

# Detecting and estimating (Neanderthal) ancestry using the *admixr* R package

A little bit of theory & practical demo/tutorial in R

Martin Petr

GLOBE Institute / GeoGenetics Center, University of Copenhagen

Max Planck Institute for Evolutionary Anthropology, Leipzig

[mp@bodkan.net](mailto:mp@bodkan.net) // @fleventy5

2021-03-31

# Demo #1

1. EIGENSTRAT file format

2. *admixr* R package

Detecting admixture  
(i.e. testing "treeness")  
using  $f$ -statistics

# 2010: Neanderthal “draft” genome

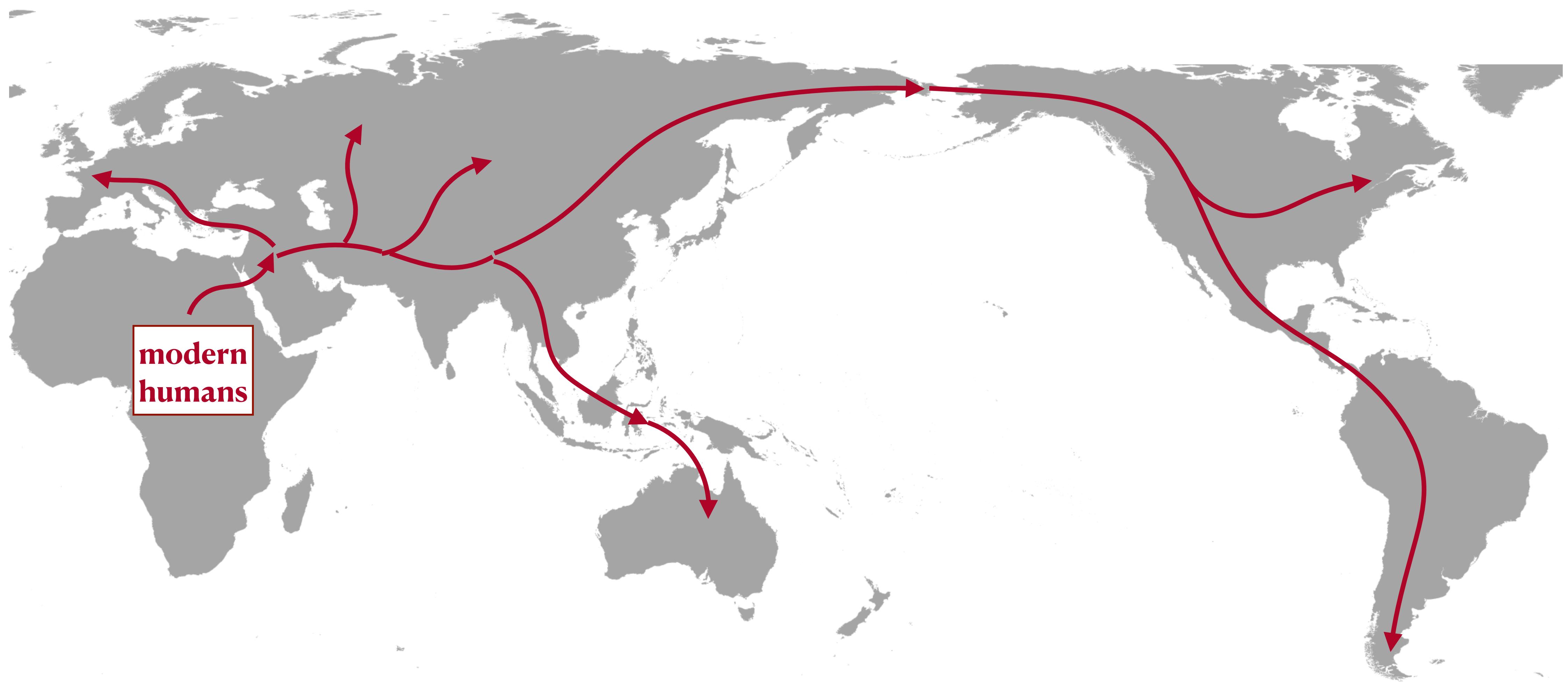


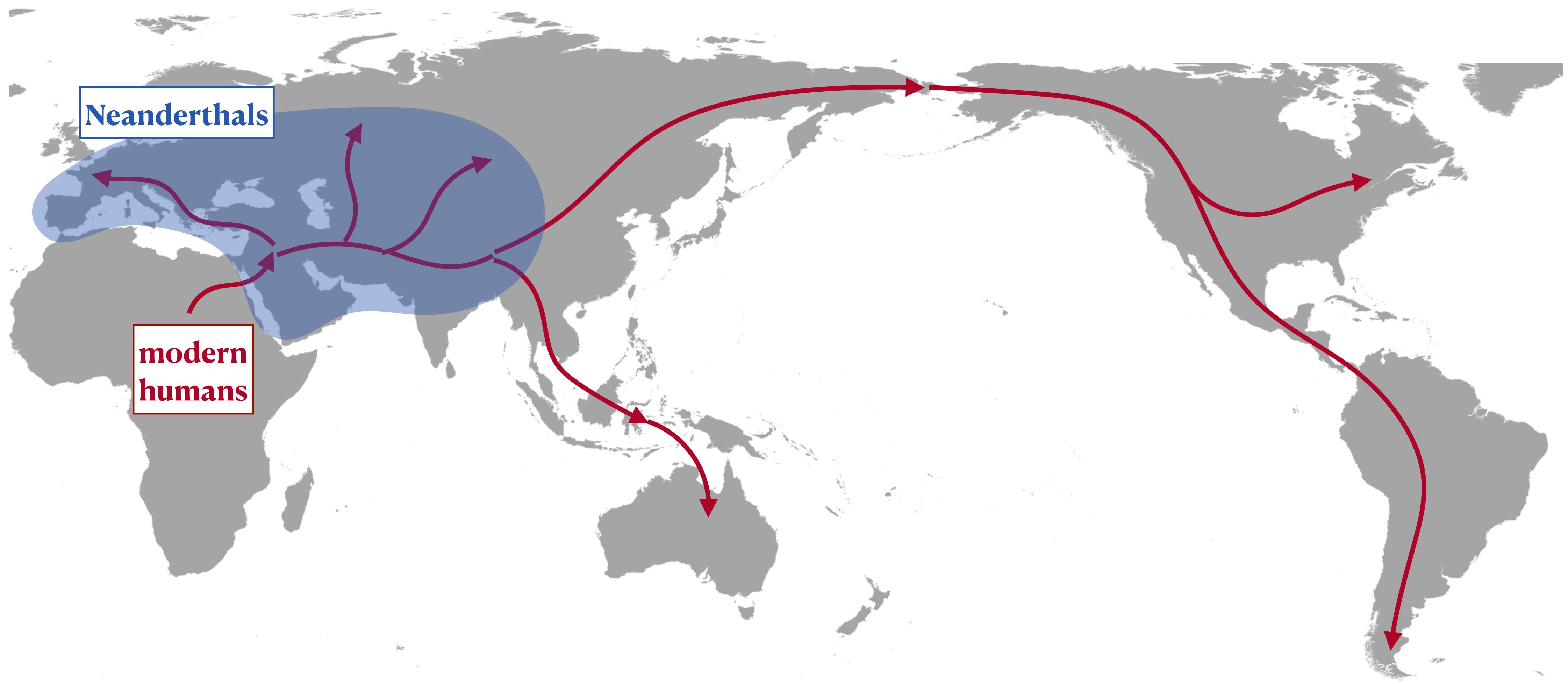
- ~1.3X coverage of the whole genome (later genomes at 30-40X)

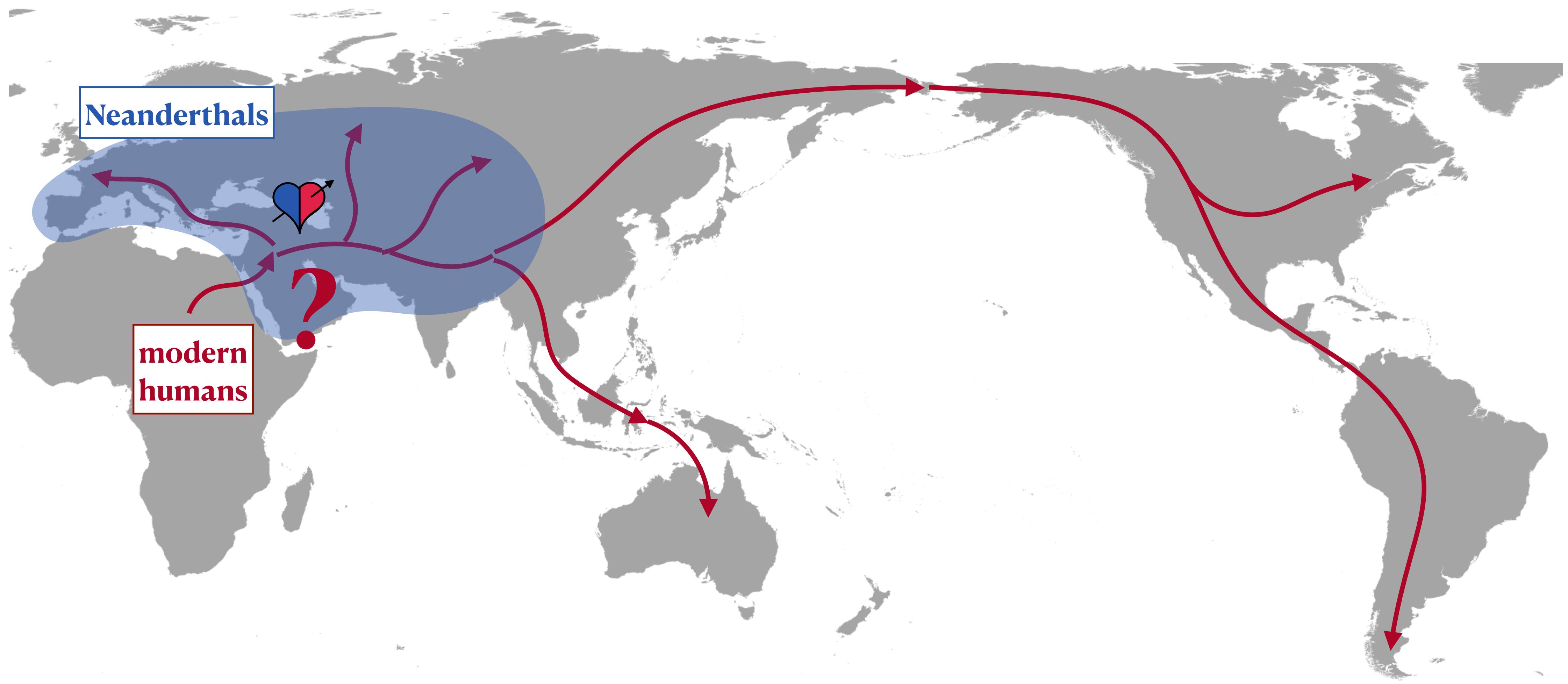




**modern  
humans**







# Detecting admixture

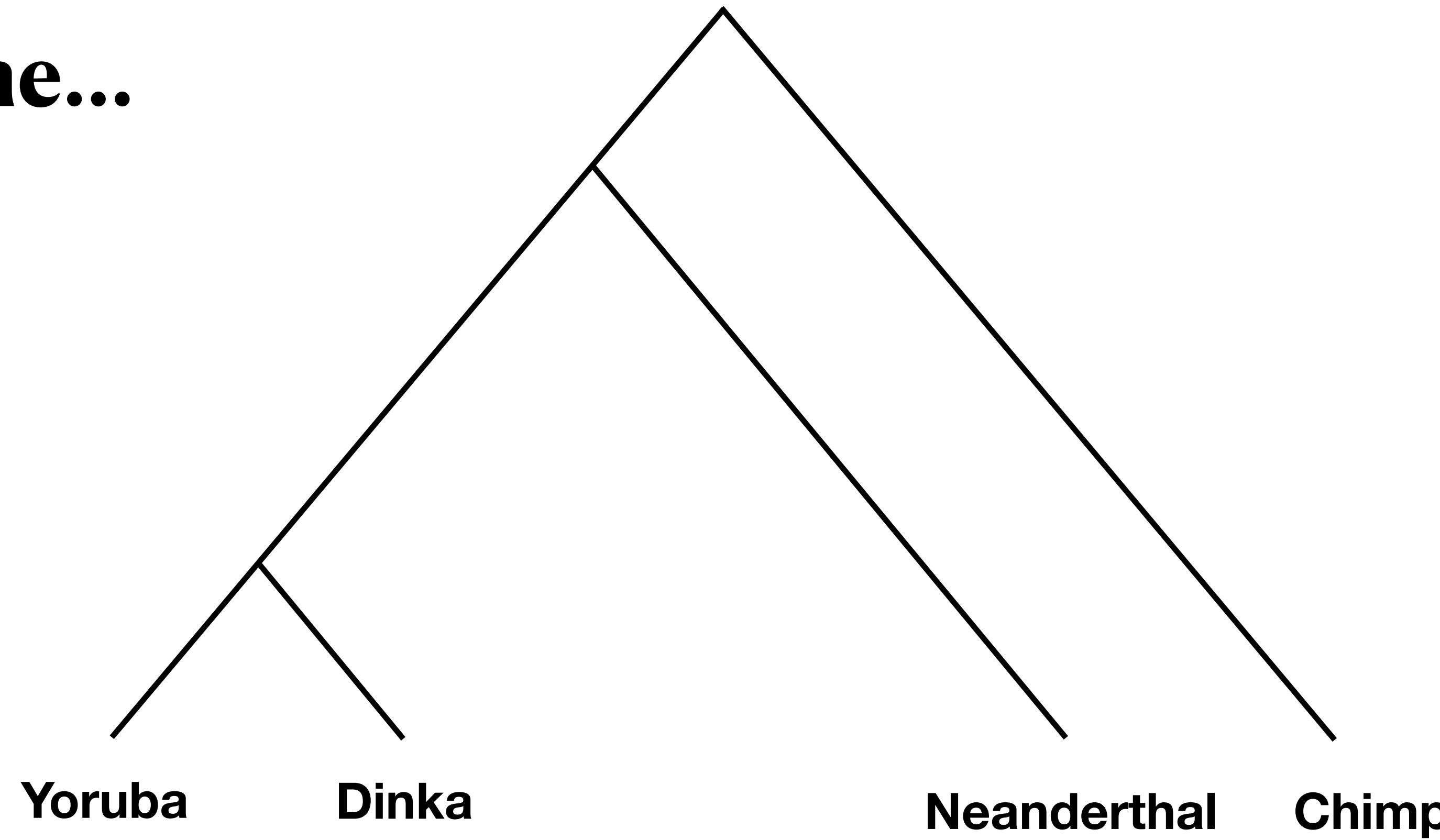
Four-population test —  $f_4$  statistic (*D* statistic)

# Detecting admixture

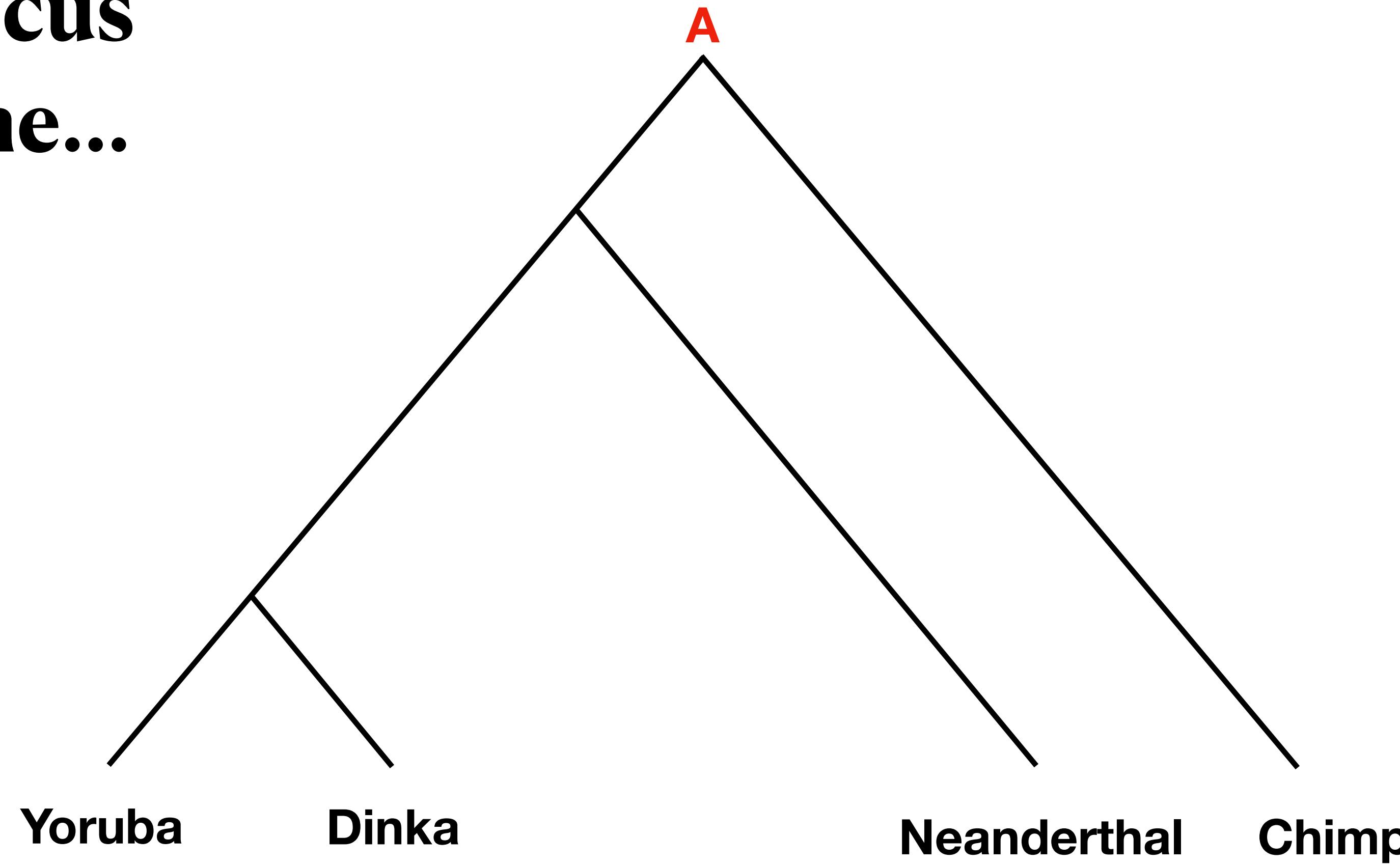
Four-population test –  $f_4$  statistic (*D statistic*)

