

ACTIVITY OUTLINE

Start with ROH estimates activity

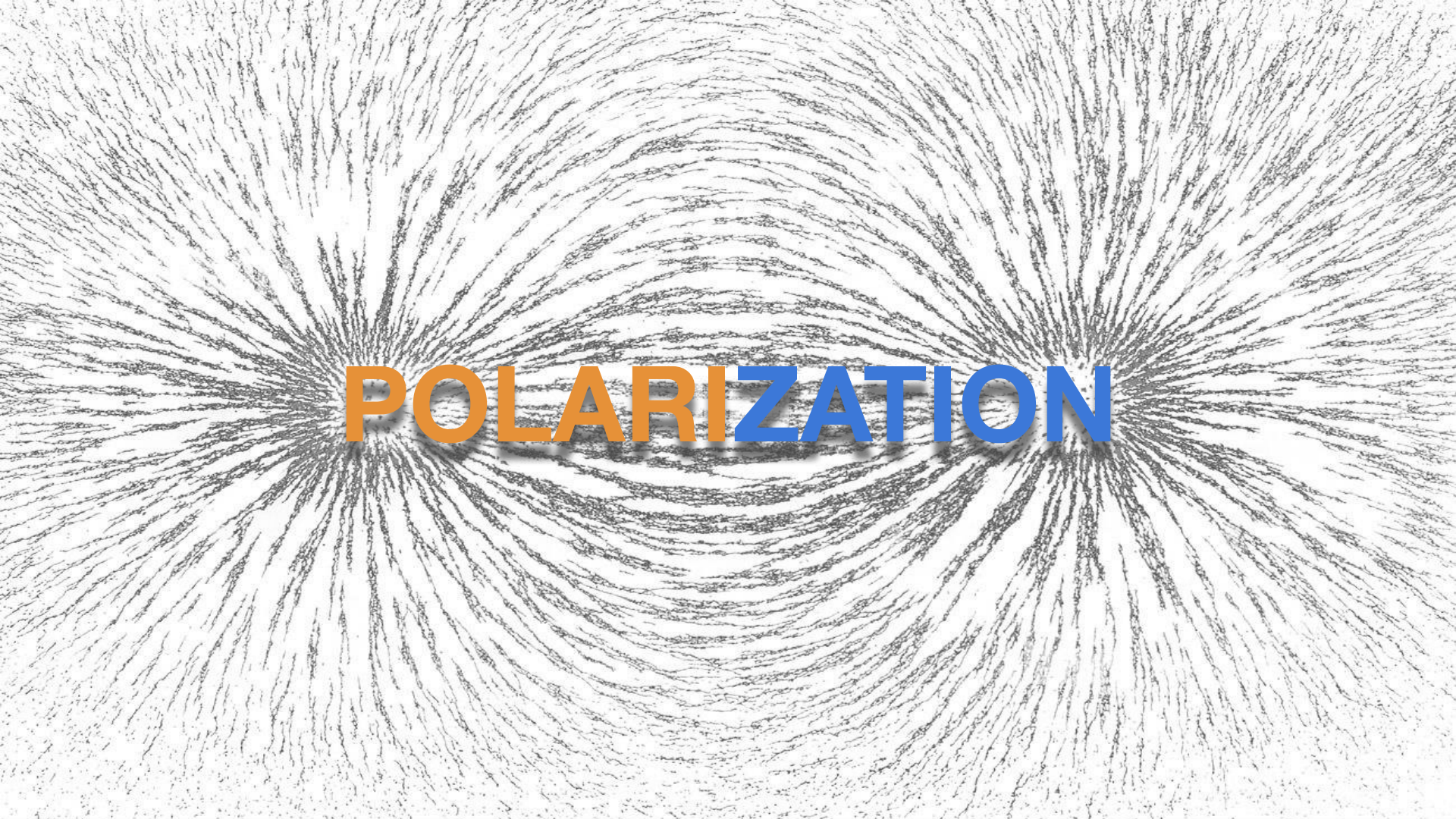
First wrap-up after ca. 45-50 minutes

Presentation of the polarization issue

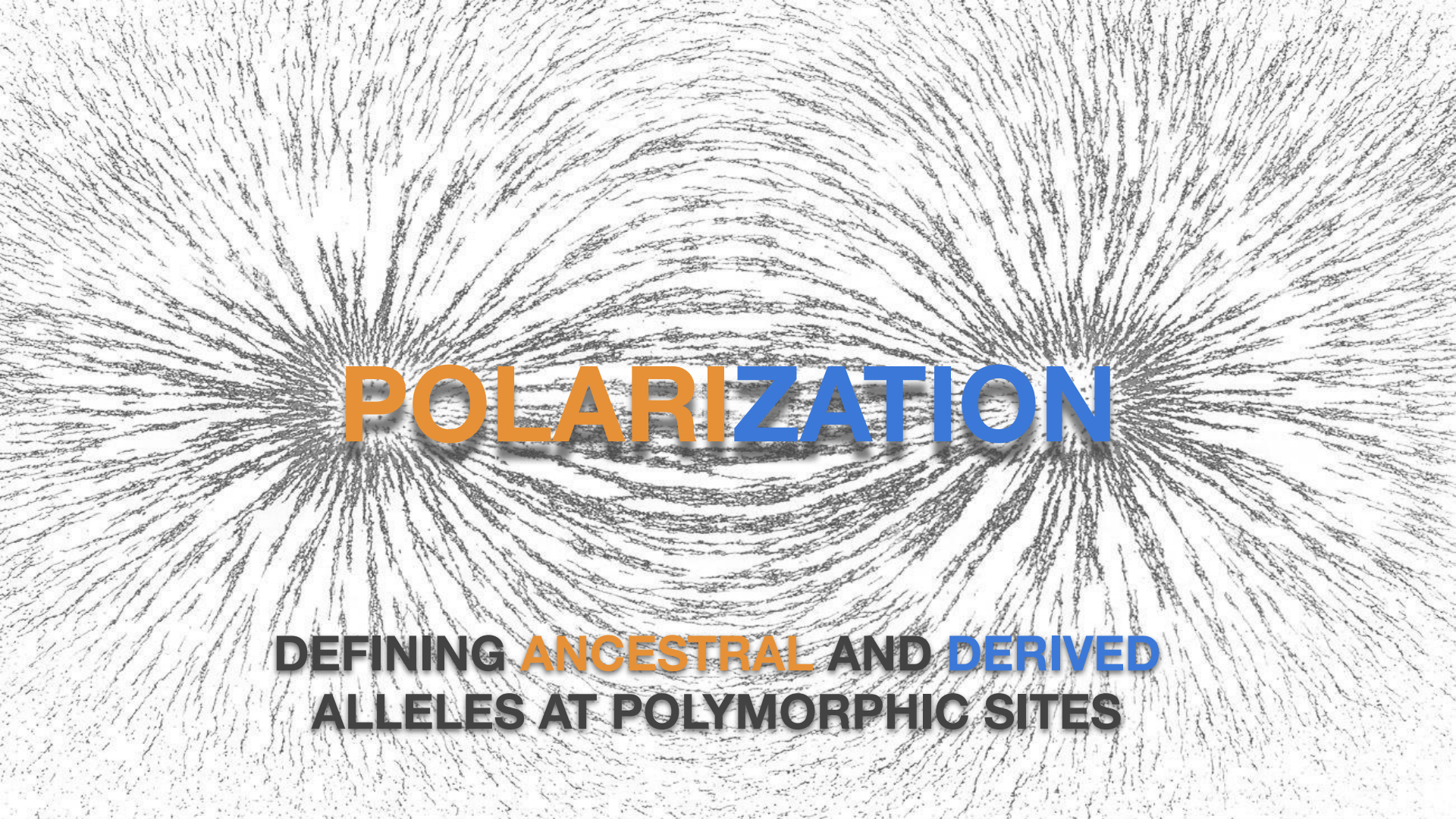
Start with GENOLOADER activity

Maybe long waiting times (10-15 mins) for calculation during this activity.
Take a break. If it runs too long we'll provide output files on the GitHub.

Second wrap-up after ca. 45-50 minutes

The background of the image is a complex, symmetrical pattern of fine, dark, wavy lines on a light gray background. These lines form a series of concentric, elongated loops that radiate from two central points, creating a visual effect similar to the interference pattern of polarized light or the magnetic field lines of a dipole magnet. The overall appearance is that of a microscopic or macroscopic view of a polarized medium.

POLARIZATION



POLARIZATION

**DEFINING ANCESTRAL AND DERIVED
ALLELES AT POLYMORPHIC SITES**

What do we need for polarization?

**TARGET
POPULATION**

What do we need for polarization?

Anything else?

**TARGET
POPULATION**

OUTGROUP

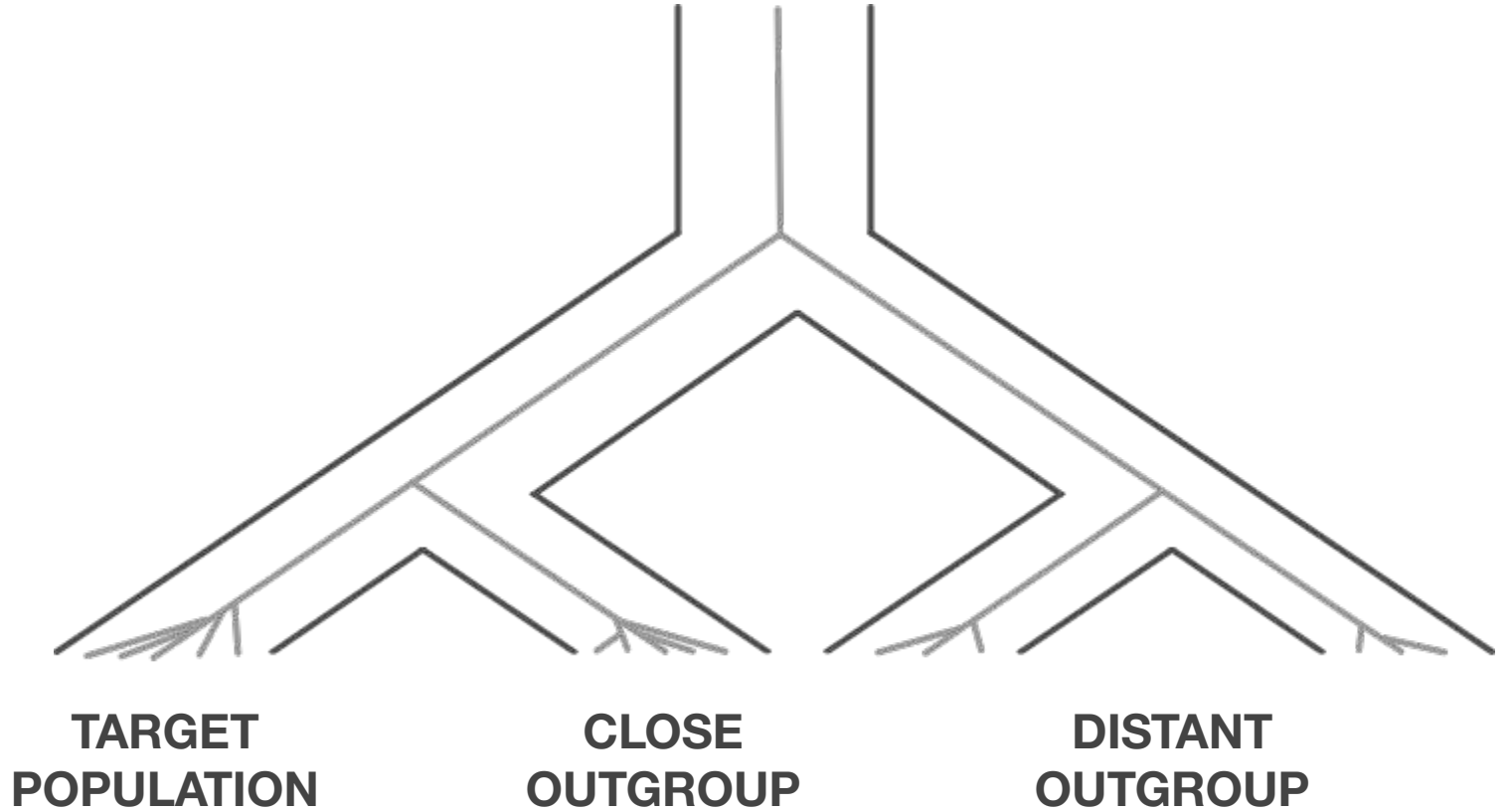
MORE OUTGROUPS!

