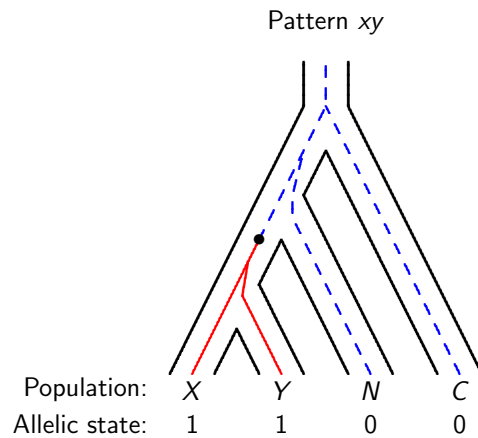


Genetic Evidence on Population Size, Subdivision, and Gene Flow

Alan R. Rogers

September 9, 2021

Notation for populations

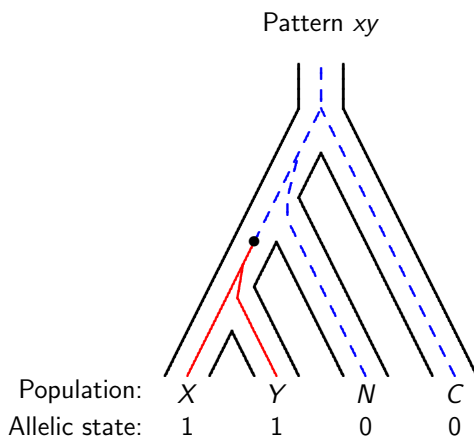


X, Y, N, and C are populations: African, European, Neanderthal, and Chimpanzee

XY: population ancestral to X and Y.

XYN: ancestral to X, Y, and N.

Nucleotide site patterns

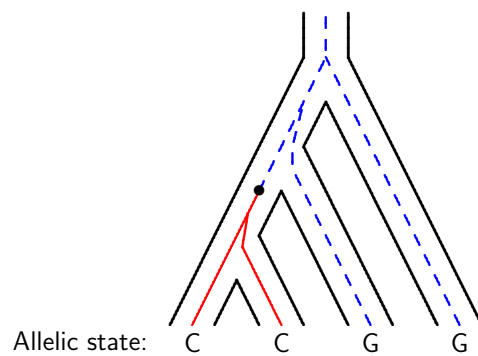


Haploid sample: 1 nucleotide from each population.

Mutation at dot would appear in samples from X, Y, not those from N, C.

Call this the *xy* site pattern.

Calling ancestral and derived alleles



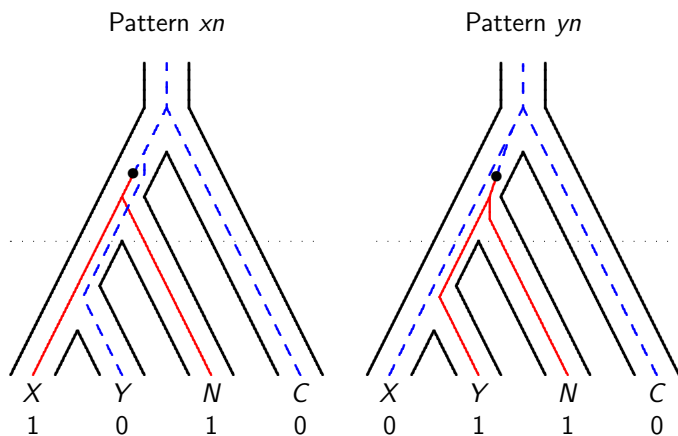
Data: C, C, G, G. Which allele is ancestral?

If C is ancestral, 2 mutations needed.

If G is ancestral, only 1 needed.

Prefer hypothesis requiring fewer mutations, because mutations are rare.