Compares the counts of BABA vs ABBA sites  
between four samples (genomes)

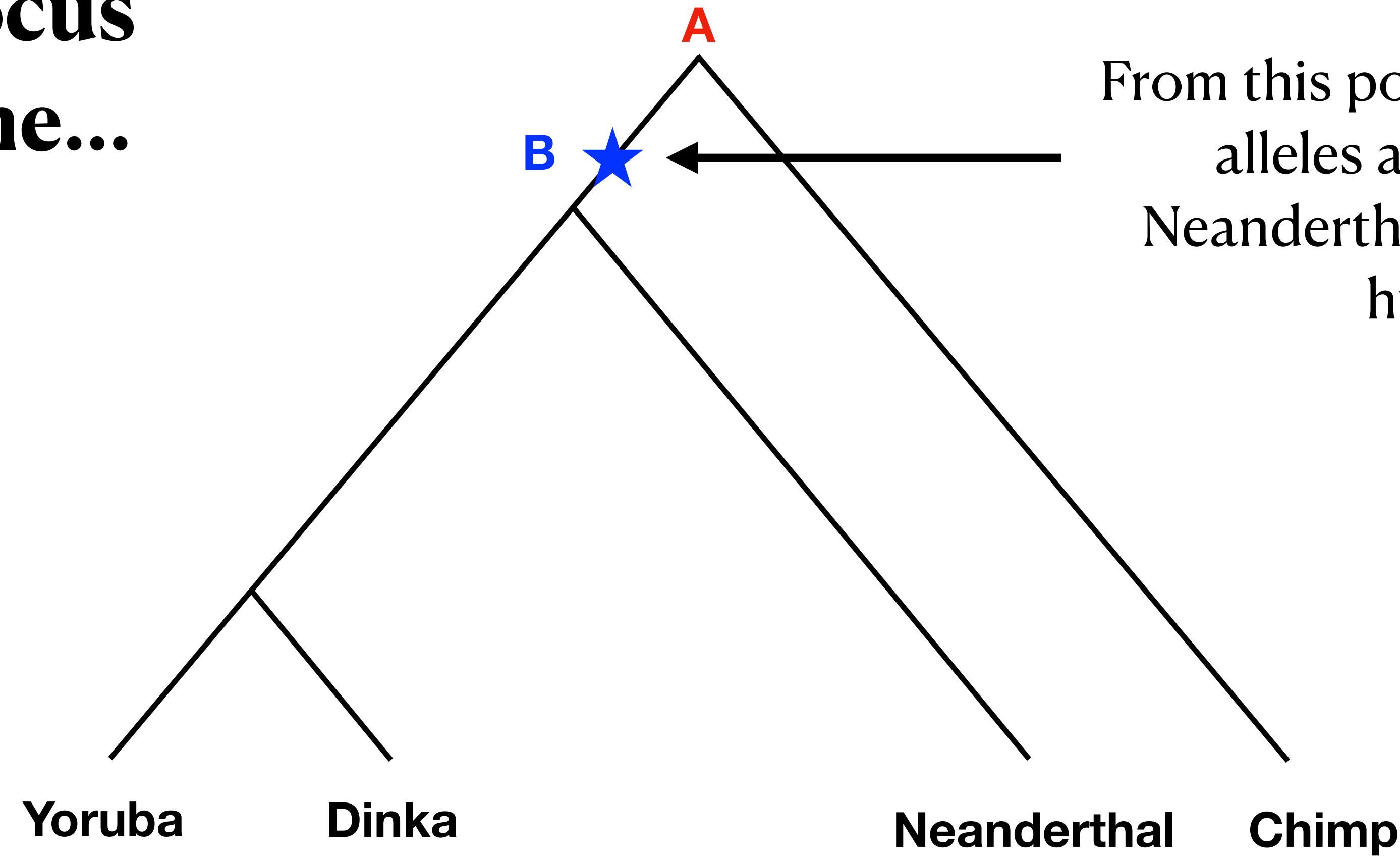
**At a single locus  
in the genome...**



**At a single locus  
in the genome...**

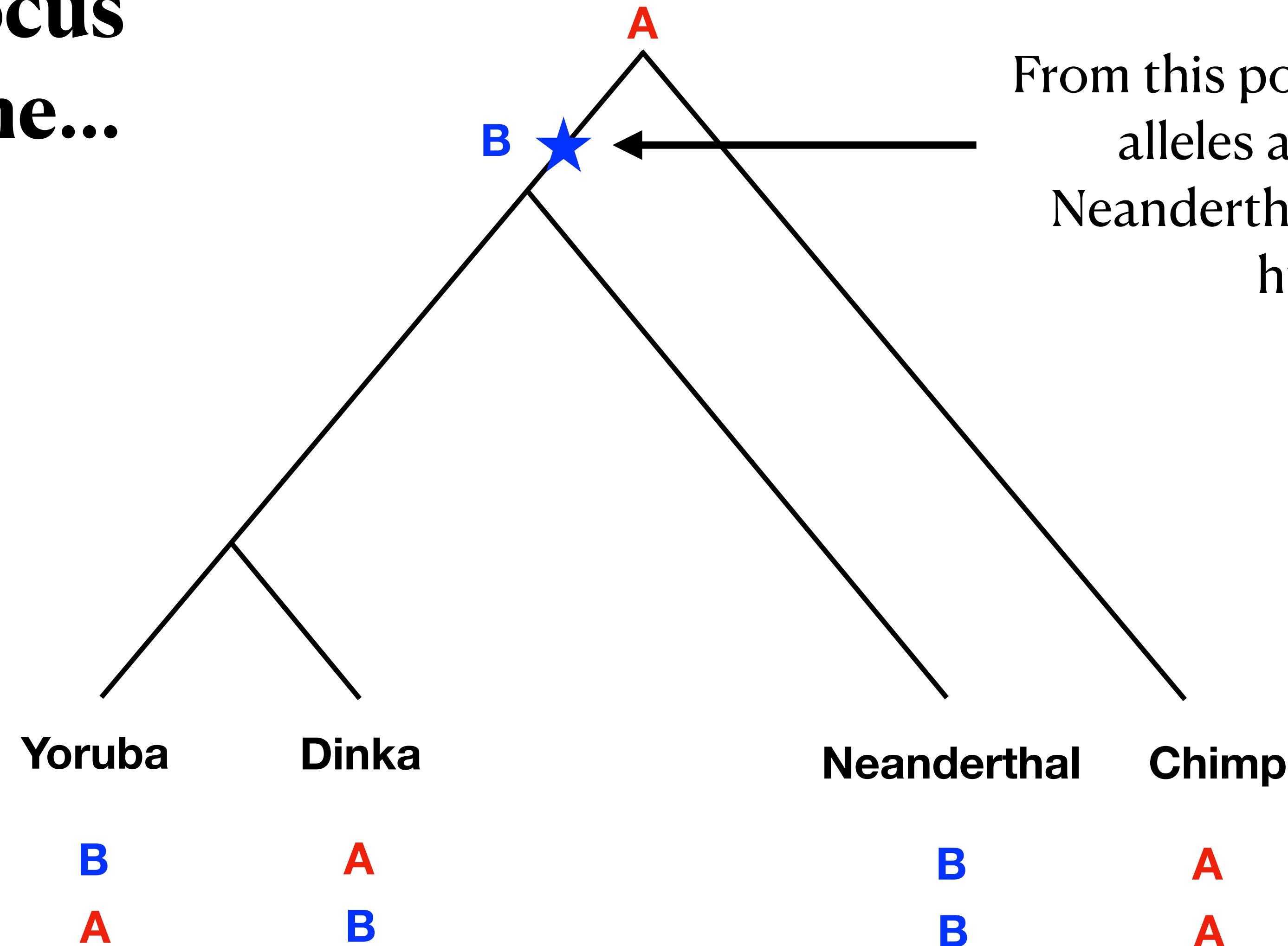


**At a single locus  
in the genome...**



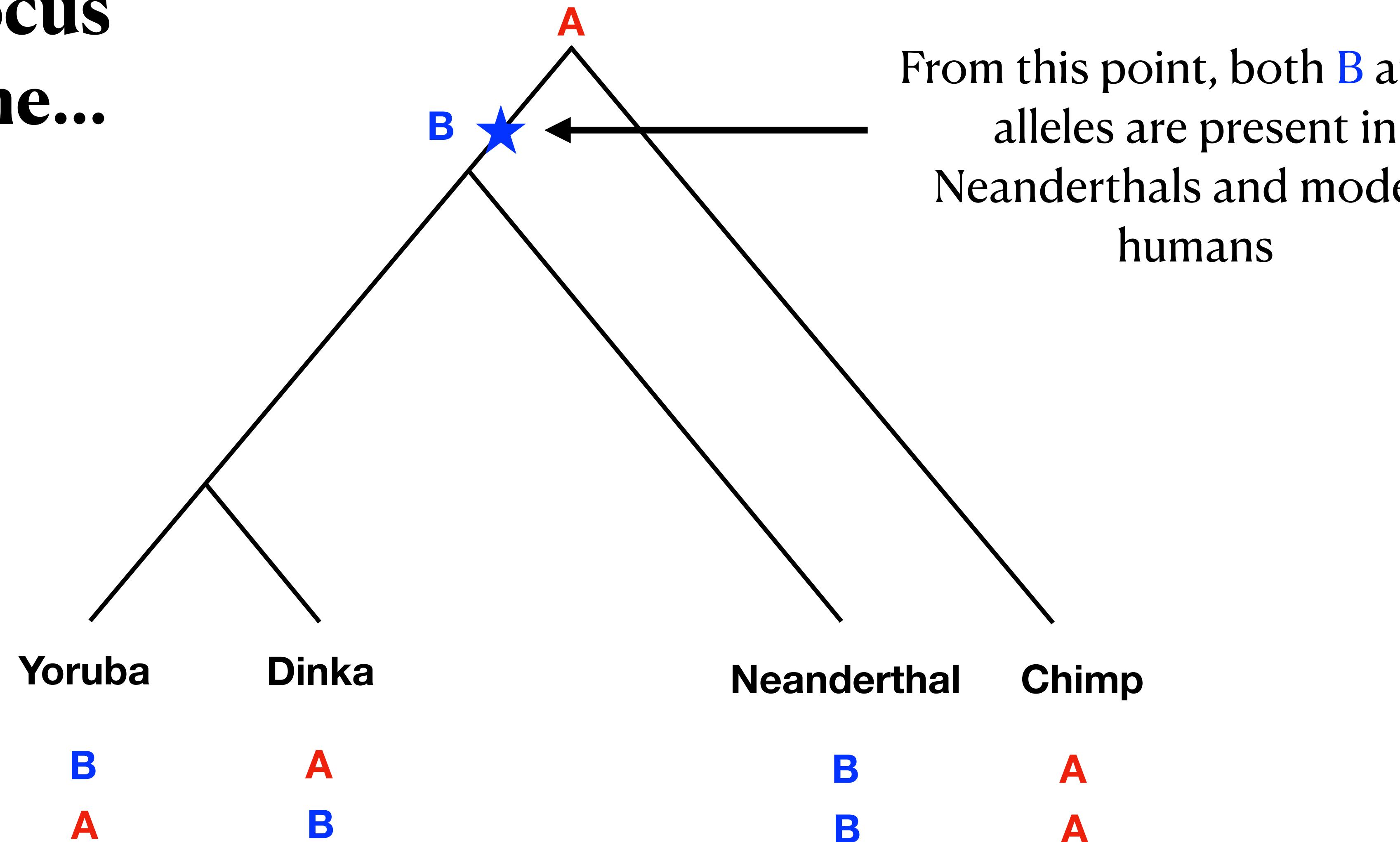
From this point, both **B** and **A** alleles are present in Neanderthals and modern humans

**At a single locus  
in the genome...**



From this point, both **B** and **A** alleles are present in Neanderthals and modern humans

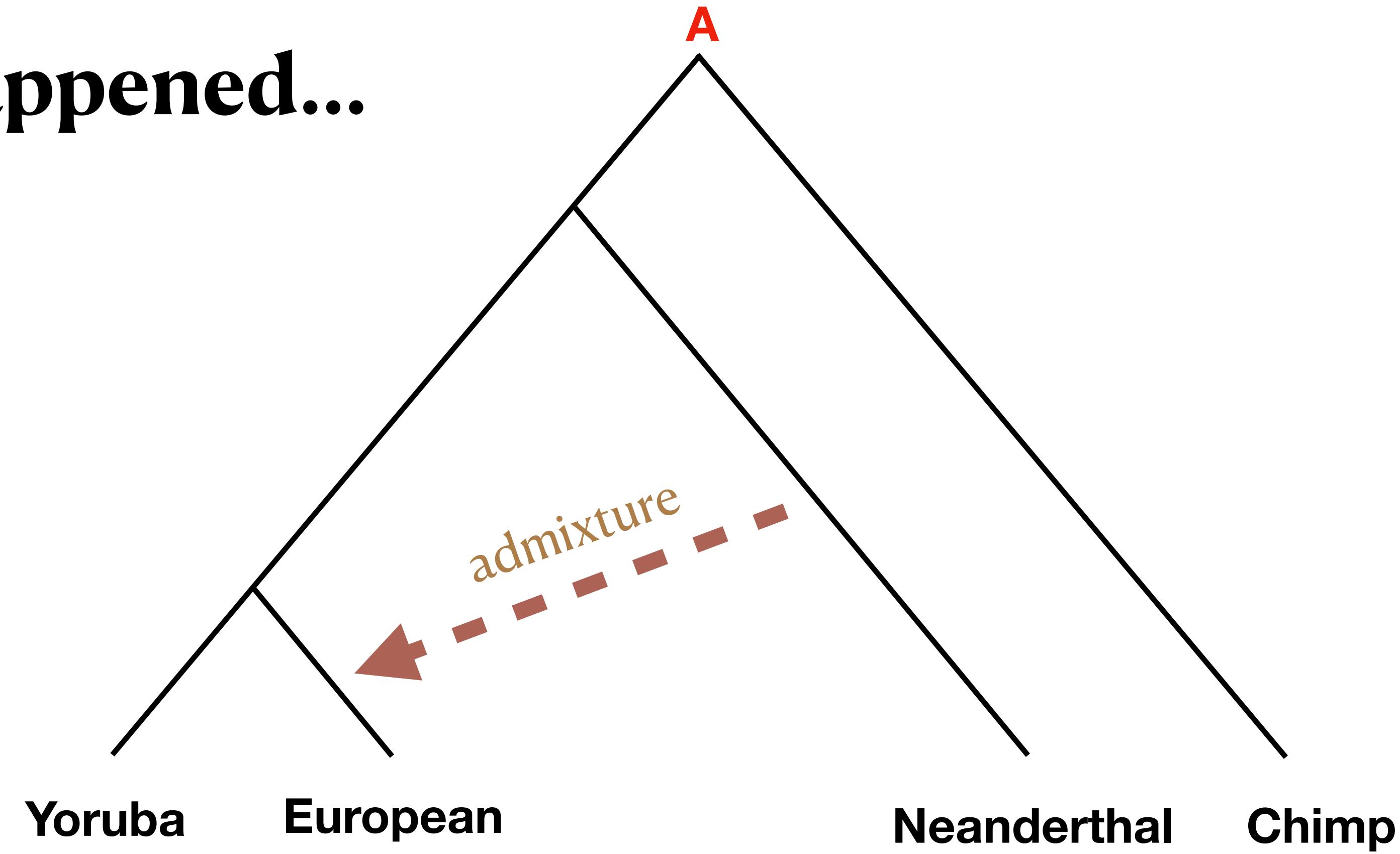
At a single locus  
in the genome...



