ACTIVITY OUTLINE

Start with ROH estimates activity

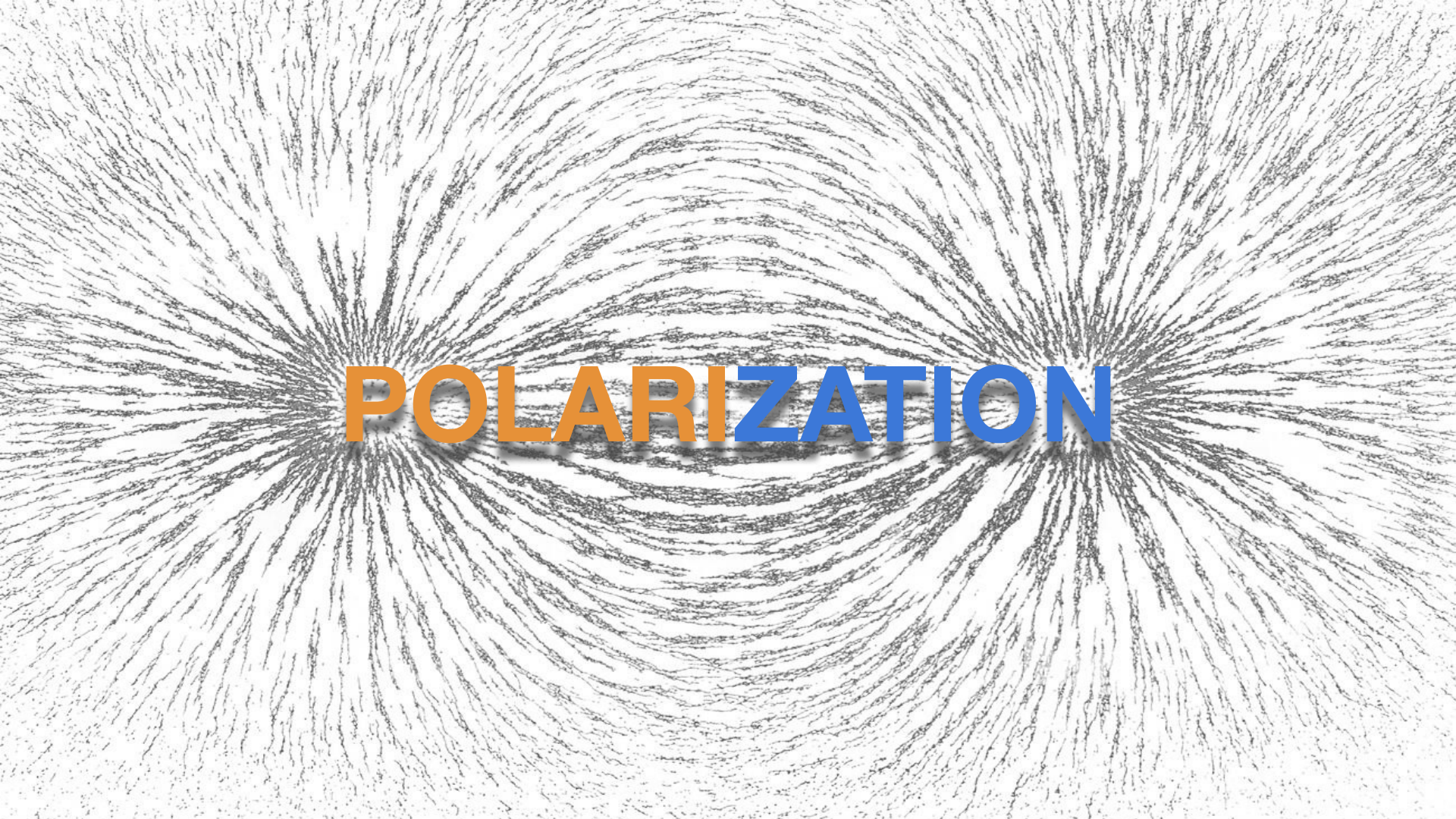
First wrap-up after ca. 45-50 minutes

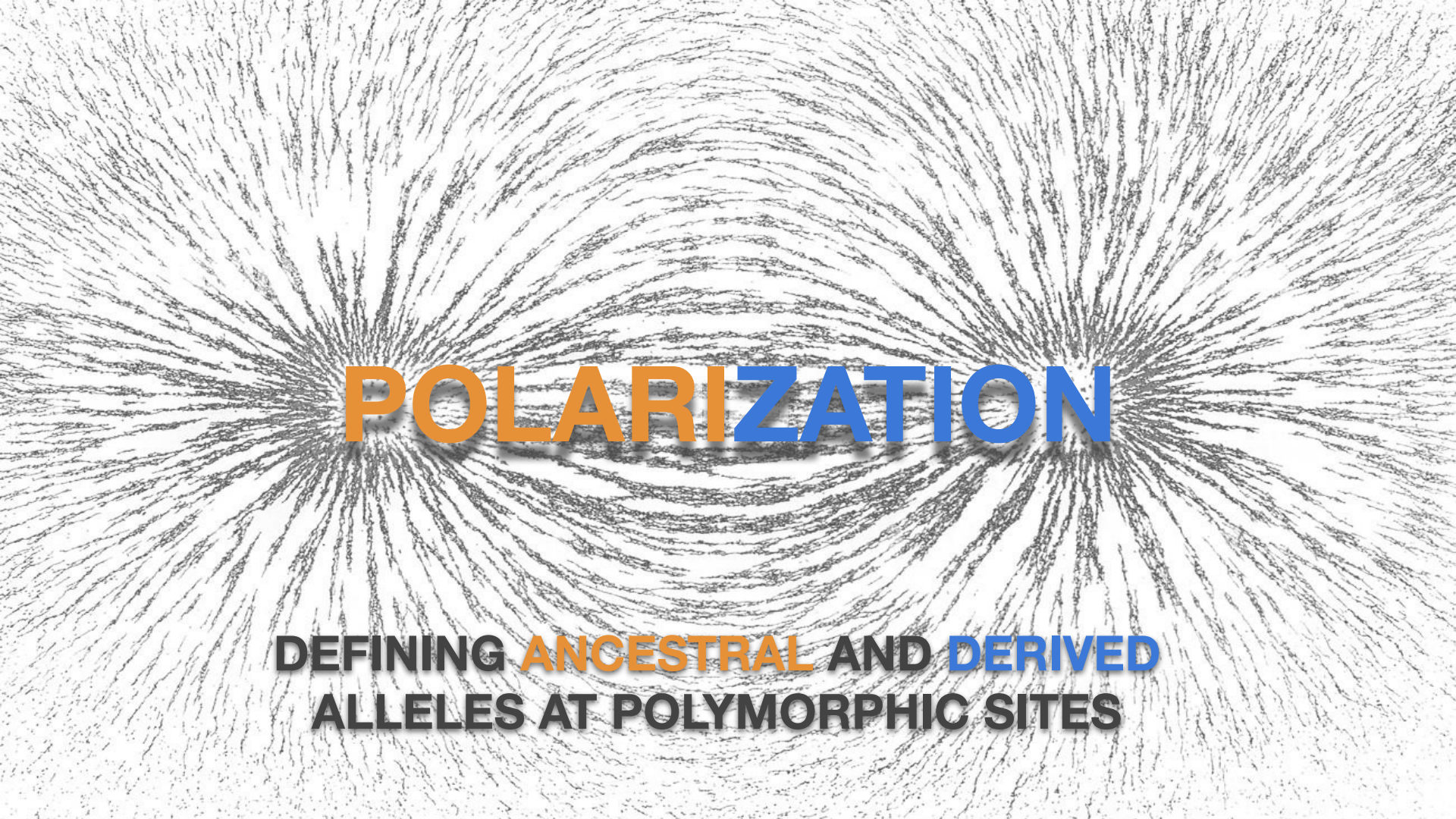
Presentation of the polarization issue

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


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

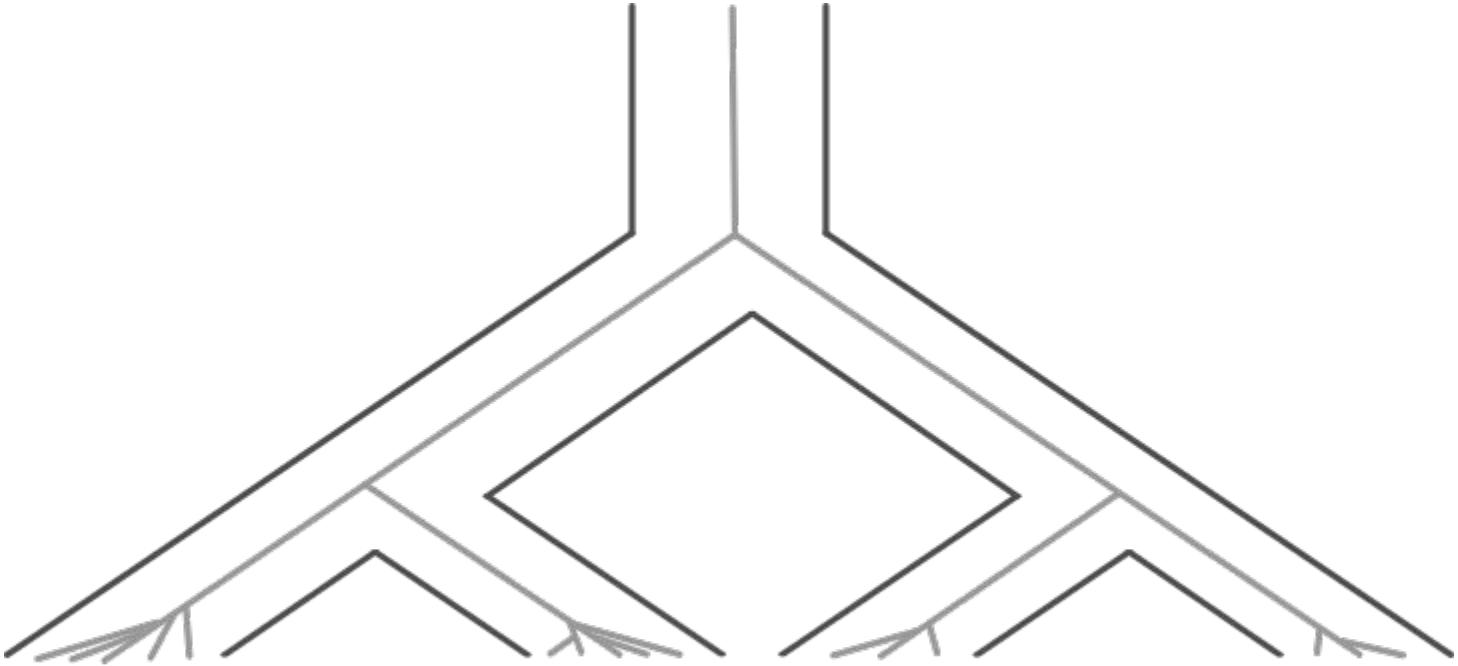
MORE OUTGROUPS!

TARGET POPULATION

OUTGROUP

OUTGROUP

MORE OUTGROUPS!