Incomplete lineage sorting



Incomplete lineage sorting again

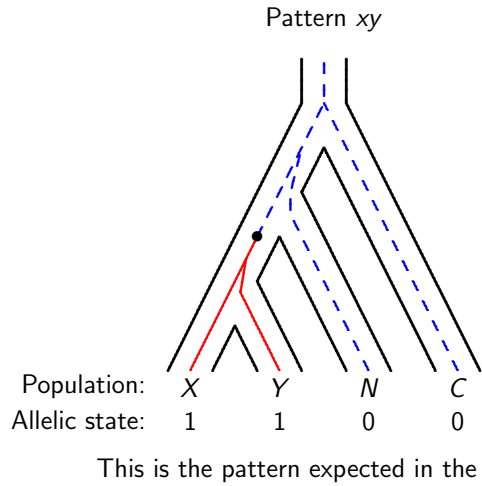
Suppose that, as we trace the ancestry of our sample backwards in time, the lineages from X and Y don't coalesce until we reach XYN.

Then there are three lineages, X, Y, and N, in the same population.

They can coalesce in any order.

Site patterns *xy*, *xn*, and *yn* are equally likely.

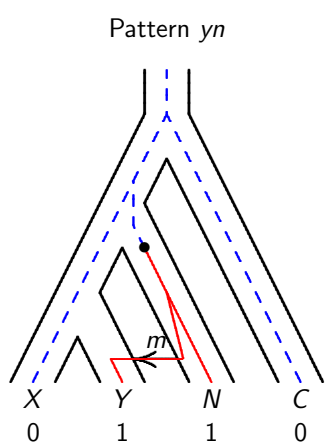
Pattern *xy* can also arise another way



The lineages from X and Y may also coalesce w/i XY, generating site pattern *xy*.
So *xy* should be more common than *xn* or *yn*.
xn and *yn* should be equally common.

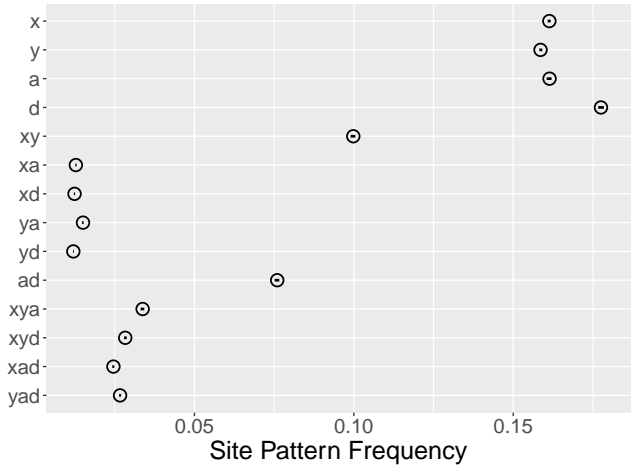
This is the pattern expected in the absence of gene flow.

The effect of gene flow

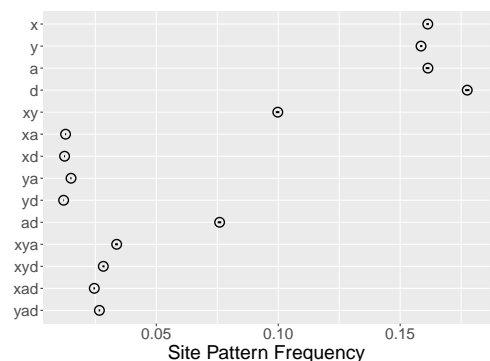


$N \rightarrow Y$ gene flow inflates the frequency of *yn*.
Also inflates frequency of *x*.
Effects are small unless the rate of gene flow is high.