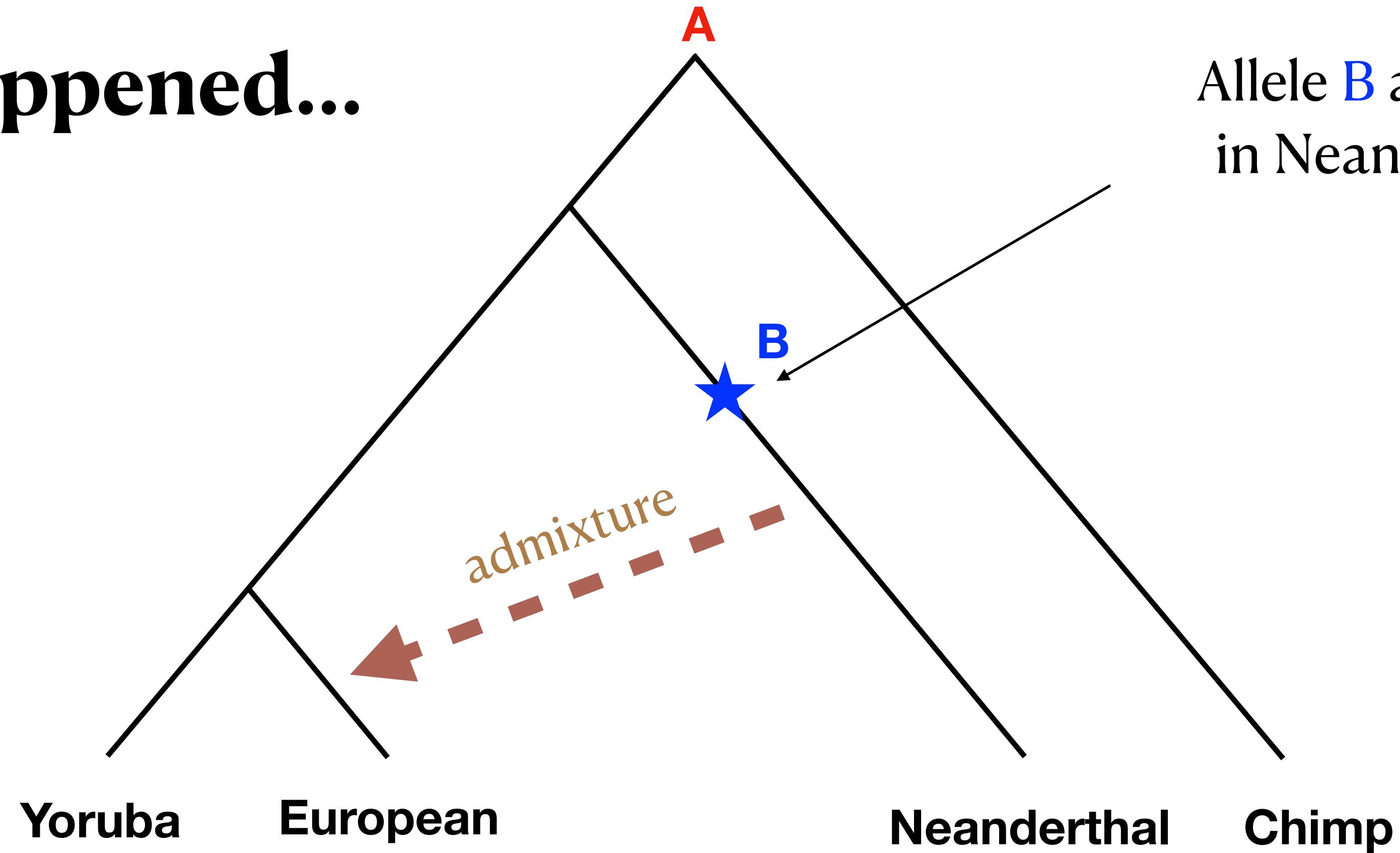
From this point, both **B** and **A** alleles are present in Neanderthals and modern humans

Data consistent with the tree:  $f_4 = (\# \text{ BABA} - \# \text{ ABBA}) / \# \text{ sites total} \sim 0$

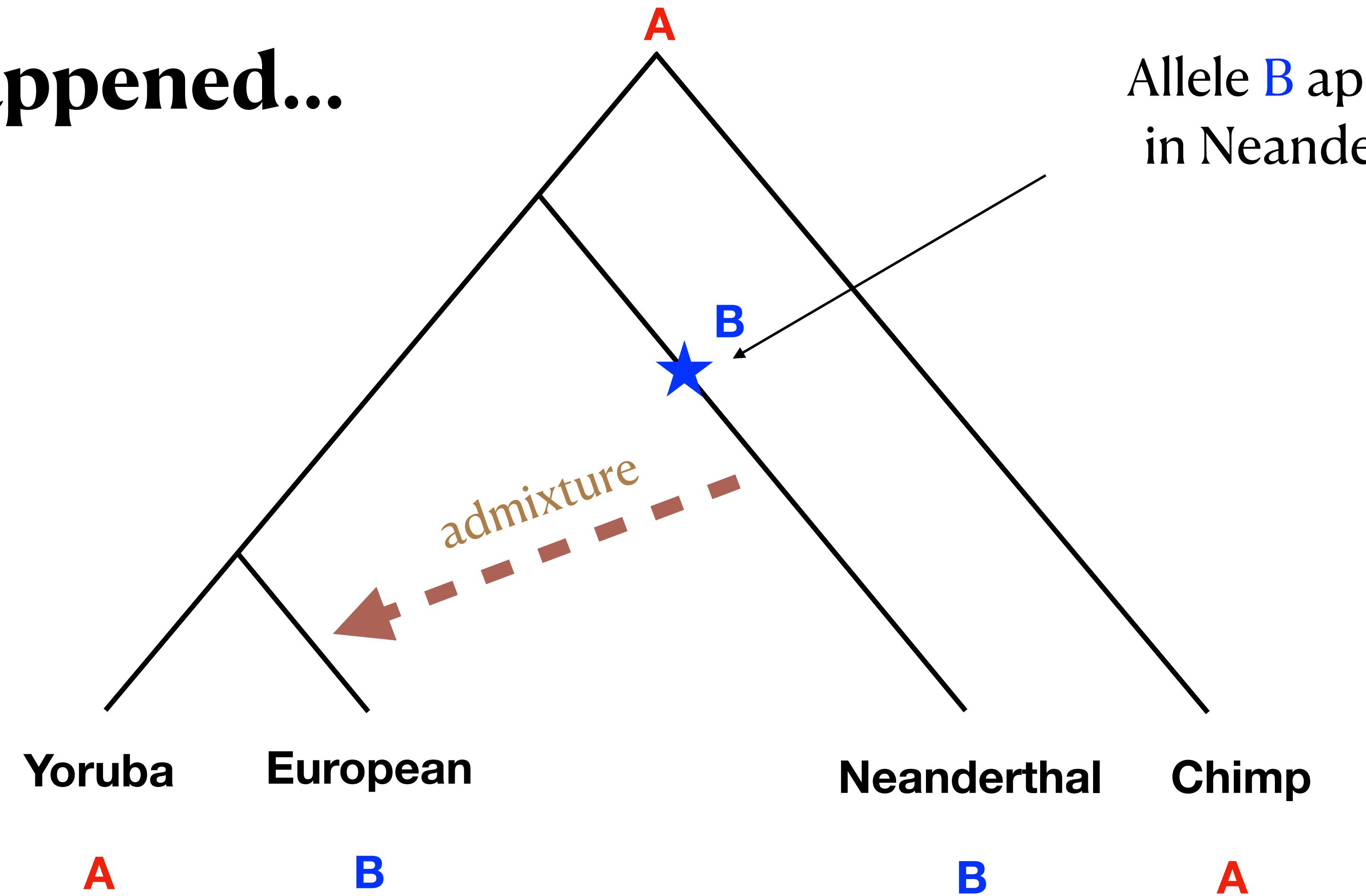
**Assume that  
admixture happened...**



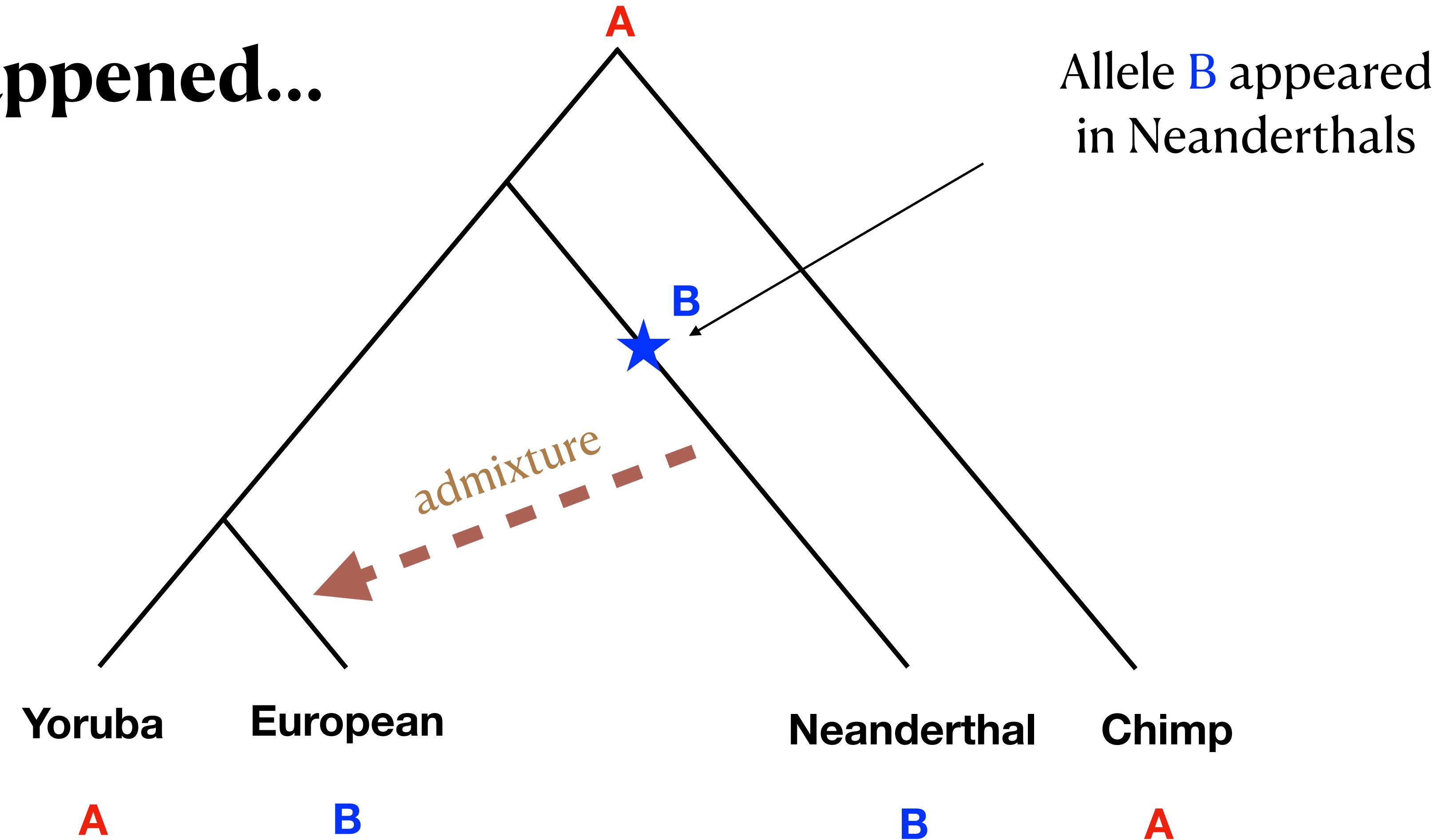
**Assume that  
admixture happened...**



**Assume that  
admixture happened...**



# Assume that admixture happened...



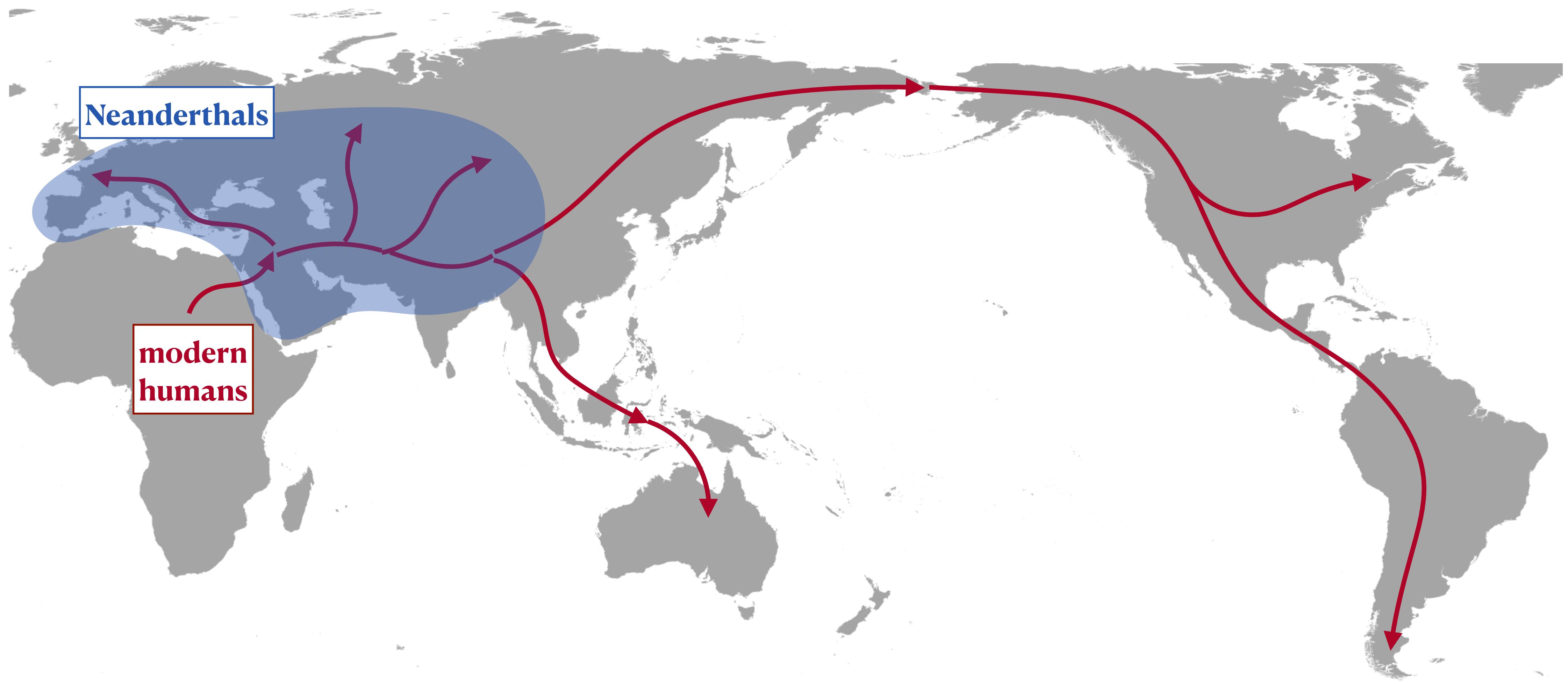
**Introgression:**  $f_4 = (\# \text{ BABA} - \# \text{ABBA}) / \# \text{ sites total} < 0$