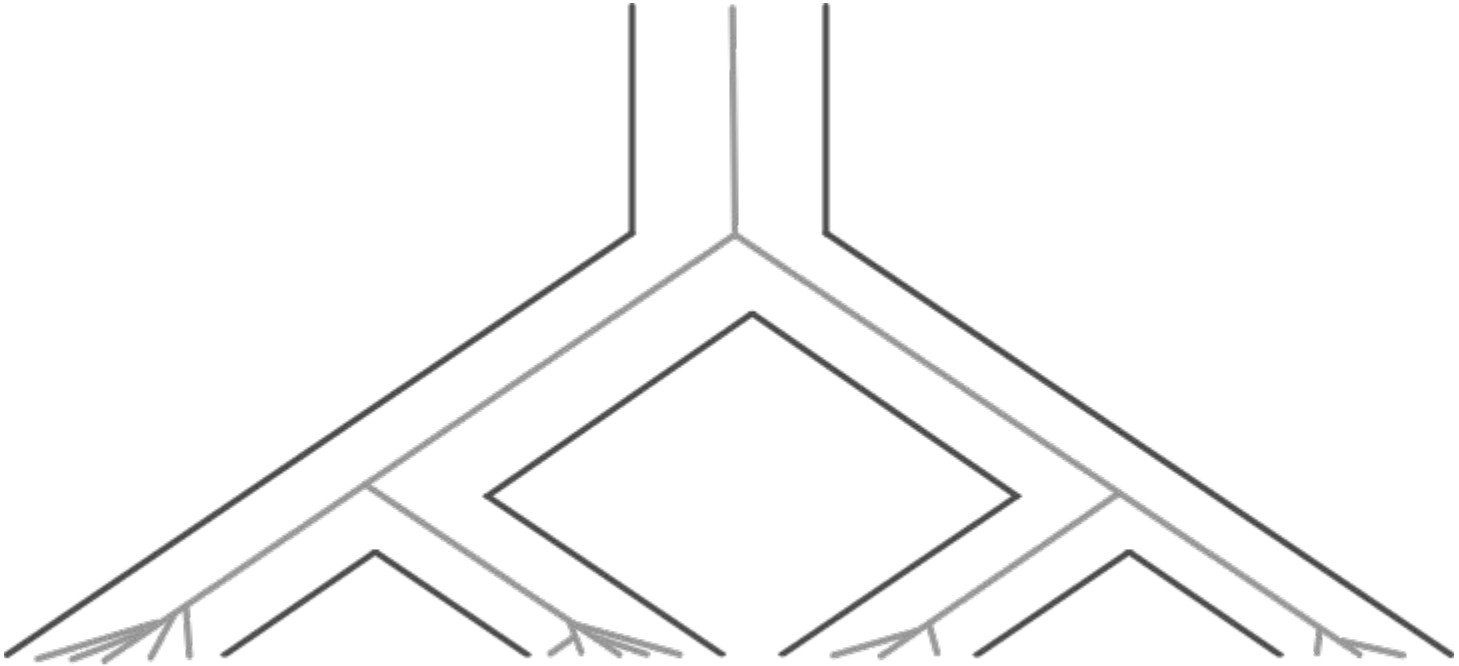


TARGET POPULATION

CLOSE OUTGROUP

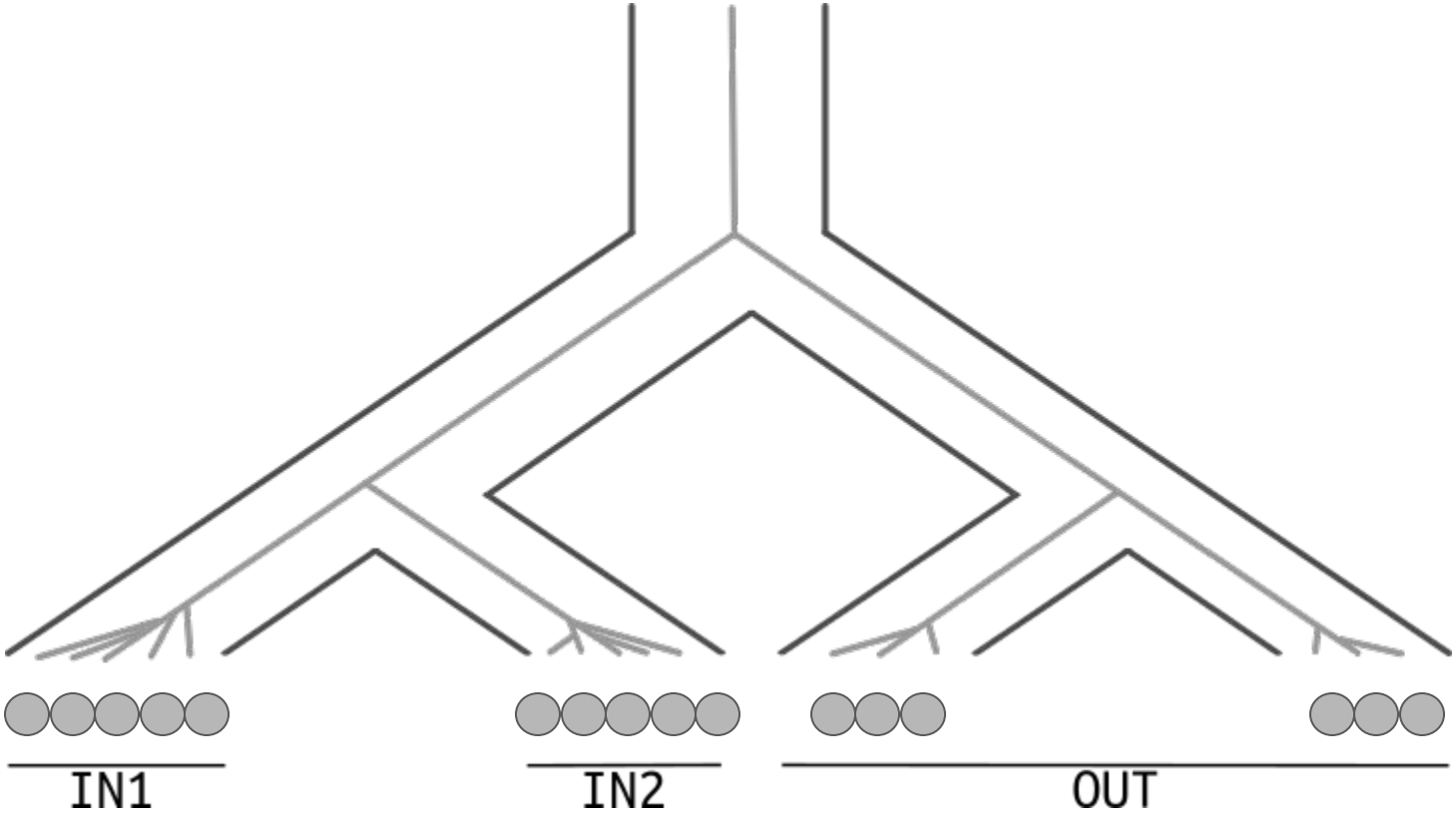
DISTANT OUTGROUP

LABELS IN THIS ACTIVITY

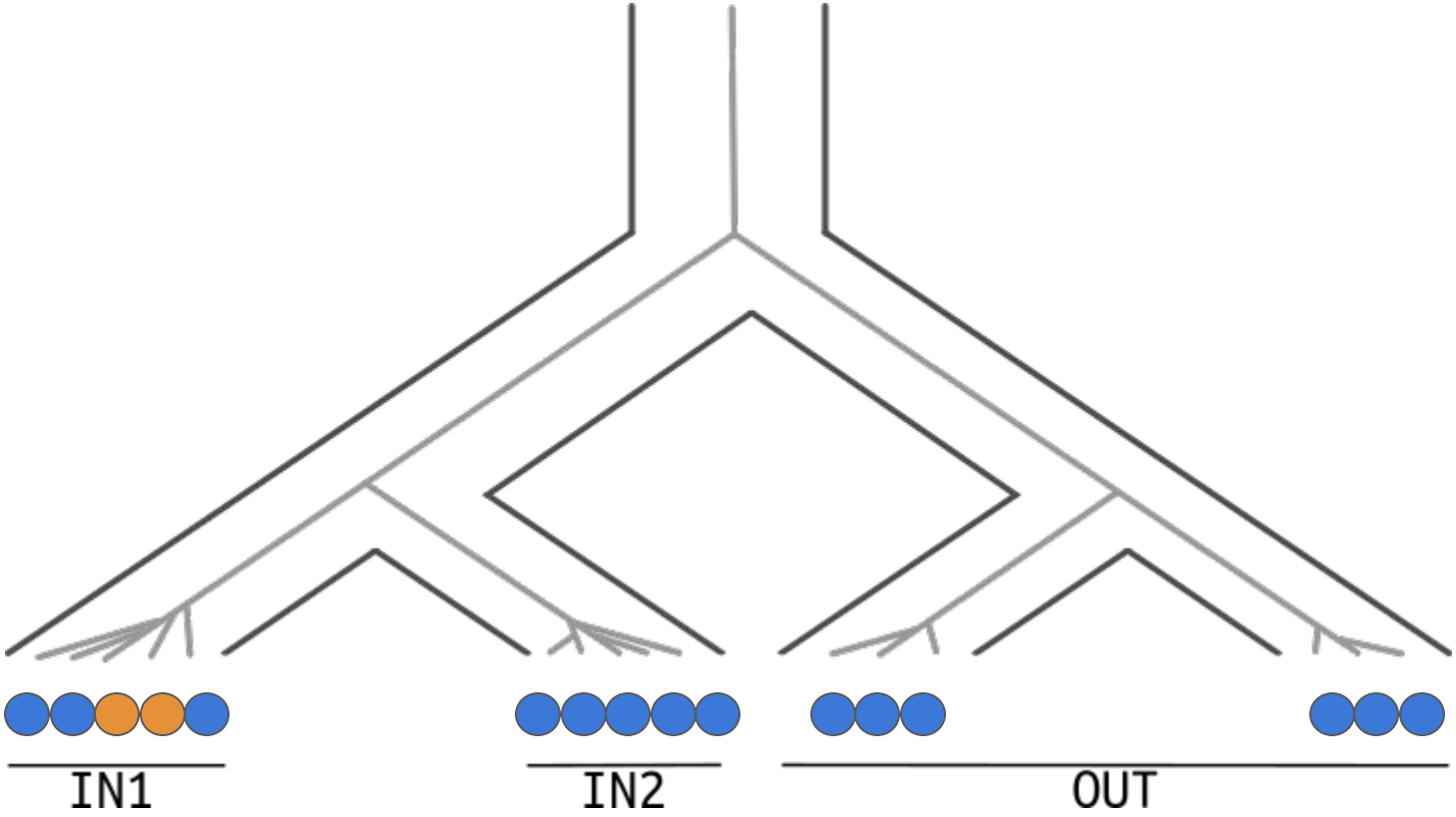


INGROUP1 INGROUP2 OUTGROUP

SOME EXAMPLES OF ALLELES CONFIGURATION



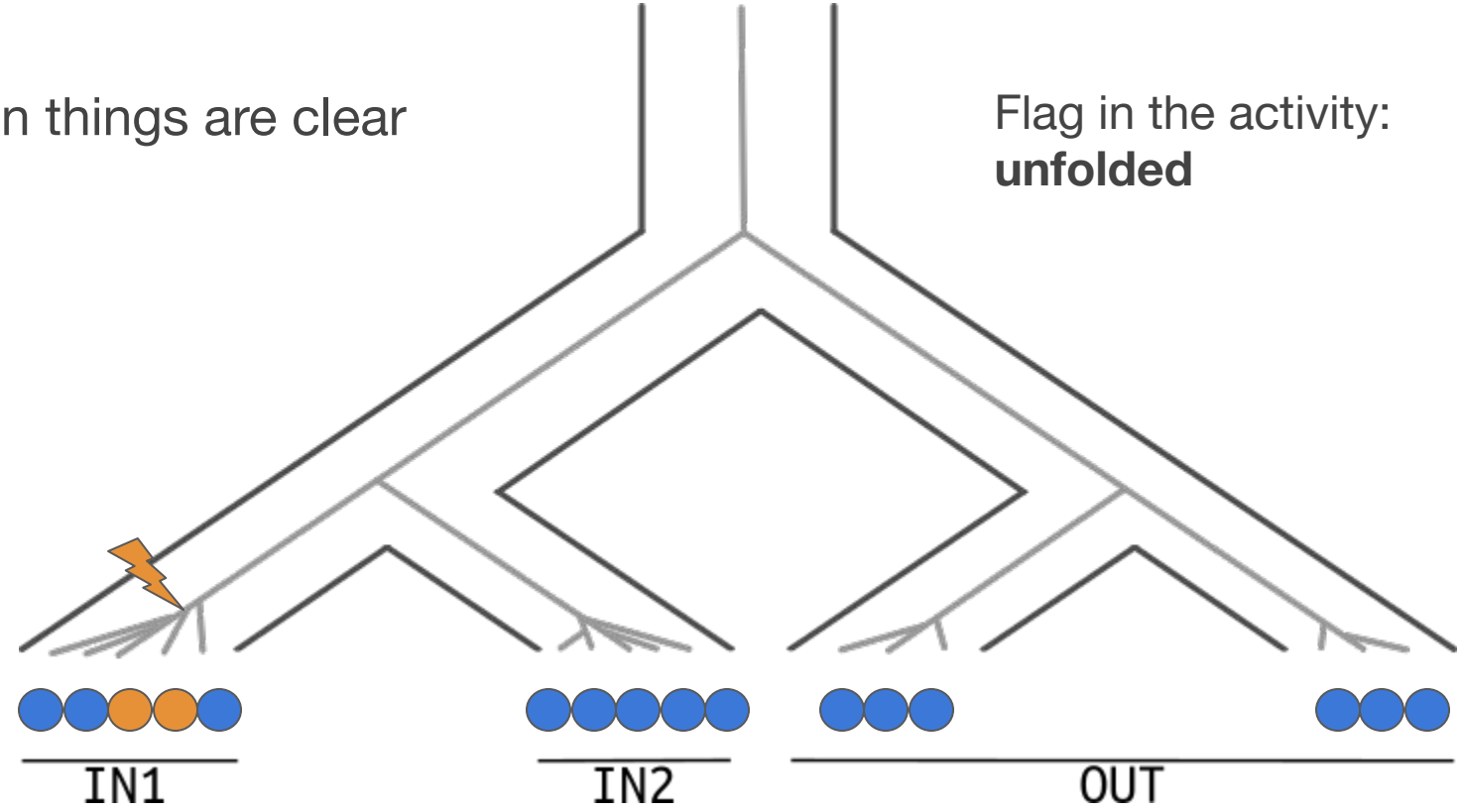
Which one is the derived allele?



Which one is the derived allele?

