

# Superarchaic Admixture

Alan R. Rogers

April 12, 2022

# Early to middle Pleistocene of Eurasia



~1.8 mya: *Homo erectus* evolves in Africa, spreads into Eurasia

~550 kya: Late Acheulean appears in Europe.

~430 kya: large-brained hominins at Sima de los Huesos

Similar fossils and tools occur earlier in Africa.

⇒ African invasion of Europe early in Middle Pleistocene.

What can genetics tell us about this period?

## Early to middle Pleistocene of Eurasia



~1.8 mya: *Homo erectus* evolves in Africa, spreads into Eurasia

~550 kya: Late Acheulean appears in Europe.

~430 kya: large-brained hominins at Sima de los Huesos

Similar fossils and tools occur earlier in Africa.

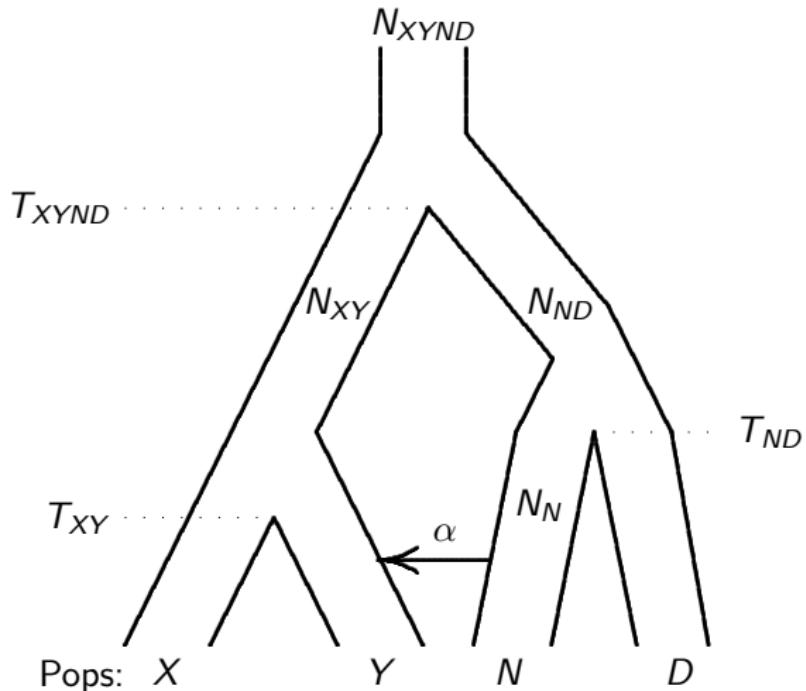
⇒ African invasion of Europe early in Middle Pleistocene.

What can genetics tell us about this period?

## *Legofit*: estimates deep population history in subdivided populations

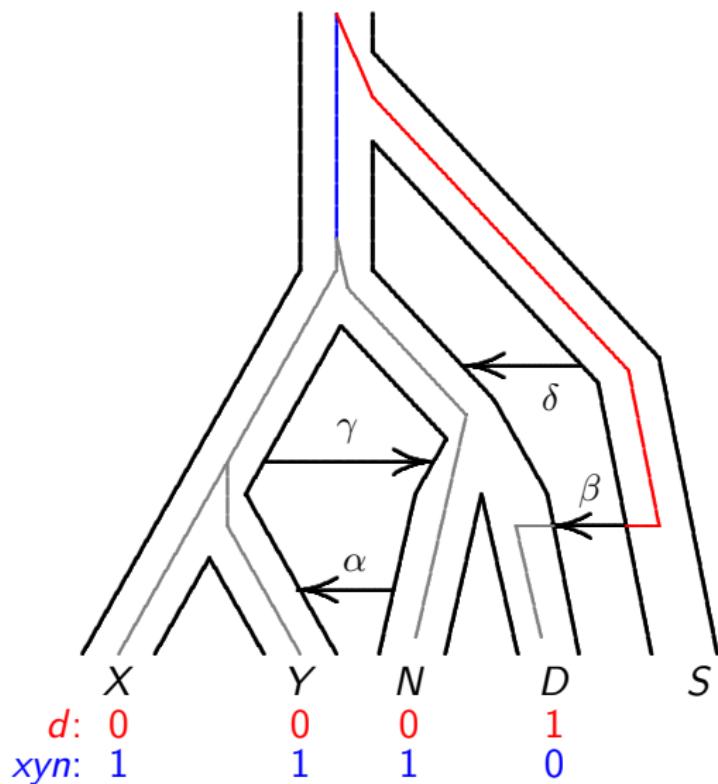
- ▶ Unaffected by recent inbreeding or changes in population size.
- ▶ Sensitive only to the distant past.
- ▶ Estimates gene flow and the sizes and separation times of ancestral populations.
- ▶ New version is orders of magnitude faster.