**TARGET
POPULATION**

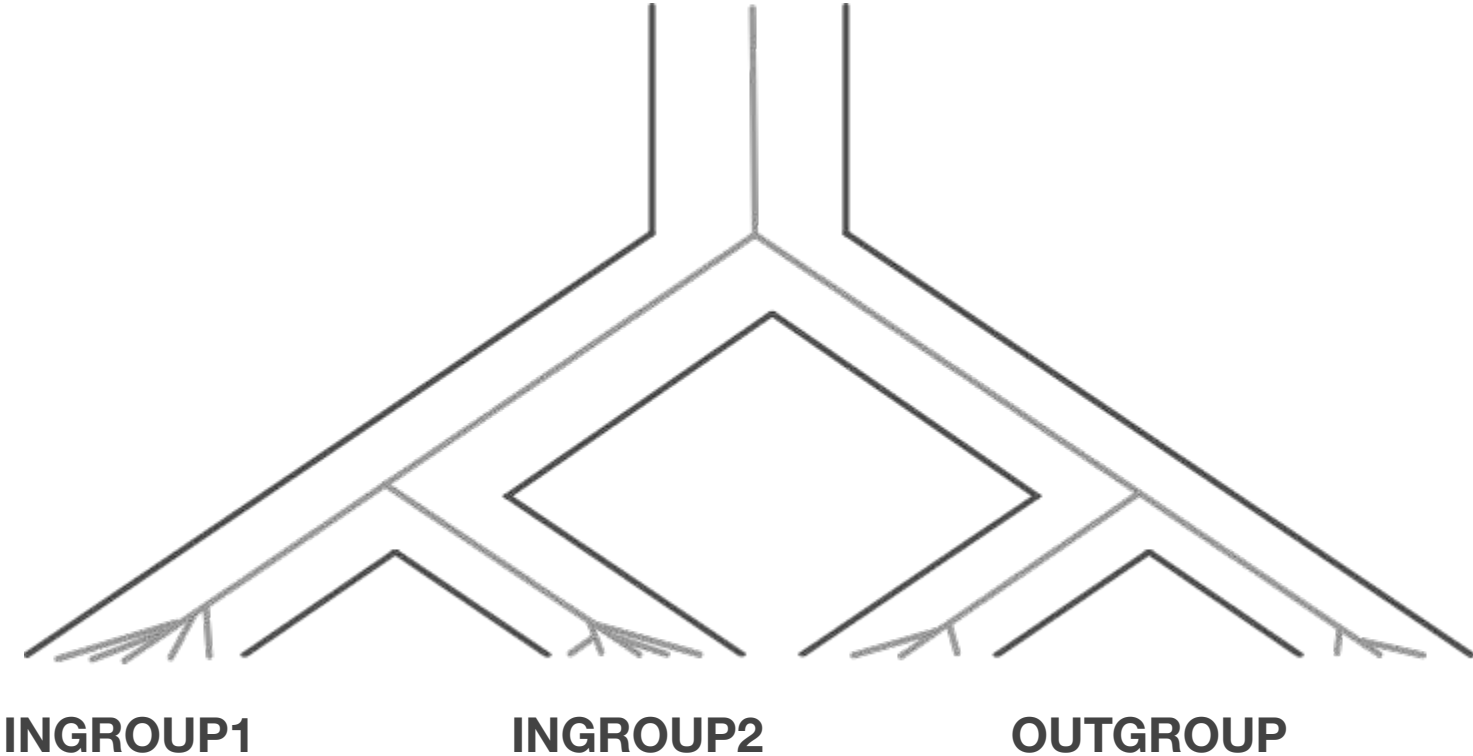
OUTGROUP

OUTGROUP

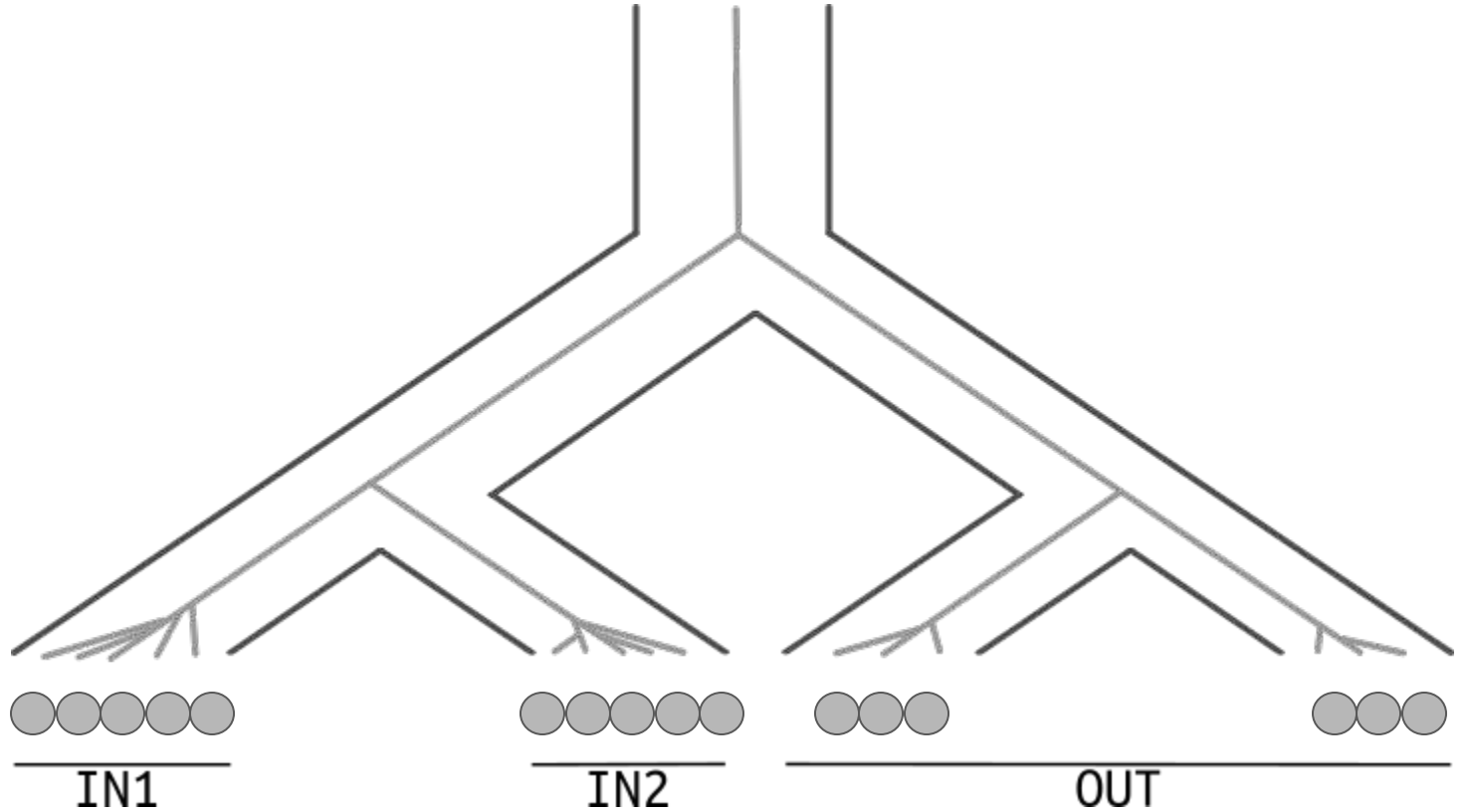
MORE OUTGROUPS!



