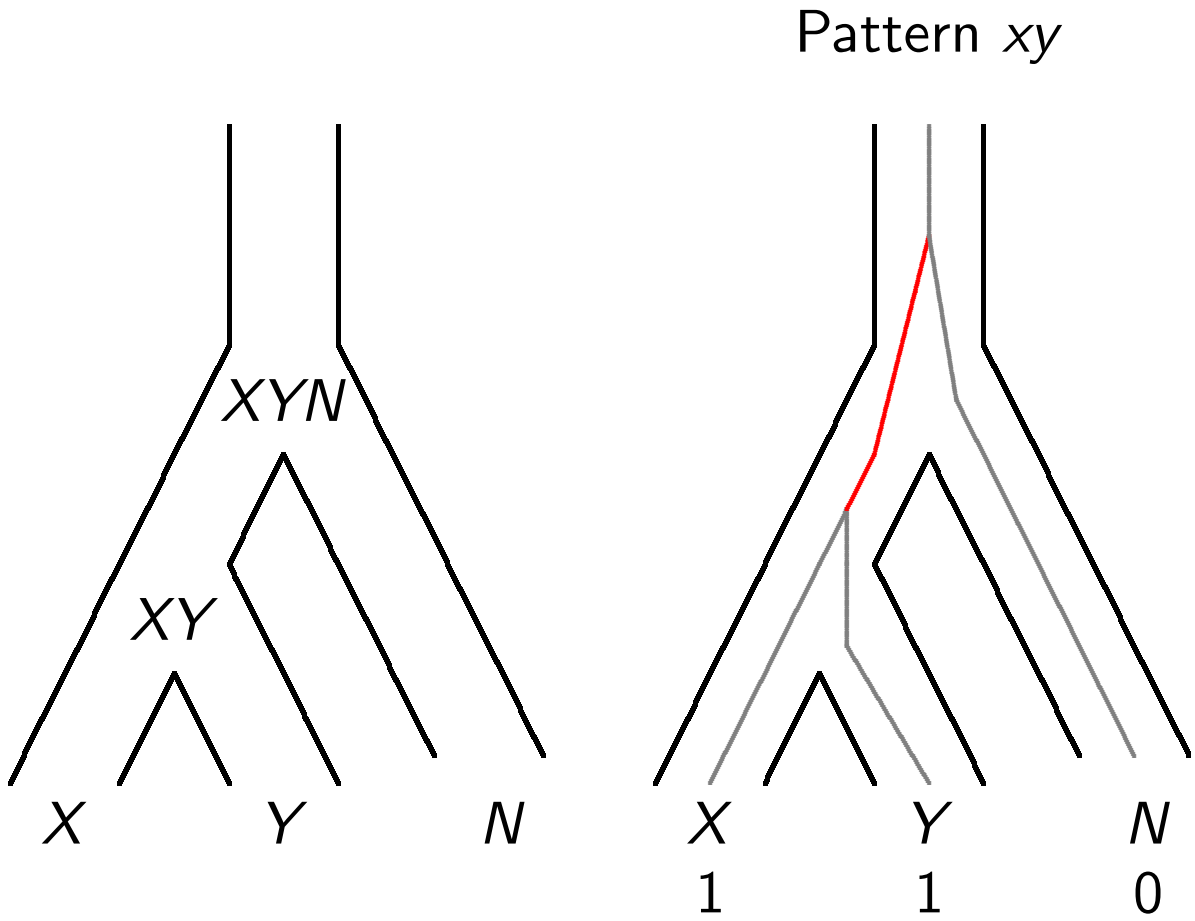
Using Genetic Data to Build Intuition about Population History