## Population network (now outdated)



$X$ , Africa;  $Y$ , Europe;  $N$ , Neanderthal;  $D$ , Denisovan

# Gene genealogies and nucleotide site patterns



Gene genealogy within population network.

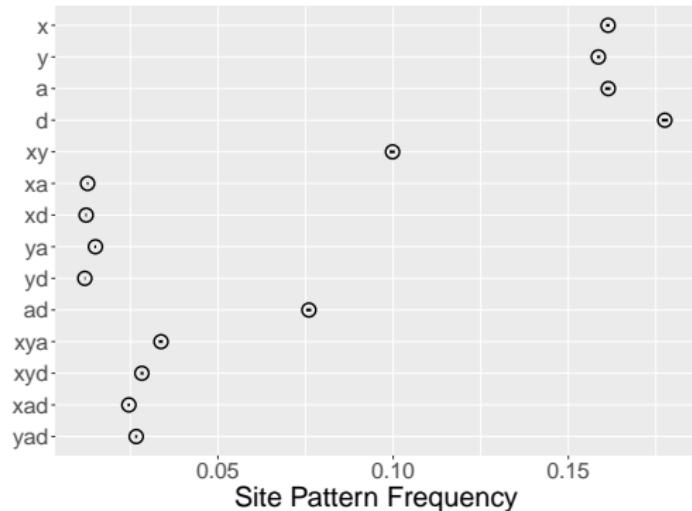
Mutation on red branch → *site pattern d*.

Blue branch → *xyn*.

0, ancestral allele; 1, derived (mutant) allele.

Data: frequencies of site patterns across autosomes

# Observed Site Pattern Frequencies (excl. Vindija)



(fraction of nucleotide sites exhibiting each pattern)

x, Africa; y, Europe; a, Altai Neanderthal; d, Denisovan.

Pattern  $xy$  is common because populations  $X$  and  $Y$  share ancestry.

Ditto  $ad$ .

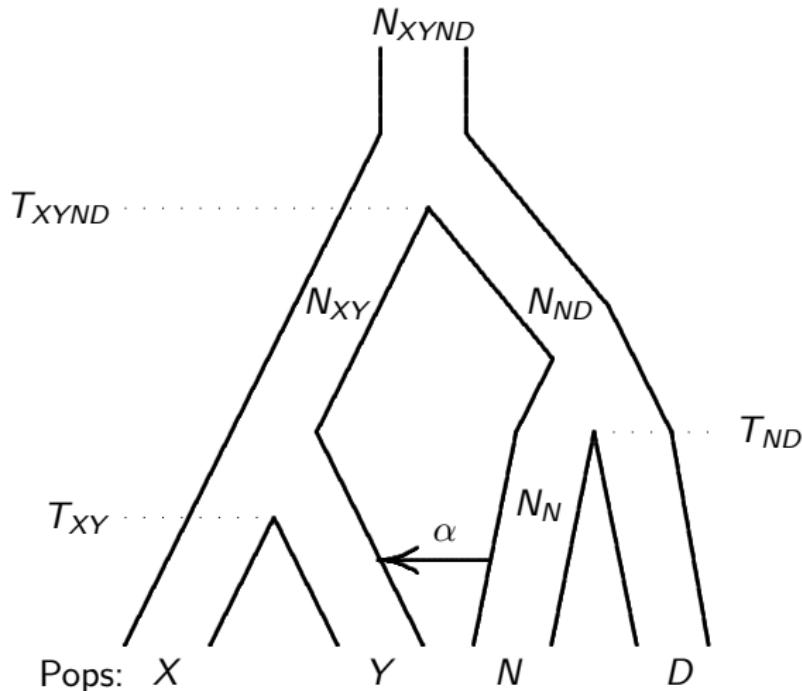
Confidence intervals are so small they look like dots.

