

2024 Phylogenomics and Population Genomics:
Inference and Applications

Population Genomics (I): Neutral and Adaptive Variation

Sebastian E. Ramos-Onsins
Centre for Research in
Agricultural Genomics (CRAG)



UNIVERSITAT DE
BARCELONA

A population and their variability

TWO main key points to take into account:

- What sample?

The sample is a likely representation of the population

- What DNA regions?

Each region has information of the past events of the population



Canadian Poultry Magazine

A population and their variability

... but HOW to choose a sample?

A population and their variability

... but HOW to choose a sample?

- Difficult task!

A population and their variability

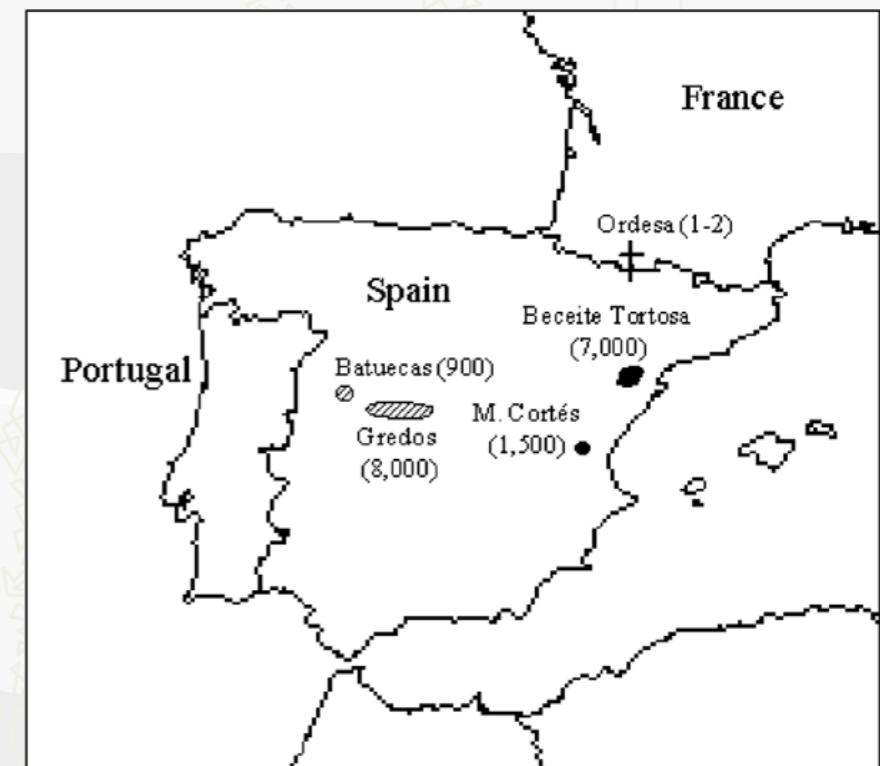
... but HOW to choose a sample?

- Difficult task!
- It depends on the ecological and geographical patterns:
 - Species distribution in patches or continuous?

A population and their variability

... but HOW to choose a sample?

- Difficult task!
- It depends on the ecological and geographical patterns:
 - Species distribution in patches or continuous?