Easy when things are clear

Flag in the activity:
unfolded



GENOLOADER

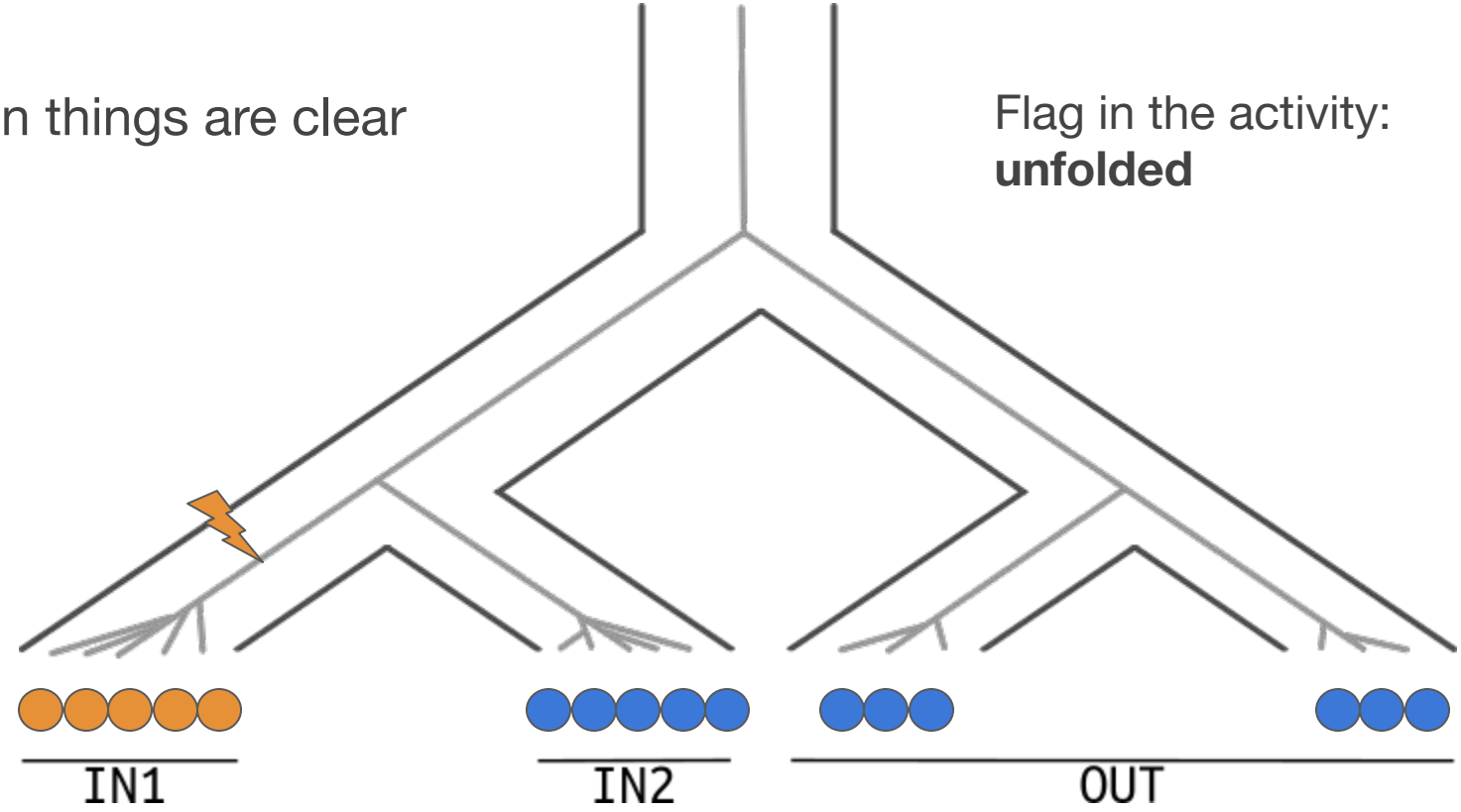
```
def get_ancestral_allele(geno_out, geno1, geno2):

    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

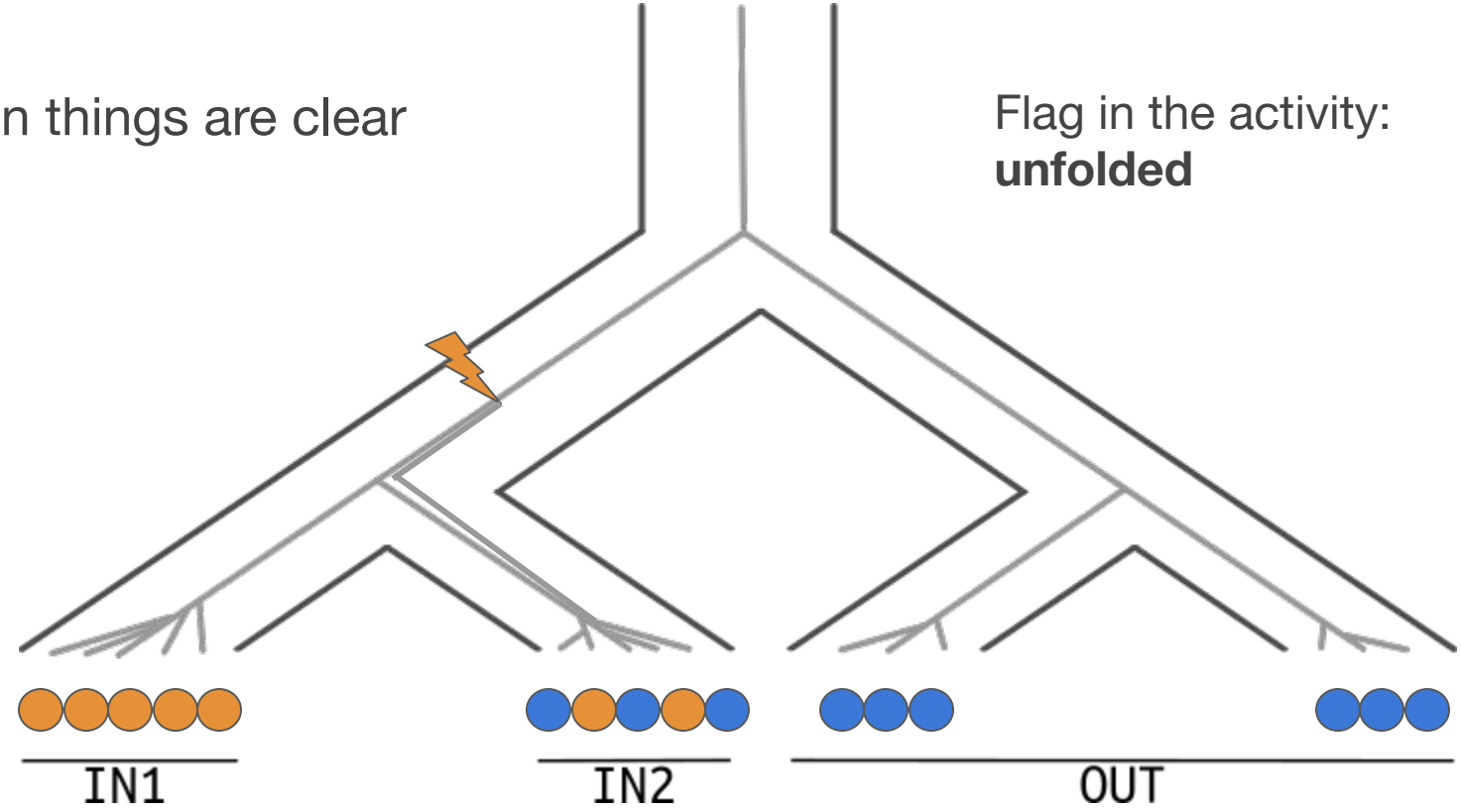
Flag in the activity:
unfolded



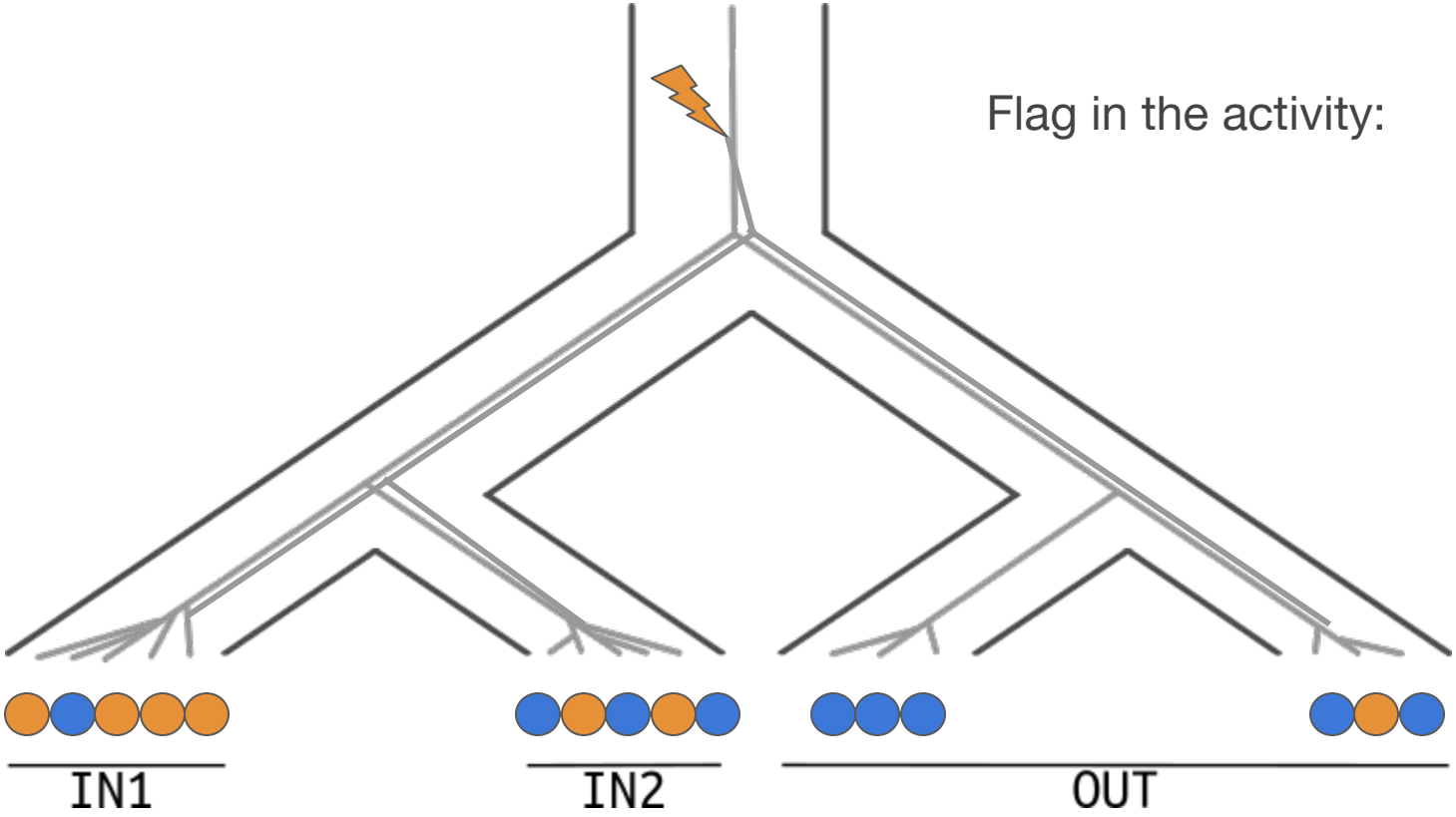
Which one is the derived allele?

Easy when things are clear

Flag in the activity:
unfolded



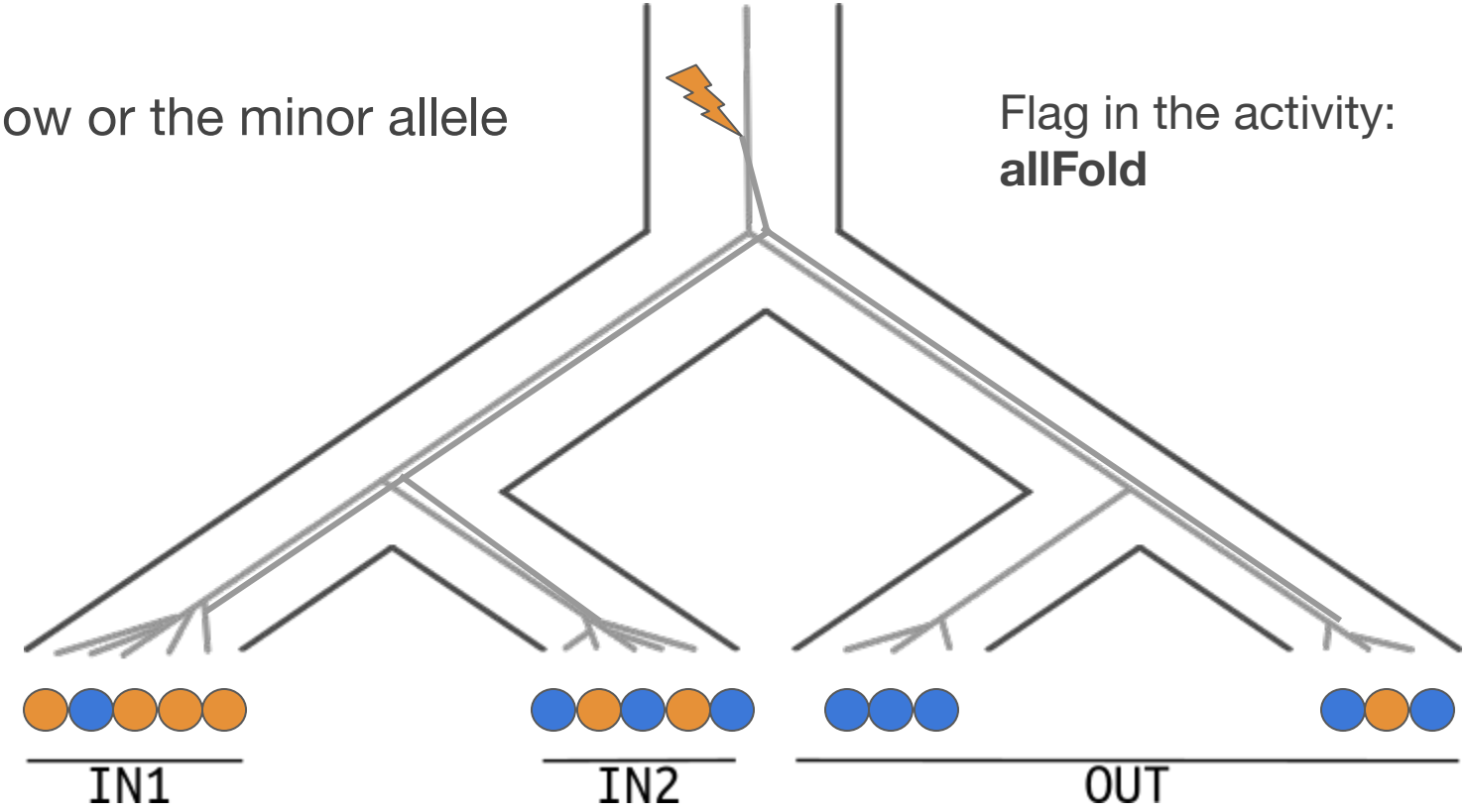