Alan R. Rogers

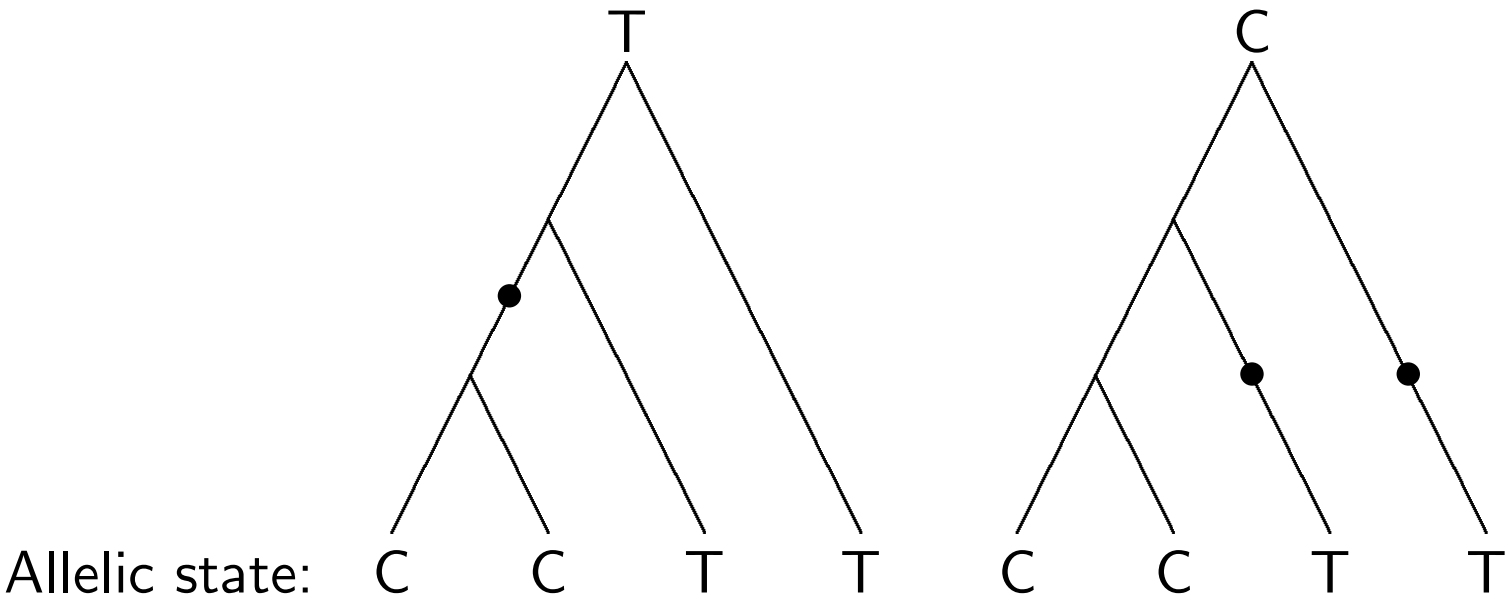
April 4, 2024

Notation



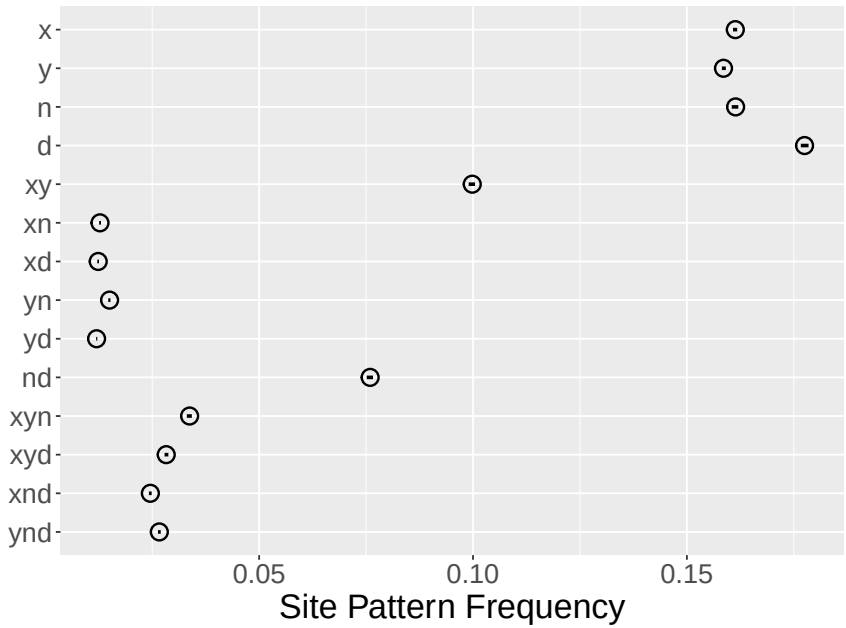
$xy \succ yn$ will mean that xy is more common than yn. $xy \sim yn$ will mean that the two are about equal in frequency.

Calling ancestral alleles



When one allele is present in the outgroup, but two are present in the ingroup, the outgroup allele is likely to be ancestral.

Observed site pattern frequencies



X is Africa; Y, Europe; N, Neanderthal, D, Denisovan.

"Dots" in circles are 95% confidence intervals.