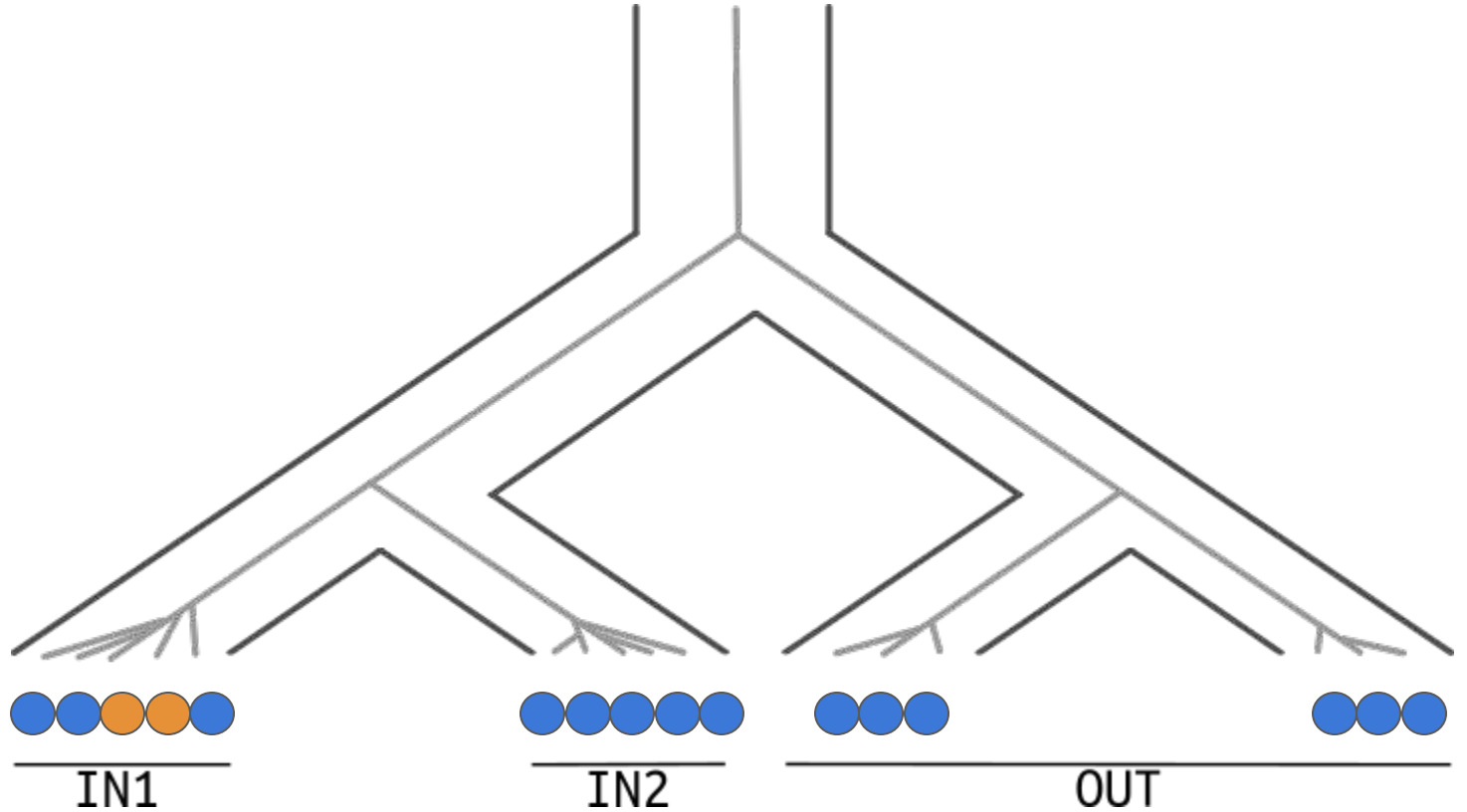
LABELS IN THIS ACTIVITY



SOME EXAMPLES OF ALLELES CONFIGURATION



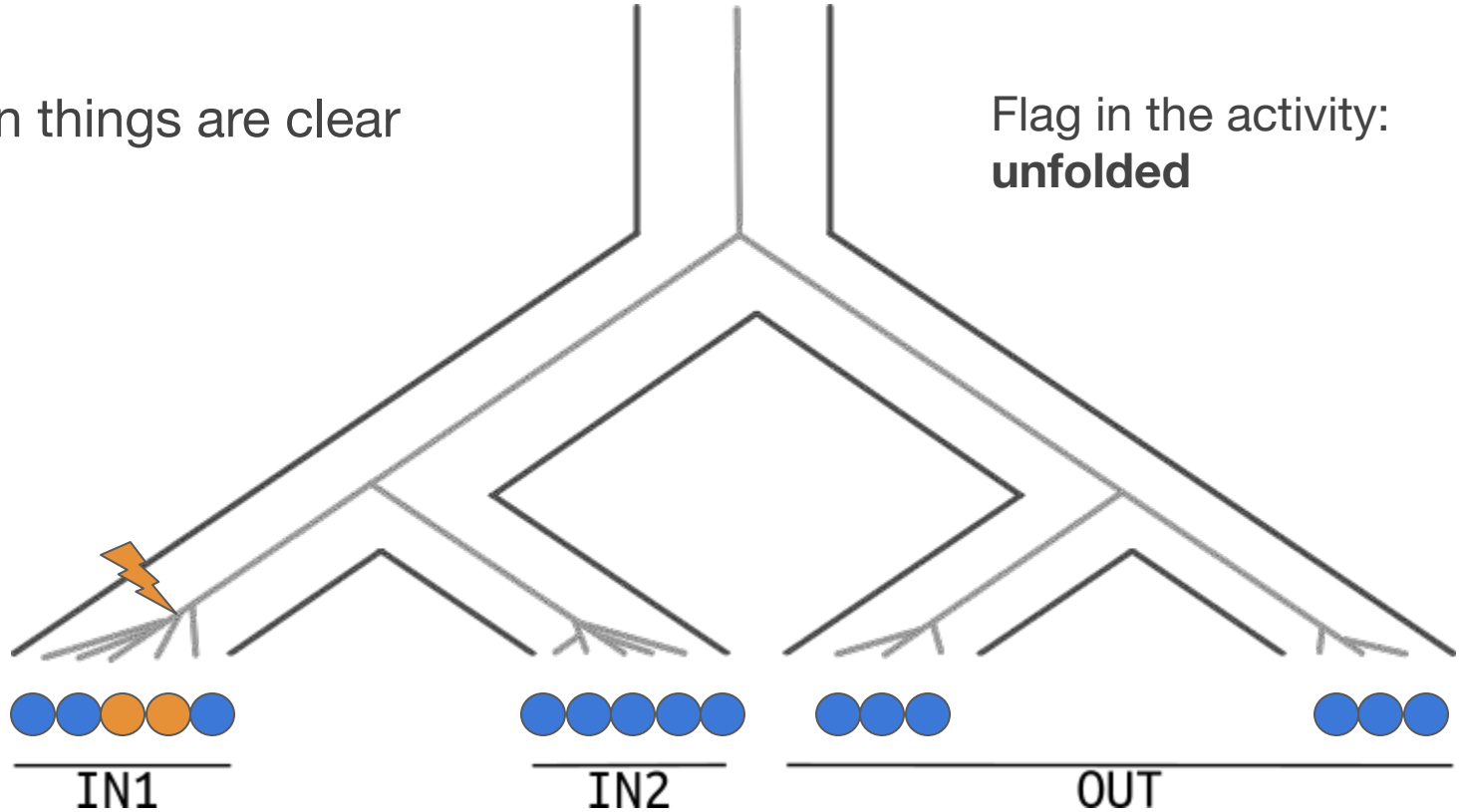
Which one is the derived allele?



Which one is the derived allele?

Easy when things are clear

Flag in the activity:
unfolded



GENOLOADER

```
def get_ancestral_allele(geno_out, geno1, geno2):

    out_geno = [int(i) for i in geno_out if i != '.']
    p1_geno = [int(i) for i in geno1 if i != '.']
    p2_geno = [int(i) for i in geno2 if i != '.']

    p12_geno = p1_geno+p2_geno
    total_alleles = p1_geno+p2_geno+out_geno

    if len(set(out_geno)) == 0 and len(set(p1_geno+p2_geno)) == 0: #all missing
        ref_allele = 9
        flag = 'allMiss'

    elif len(set(p1_geno+p2_geno)) == 0: #ingroup pops missing
        ref_allele = int(stats.mode(out_geno)[0])
        flag = 'inMiss'

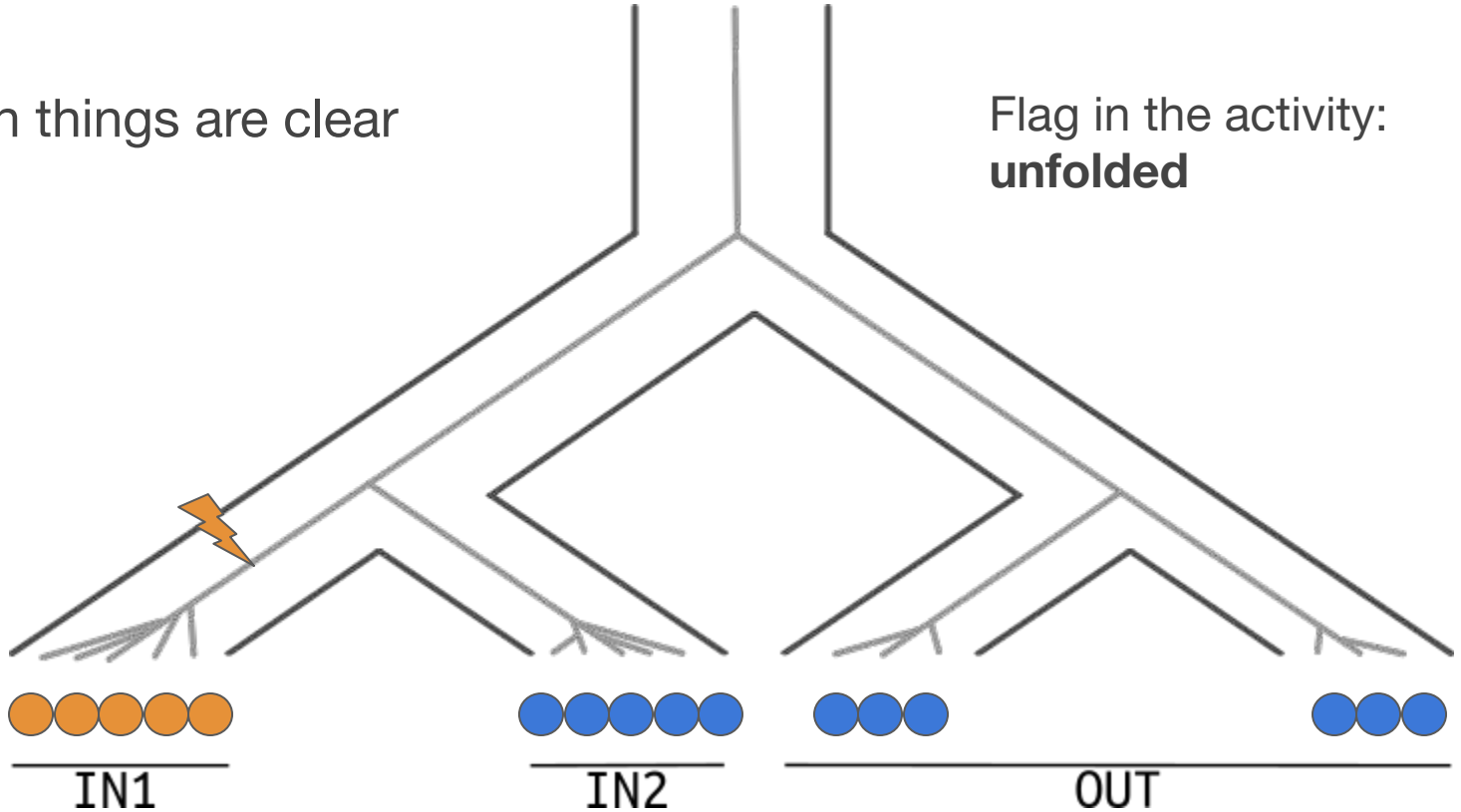
    elif len(set(out_geno)) == 0 and len(set(p1_geno+p2_geno)) > 0: #outgroup missing
        if len(set(p1_geno)) == 0:
            ref_allele = int(stats.mode(p2_geno)[0]) #pop1 missing
            flag = 'in2Fold'
        elif len(set(p2_geno)) == 0: #pop2 missing
            ref_allele = int(stats.mode(p1_geno)[0])
            flag = 'in1Fold'
        elif len(set(p1_geno)) == 1 and len(set(p2_geno)) == 1:
            ref_allele = p1_geno[0] #pop1 allele sets as reference => conservative
            flag = 'InFixOutMiss'
        elif len(set(p1_geno)) == 2 and len(set(p2_geno)) == 1:
            ref_allele = p2_geno[0]
            flag = 'unfoldOutMiss'
        elif len(set(p1_geno)) == 1 and len(set(p2_geno)) == 2:
            ref_allele = p1_geno[0]
            flag = 'unfoldOutMiss'
        elif len(set(p1_geno)) == 2 and len(set(p2_geno)) == 2:
            ref_allele = int(stats.mode(p1_geno+p2_geno)[0]) ### not random! Will always be 0 in case of 50/50
            flag = 'inFold'
```

Flags definition

Which one is the derived allele?