Which one is the derived allele?



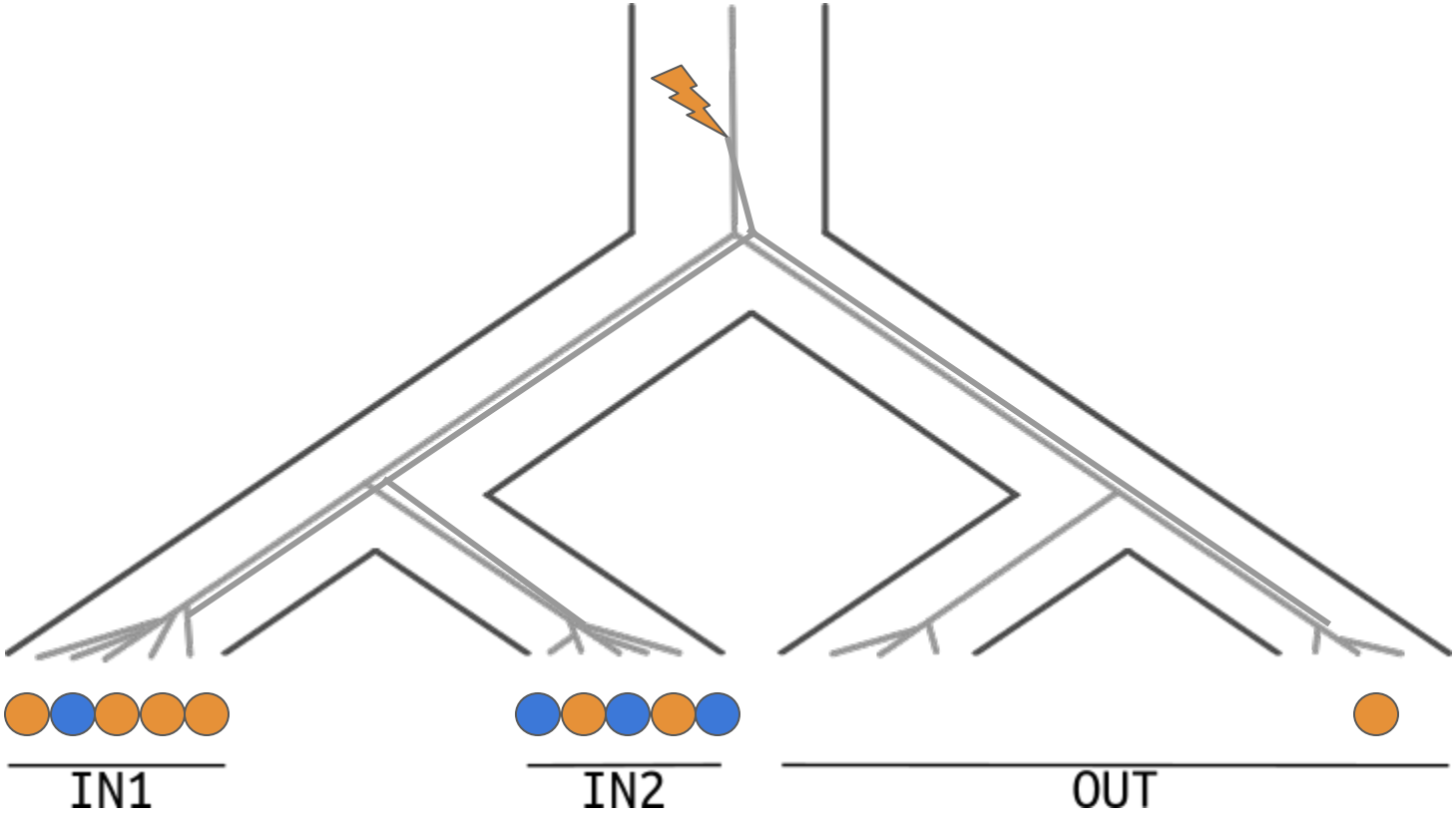
Which one is the derived allele?

I don’t know or the minor allele

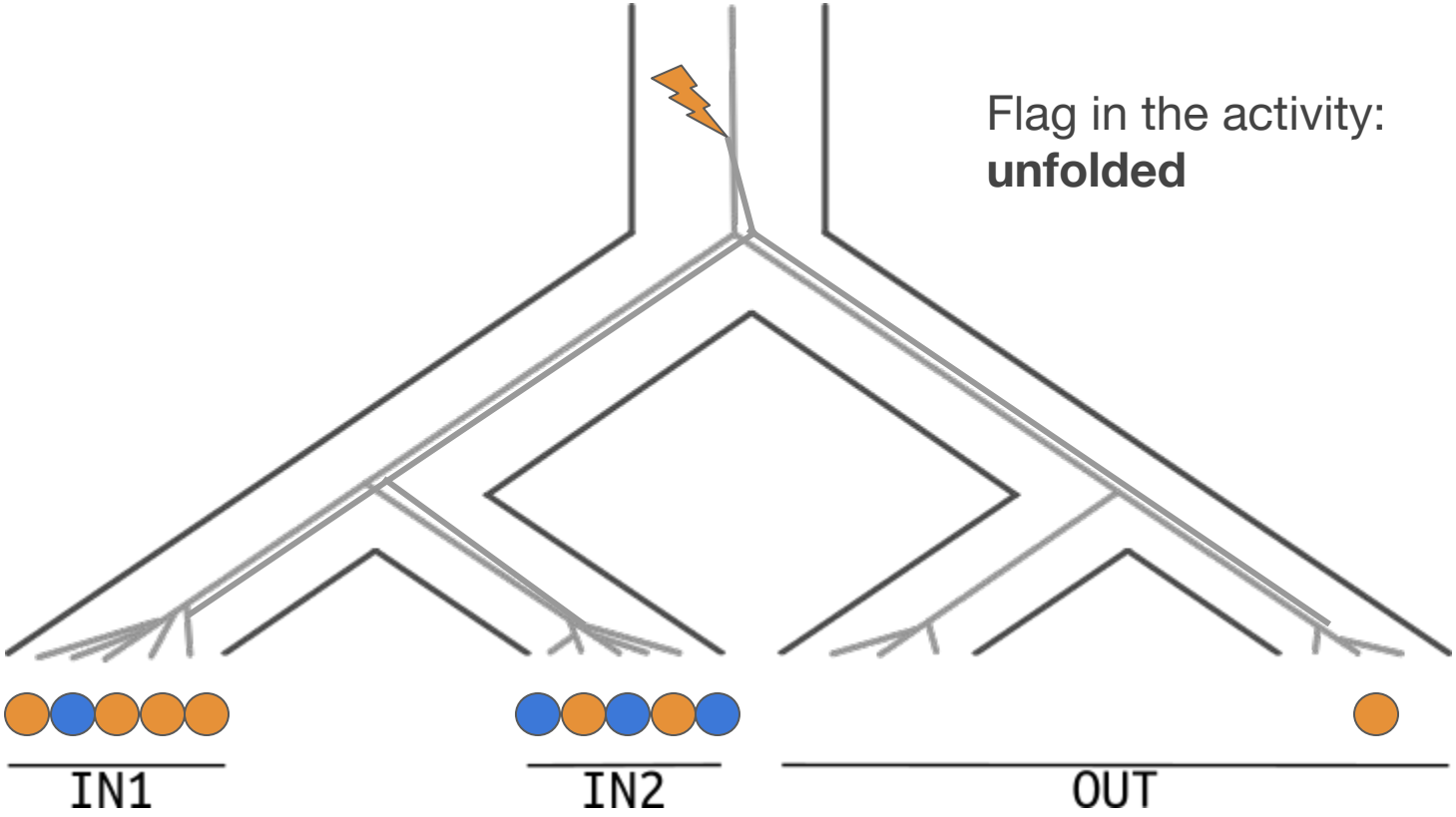
Flag in the activity:
allFold



Keep track of missing data



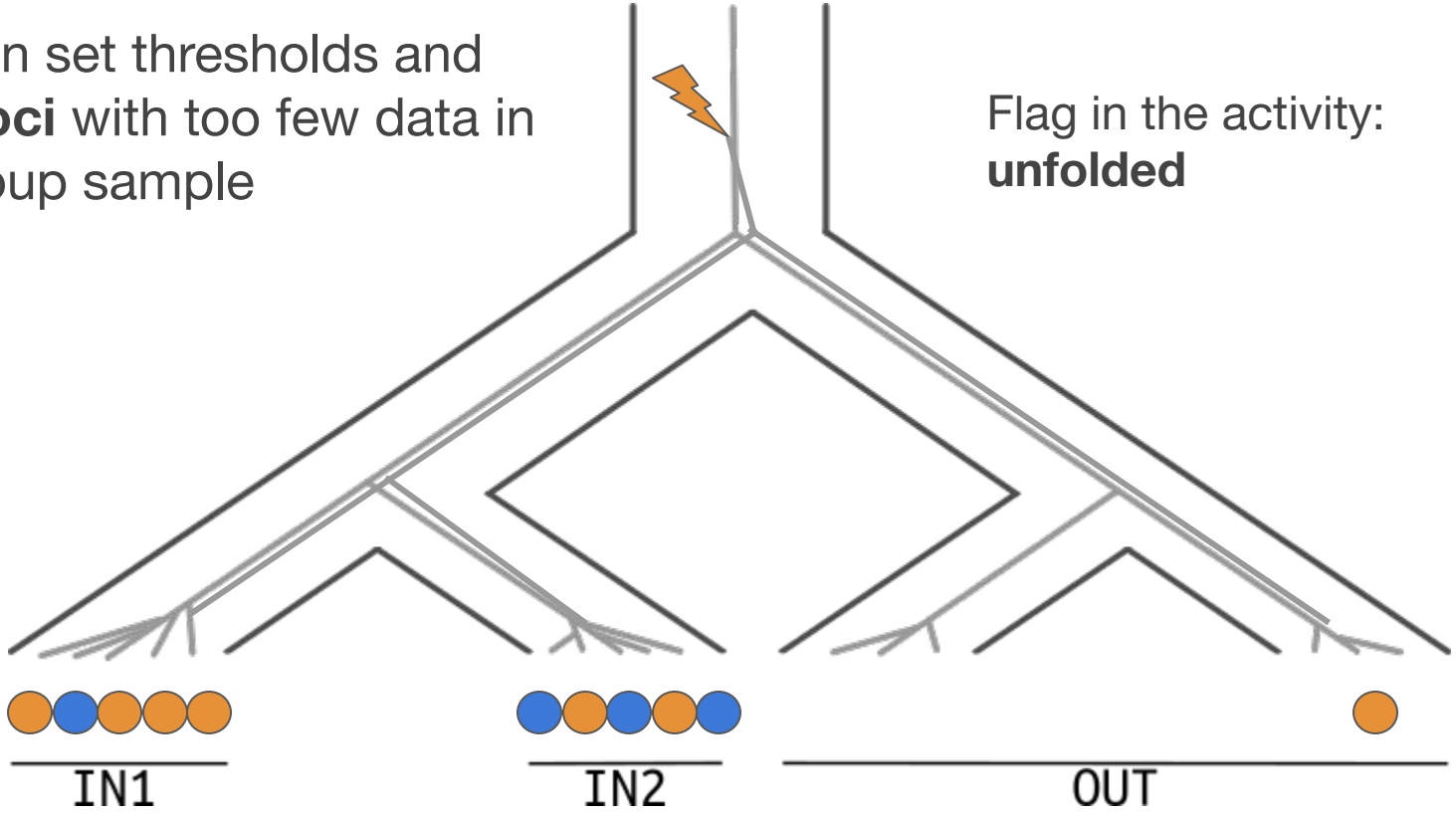
Keep track of missing data



Keep track of missing data

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

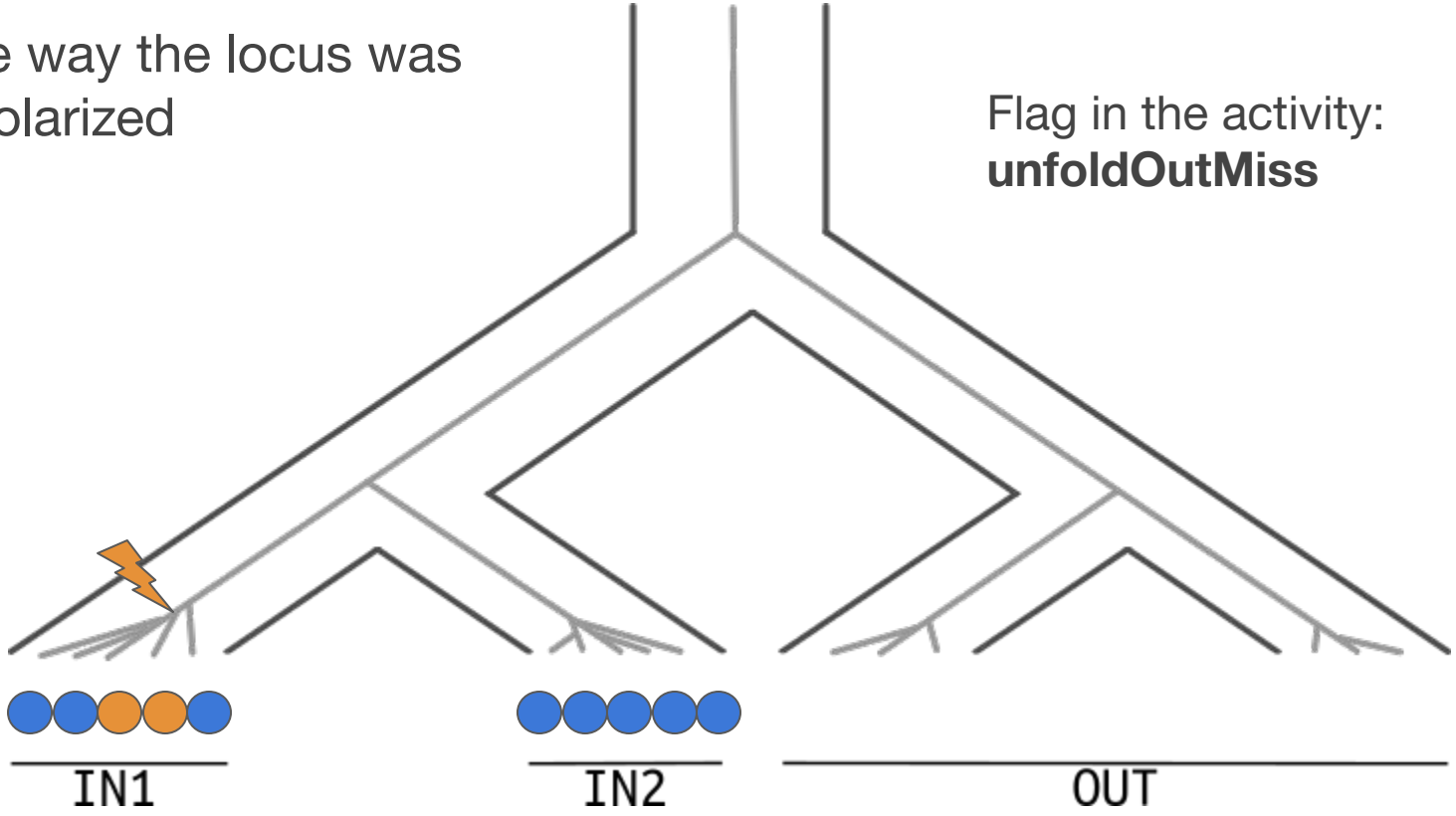
Flag in the activity: **unfolded**