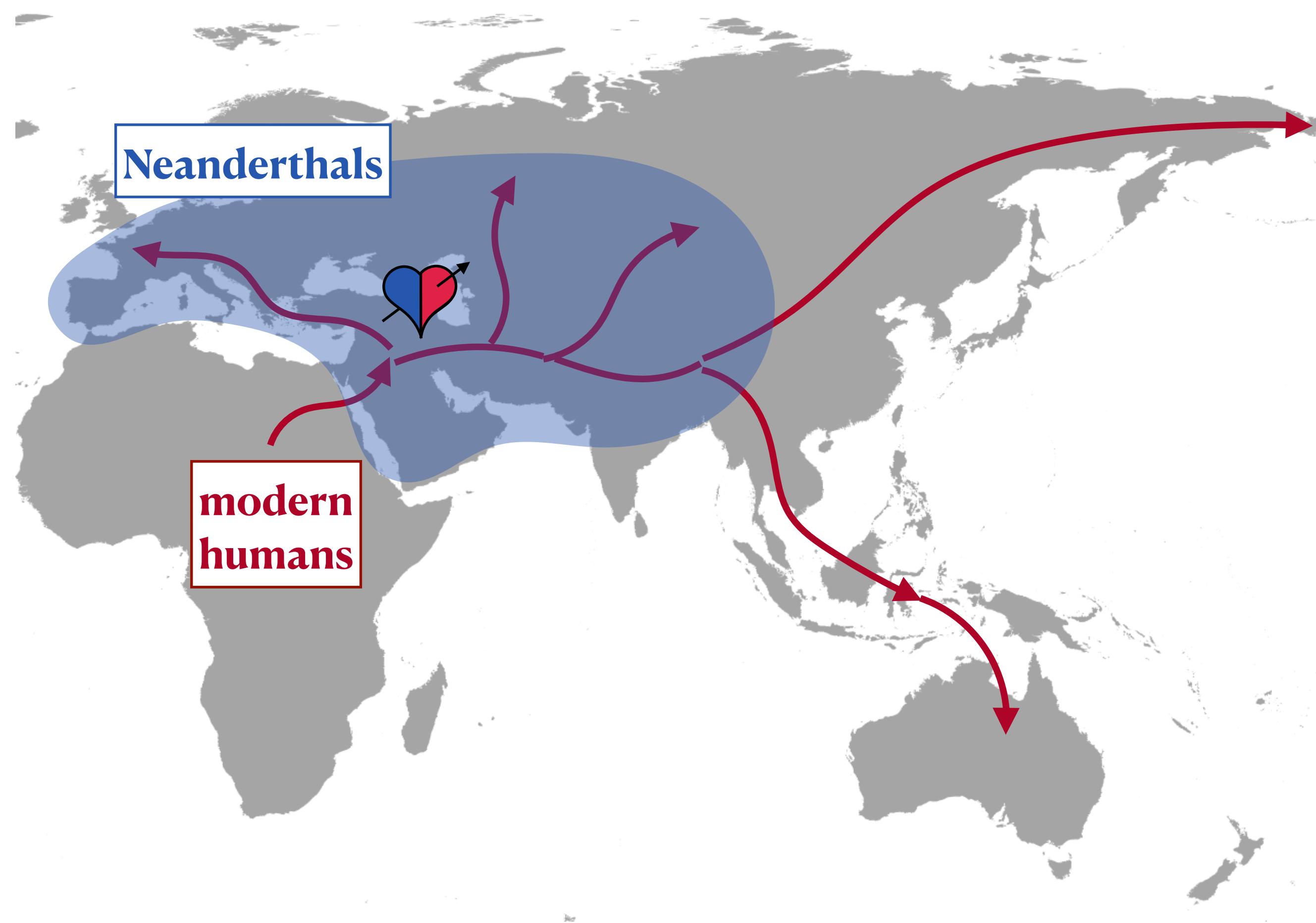
(the sign can flip if we swap Yoruba and European)



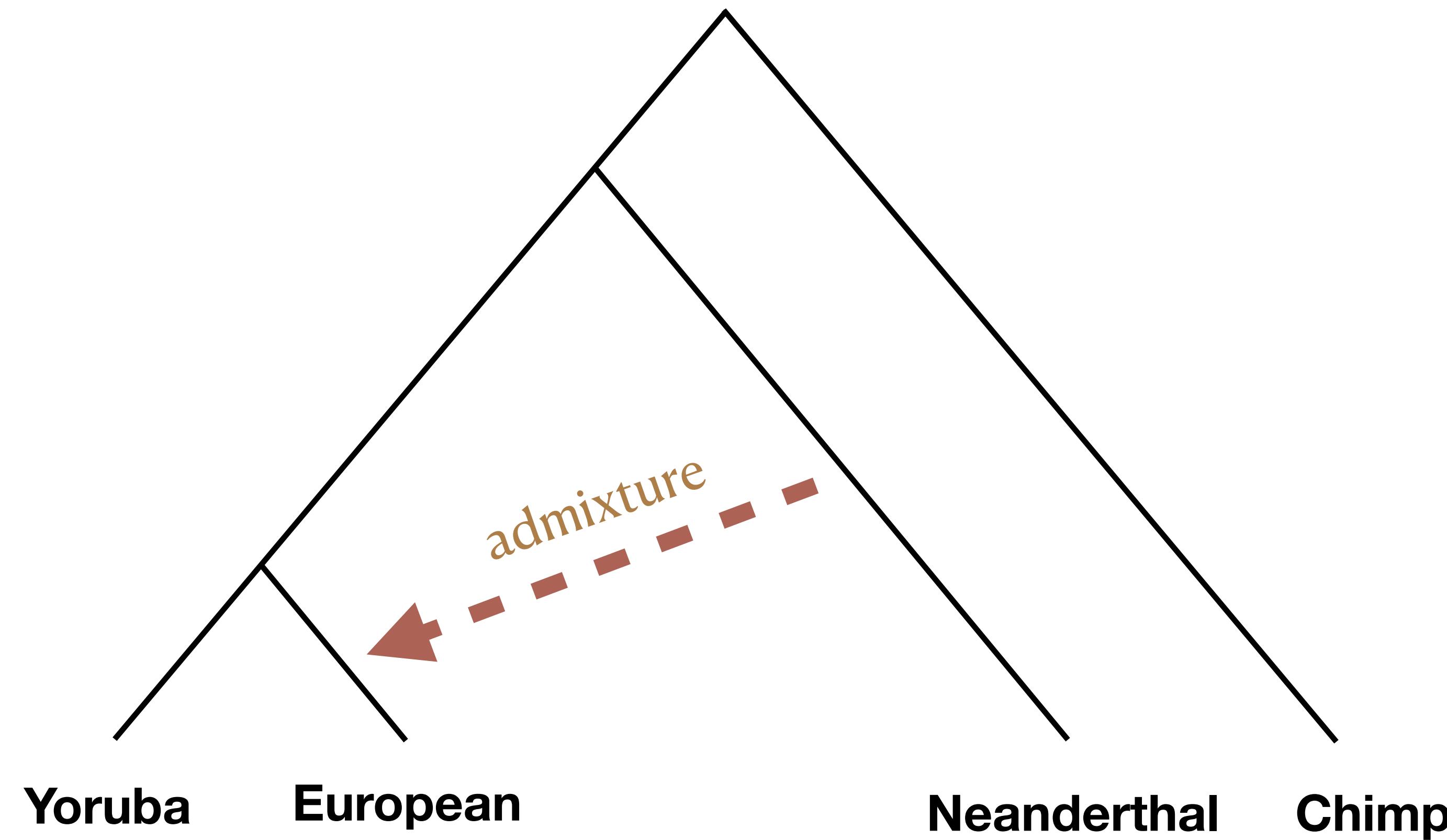
# Demo #2

Performing admixture tests using *admixr*



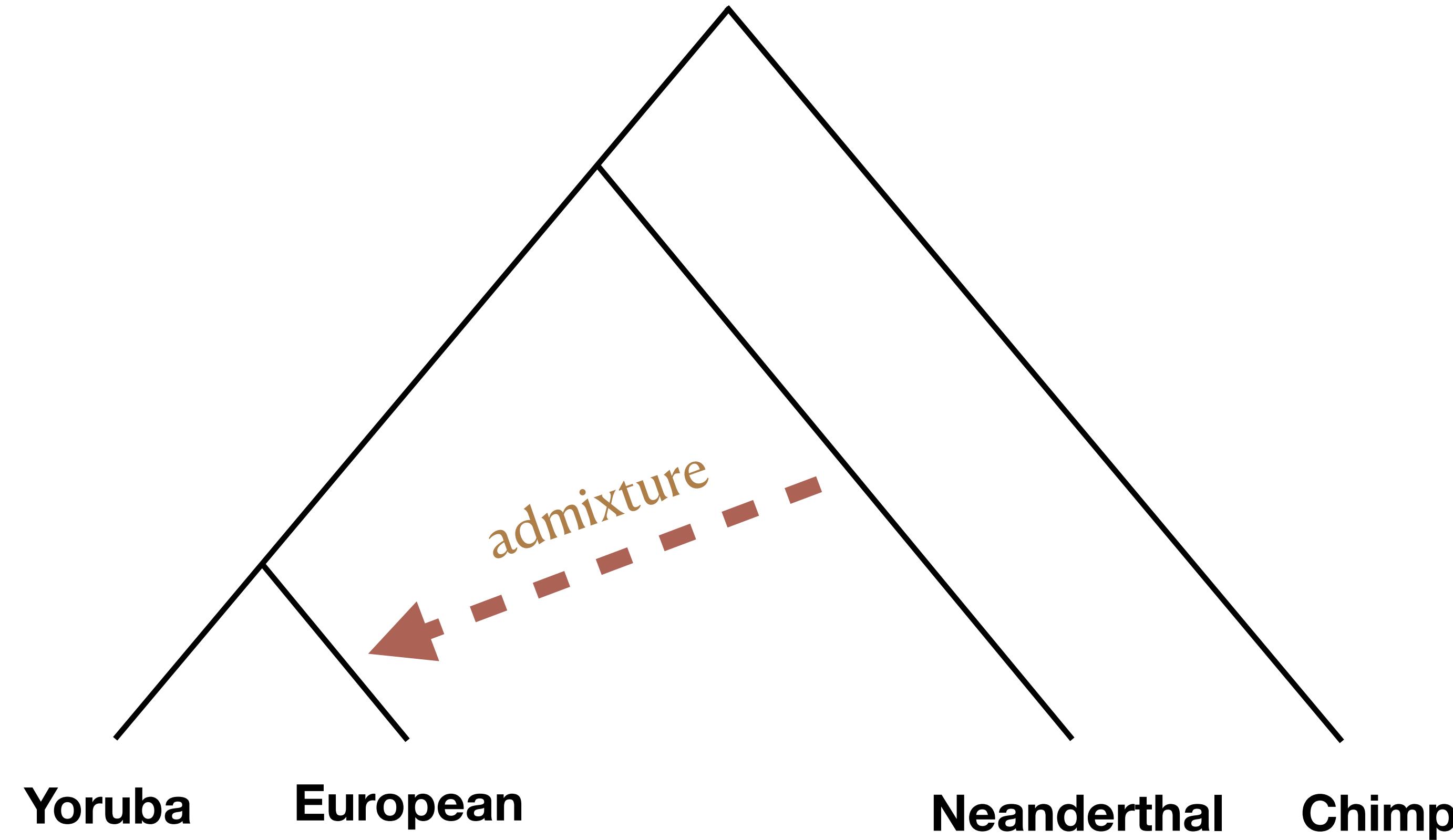


# Detecting admixture



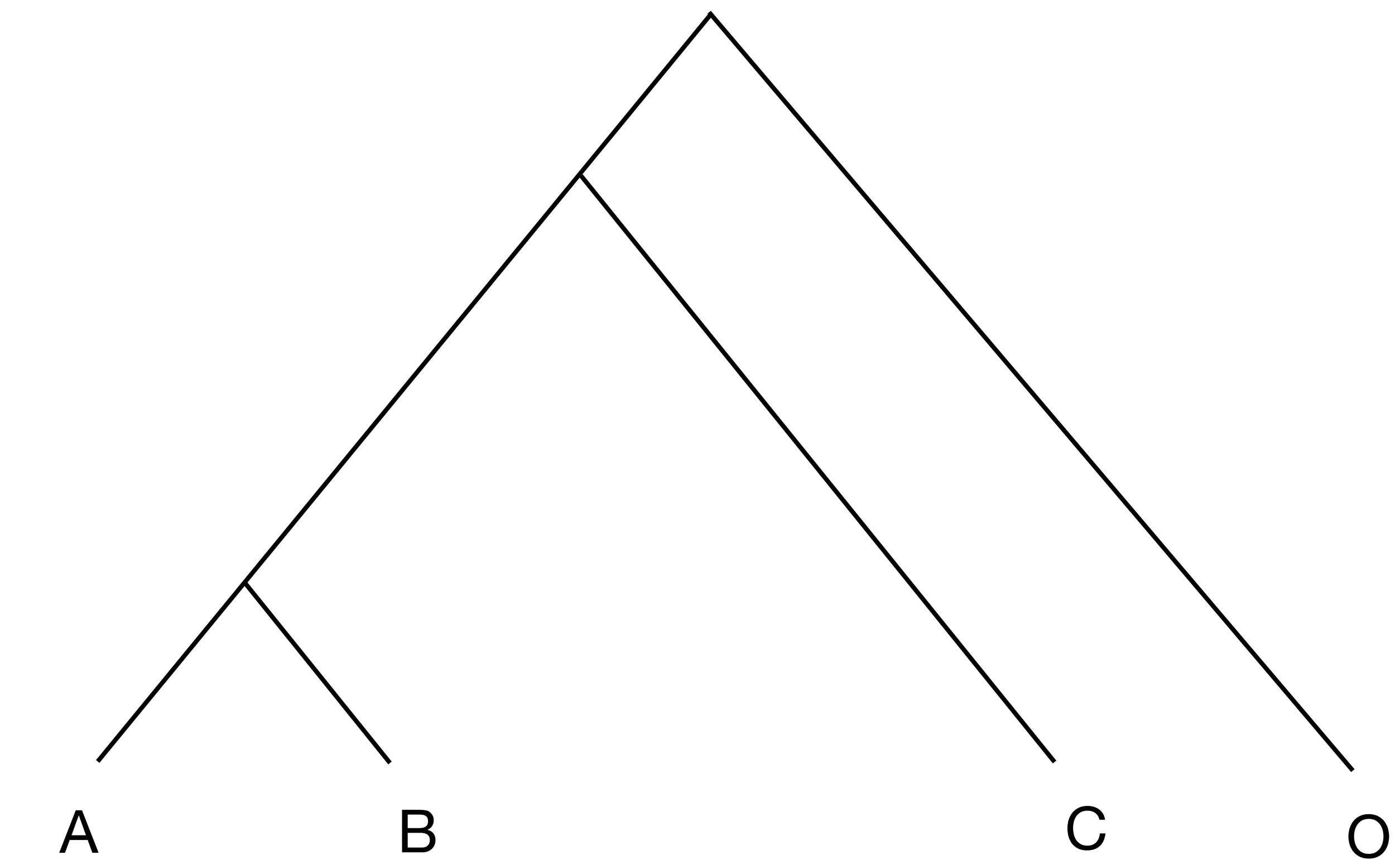
**Introgression:**  $f_4 = (\# \text{ BABA} - \# \text{ ABBA}) / \# \text{ sites total} < 0$

# Detecting admixture proportion?



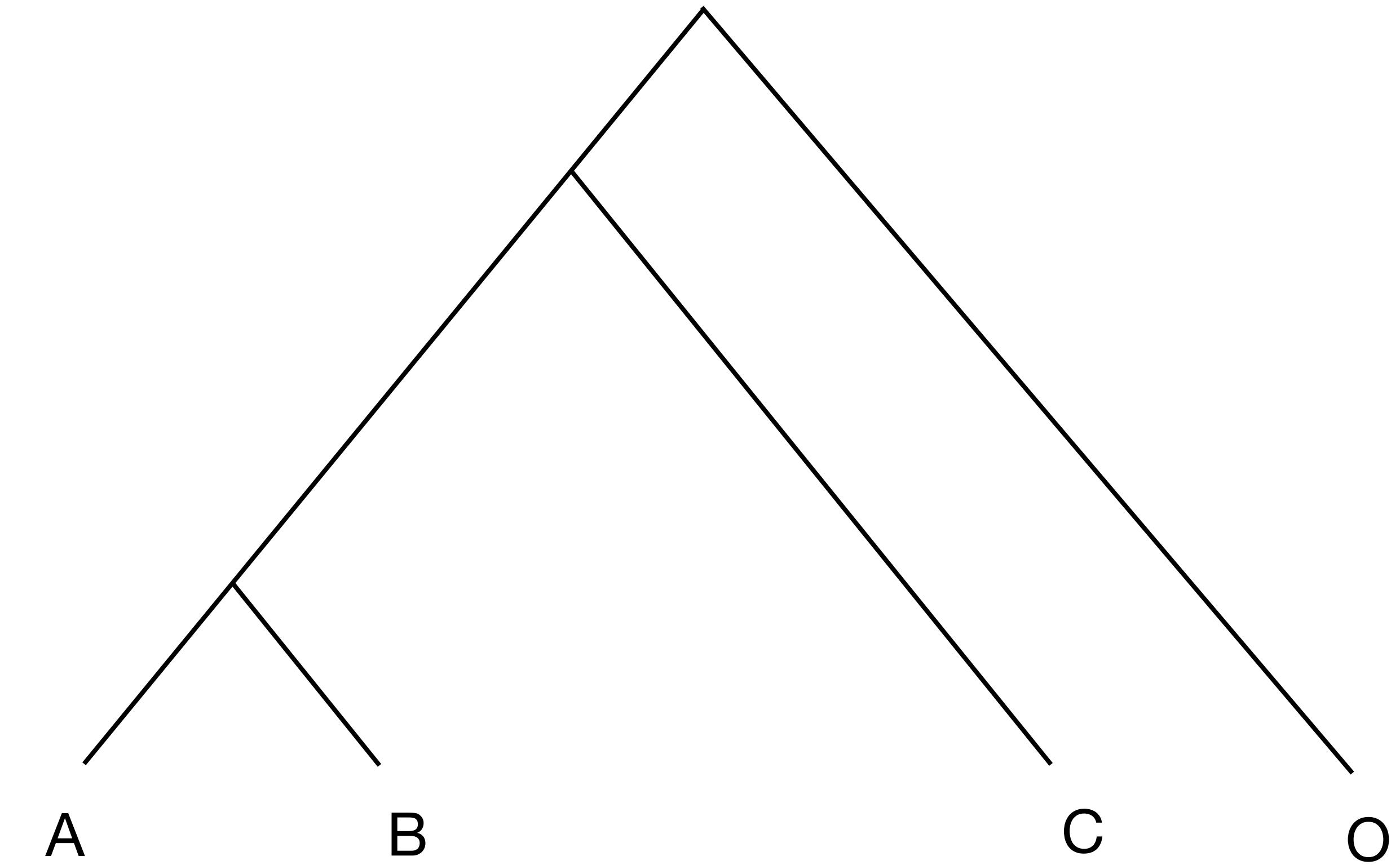
**Introgression:**  $f_4 = (\# \text{ BABA} - \# \text{ ABBA}) / \# \text{ sites total} < 0$

