic data, Keinan and Clark (p. 740) examined the effects of population growth on our ability to detect rare genetic variants, those hypothesized to be most likely associated with disease. It appears that rapid recent growth increases the load of rare variants and is likely to play an important role in the individual genetic burden of complex disease risk.

Abstract

Human populations have experienced recent explosive growth, expanding by at least three orders of magnitude over the past 400 generations. This departure from equilibrium skews patterns of genetic variation and distorts basic principles of population genetics. We characterized the empirical signatures of explosive growth on the site frequency spectrum and found that the discrepancy in rare variant abundance across demographic modeling studies is mostly due to differences in sample size. Rapid recent growth increases the load of rare variants and is likely to play a role in the individual genetic burden of complex disease risk. Hence, the extreme recent human population growth needs to be taken into consideration in studying the genetics of complex diseases and traits.

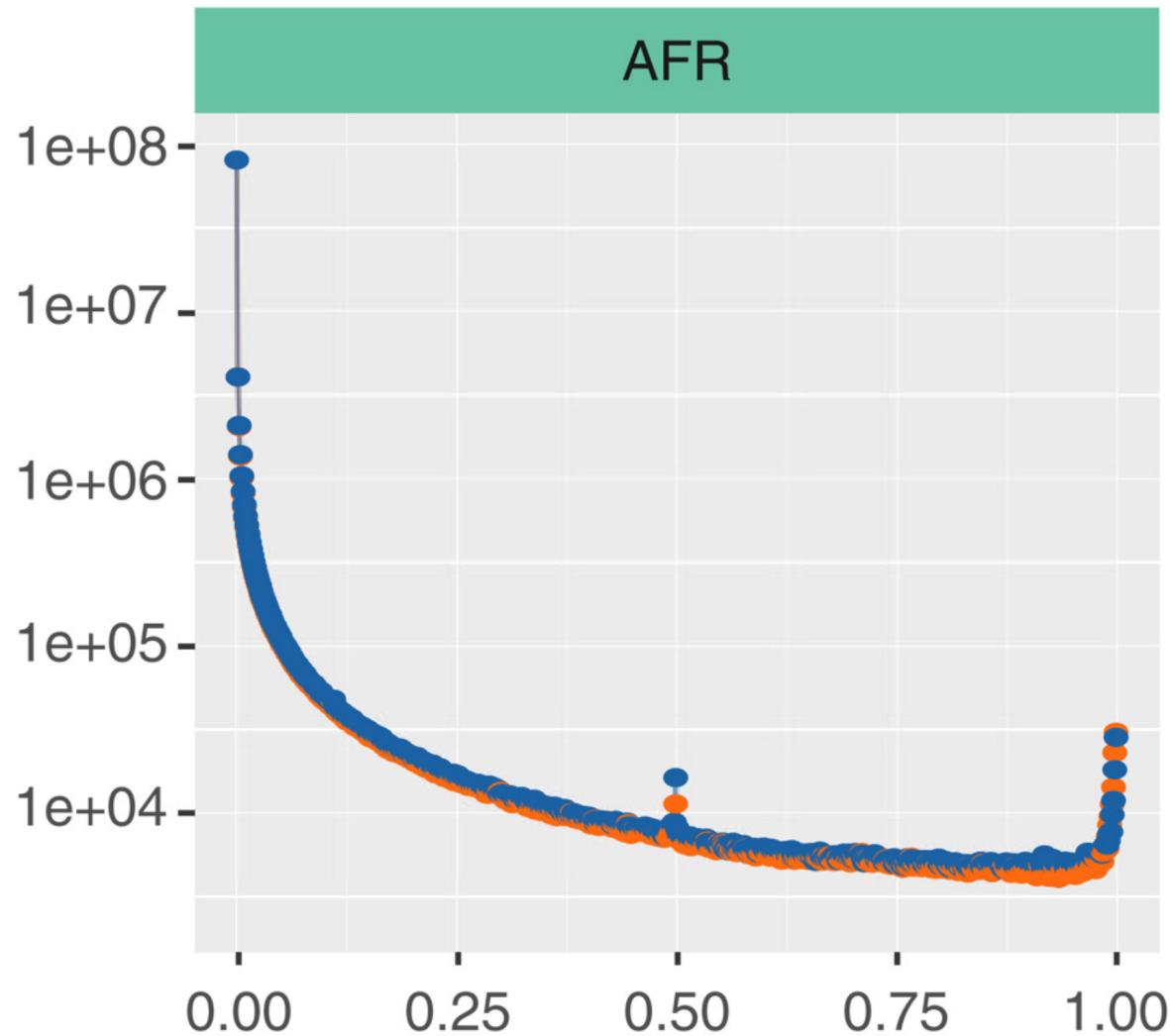
This is a visualization from OurWorldinData.org.



IYDE) and the United Nations.

Licensed under CC-BY-SA by the author Max Roser.

Allele frequency



Site frequency spectrum of variants in ~500 AFR genomes from the 1000 Genomes collection

Note the huge fraction of rare alleles (occurring in less than 5% of genomes)