Keep track of missing data

Or flag the way the locus was actually polarized

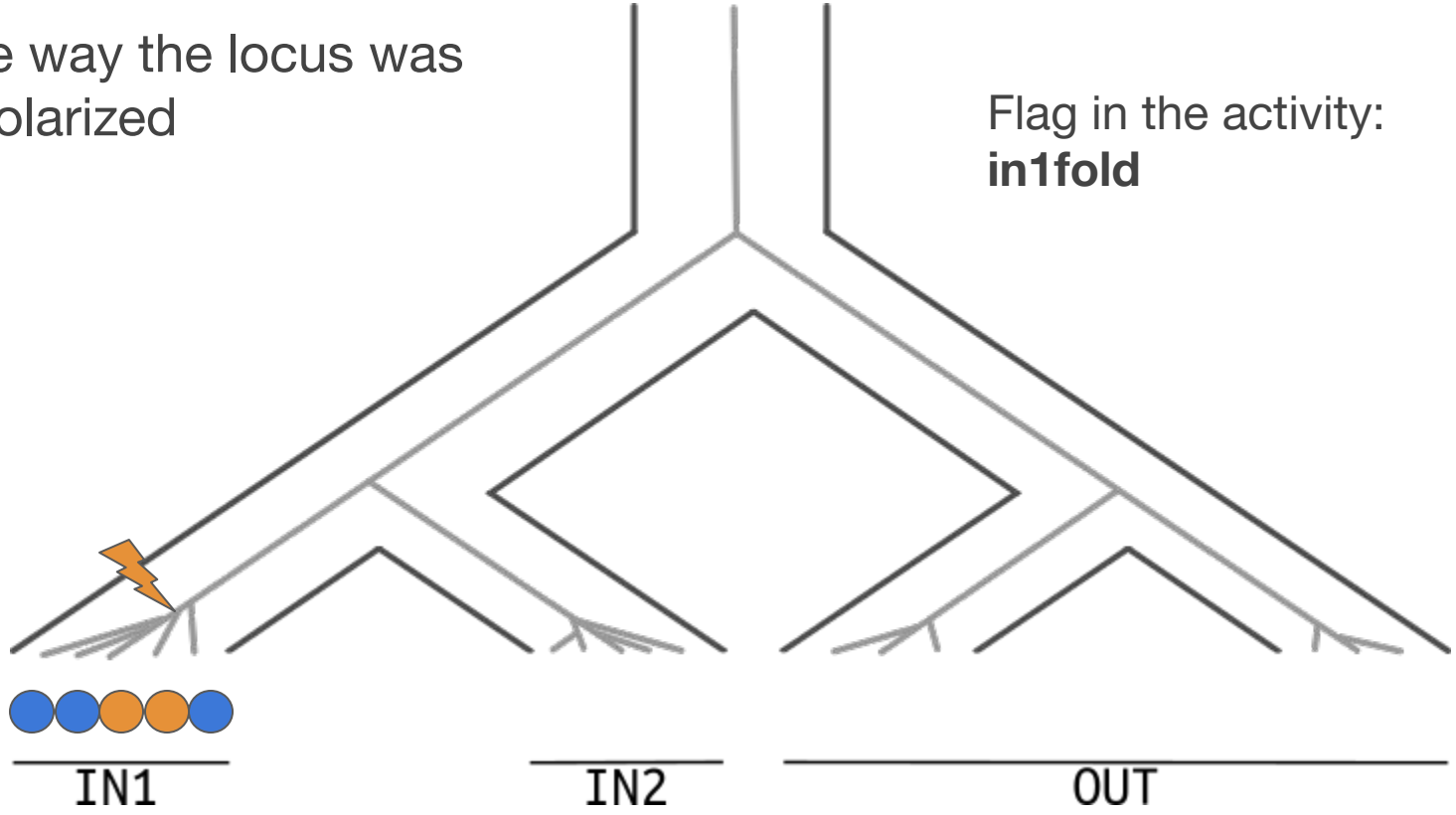
Flag in the activity:
unfoldOutMiss



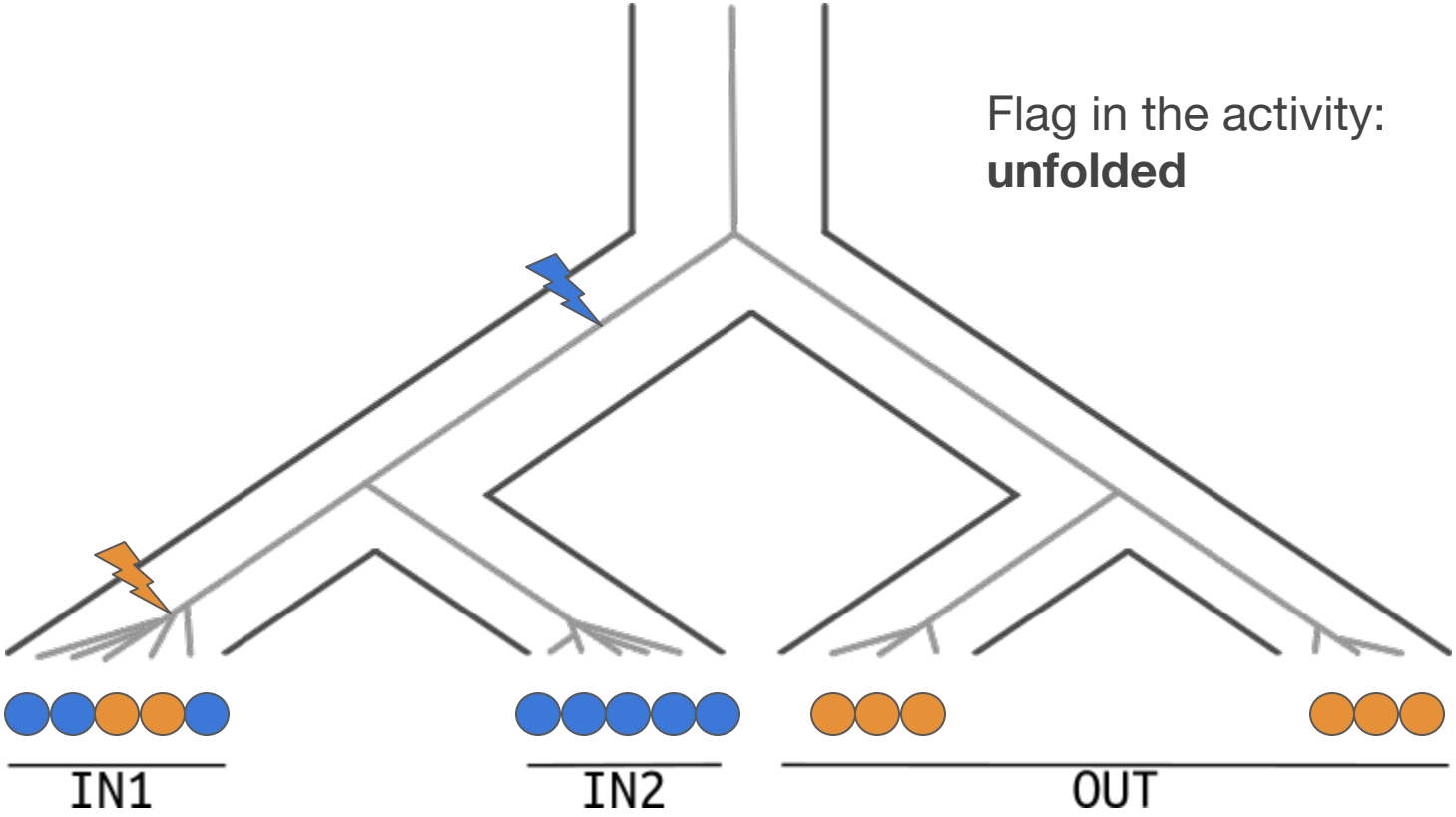
Keep track of missing data

Or flag the way the locus was actually polarized

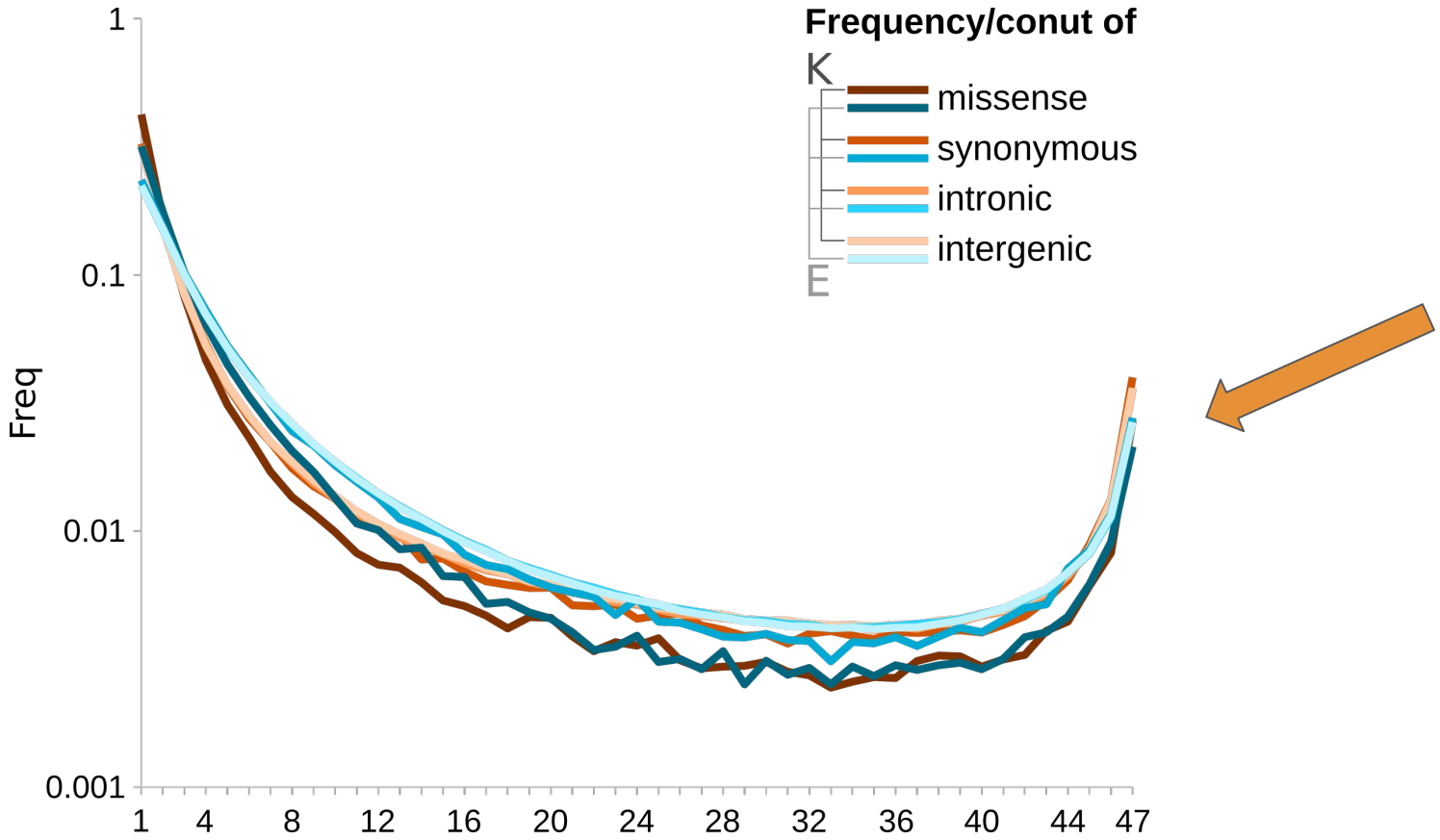
Flag in the activity:
in1fold



And recurrent mutations mess things up!

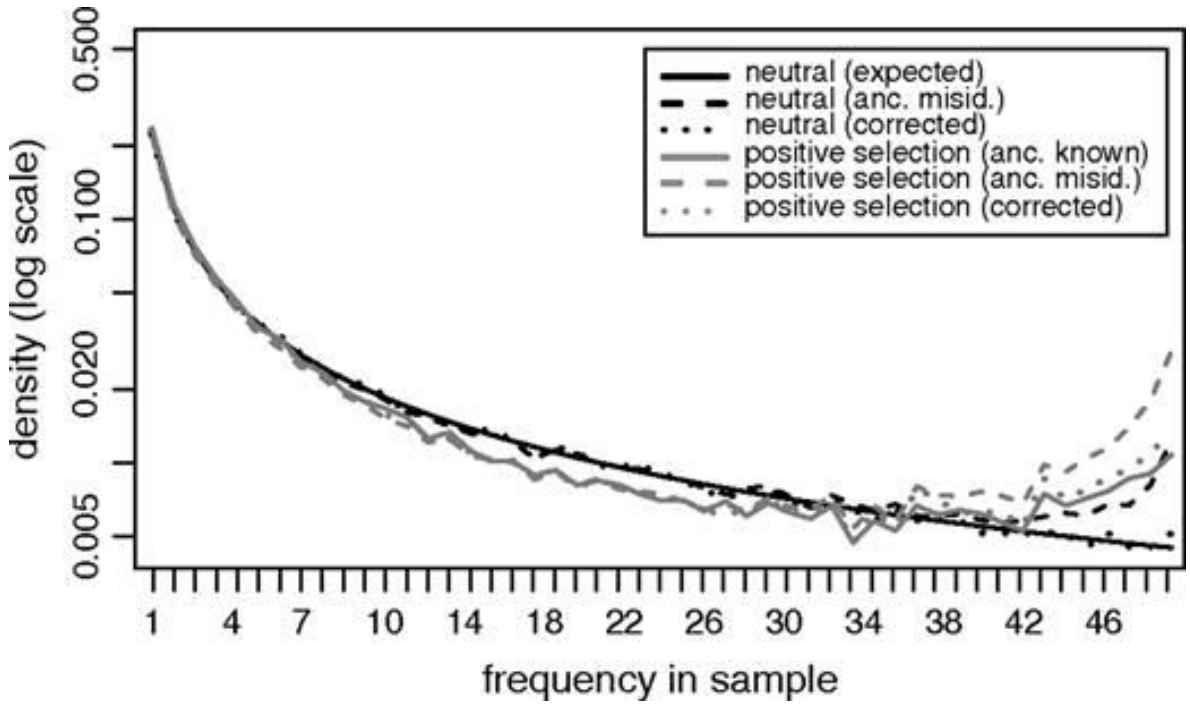


And recurrent mutations mess things up!