# $f_4$ -ratio statistic: two possible paths for an admixed sample



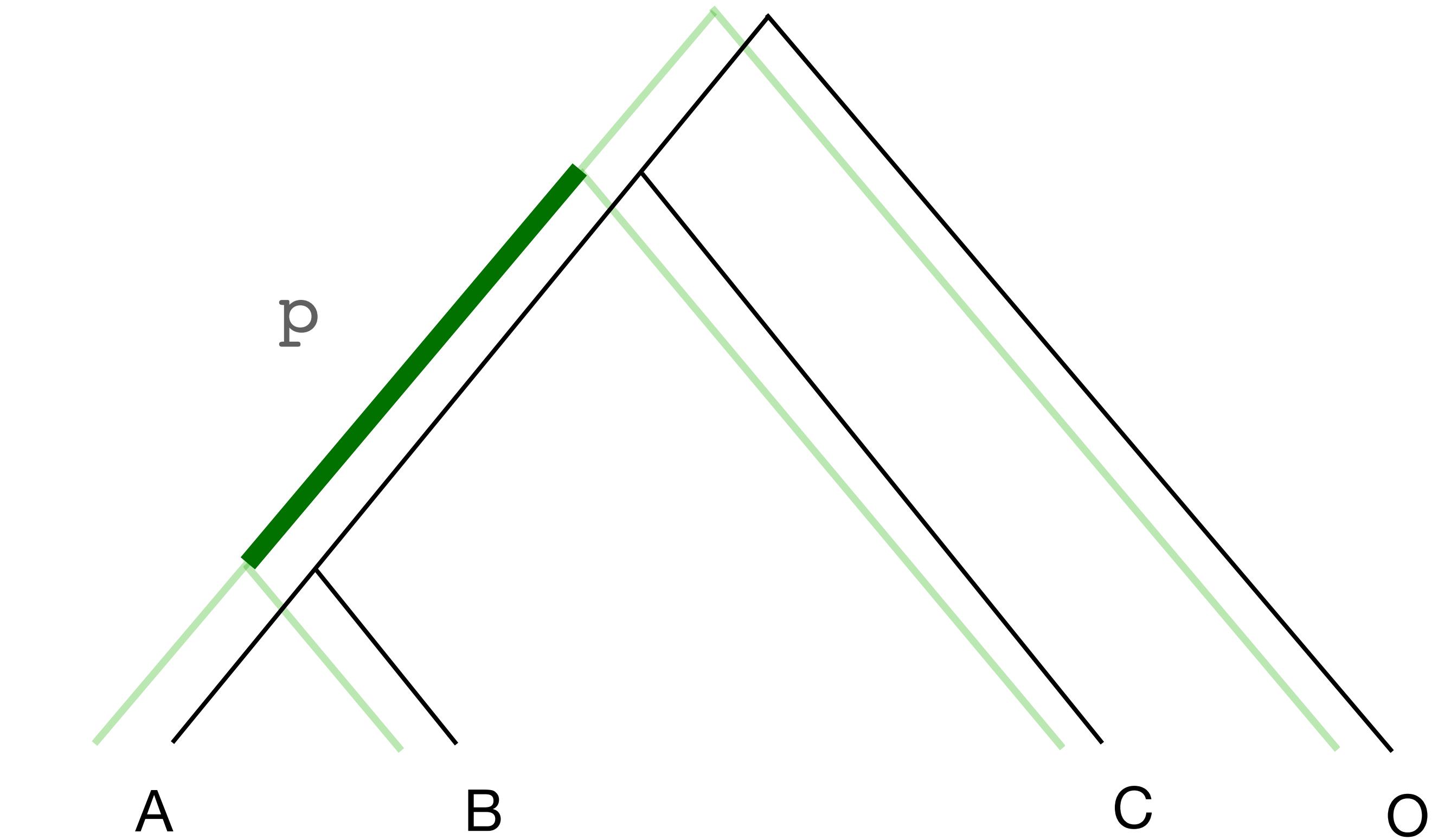
# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$f_4(C, B; A, O)$$



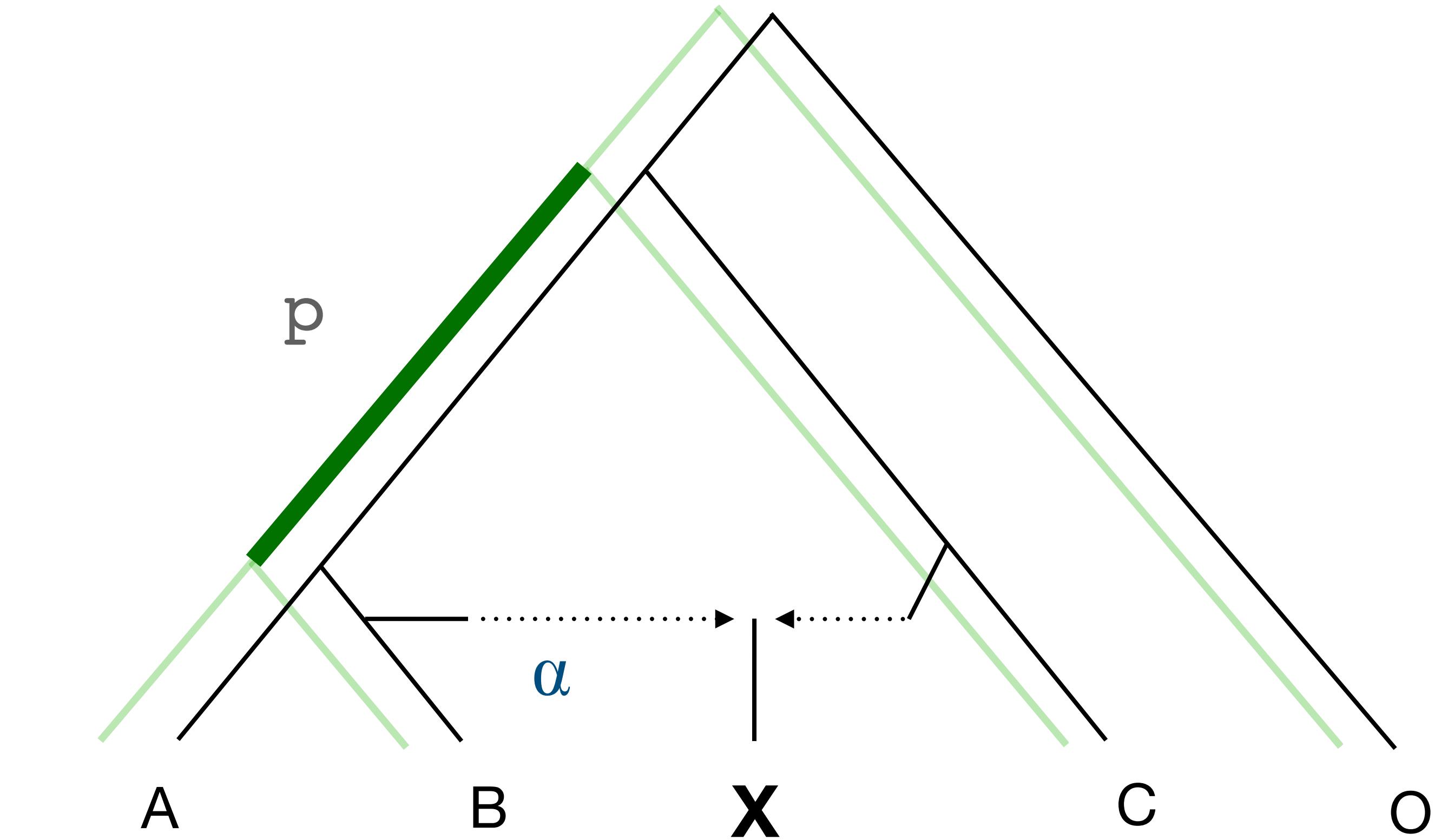
# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$



# $f_4$ -ratio statistic: two possible paths for an admixed sample

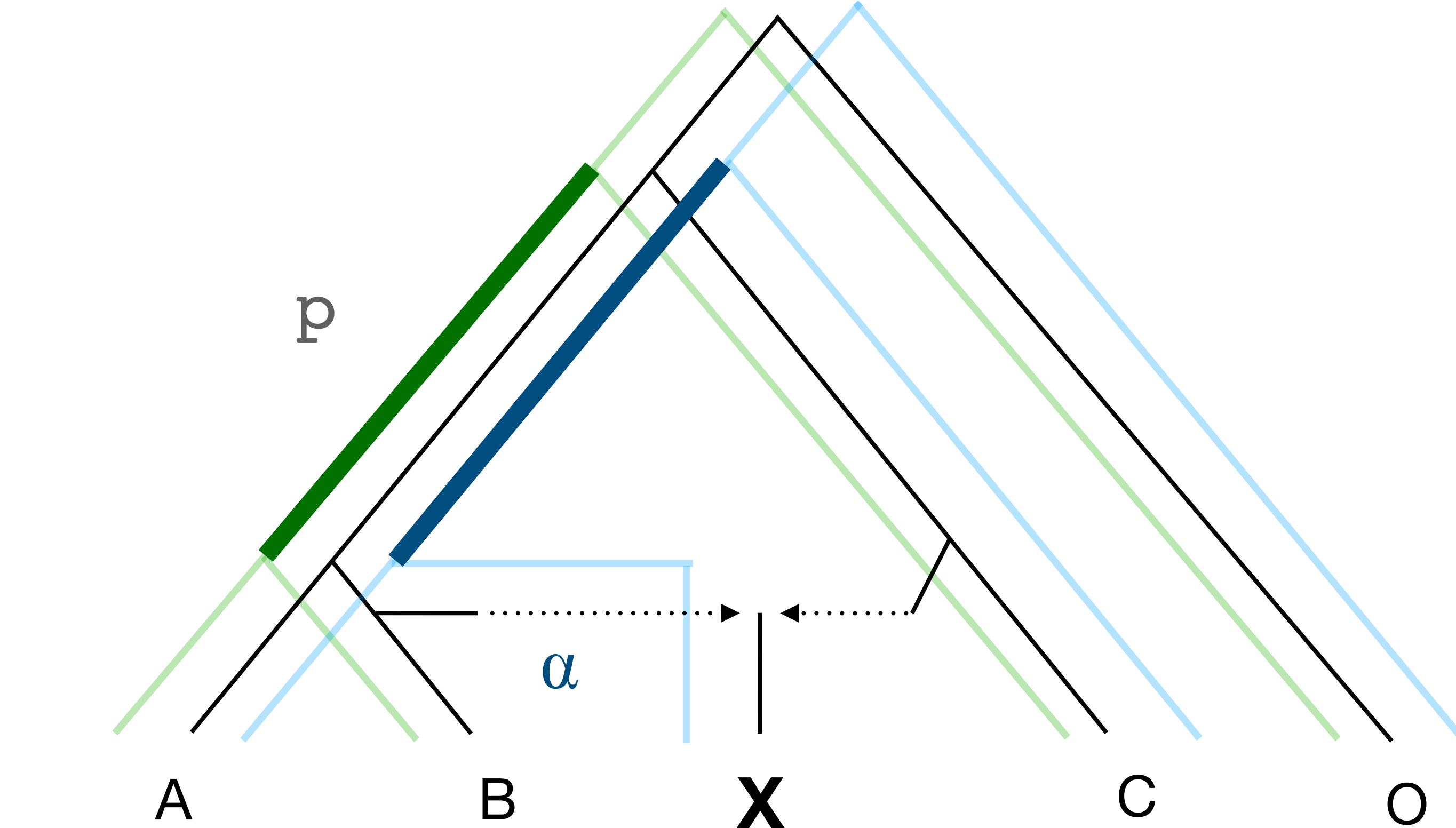
$$p = f_4(C, B; A, O)$$



# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$

$$\alpha p = f_4(C, X; A, O)$$



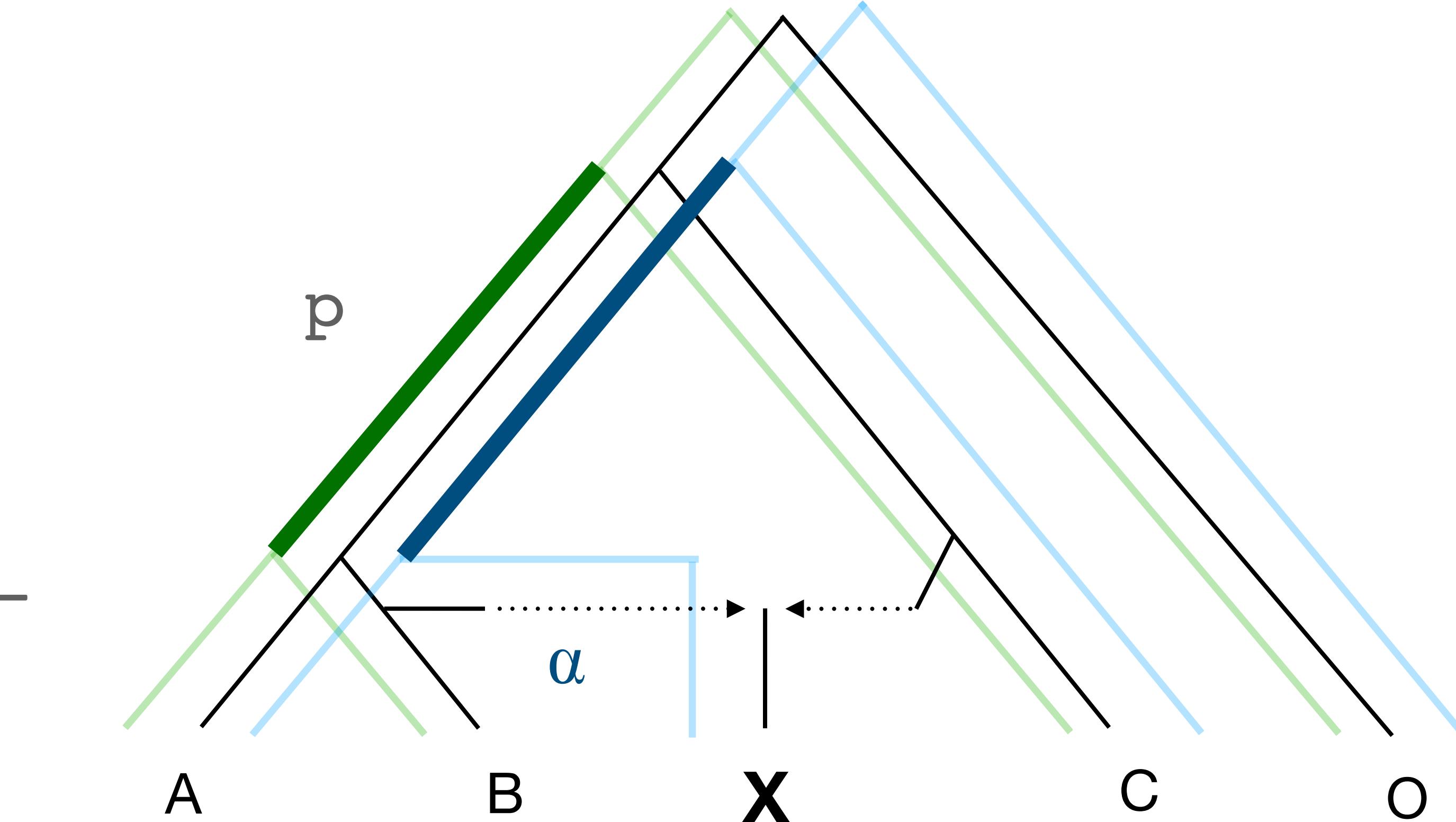
# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$