$x \succ y$

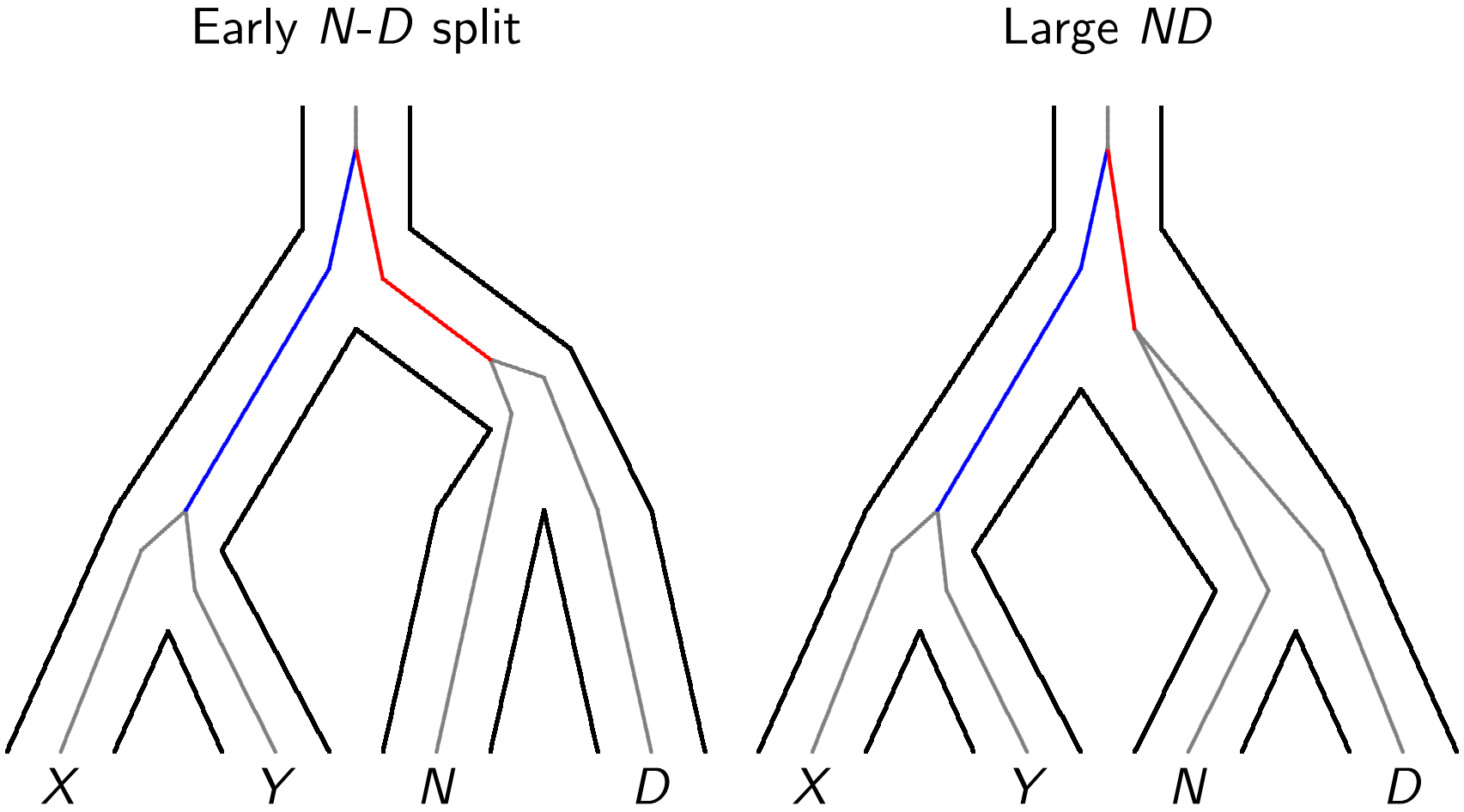
$d \succ$ all other singletons.

xy, $nd \succ$ other doubletons.

$xy \succ nd$

$yn \succ$ other rare doubletons; $xyn \succ$ other tripletons

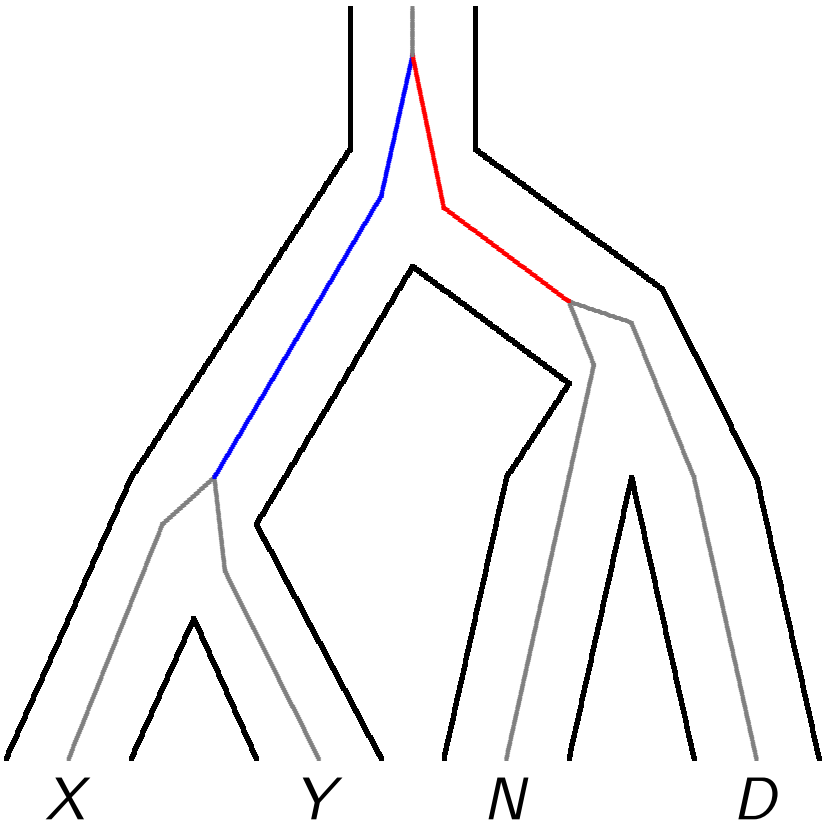
Why are xy, $nd \succ$ other doubletons?