Easy when things are clear

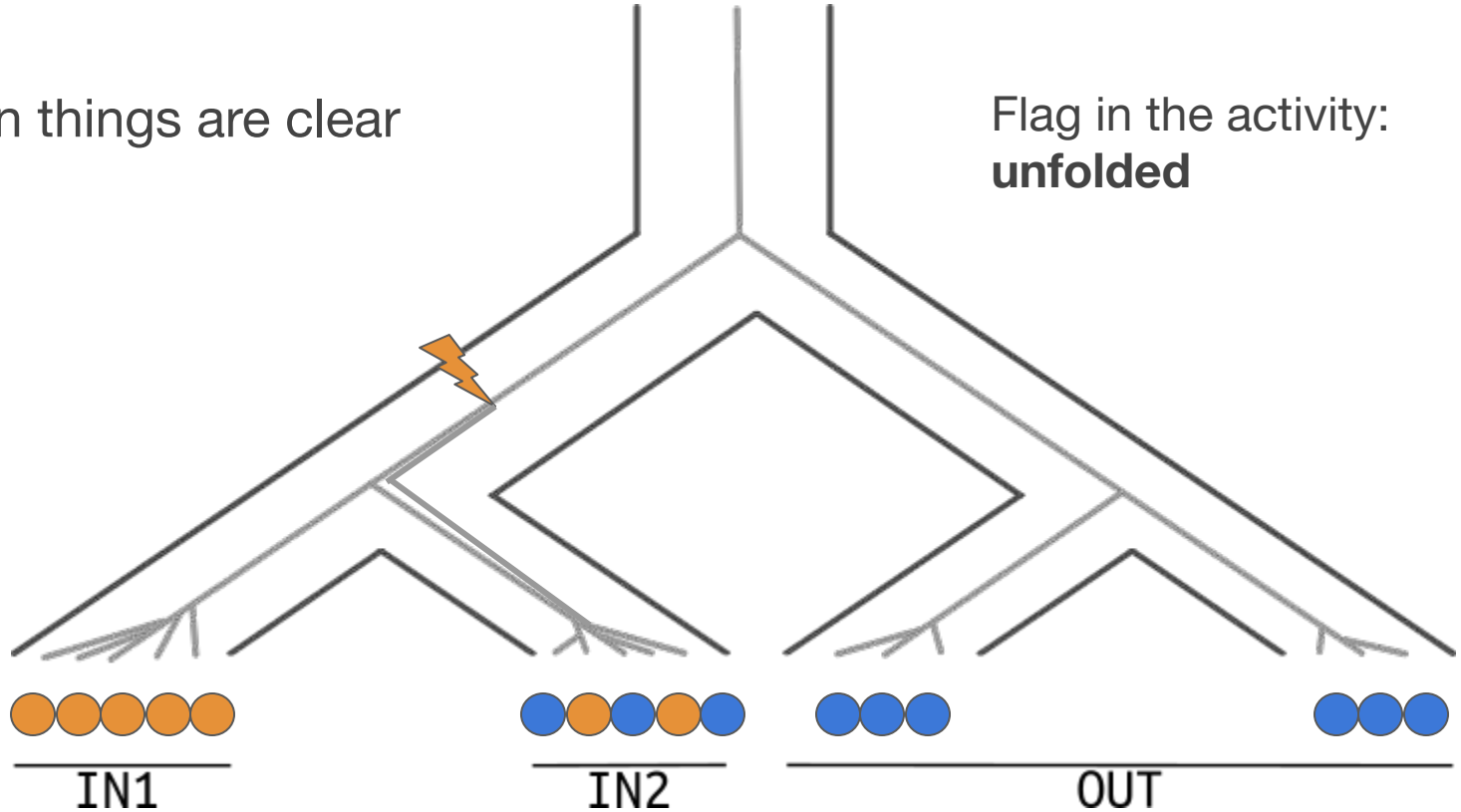
Flag in the activity:
unfolded



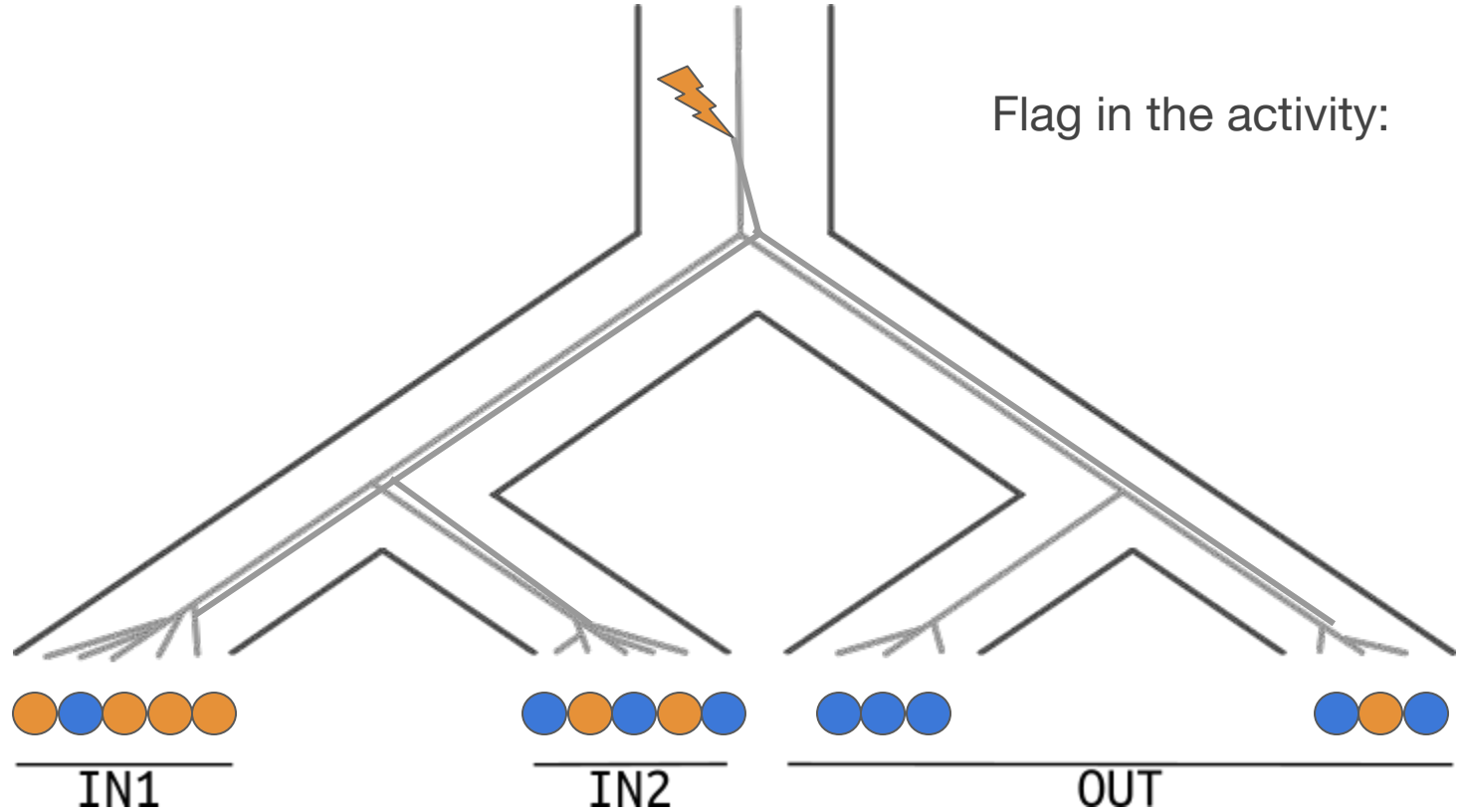
Which one is the derived allele?

Easy when things are clear

Flag in the activity:
unfolded



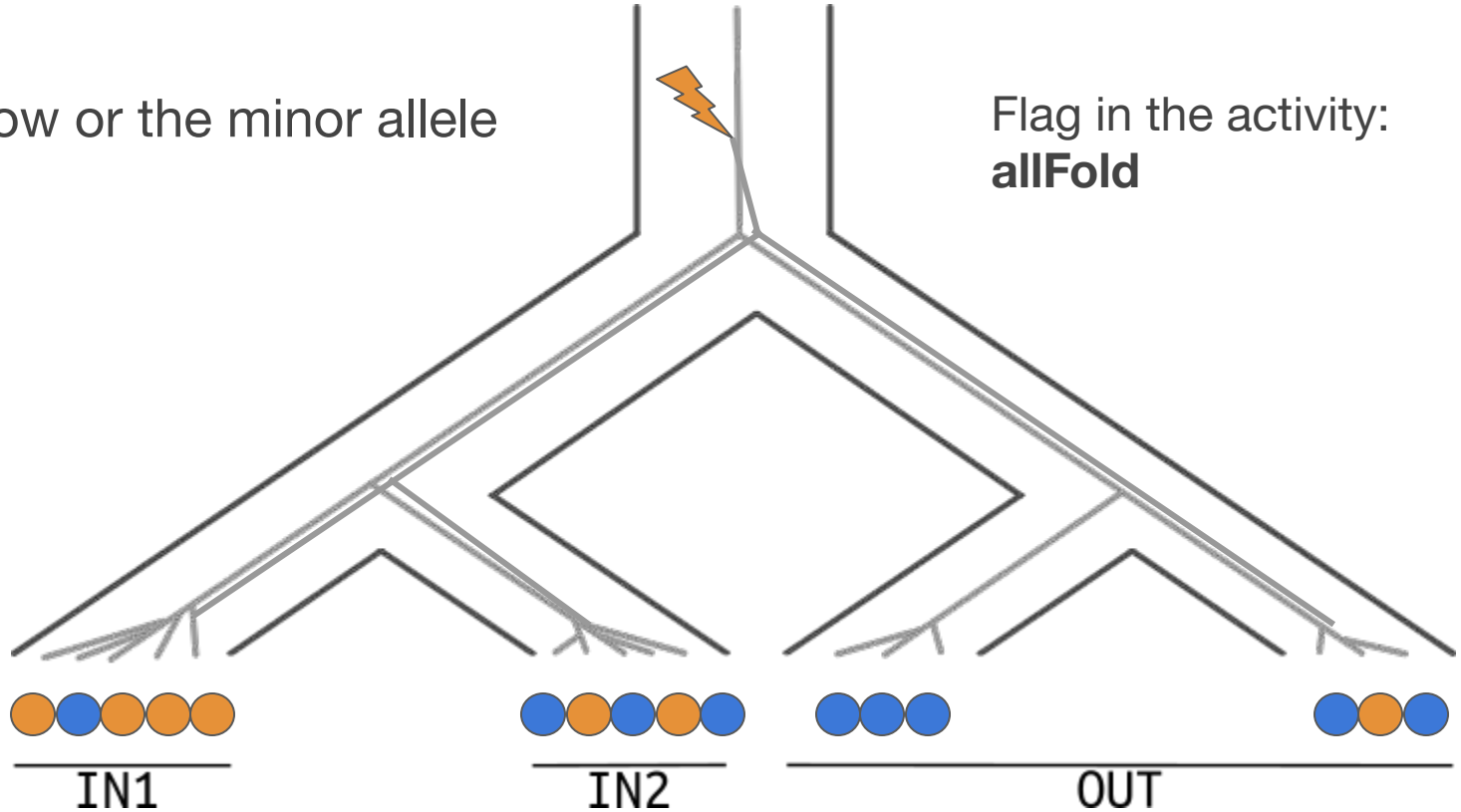
Which one is the derived allele?



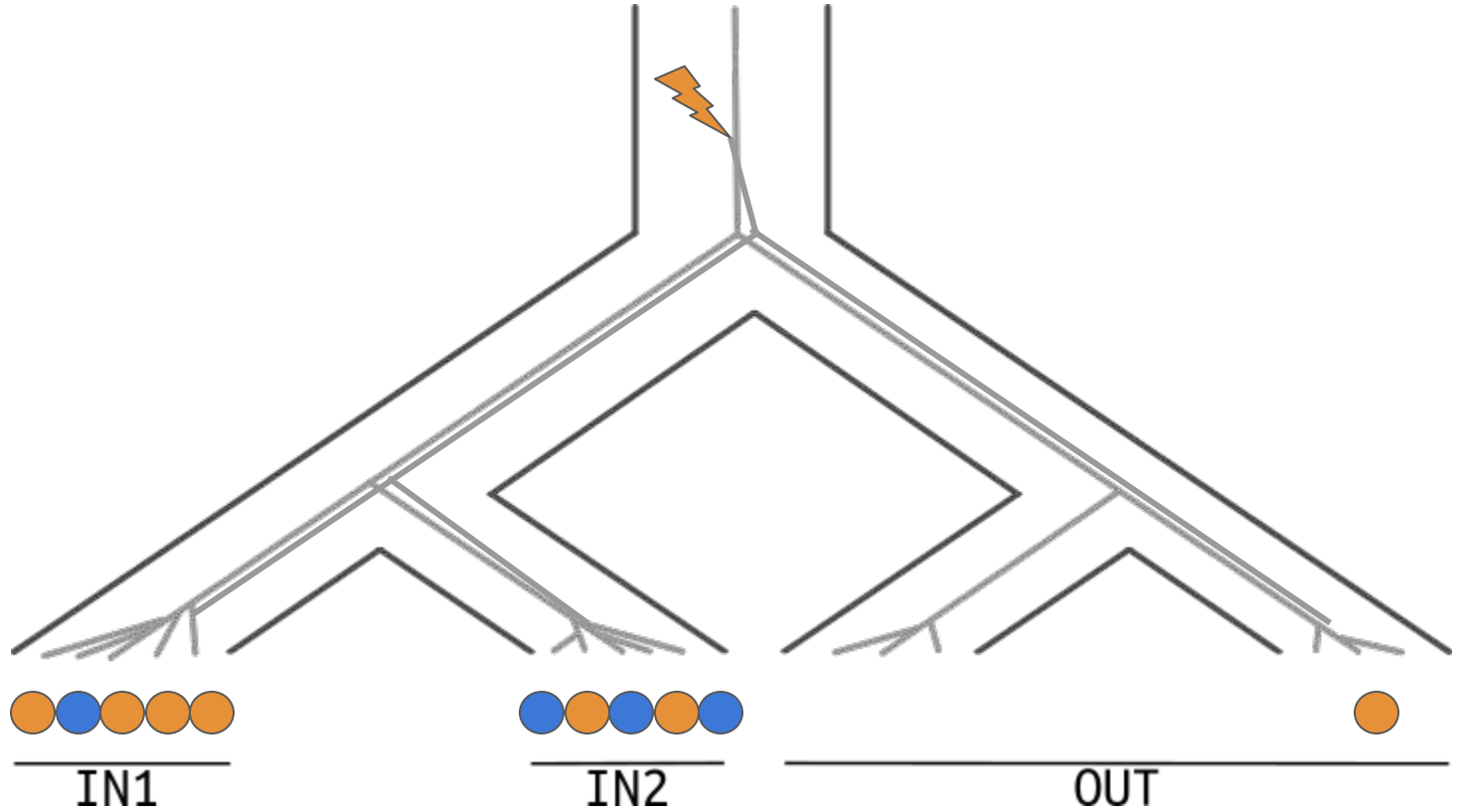
Which one is the derived allele?