Goal: infer history from similar data, but including Vindija.

# Estimation

1. Maximize composite likelihood, a function of sizes and separation times of populations, and rates of gene flow.
2. Old Legofit used simulations to estimate likelihood. New algorithm is deterministic.
3. Uncertainties by moving-blocks bootstrap.

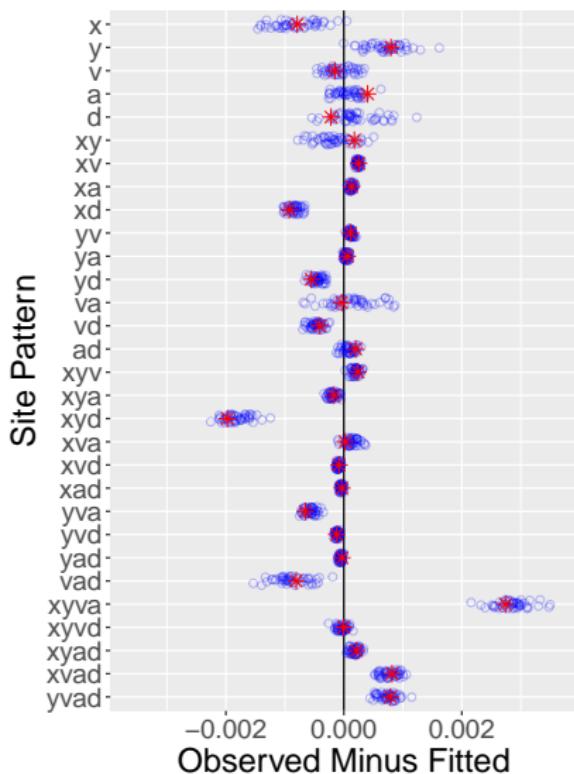
In 2017, we fit model  $\alpha$  to the data



$X$ , Africa;  $Y$ , Europe;  $N$ , Neanderthal;  $D$ , Denisovan

# Residual error from model $\alpha$

Model  $\alpha$



Red asterisks: fitted model. Blue circles: bootstrap replicates.

If model fit well, all points would be near 0.

Discrepancies show that something is missing from the model. What?

## Ideas from the literature

- $\beta$  Gene flow from a “superarchaic” population into Denisovans  
(Prüfer et al 2014)
- $\gamma$  Gene flow from early modern humans into Neanderthals  
(Kuhlwilm et al 2016)

These improved the fit but were still unsatisfactory.

What else is missing?

## Think back to what I said about the Middle Pleistocene



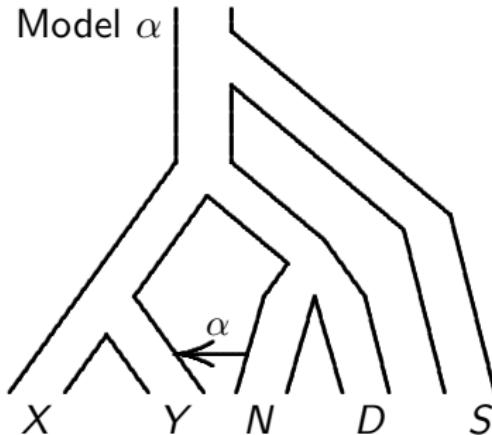
~600 kya Eurasia invaded by large-brained hominins, who probably came from Africa.

Ancestors of Neanderthals and Denisovans: let's call them "neandersovans."

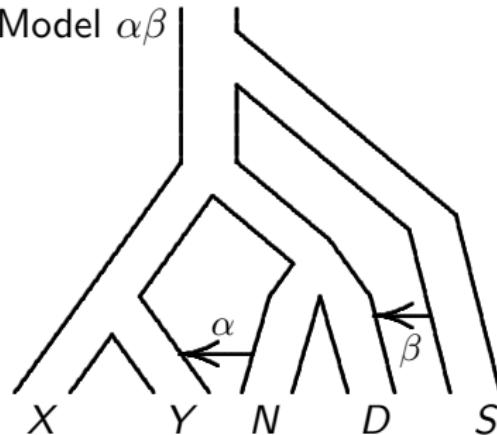
But Eurasia had been inhabited since ~2 my ago by "superarchaics."