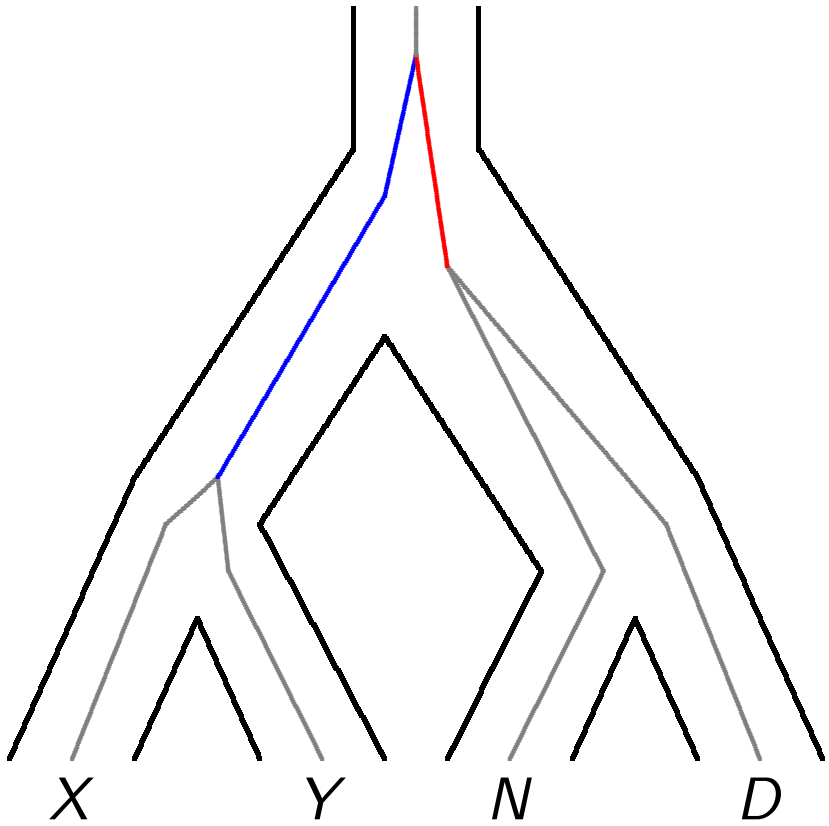
(X, Y) and (N, D) are pairs of closely related populations. Close relatives share ancestors, and mutations in these ancestors generate xy and nd.

Why is $xy \succ nd$?

Early N-D split

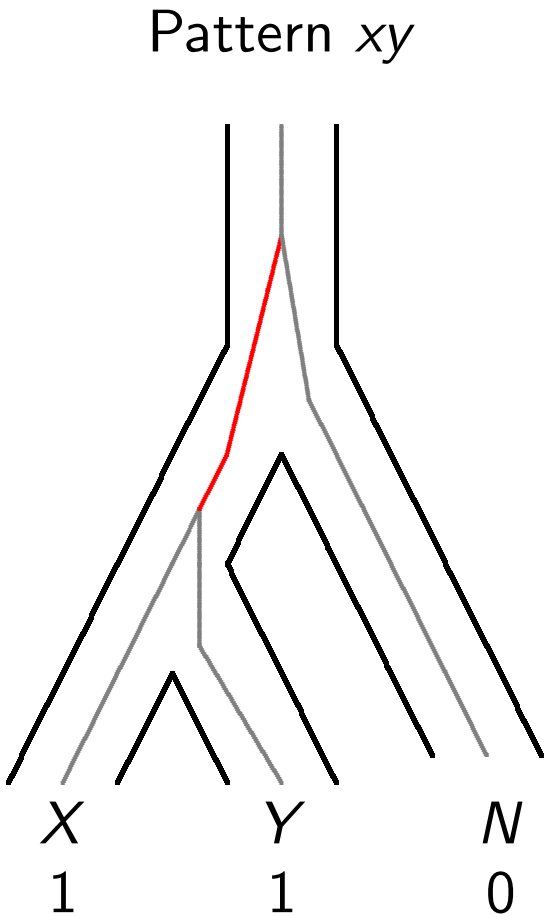


Large ND



Either N and D split earlier than X and Y, or ND was larger than XY.

Where do counterintuitive site patterns come from?