I don't know or the minor allele

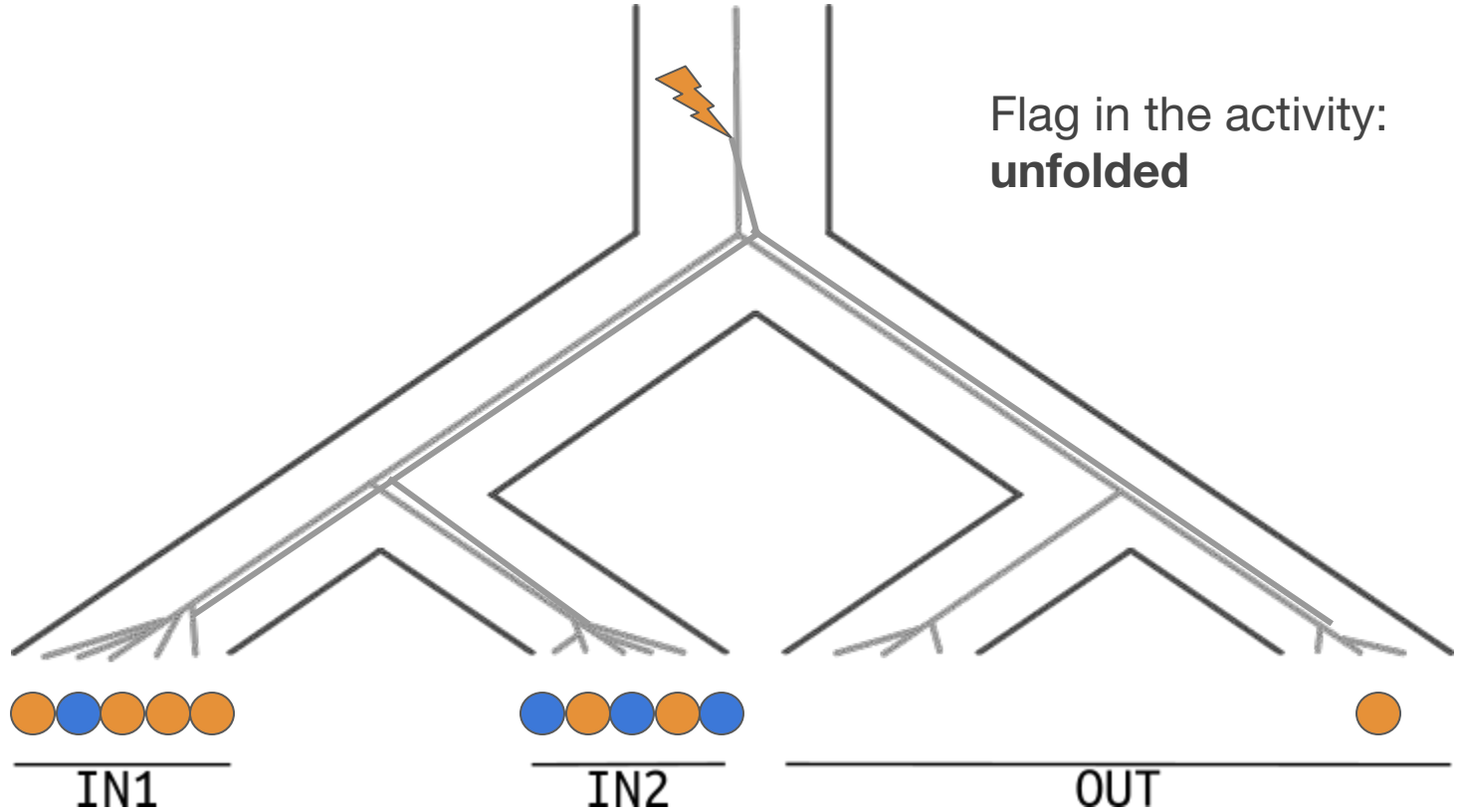
Flag in the activity:
allFold



Keep track of missing data



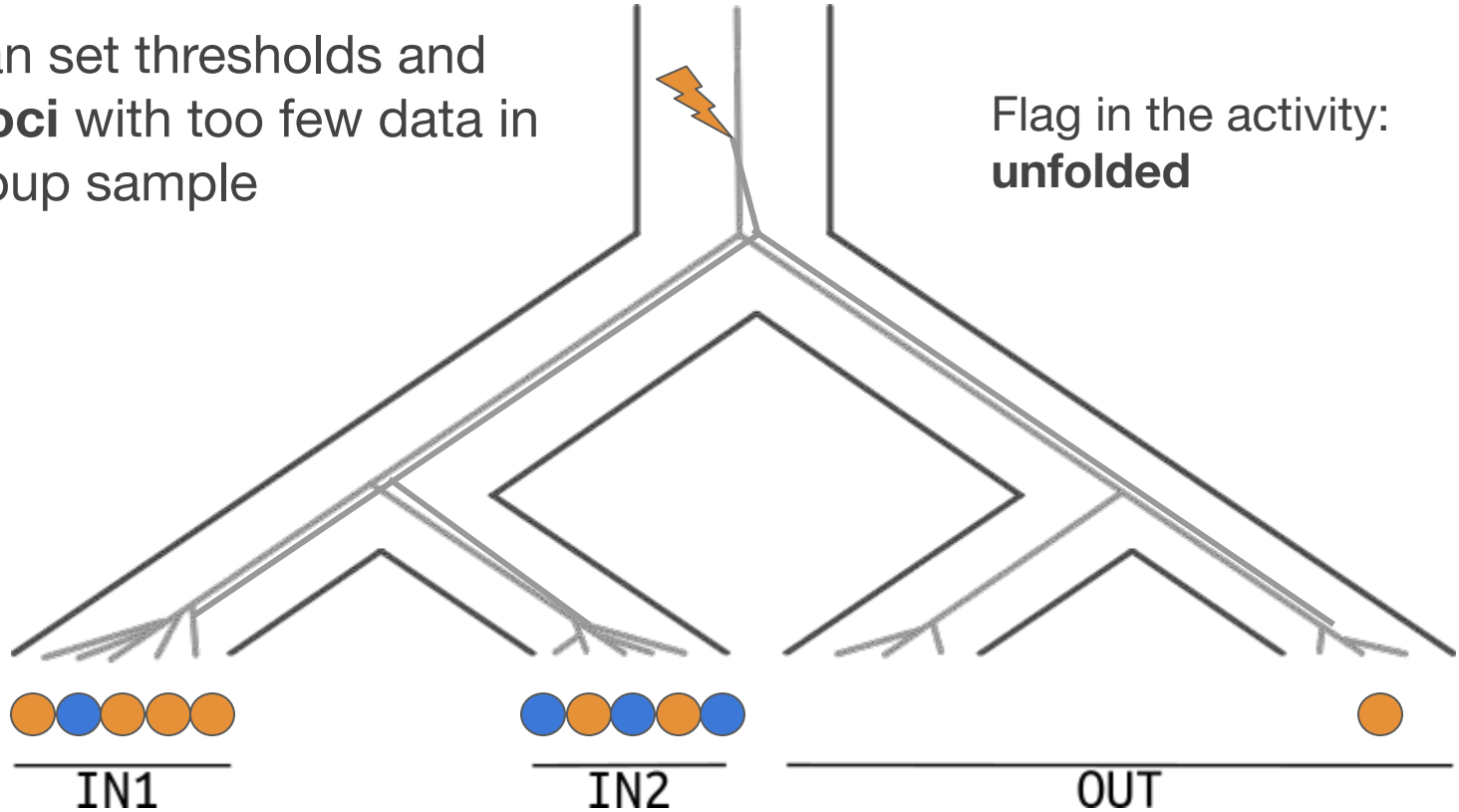
Keep track of missing data



Keep track of missing data

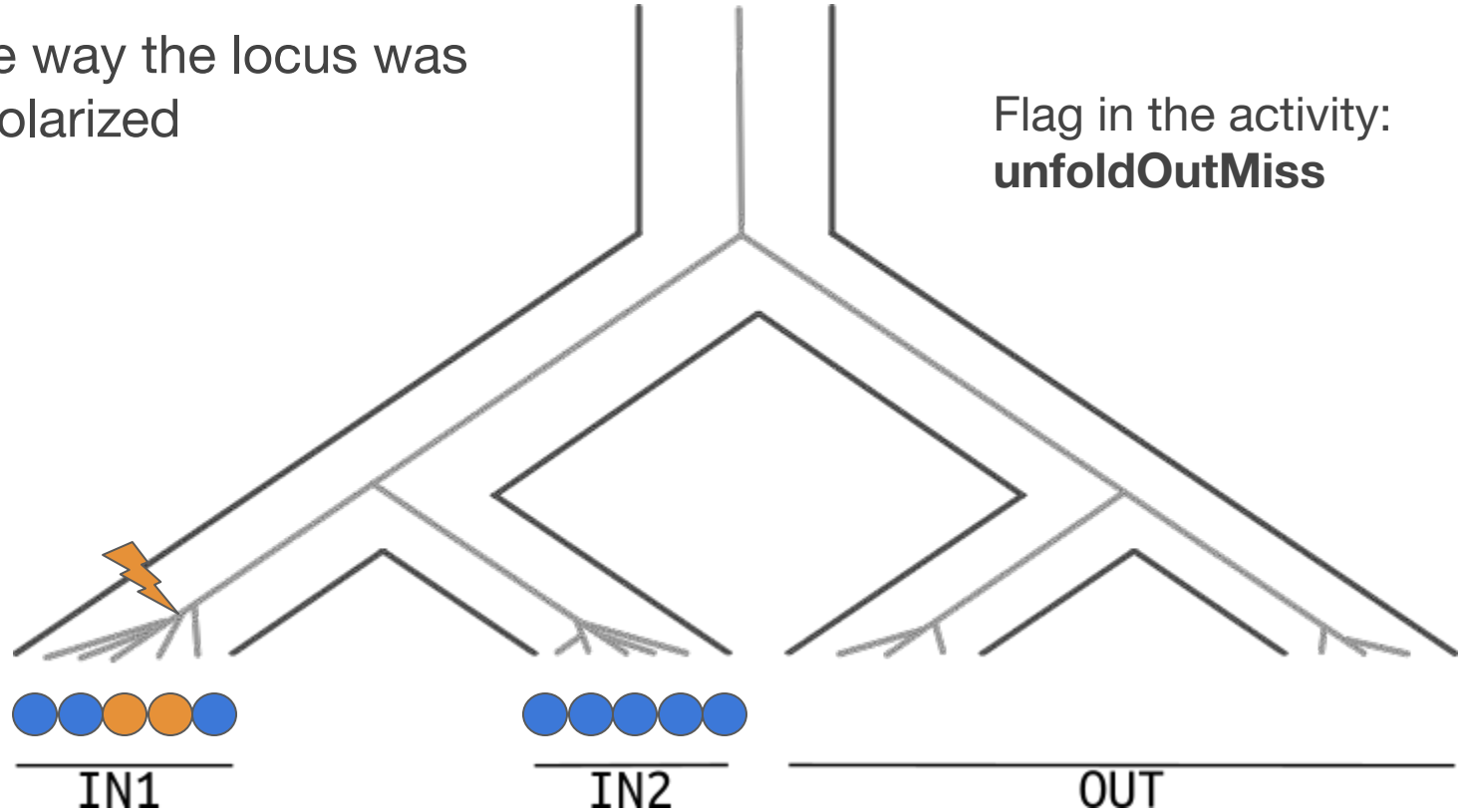
But we can set thresholds and **discard loci** with too few data in the outgroup sample