Neandersovans would have met, and maybe interbred with, superarchaics. Suggests a fourth episode of admixture.

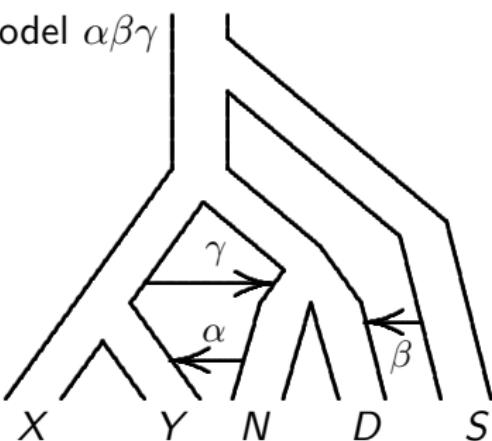
Model  $\alpha$



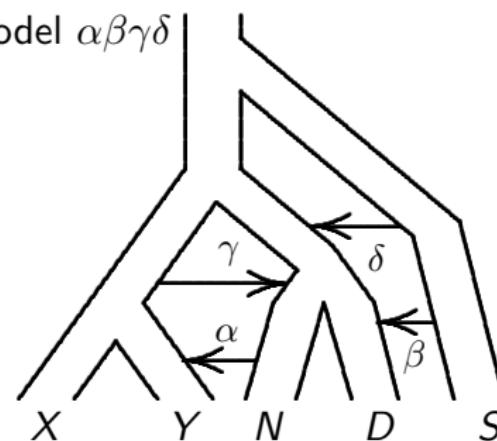
Model  $\alpha\beta$



Model  $\alpha\beta\gamma$

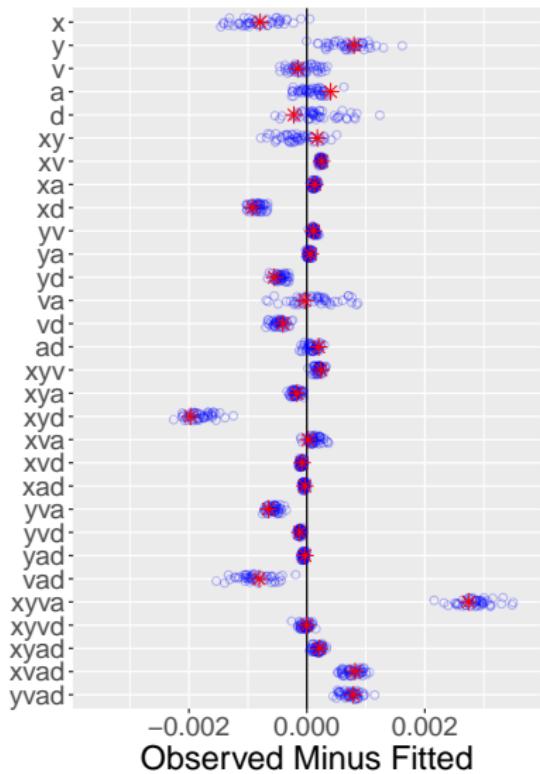


Model  $\alpha\beta\gamma\delta$



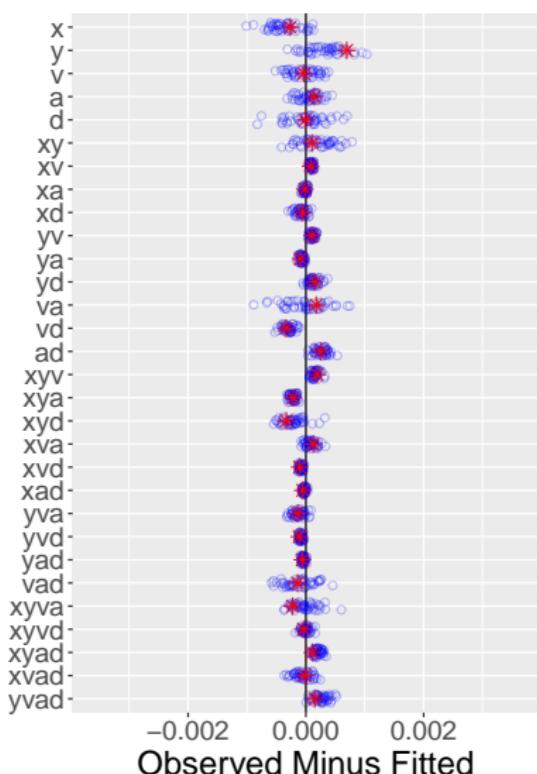
# Model $\alpha$

Site Pattern



# Model $\alpha\beta\gamma\delta$

Site Pattern



Key: \*, real data; o, bootstrap replicates.