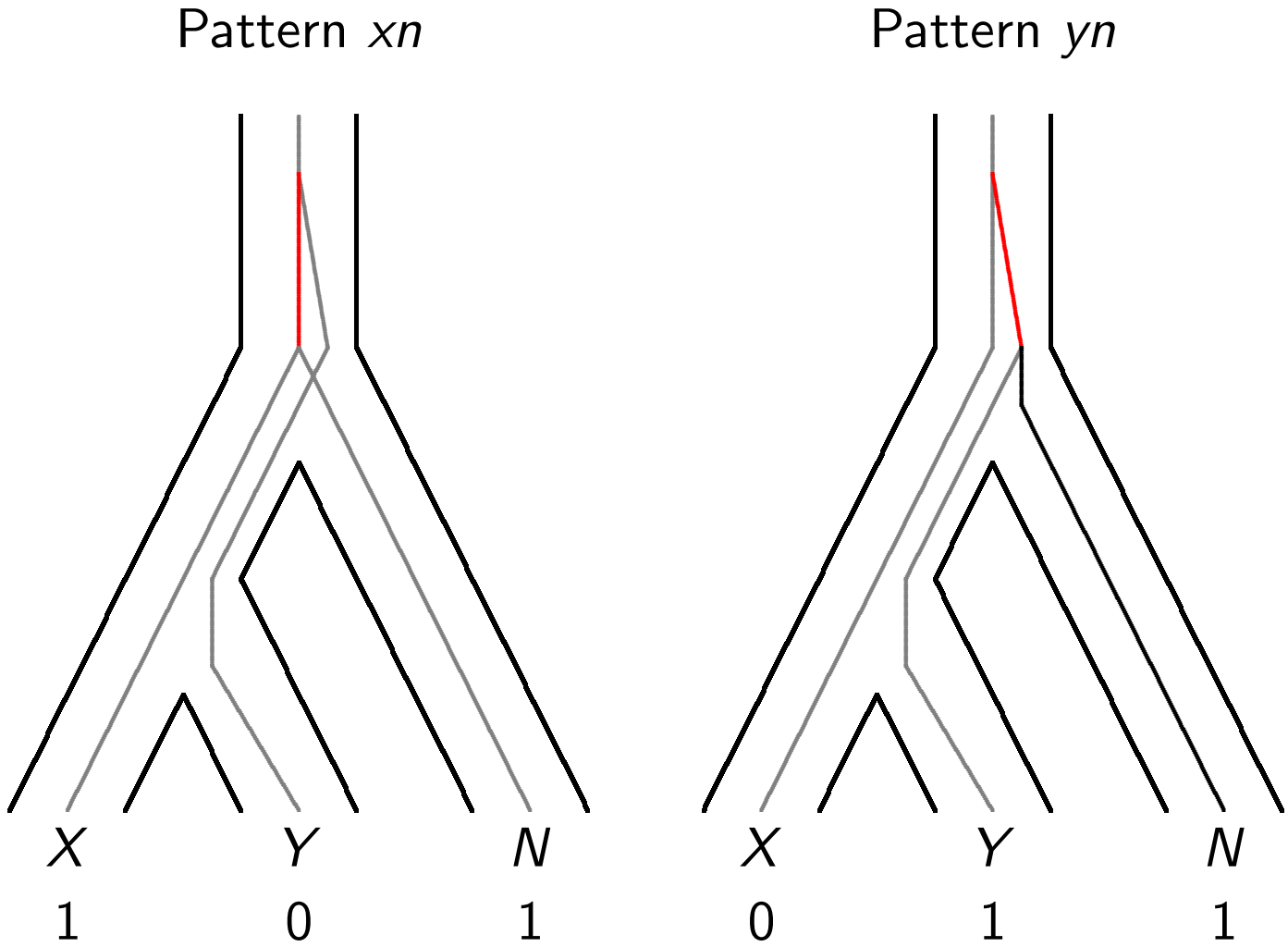


It's easy to understand where *xy* comes from.

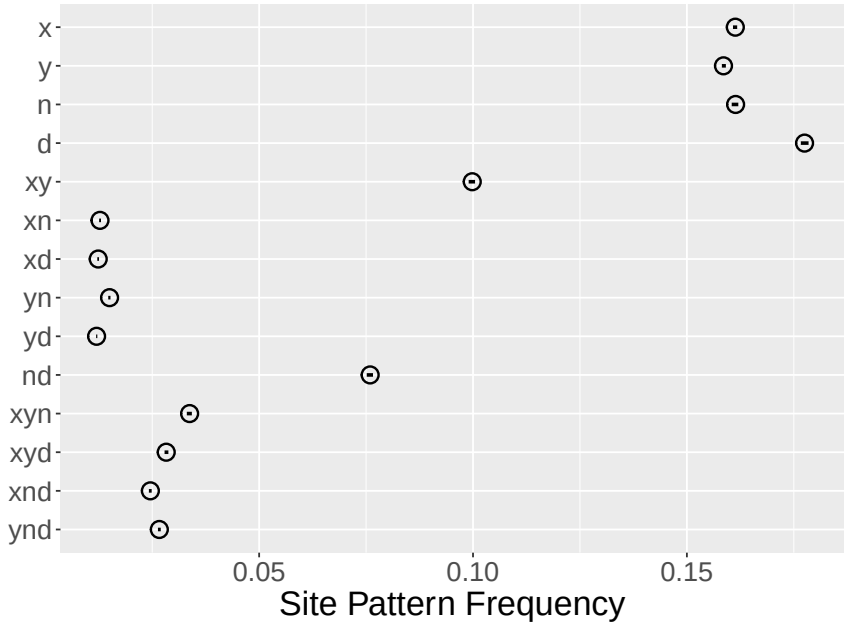
But why do *xn* and *yn* exist?