Flag in the activity:
unfolded



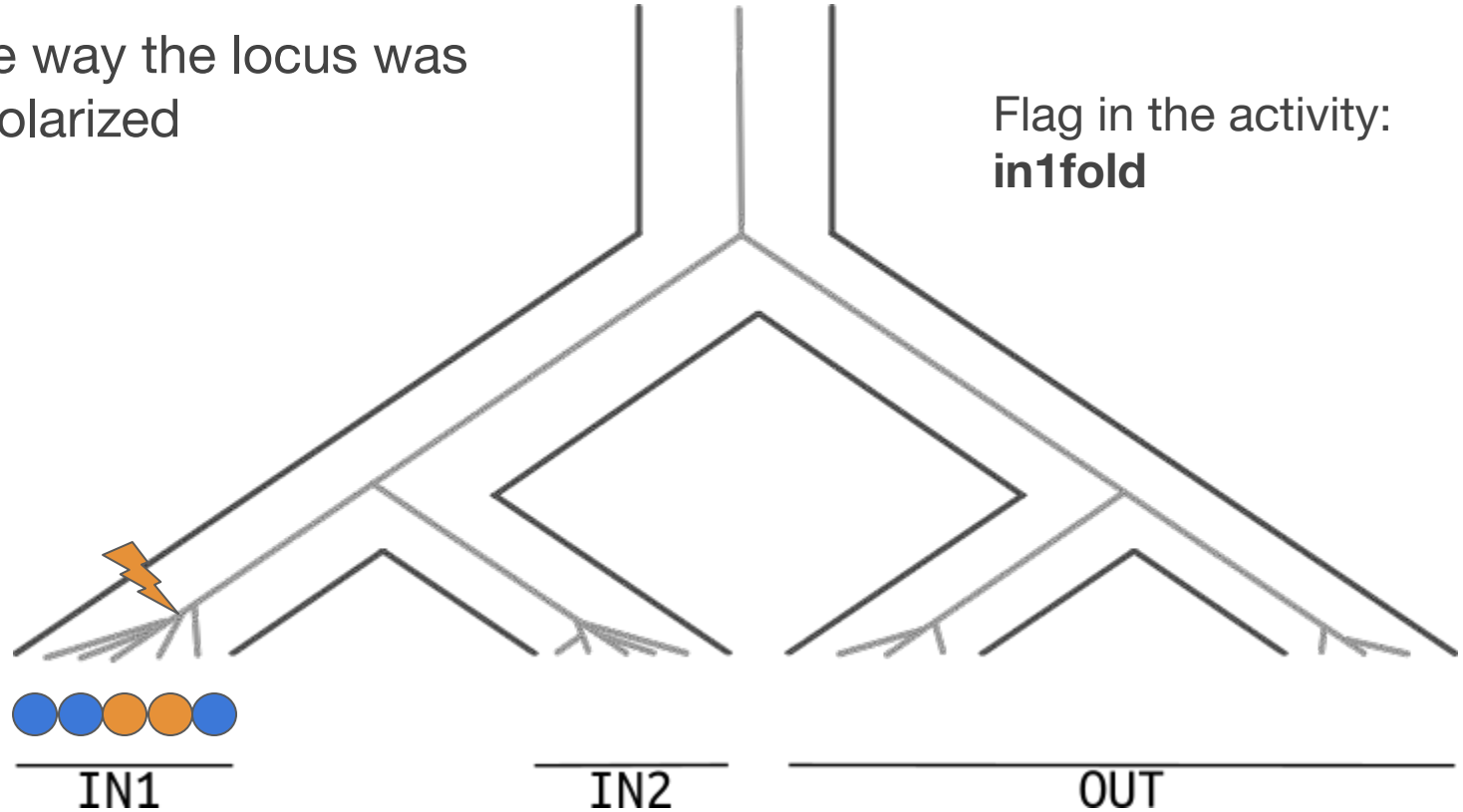
Keep track of missing data

Or flag the way the locus was actually polarized

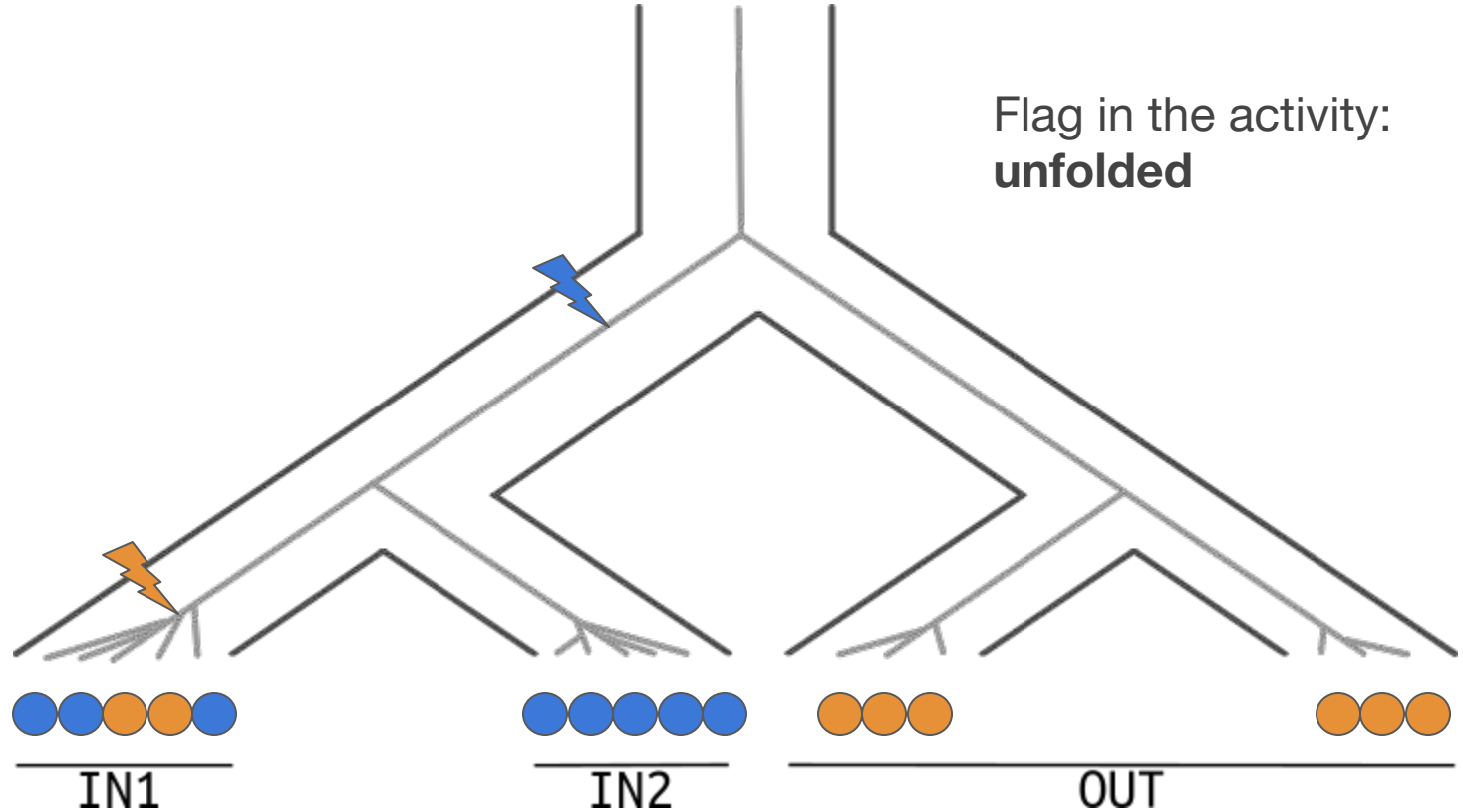


Keep track of missing data

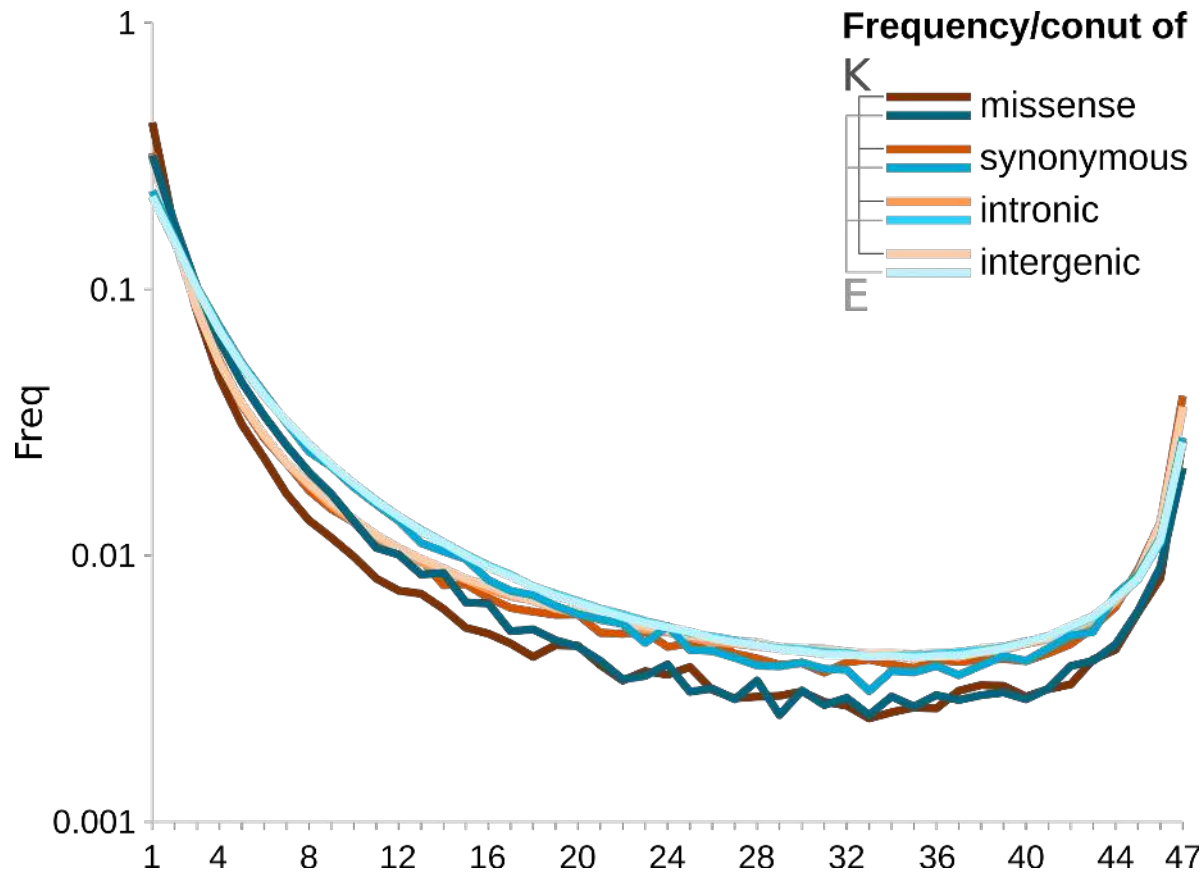
Or flag the way the locus was actually polarized



And recurrent mutations mess things up!