$$\alpha p = f_4(C, X; A, O)$$

$$f_4(C, X; A, O)$$

$$f_4(C, B; A, O)$$

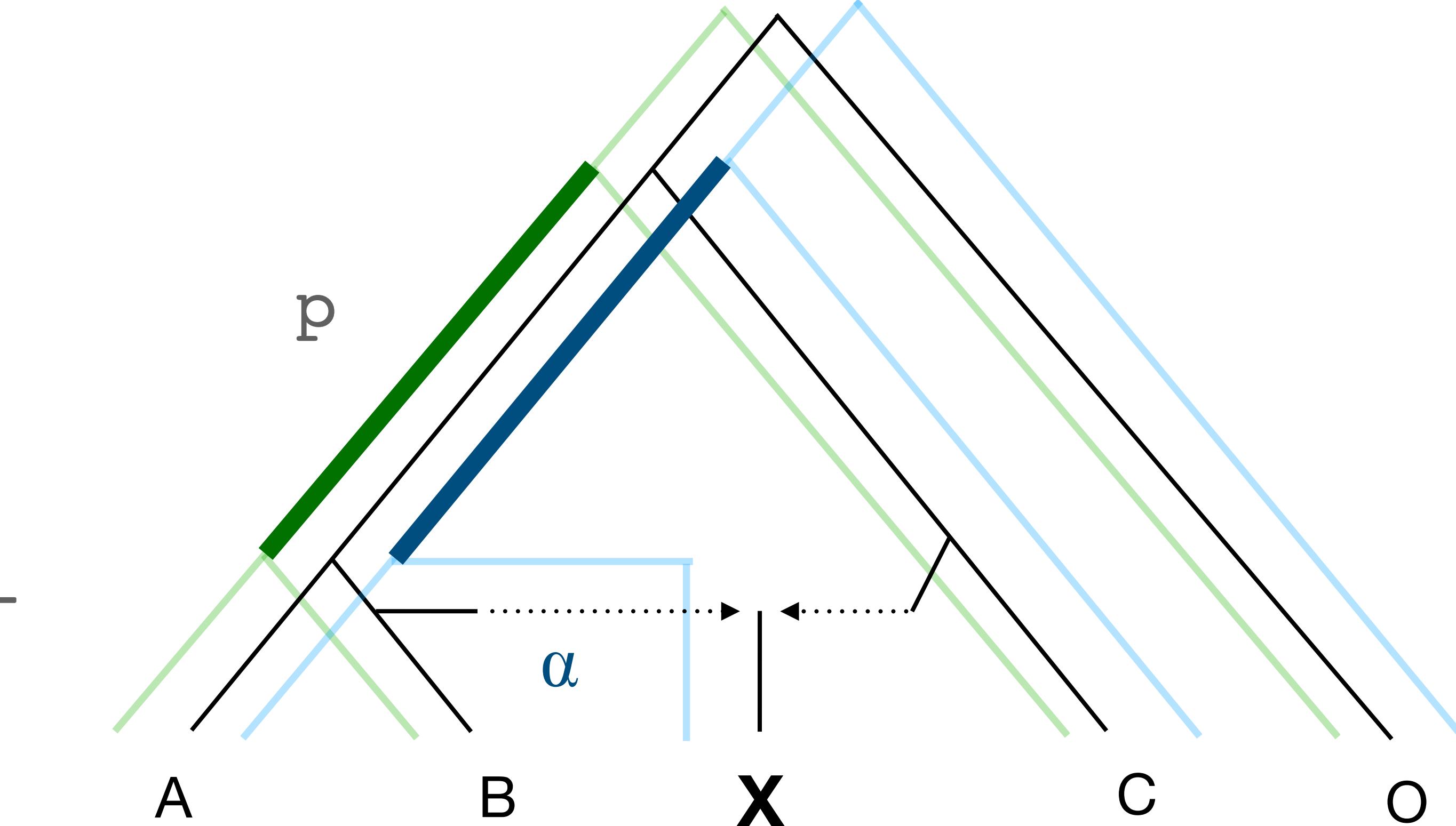


# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$

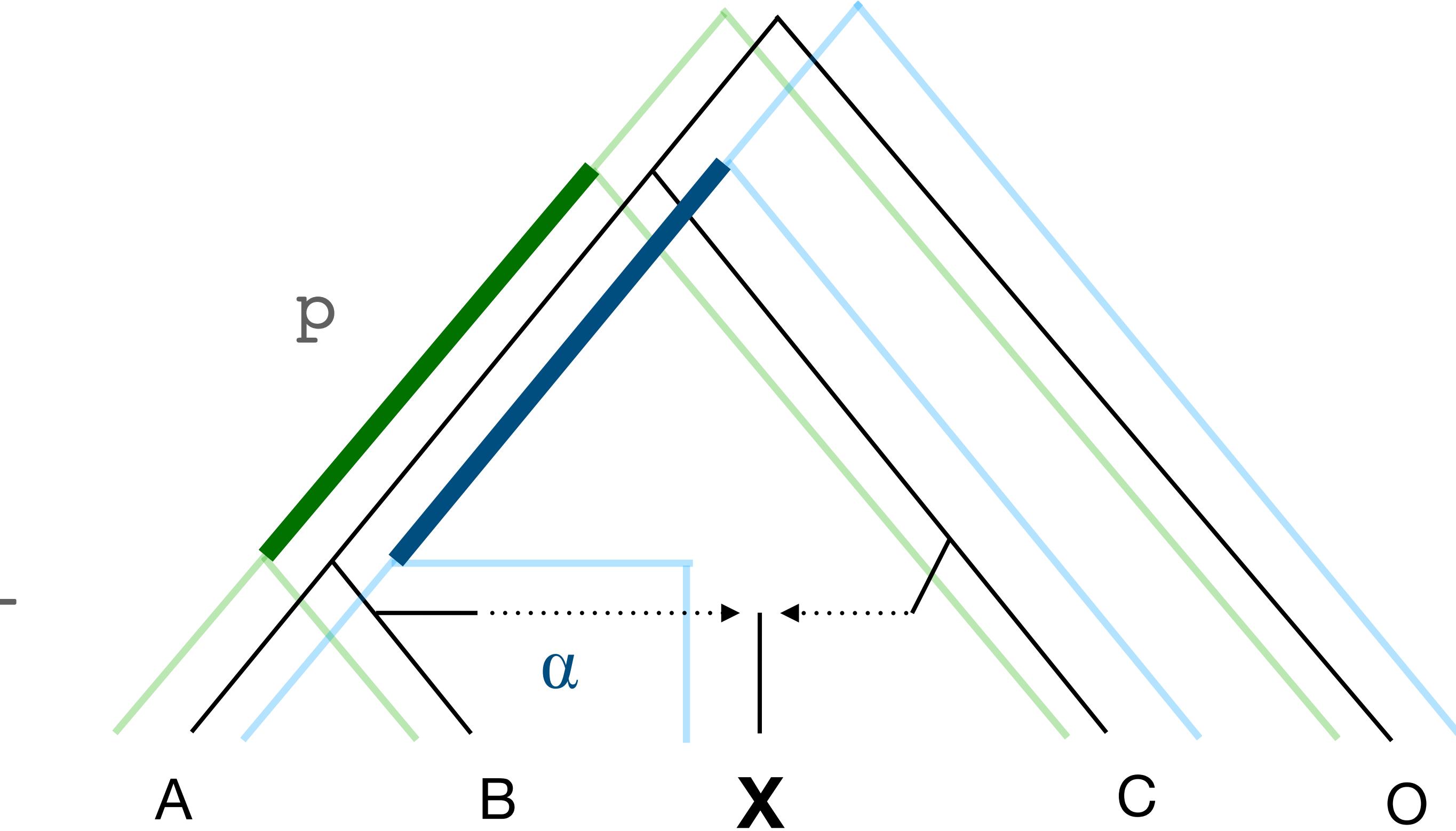
$$\alpha p = f_4(C, X; A, O)$$

$$\frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$



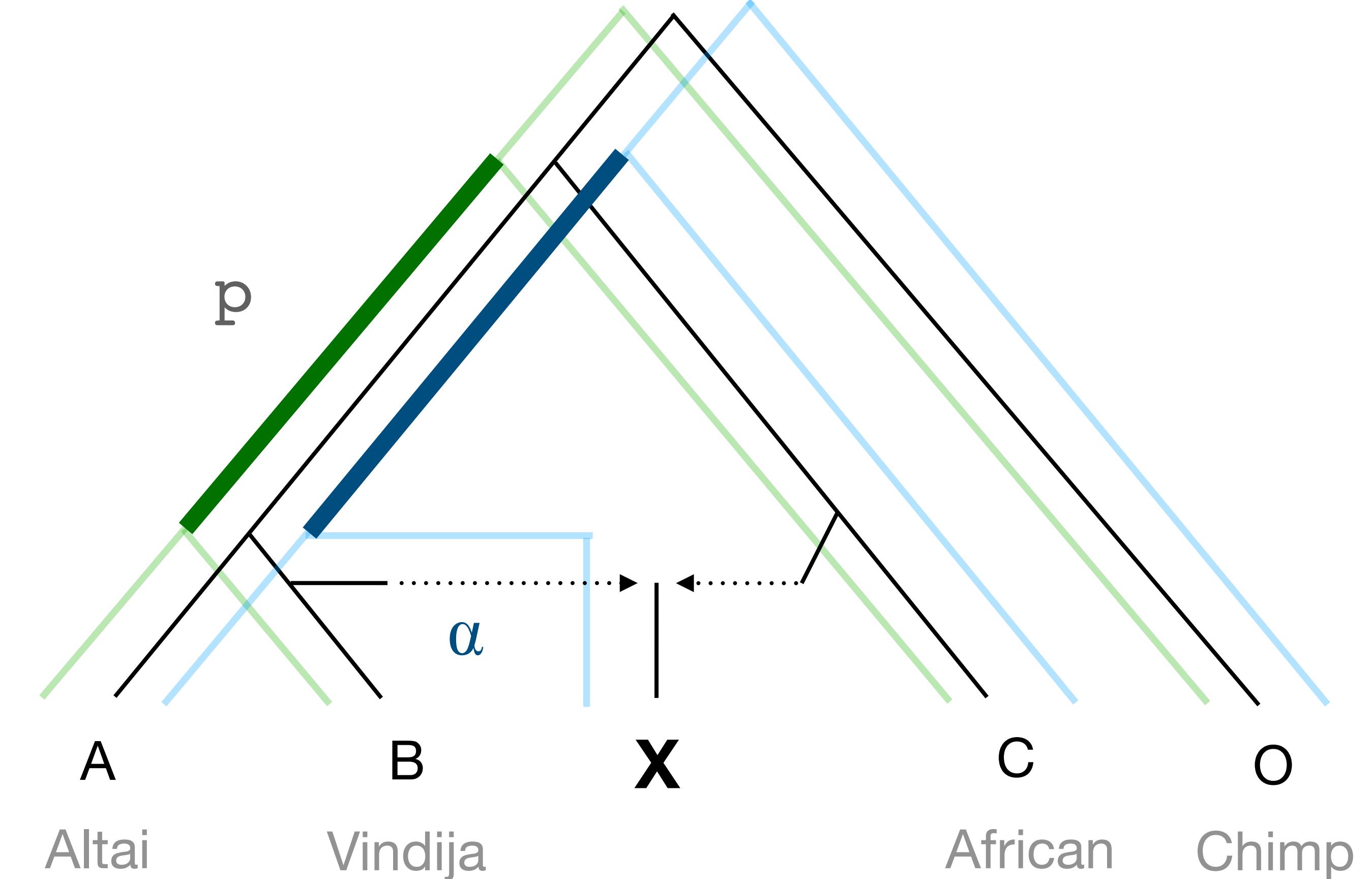
# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$
$$\alpha p = f_4(C, X; A, O)$$
$$\alpha = \frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$



# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$
$$\alpha p = f_4(C, X; A, O)$$
$$\alpha = \frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$



# $f_4$ -ratio statistic: two possible paths for an admixed sample

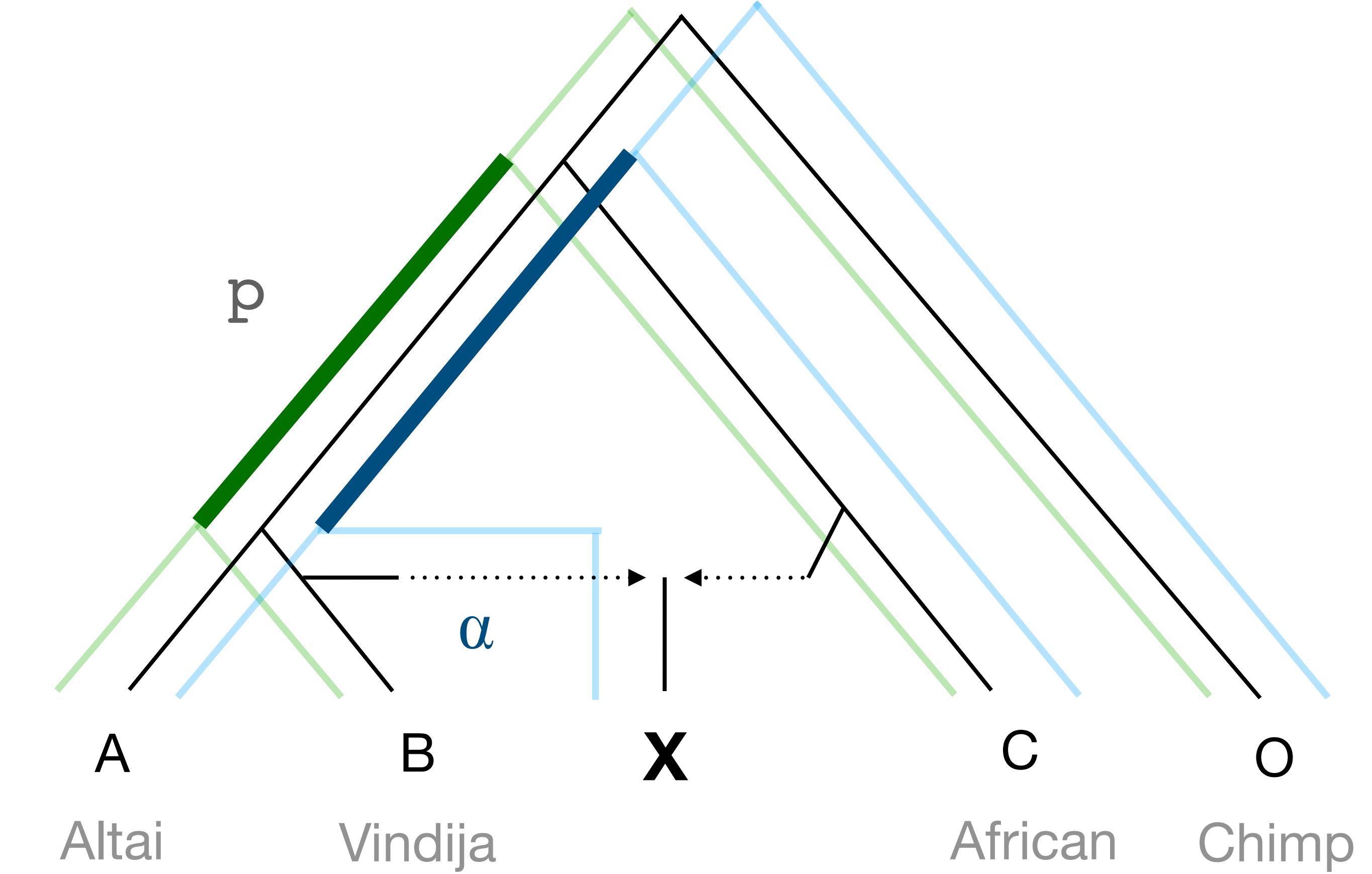
$$p = f_4(C, B; A, O)$$
$$\alpha p = f_4(C, X; A, O)$$
$$\alpha = \frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$

$f_4(\text{African}, X; \text{Altai}, \text{Chimp})$

(C)