One answer is *incomplete lineage sorting*. . .

Incomplete lineage sorting



Two puzzles



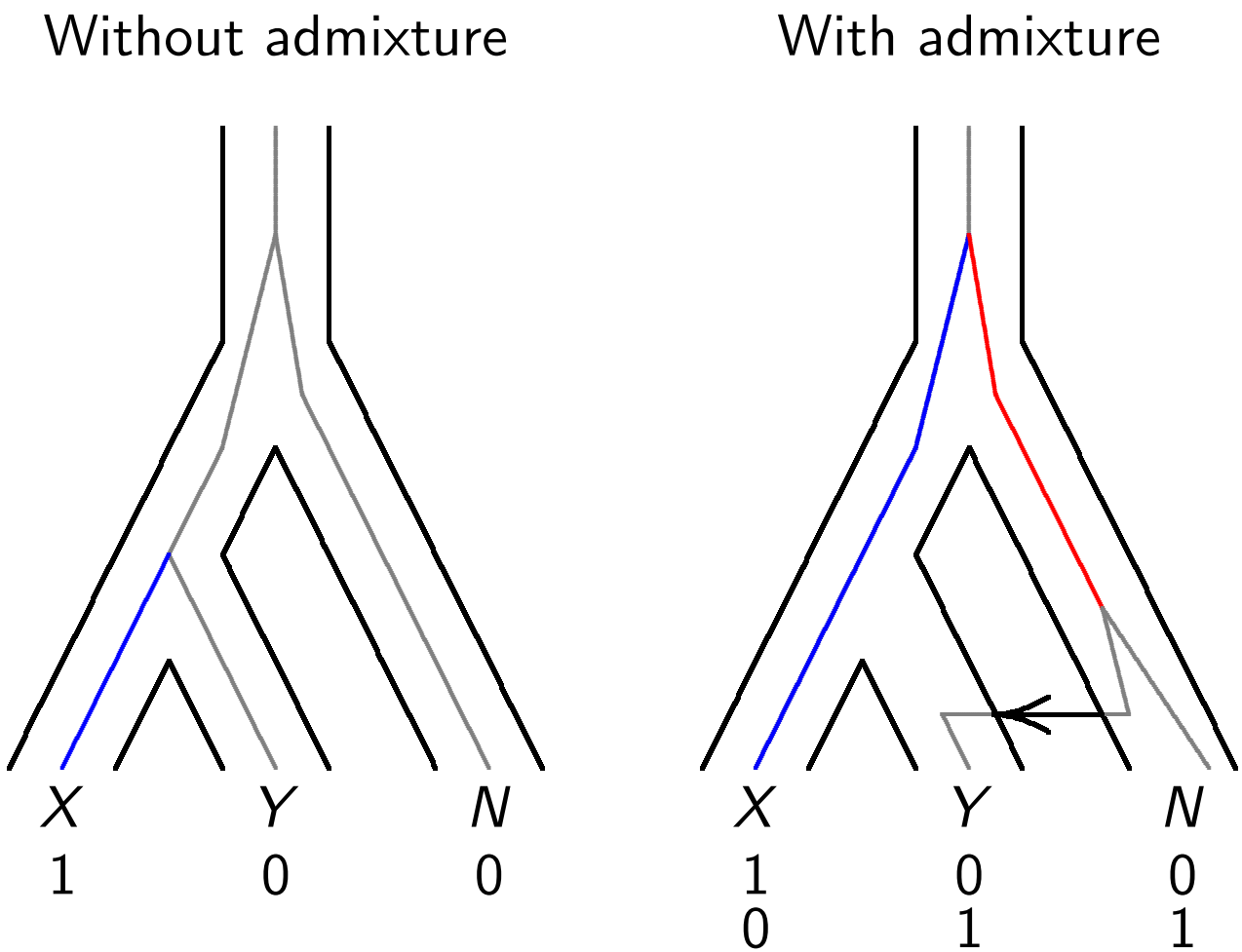
Incomplete lineage sorting predicts that $yn \sim xn$, and $yd \sim xd$.

It's true that $yd \sim xd$.