## Model selection and model averaging

Model selection by **bepe**, the bootstrap estimate of predictive error (Efron & Tibshirani 1993). Prefer model with smallest bepe value. Avoids overfitting.

Model averaging by **booma**, bootstrap model averaging (Buckland, Burnham, and Augustin, 1997). Weight of  $i$ th model is fraction of bootstrap replicates in which that model is best. Parameter estimates are weighted averages of per-model estimates. Addresses identifiability problems.

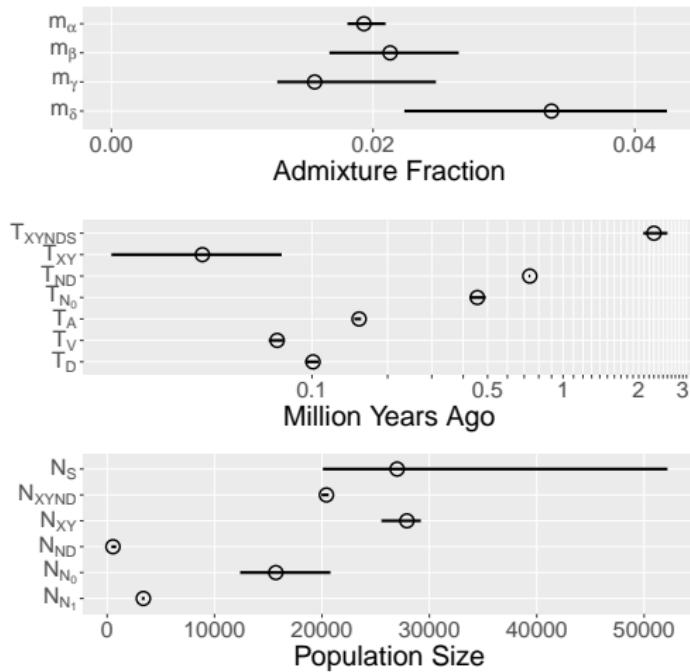
## Evaluating the models

Model	bepe	weight
$\alpha$	$1.16 \times 10^{-6}$	0
$\alpha\delta$	$0.87 \times 10^{-6}$	0
$\alpha\gamma$	$0.62 \times 10^{-6}$	0
$\alpha\gamma\delta$	$0.44 \times 10^{-6}$	0
$\alpha\beta$	$0.18 \times 10^{-6}$	0
$\alpha\beta\gamma$	$0.17 \times 10^{-6}$	0
$\alpha\beta\delta$	$0.15 \times 10^{-6}$	0.16
$\alpha\beta\gamma\delta$	$0.13 \times 10^{-6}$	0.84

Reject models with weight zero: their disadvantage is large compared with variation in repeated sampling.

Strong support for two episodes of superarchaic admixture ( $\beta$  and  $\delta$ ); qualified support for admixture ( $\gamma$ ) from early moderns into Neanderthals.

## Parameter estimates



Superarchaic population separated  $\sim 2$  mya. It was large—between 20,000 and 50,000—or deeply subdivided.

neandersovan population ( $N_{ND}$ ) was tiny, and split early ( $T_{ND} > 700$  kya) to form Neanderthals and Denisovans.

$\sim 3\%$  admixture into neandersovans from superarchaics.

## Summary

Superarchaics separated from other hominins ~2 mya. They may represent the earliest Eurasians. Their population was either large or deeply subdivided.

~750 kya, neandersovans separated from an African population, expanded into Eurasia, endured a bottleneck, interbred with superarchaics, and then (~730 kya) split into eastern and western subpopulations (Denisovans & Neanderthals).

## Acknowledgements

**Collaborators:** Nathan Harris, Alan Achenbach, Kiela Gwin, Mitchell Lokey, Daniel Tabin.

**Support:** NSF BCS 1638840; NSF BCS 1945782; Center for High Performance Computing, U. of Utah.