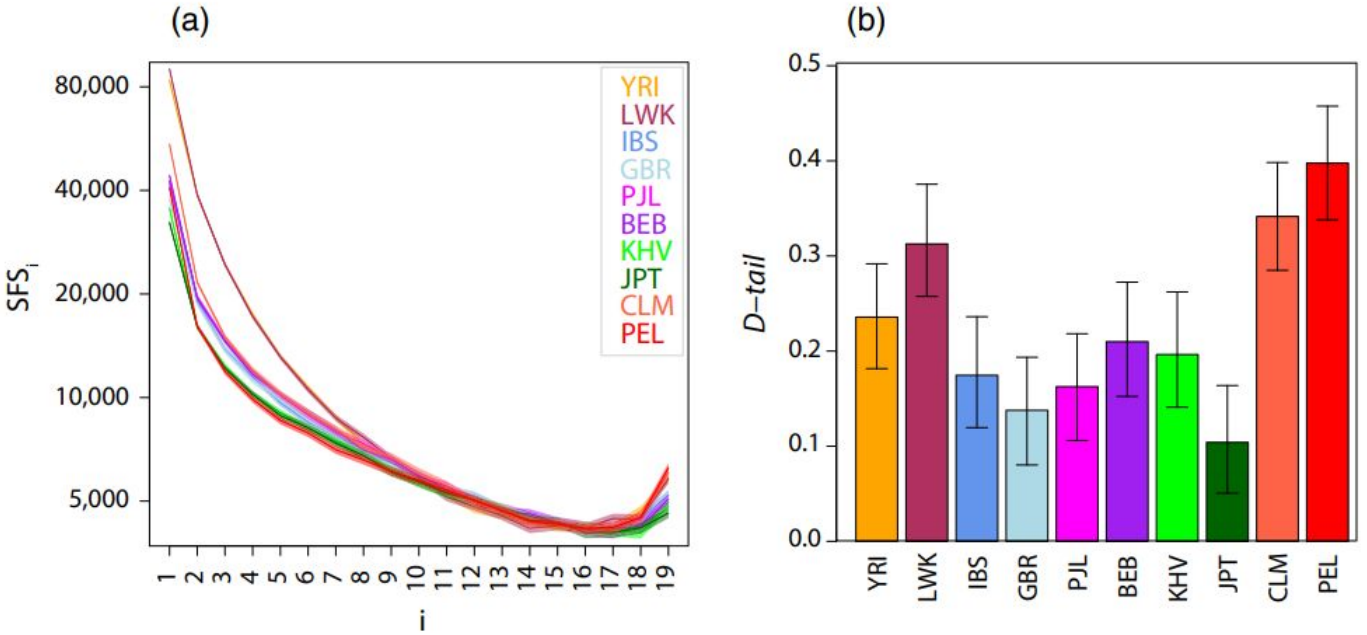
And recurrent mutations mess things up!

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)

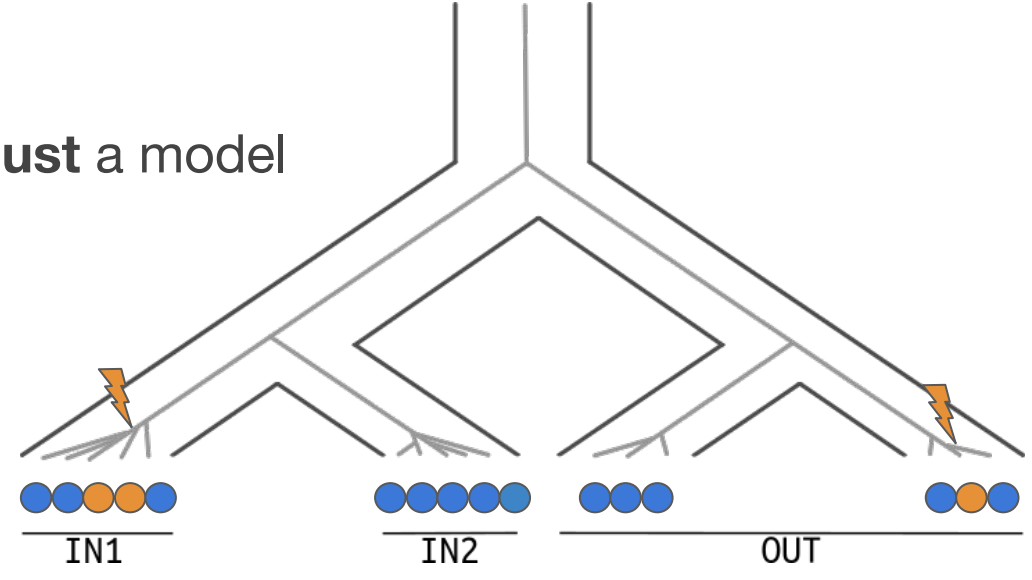


Use only the sites you can **confidently** polarize depending on their configuration and on the **missing data** in ingroup and outgroup**s**

POLARIZATION

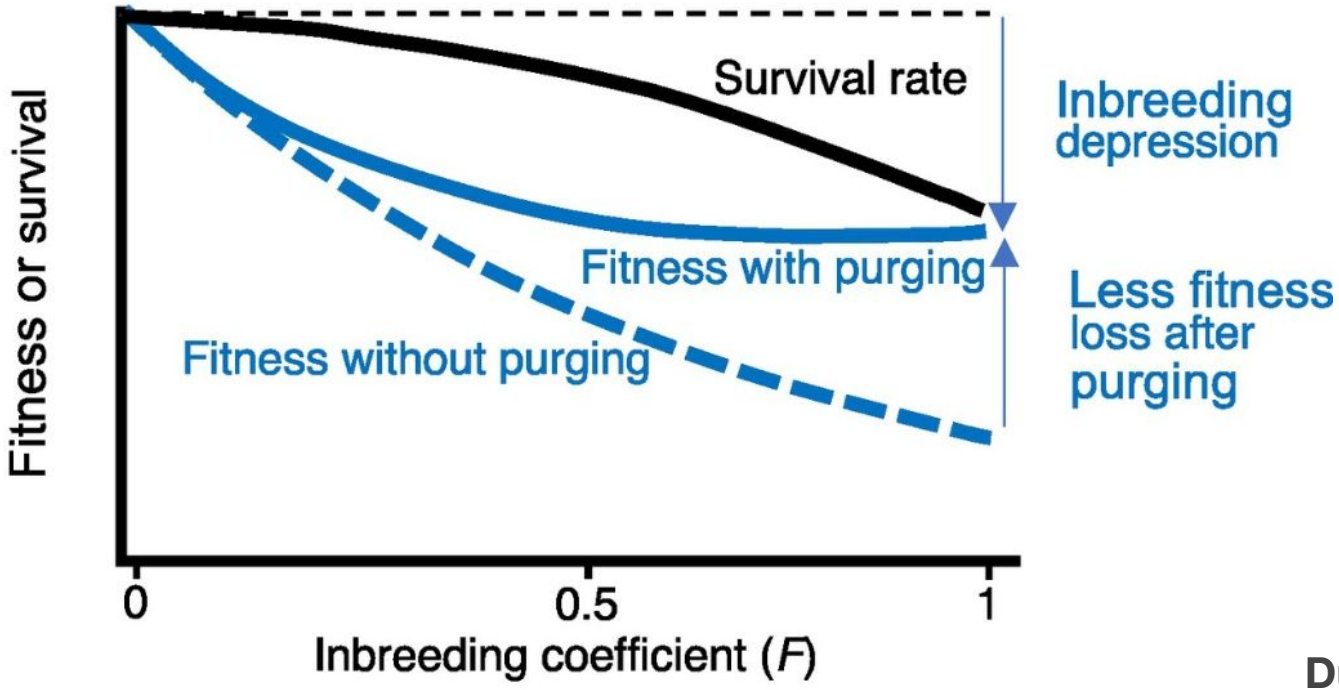
Use only the sites you can **confidently** polarize depending on their configuration and on the **missing data** in ingroup and outgroup**s**