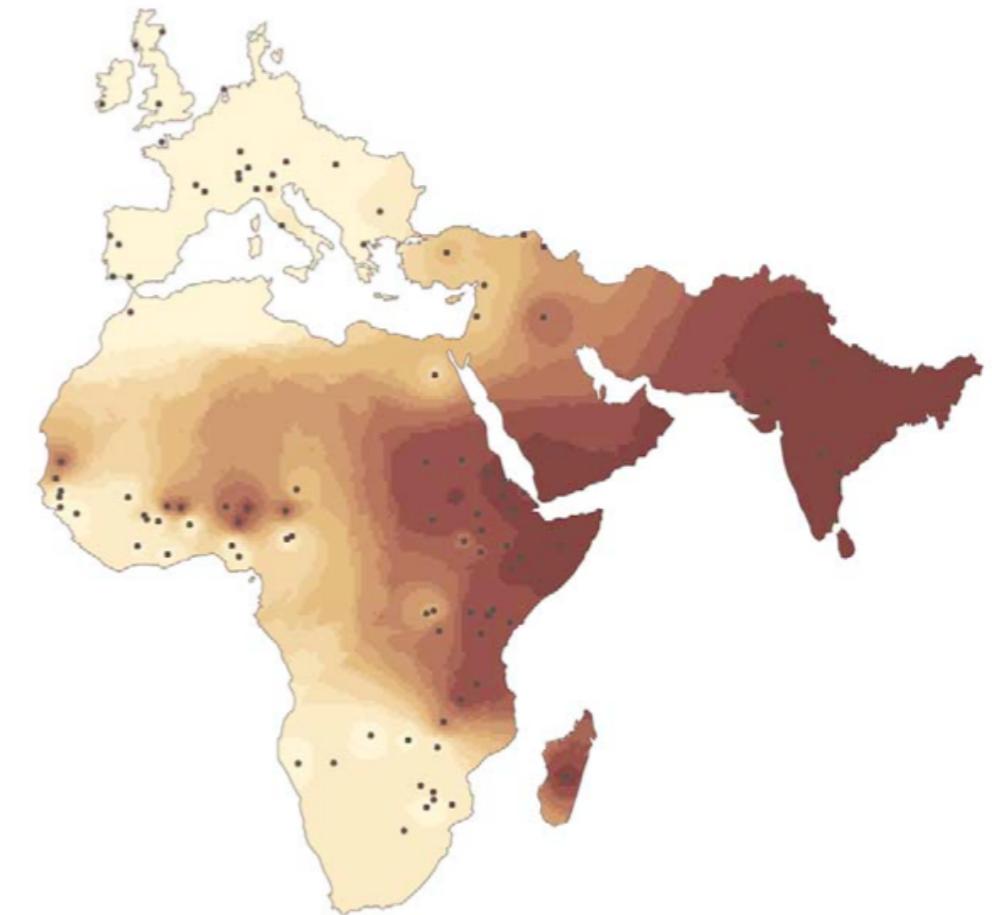


Capra pyrenaica hispanica
Capra pyrenaica victoriae
+ **Capra pyrenaica pyrenaica**

A population and their variability

... but HOW to choose a sample?

- Difficult task!
- It depends on the ecological and geographical patterns:
 - Species distribution in patches or continuous?