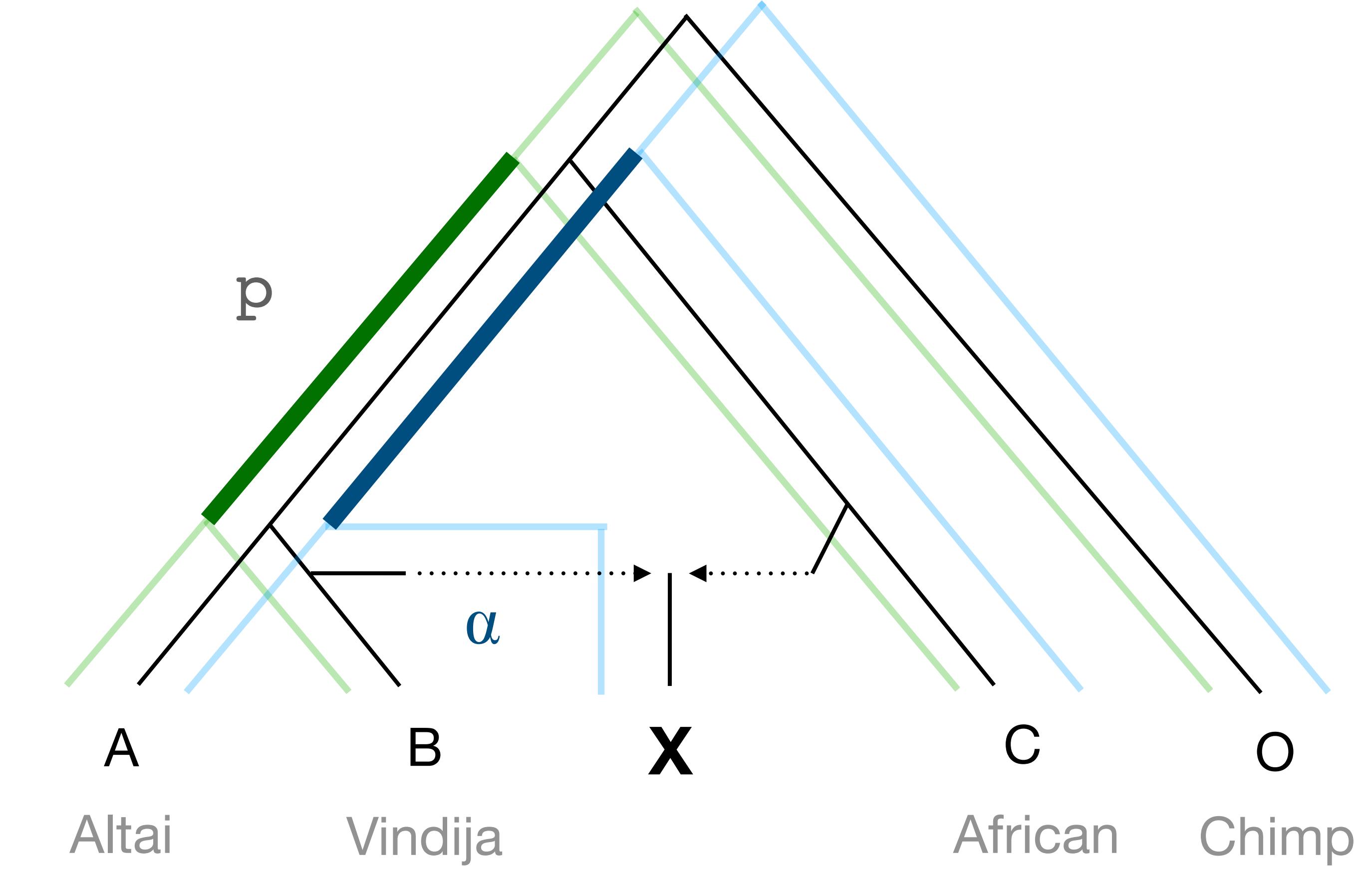
(A)

(O)



# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$
$$\alpha p = f_4(C, X; A, O)$$
$$\alpha = \frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$



$$f_4(\text{African}, X; \text{Altai}, \text{Chimp}) / f_4(\text{African}, \text{Vindija}; \text{Altai}, \text{Chimp})$$

(C)

(A)

(O)

(C)

(B)

(A)

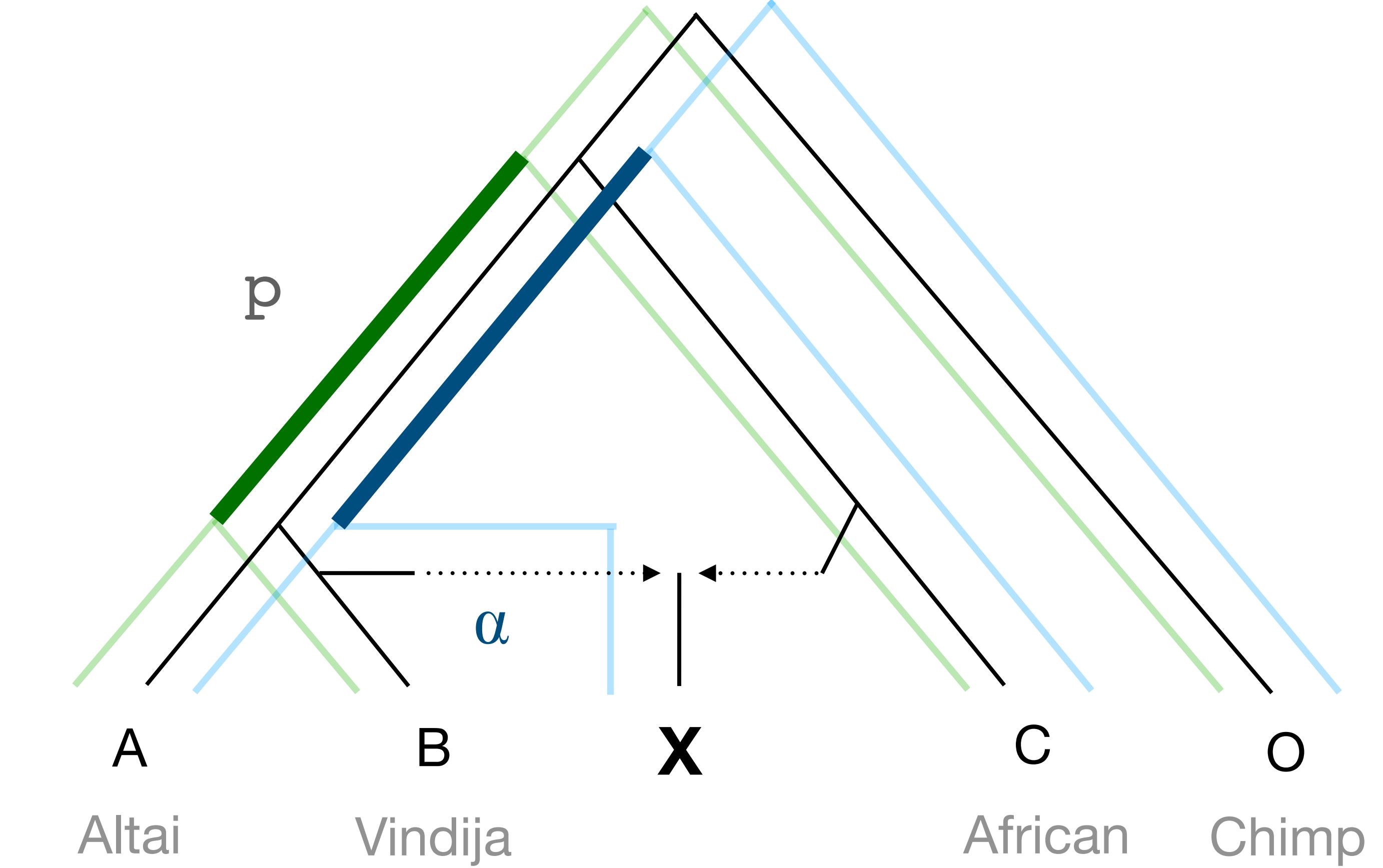
(O)

# $f_4$ -ratio statistic: two possible paths for an admixed sample

$$p = f_4(C, B; A, O)$$

$$\alpha p = f_4(C, X; A, O)$$

$$\alpha = \frac{\alpha p}{p} = \frac{f_4(C, X; A, O)}{f_4(C, B; A, O)}$$



$$\alpha = f_4(\text{African}, X; \text{Altai}, \text{Chimp}) / f_4(\text{African}, \text{Vindija}; \text{Altai}, \text{Chimp})$$

(C)

(A)

(O)

(C)

(B)

(A)

(O)

$f_4(\text{African, X; Altai, Chimp}) / f_4(\text{African, Vindija; Altai, Chimp})$

In the BABA vs ABBA context, we are comparing  
the rate of allele **sharing of X with one Neanderthal**,  
with  
the rate of **sharing between two Neanderthals**

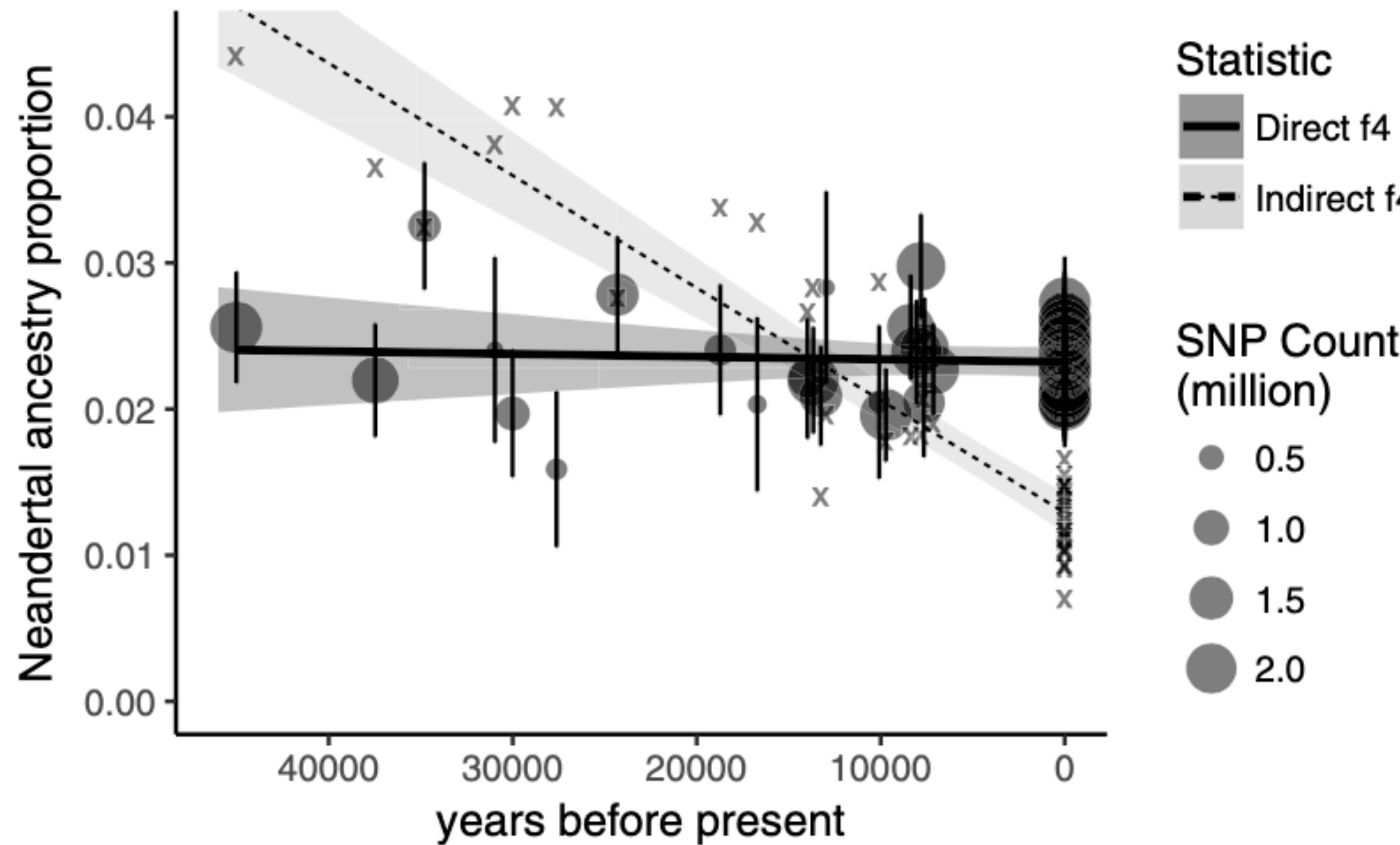
$$f_4(\text{African, X; Altai, Chimp}) / f_4(\text{African, Vindija; Altai, Chimp})$$