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



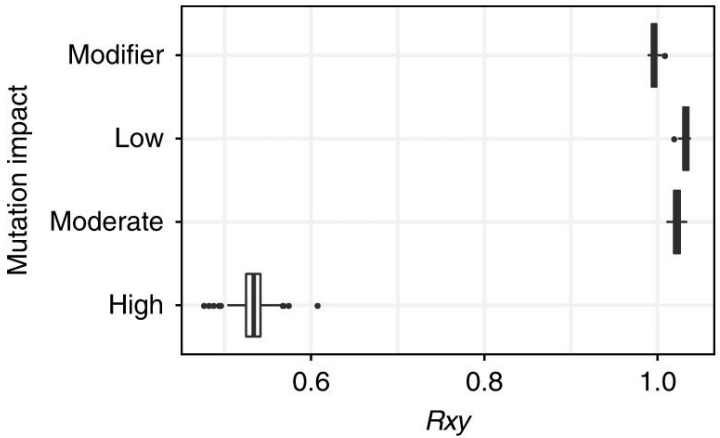
ONE MORE THING:
DO YOU KNOW WHAT PURGING MEANS WHEN TALKING OF GENETIC LOAD?

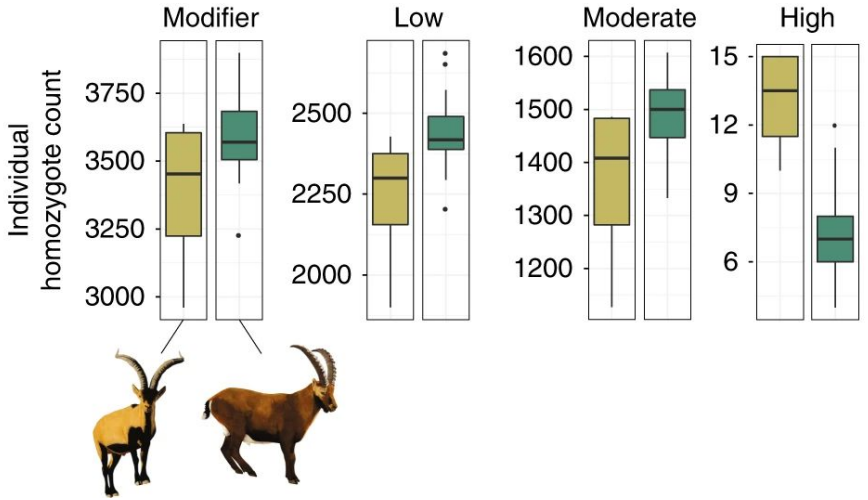
PURGING
DELETERIOUS MUTATIONS



Dussex et al 2023

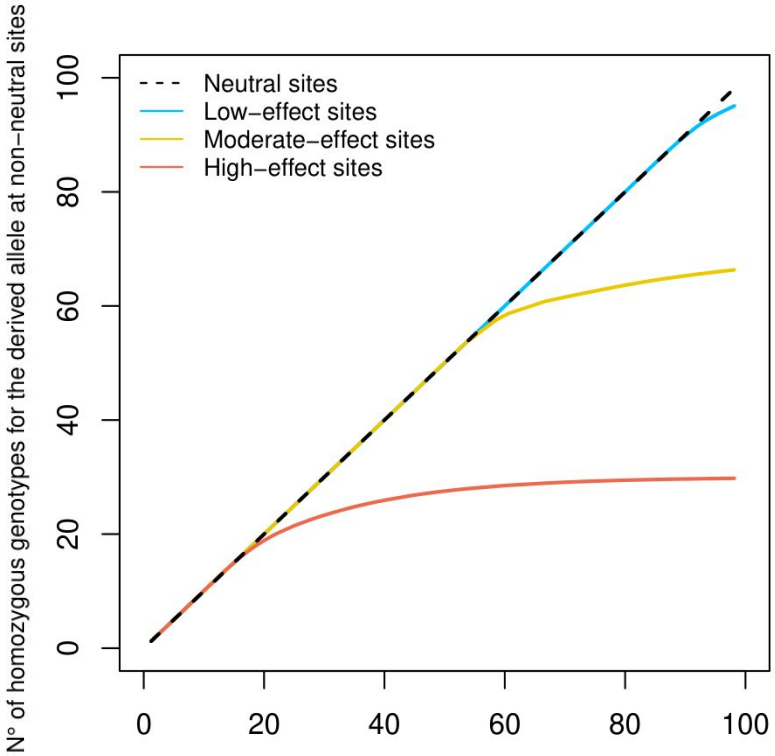
PURGING
DELETERIOUS MUTATIONS





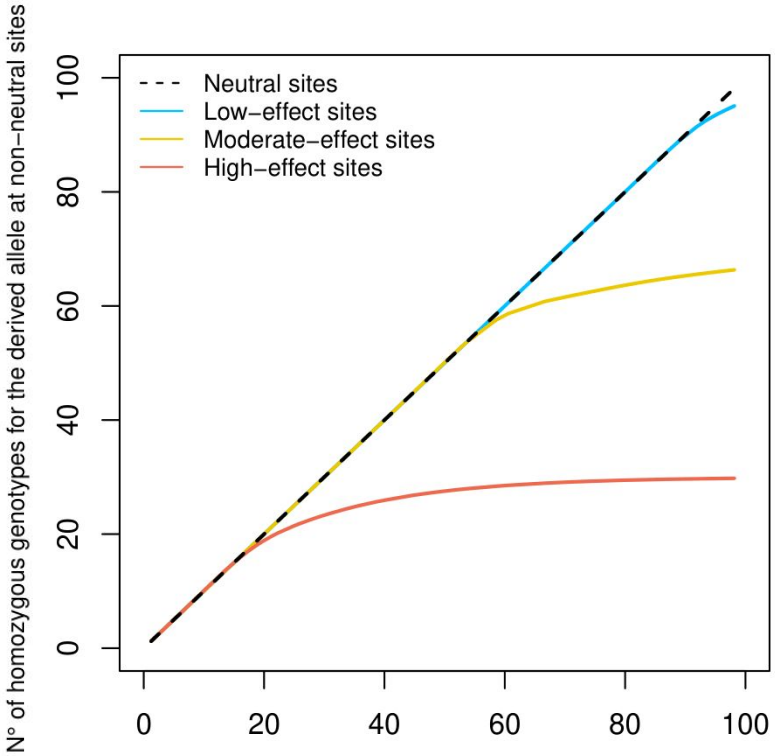
Grossen et al 2020