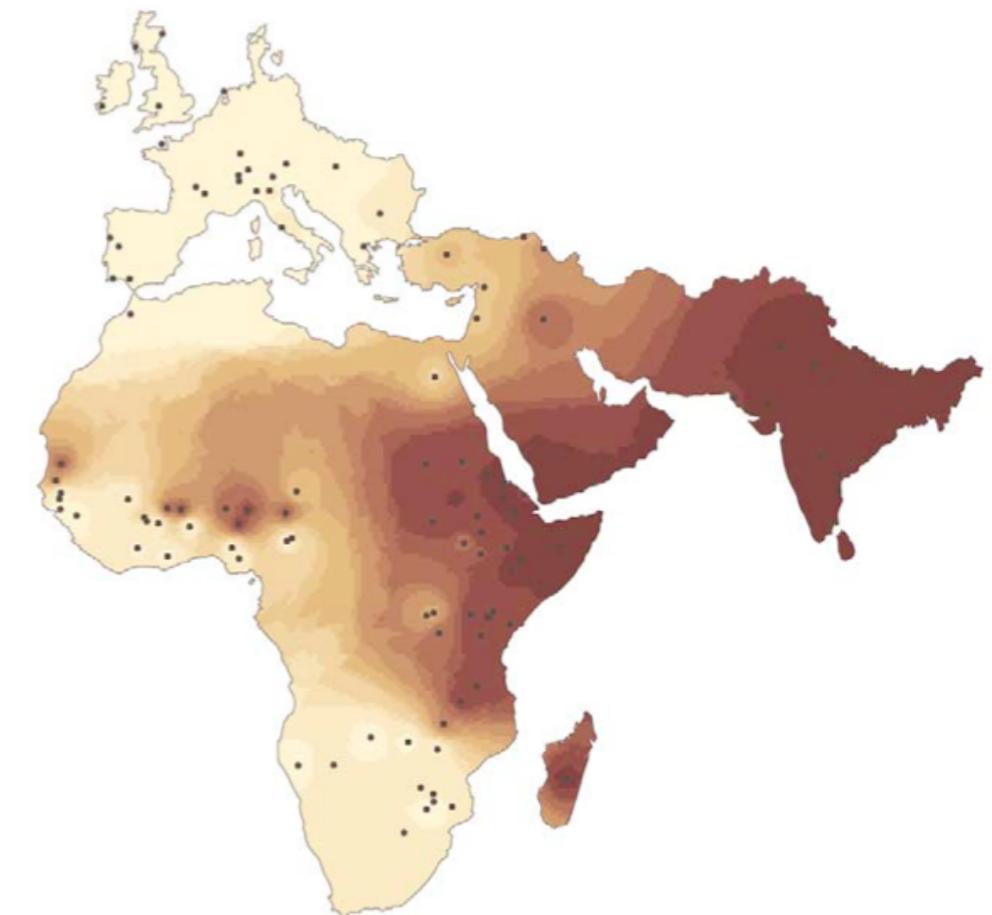


Zeder *et al.* (2006); Freeman *et al.* (2005)

A population and their variability

... but HOW to choose a sample?

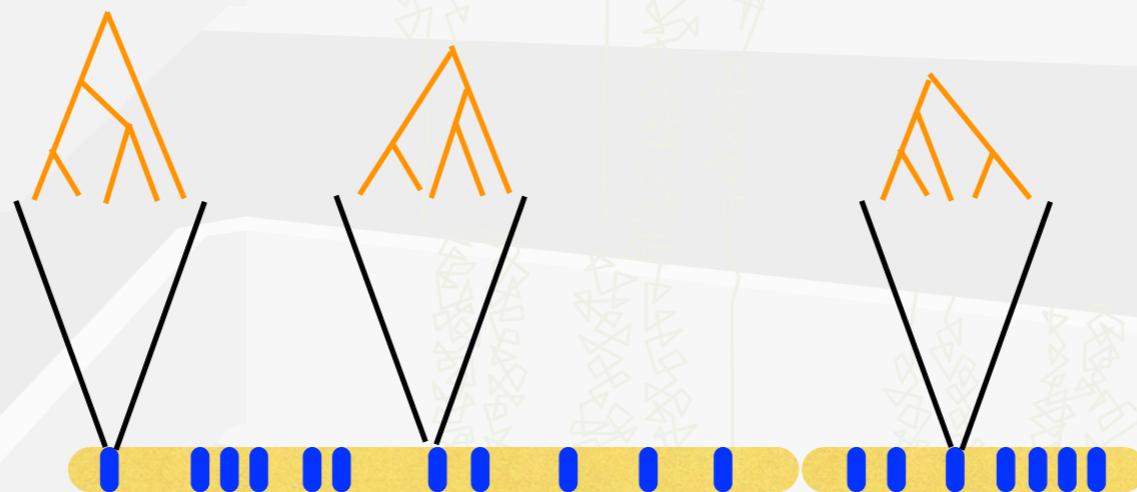
- Difficult task!
- It depends on the ecological and geographical patterns:
 - Species distribution in patches or continuous?
- Also on the objectives of the study
 - Local or whole species study?