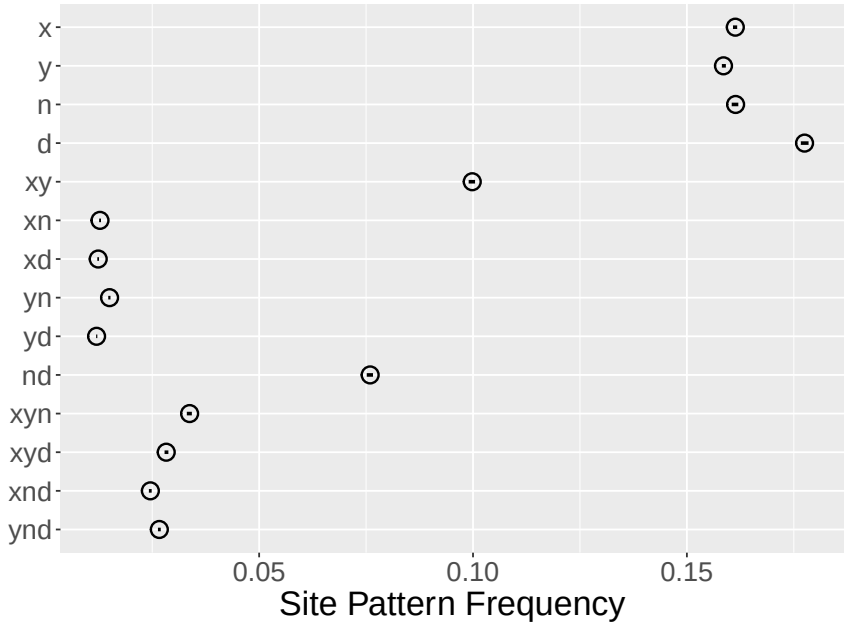
However, $yn \succ xn$: why?

Also, $x \succ y$: why?

$N \rightarrow Y$ admixture inflates yn and x



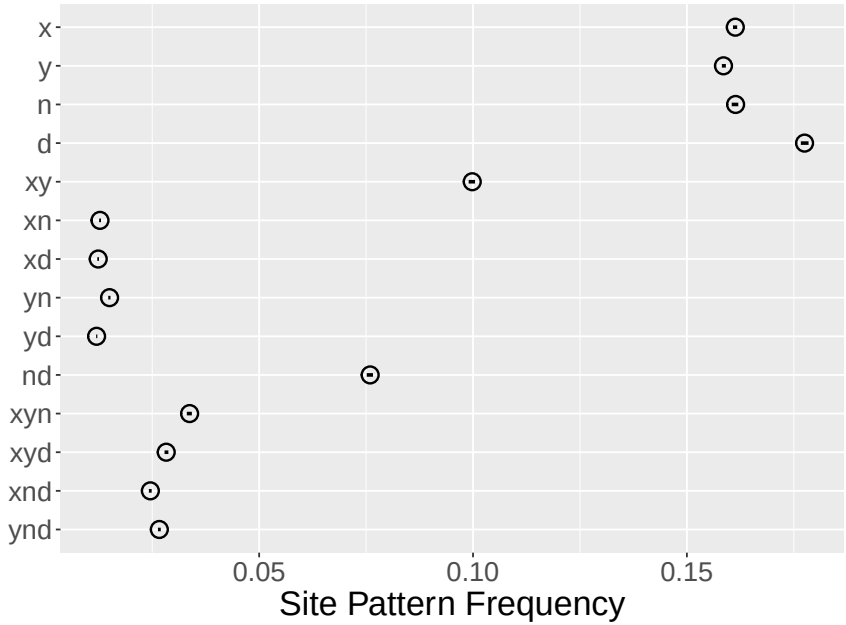
Evidence for Neanderthal admixture into Eurasians



$yn \succ xn$ and $x \succ y$

This suggest admixture from Neanderthals into Eurasians.