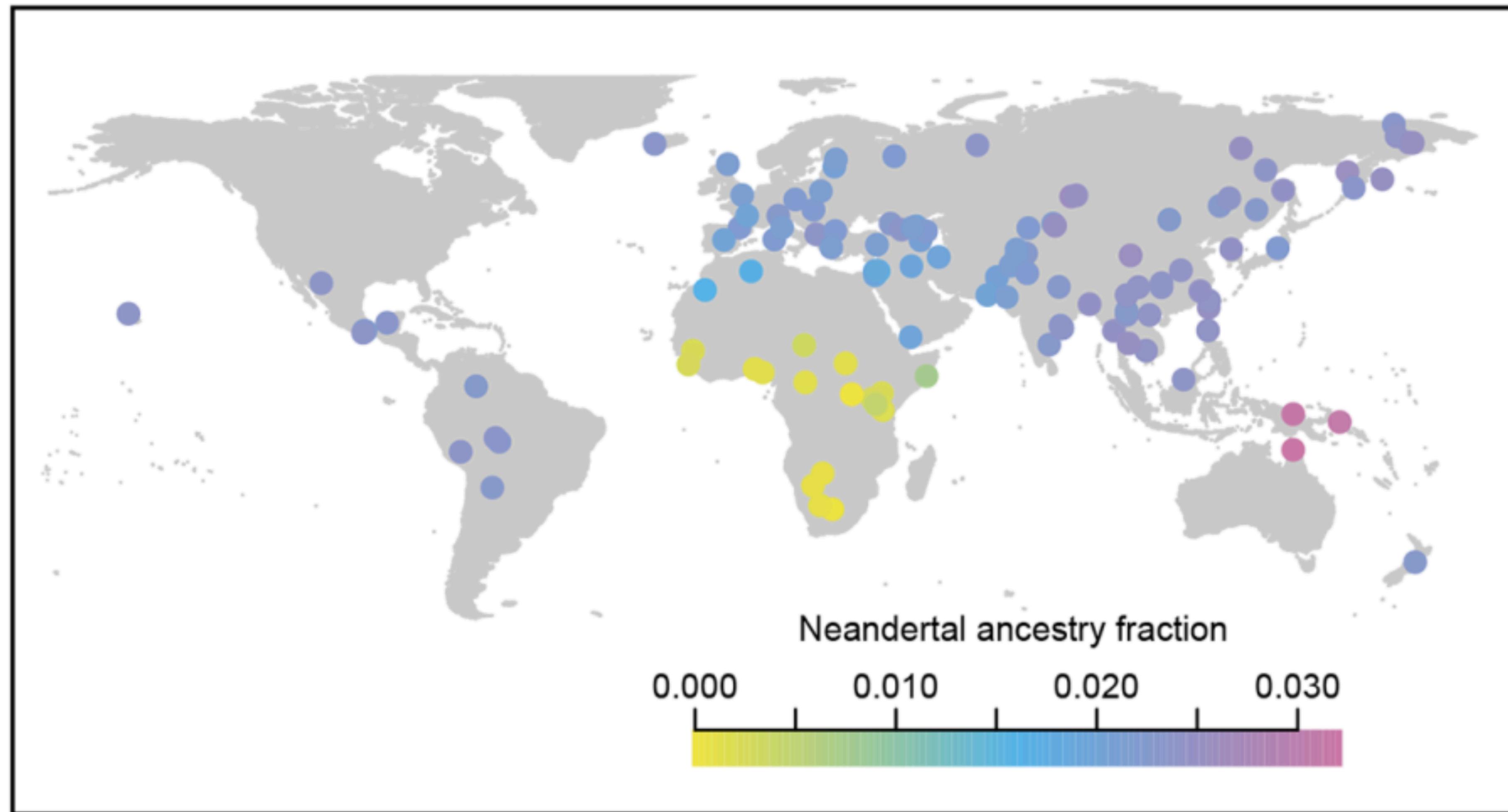
# Demo #3

Estimating Neanderthal ancestry  
proportions using *admixr*

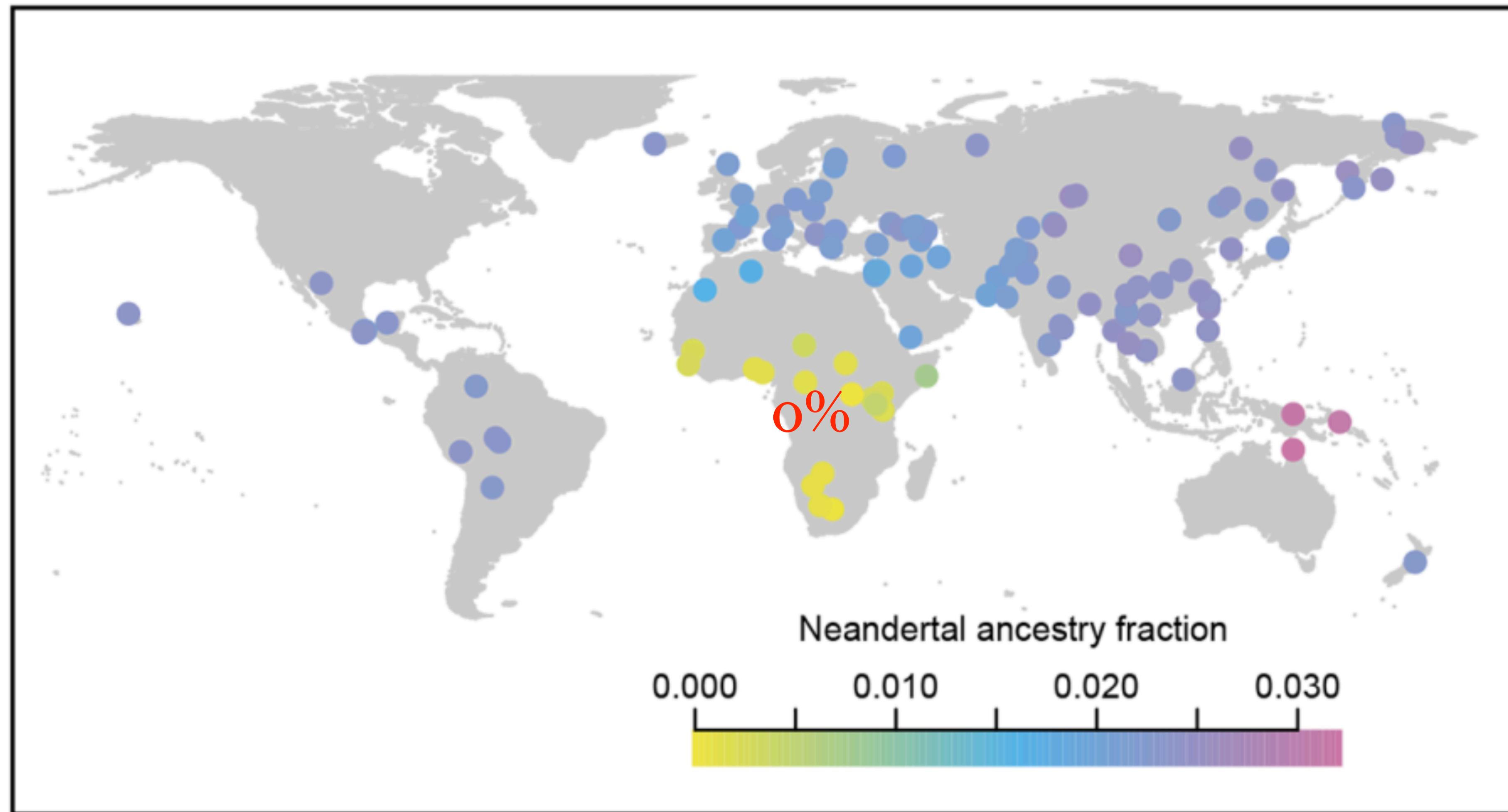
# Neanderthal ancestry in Europe over time



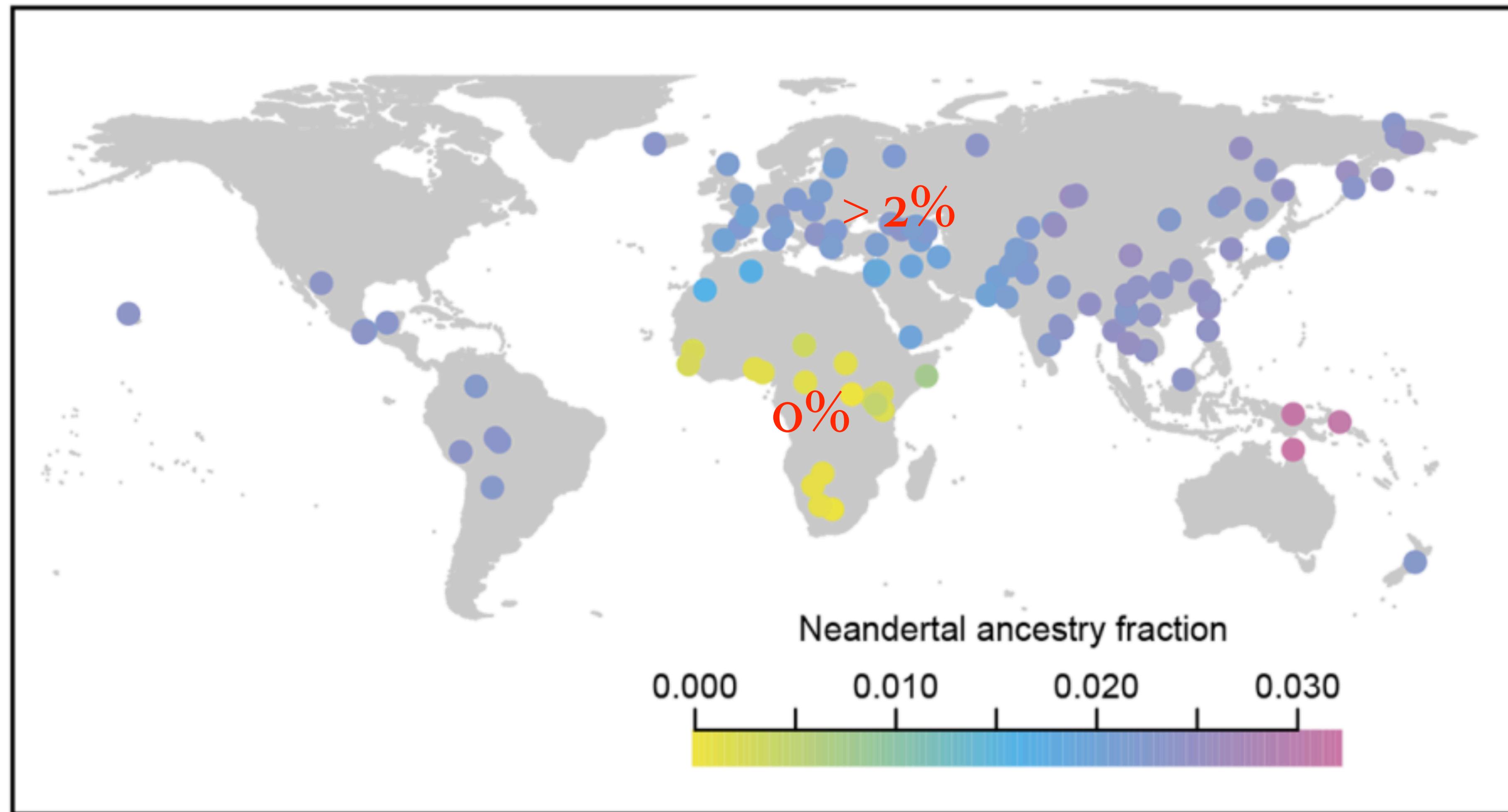
# Geographic distribution of Neanderthal ancestry in the world



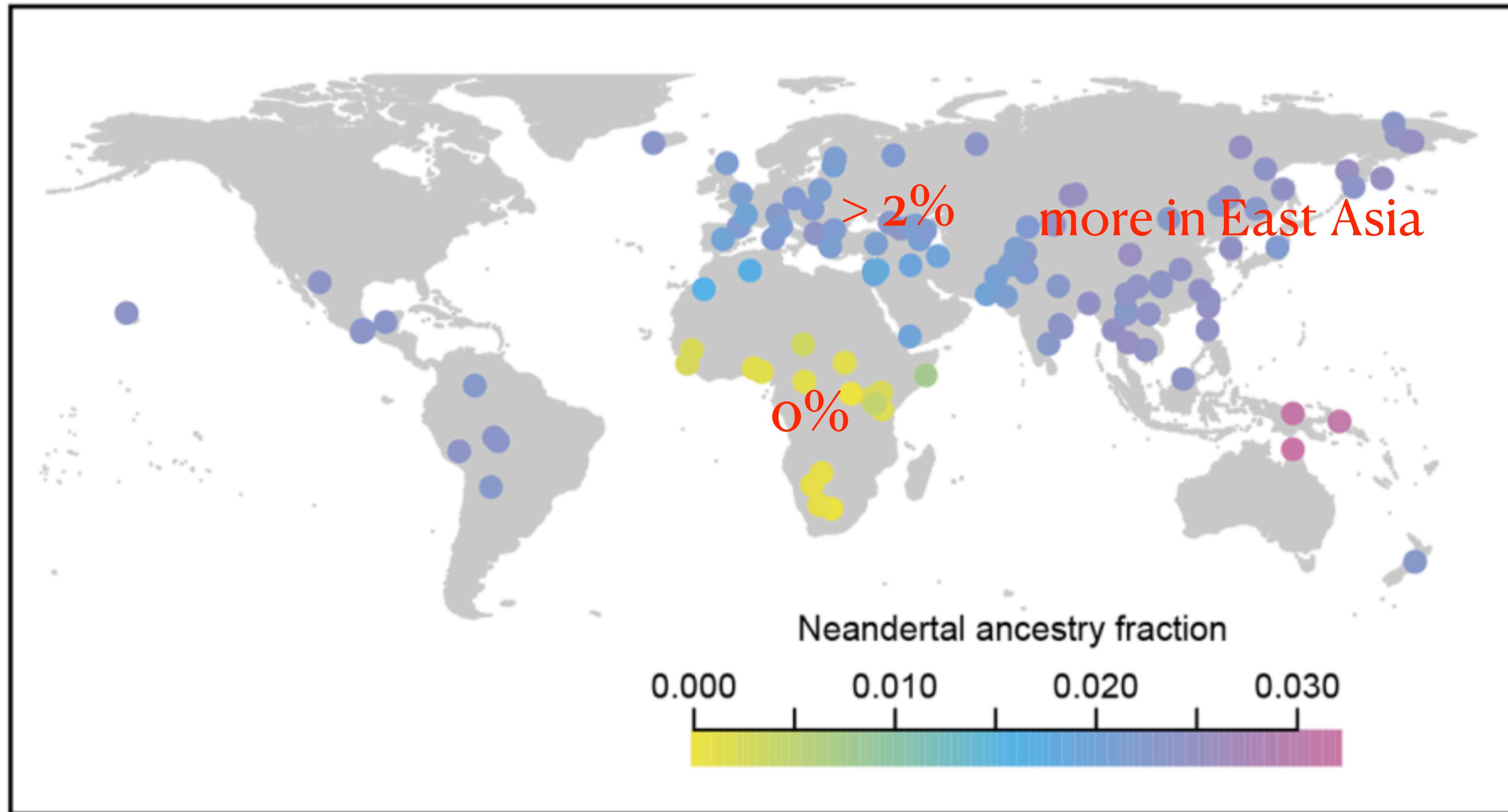
# Geographic distribution of Neanderthal ancestry in the world



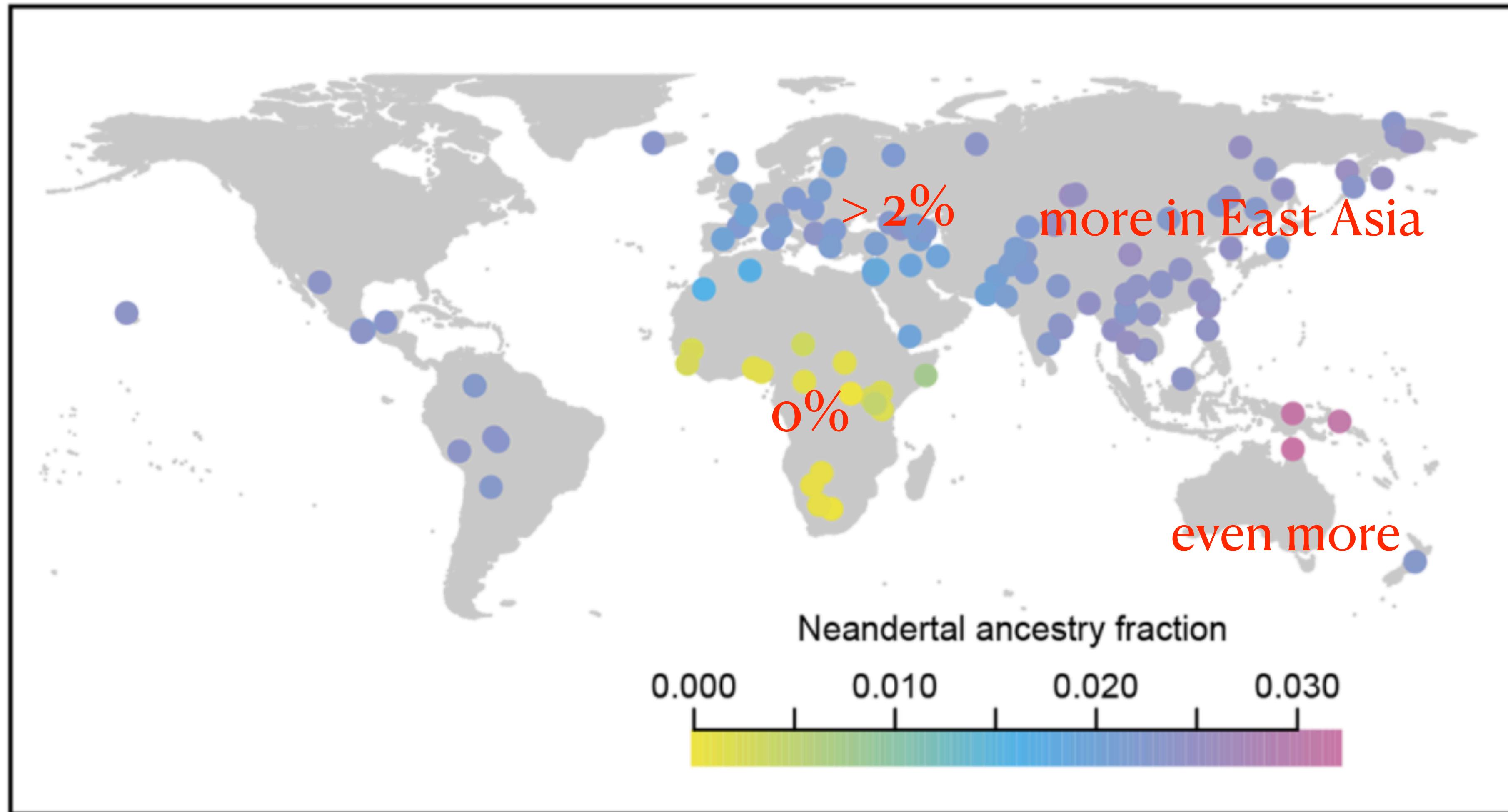
# Geographic distribution of Neanderthal ancestry in the world



# Geographic distribution of Neanderthal ancestry in the world



# Geographic distribution of Neanderthal ancestry in the world



# Bonus demo #4

Using ADMIXTOOLS  
the hardcore way

# Exercises?

1. Try changing the order of samples in a simple  $f_4$  or  $D$  calculation and convince yourselves that many orderings return the same result, or differ up to a sign. Try to guess which ones will before running the code.
2. Run the "outgroup"  $f_3$  statistic on the data we used in this lecture. You can use the example code from the "official" *admixr* tutorial [1].
3. What happens when you change the African in the  $f_4$ -ratio calculation (use Africans listed in the data/samples.tsv table and replace the Dinka).
  - Do you see the same amount of Neanderthal ancestry in, say, French individual even with different Africans? Compare what you found with the results shown in Fig. S4 of [2].
4. Try to reproduce Fig. S2A or Fig. S6 from [2] using  $f_4$  or  $D$  statistic.

[1] <https://bodkan.net/admixr/articles/tutorial.html#f3-statistic-1>

[2] Petr *et al.* (PNAS 2019)

# References

**Ancient Admixture in Human History**, Patterson N. *et al.* (GENETICS 2012)

**Admixture, Population Structure, and F-Statistics**, Peter B. (GENETICS 2016)

**Applying f<sub>4</sub>-statistics and admixture graphs**, Lipson M. (Mol Ecol Resour. 2020)

***admixr*—R package for reproducible analyses using ADMIXTOOLS**, Petr *et al.* (Bioinformatics 2019)

& the tutorial at <https://www.bodkan.net/admixr>

**R for Data Science**, Wickham and Grolemund - <https://r4ds.had.co.nz/> (O'Reilly 2017)

# Slides, R Markdown notebooks and data

<https://github.com/bodkan/admixr-workshop>

slides: slides . {key, pdf}

demo notebook: demo . Rmd