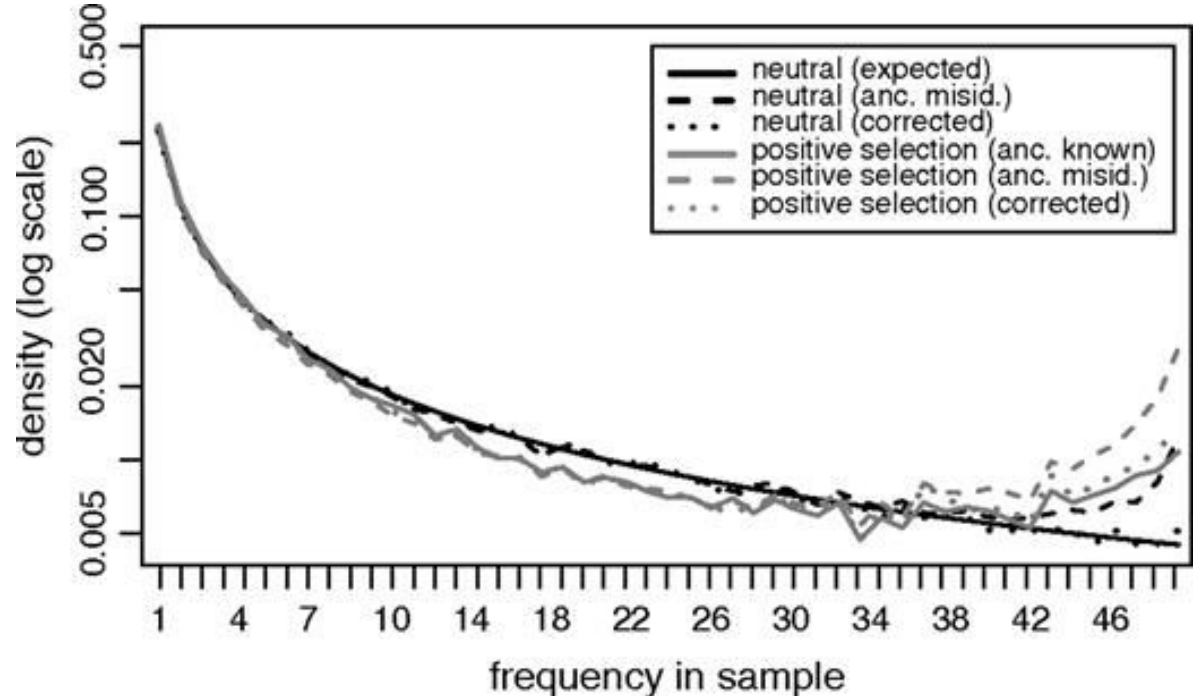


And recurrent mutations mess things up!



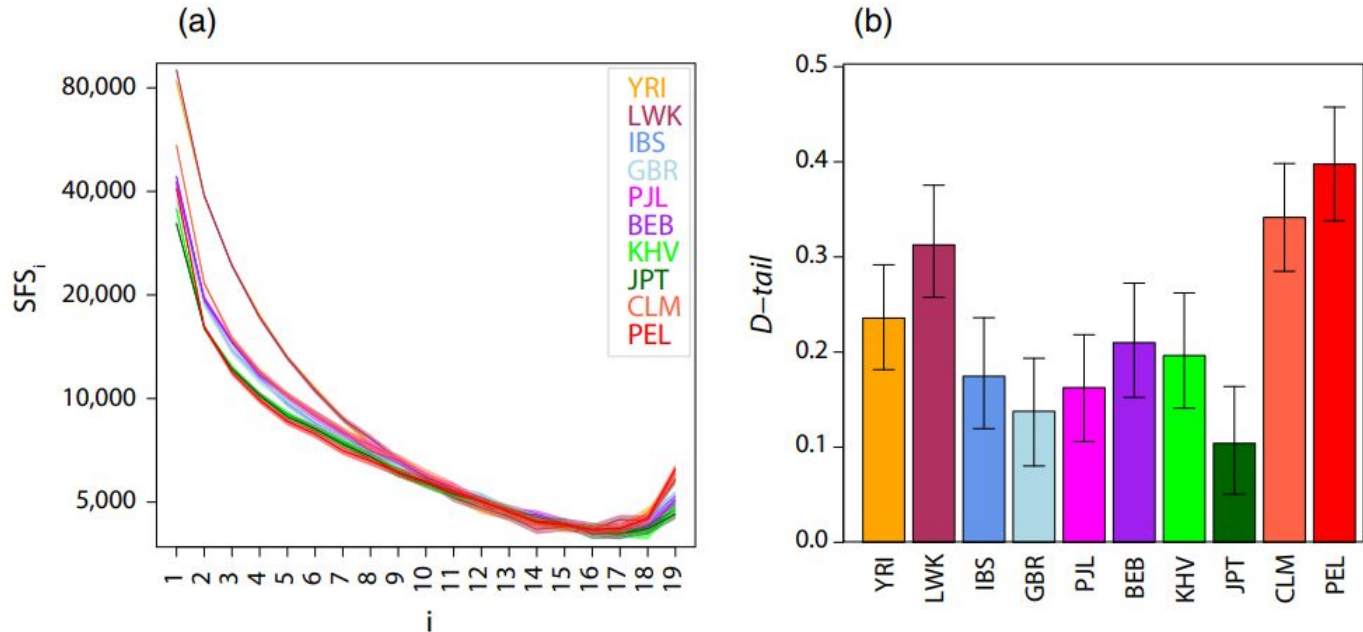
And recurrent mutations mess things up!

It can be corrected by using
context-dependent mutation models
(Hernandez et al 2007)



Apparent mispolarizations

Due to migration from a 'ghost' population
(Marchi and Excoffier 2020)





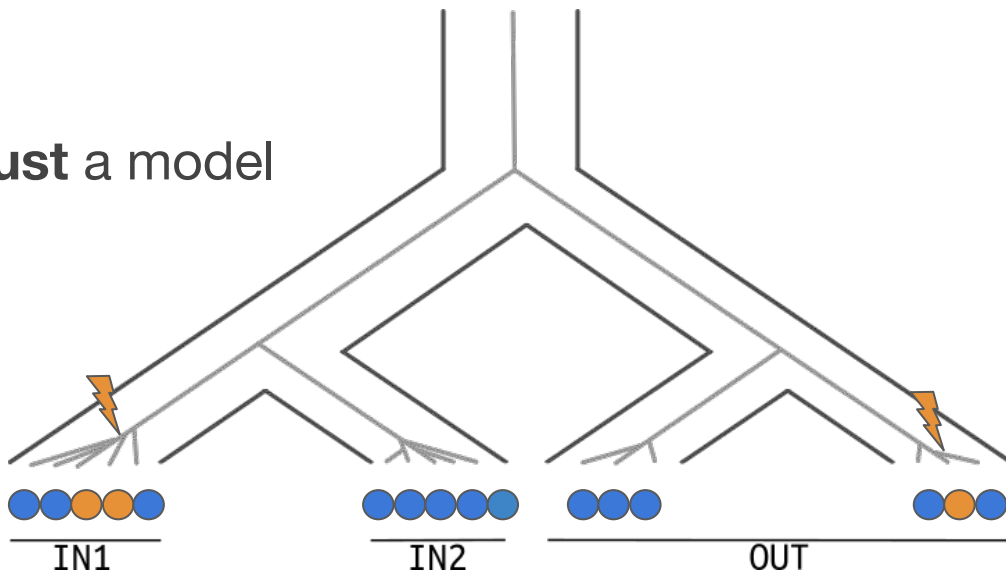
POLARIZATION

Use only the sites you can **confidently** polarize depending on their configuration and on the **missing data** in ingroup and outgroups

POLARIZATION

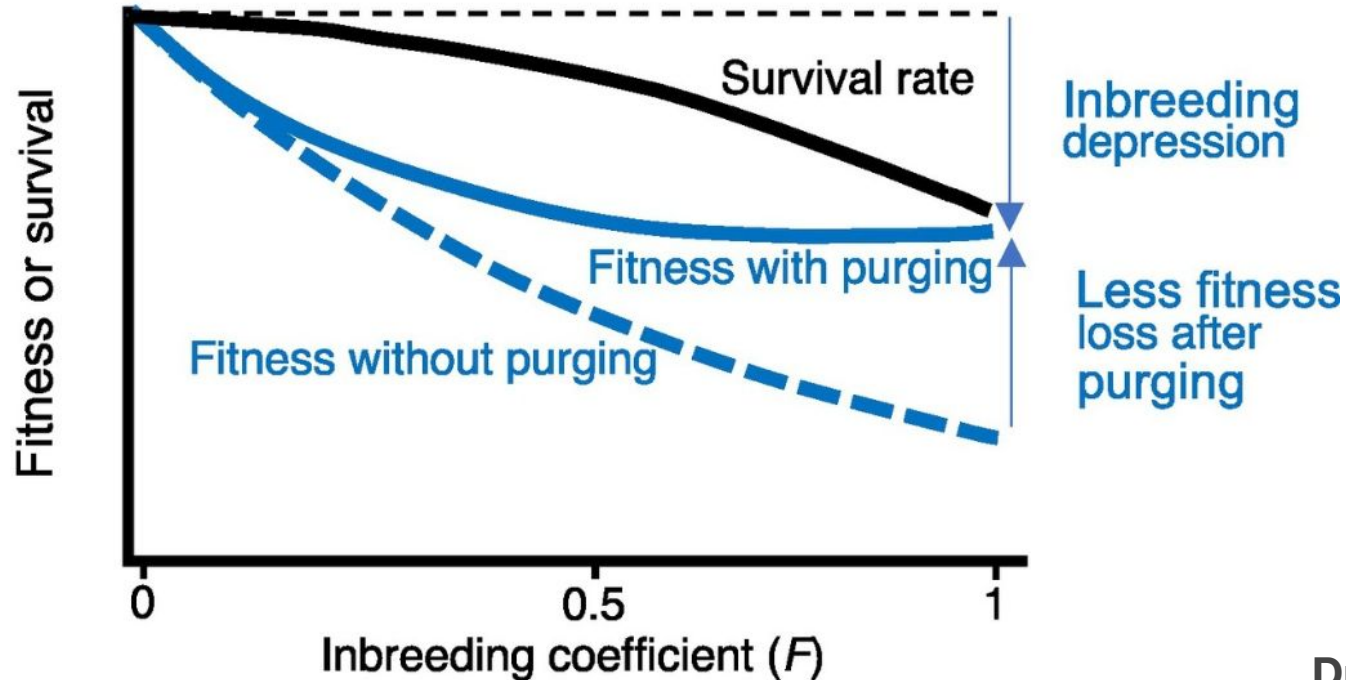
Use only the sites you can **confidently** polarize depending on their configuration and on the **missing data** in ingroup and outgroups