Observed site pattern frequencies

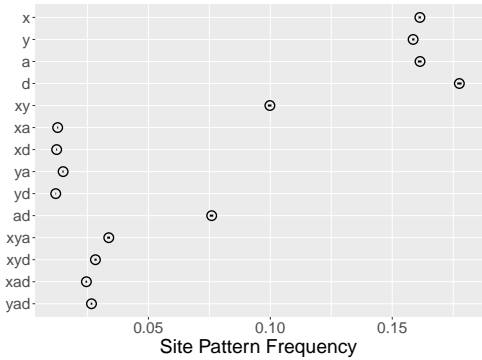


Doubletons confirm population relationships



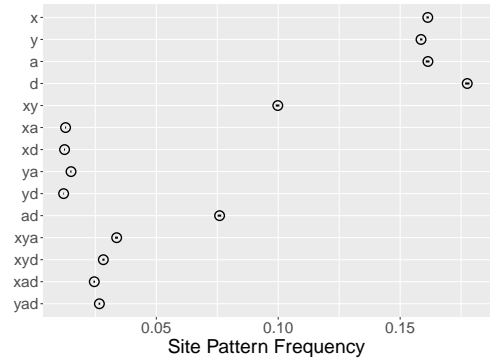
xy is common because X and Y are close relatives.
nd is common because N and D are close relatives.
Close relatives share ancestors; mutations in these ancestors generate these patterns.

X and Y separated more recently than N and D



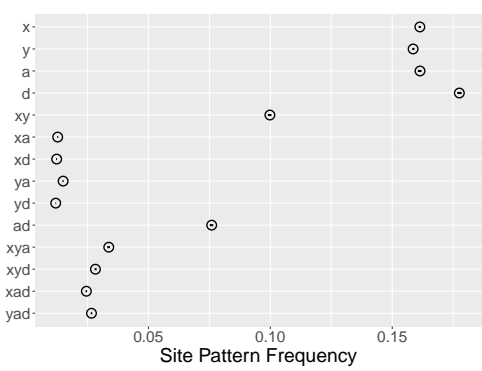
xy is more common than *nd*.
Implies that X and Y had longer period of shared ancestry.
Implies that X and Y separated more recently.

$N \rightarrow Y$ gene flow



ya is more common than *xa*.
x is more common than *y*.
Effects small but significant.
Signature of $N \rightarrow Y$ gene flow.

Puzzling excess of *d* site pattern



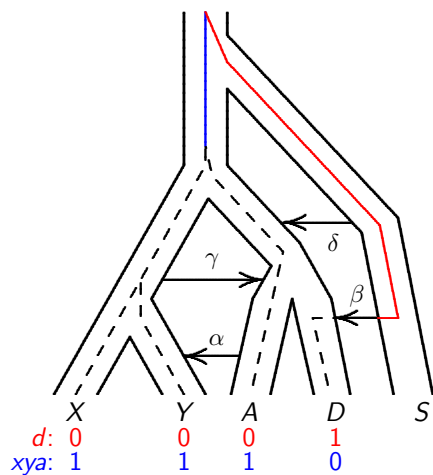
d is most common site pattern

Suggests Denisovan fossil is young and *N-D* separation old.

But our 2017 analysis of this hypothesis led to absurd result: Denisovan fossil only 4000 y old.

Something was missing from our model

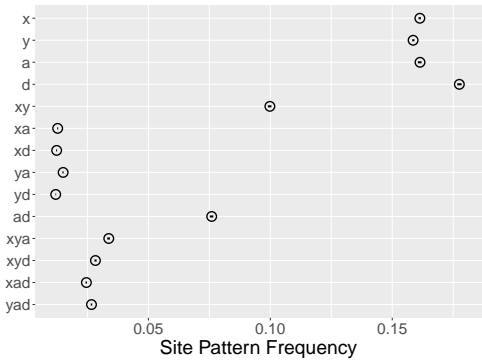
Admixture from superarchaics into Denisovans



S is a "superarchaic" hominin, distantly related to all others.

S → *D* gene flow inflated frequency of *d* and *xya*.

Superarchaic gene flow into *D*



d is most common singleton

xya is most common tripleton

Signature of *S* → *D* gene flow (Prüfer et al 2014).

Conclusions

1. Europeans and Africans are close relatives.
 2. So are Neanderthals and Denisovans.
 3. Europeans and Africans separated more recently than did Neanderthals and Denisovans.
 4. Neanderthals contributed genes to Europeans
 5. Superarchaics contributed genes to Denisovans.
- None of this is new, but we got there just by looking at the data.