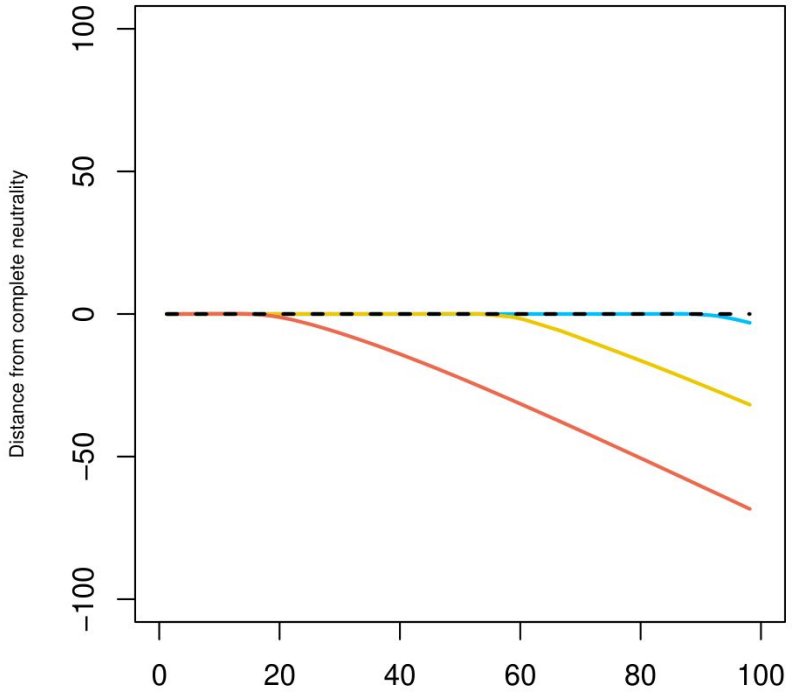
PURGING
DELETERIOUS MUTATIONS



Giannelli et al *in prep*

PURGING
DELETERIOUS MUTATIONS

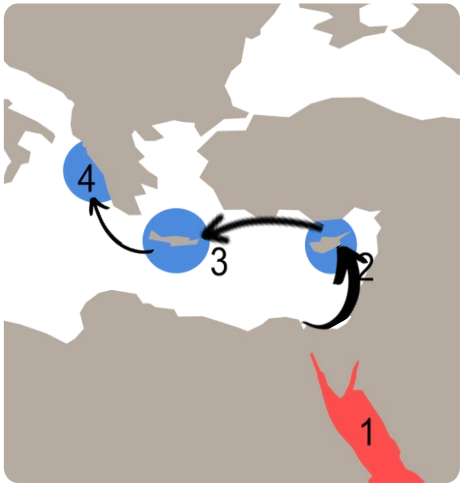




N° of homozygous genotypes for the derived allele at neutral sites

Giannelli et al *in prep*

PURGING
DELETERIOUS MUTATIONS



Giannelli et al *in prep*

QUESTIONS?

QUESTIONS?

INTERESTED IN A WORKSHOP FOCUSED ON CONSERVATION GENOMICS?

Here is a short questionnaire to let us know:
https://forms.gle/E2iAMHjwbbR5aamQ8