The infinite site model is **just** a model

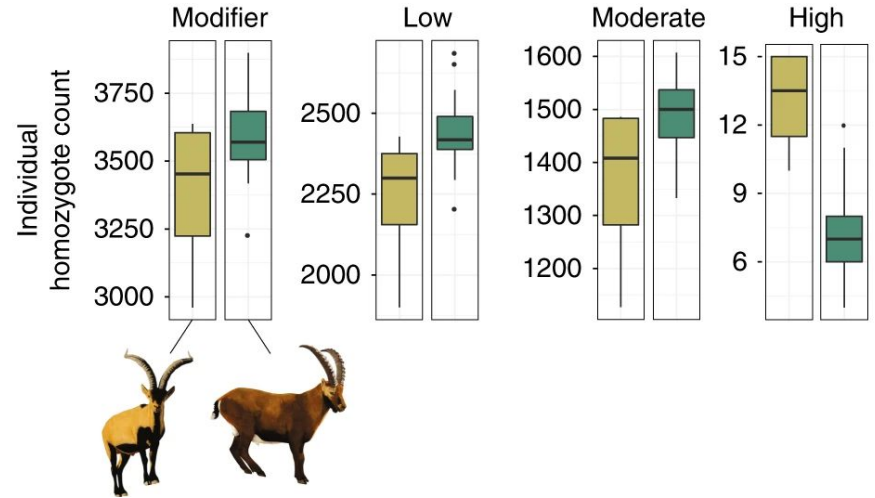
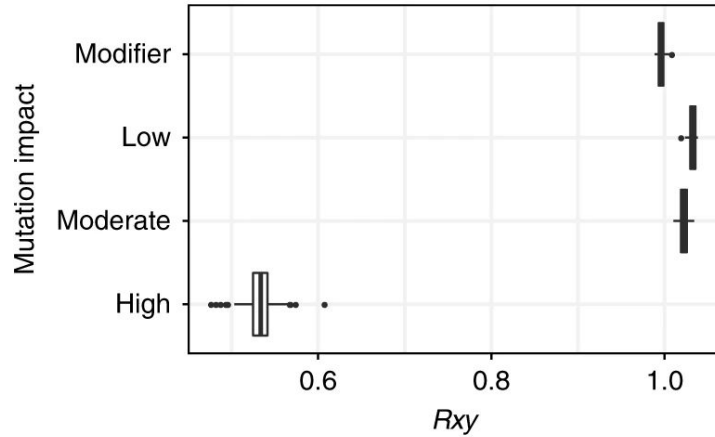


ONE MORE THING:
DO YOU KNOW WHAT **PURGING** MEANS
WHEN TALKING OF GENETIC LOAD?

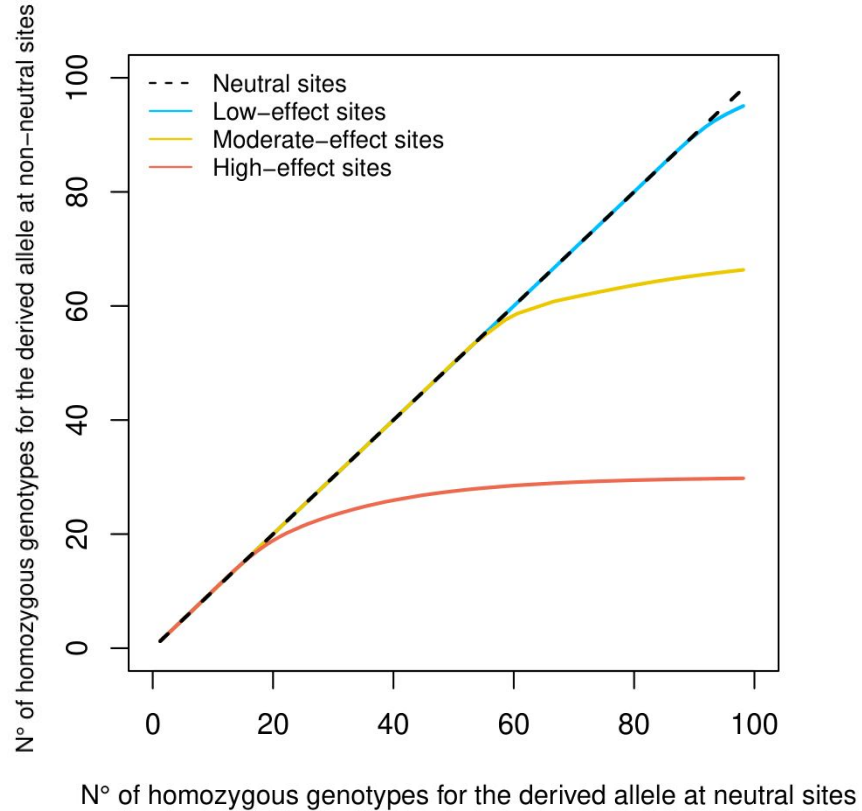
PURGING DELETERIOUS MUTATIONS



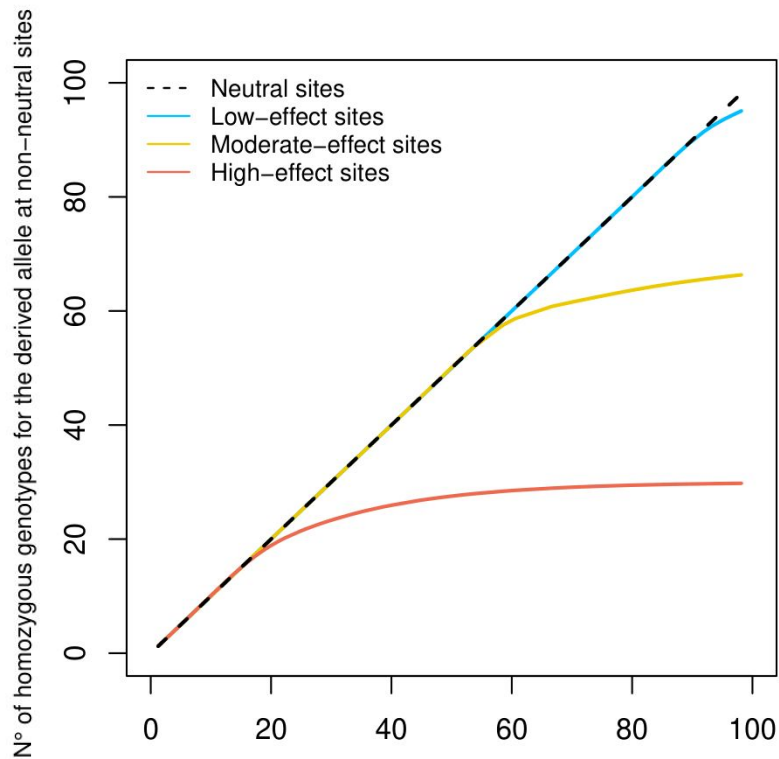
PURGING DELETERIOUS MUTATIONS



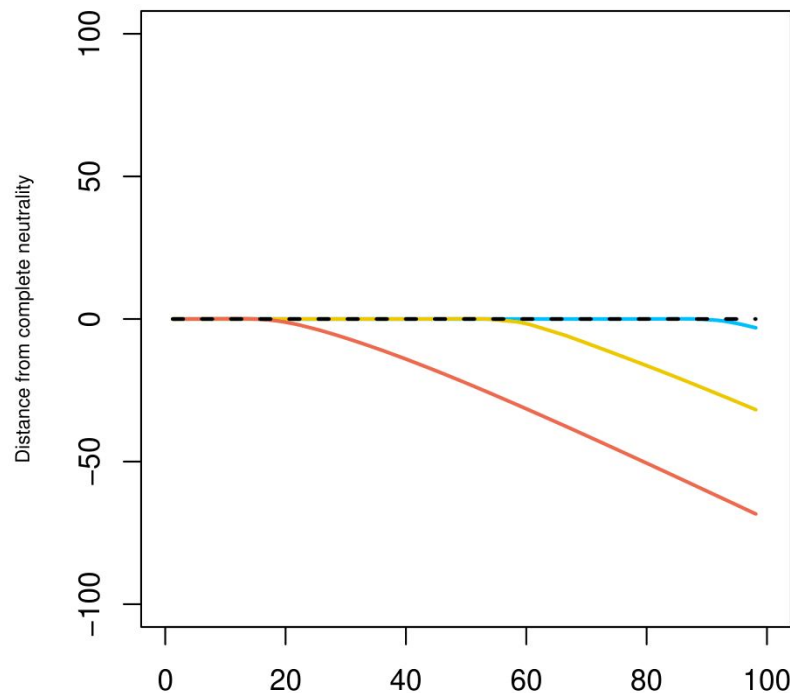
PURGING DELETERIOUS MUTATIONS



PURGING DELETERIOUS MUTATIONS



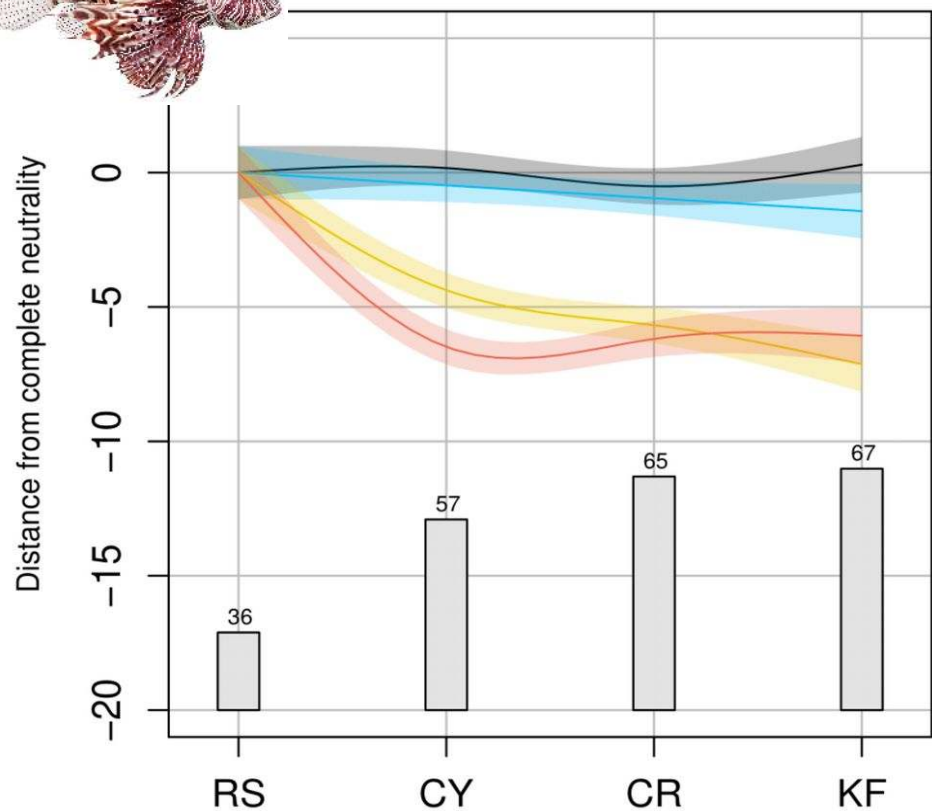
N° of homozygous genotypes for the derived allele at neutral sites



Giannelli et al *in prep*

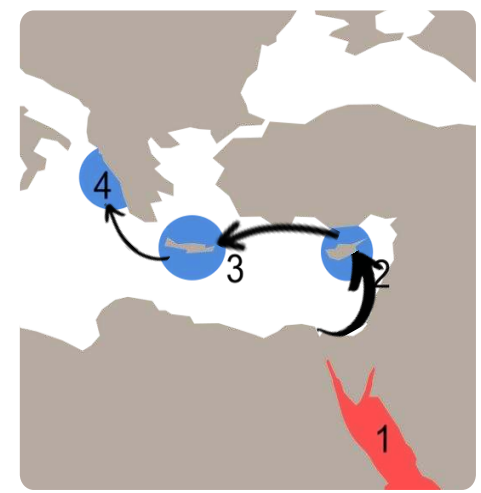


Pterois miles



- Neutral
- Low effect
- Moderate effect
- High effect

PURGING DELETERIOUS MUTATIONS



QUESTIONS?



QUESTIONS?

**INTERESTED IN A WORKSHOP
FOCUSED ON CONSERVATION
GENOMICS?**

Here is a short questionnaire to let us know:
<https://forms.gle/E2iAMHjwbbbR5aamQ8>