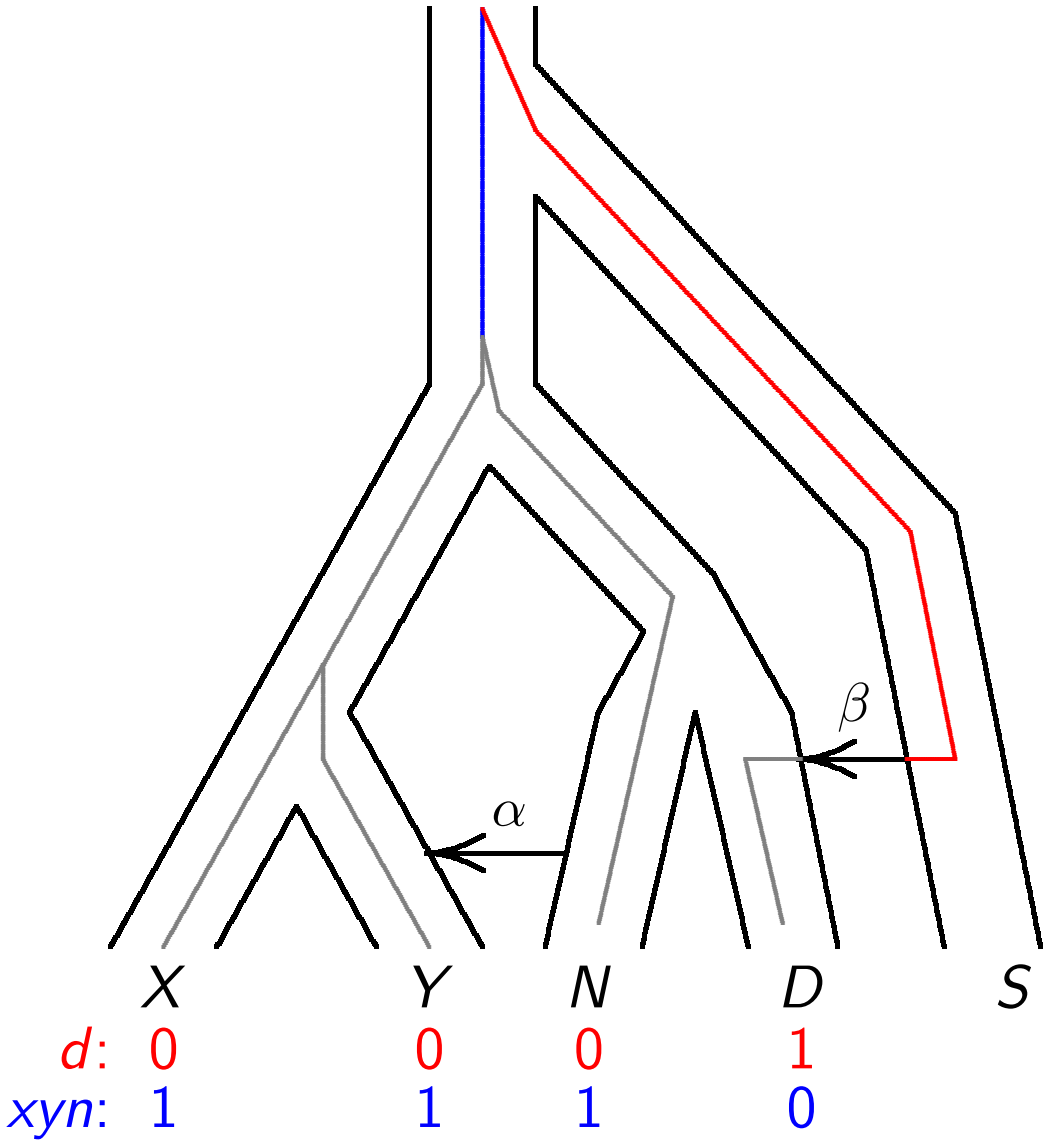
Another puzzle



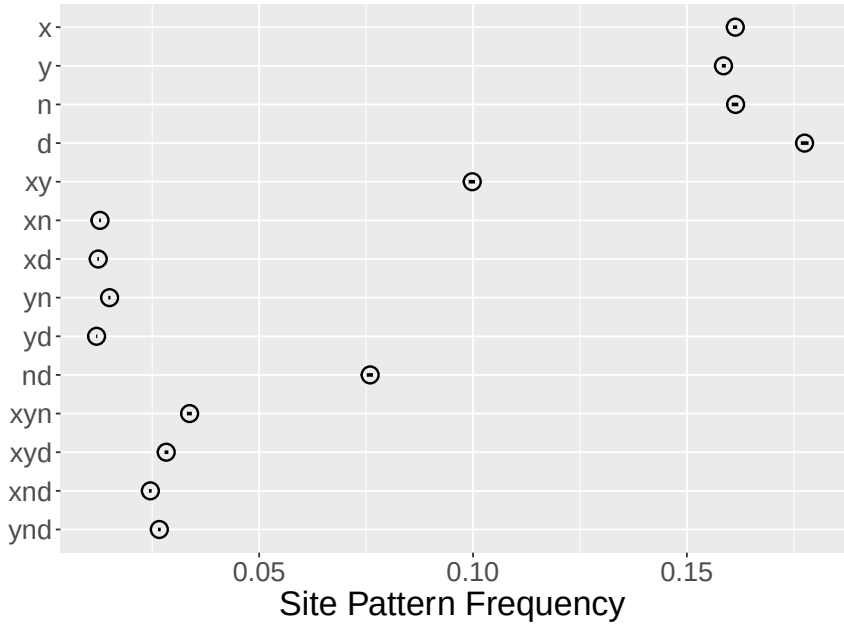
Note that $d \succ$ other singletons and that $xyn \succ$ other tripletons.

Why should this be?

$S \to D$ gene flow inflates d and xyn



Observed site pattern frequencies



High frequencies of d and xyn suggest $S \rightarrow D$ admixture.

Conclusions

We have used no formal model; we have tested no hypotheses. All we've done is to look at the data. Yet this informal analysis has been productive. It suggests

1. modern Europeans and Africans are closely related,
2. so were Neanderthals and Denisovans,
3. the separation between Europeans and Africans was more recent than that between Neanderthals and Denisovans,
4. Neanderthals contributed DNA to Europeans, and
5. superarchaics contributed DNA to Denisovans.