Zeder *et al.* (2006); Freeman *et al.* (2005)

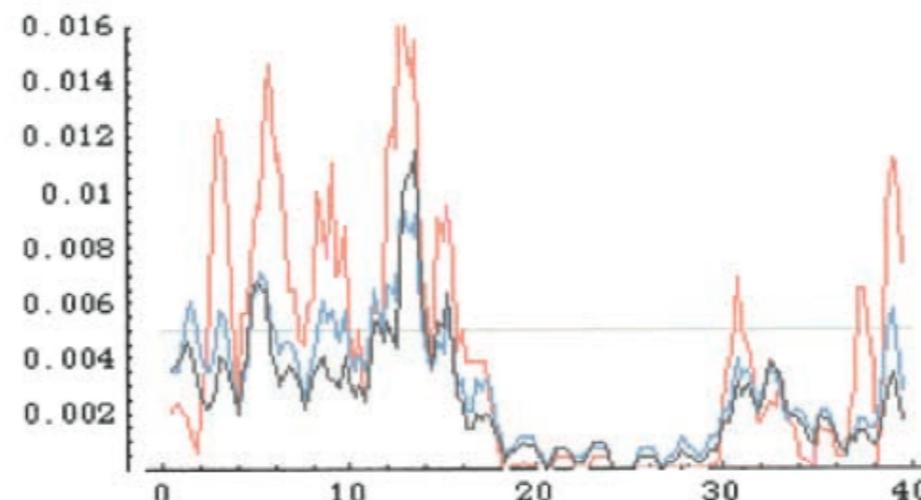
A population and their variability

- What DNA regions?
- If we want to describe the population history:
 - Better to work with short fragments across all the genome (neutral)



A population and their variability

- What DNA regions?
- If we want to describe the population history:
Better to work with short fragments across all the genome (neutral)
- If we are looking for Adaptive processes, positive selection:
Better to work with long linked regions and focus on functional positions



Analysis of Nucleotide Variability

Look at levels and patterns of variability

| sel=0 | 281 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
|-------------|-----|---|
| AY5Brek | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYLW2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYP2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYSBlackJ | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| DQ3Kwb | | AACCTCTAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| DQ0Kwb | | AACCTCTAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB1riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB2riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB5riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChYun | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChFu | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChHai | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWbdb6 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb26 | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFwb371 | | AACCTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb48 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb47 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb65 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EUCHiN | | AACCTCTAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbViet | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| TypeIILanyu | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| TypeIILanyu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYMeish1 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYMeish2 | | AACCTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFXiang | | GATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYJNP13 | | AACCTCCAGTCAACATGCGTATCACCACCAACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| Phaco | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |

Analysis of Nucleotide Variability

Look at levels and patterns of variability

| sel=0 | 281 | |
|-------------|---|--|
| AY5Brek | AATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYLW2 | AATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYP2 | AATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYSBlackJ | AACCTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| DQ3Kwb | AACTCTAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| DQ0Kwb | AACTCTAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB1riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB2riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB5riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChYun | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChFu | AACCTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChHai | AATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWbdb6 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb26 | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFwb371 | AACCTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb48 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb47 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb65 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EUCHiN | AACTCTAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbViet | AACCTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| TypeIILanyu | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| TypeIILanyu | AATTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYMeish1 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYMeish2 | AACTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFXiang | GATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYJNP13 | AACTCCAGTCAACATGCGTATCACTCCACTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| Phaco | AATTCCAGTCAACATGCGTATCACTCCATTAGATACGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |

Analysis of Nucleotide Variability

Look at levels and patterns of variability

| sel=0 | 281 | |
|-------------|---|--|
| AY5Brek | AATTCCAGTCAACATGCGTATCACTCCACTTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYLW2 | AATTCCAGTCAACATGCGTATCACTCCACTTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYP2 | AATTCCAGTCAACATGCGTATCACTCCACTTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYSBlackJ | AACTCCAGTCAACATGCGTATCACTCCACTTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| DQ3Kwb | AACTCTAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| DQ0Kwb | AACTCTAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB1riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB2riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AB5riuk | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChYun | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChFu | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbChHai | AATTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWbdb6 | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb26 | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFwb371 | AACTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb48 | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb47 | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| ChWb65 | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EUCHiN | AACTCTAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFWbViet | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| TvpeIIlanyu | AACTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| TvpeIILanyu | AATTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYMeish1 | AACCTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYMeish2 | AACCTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| EFXiang | GATTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| AYJNP13 | AACTCCAGTCAACATGCGTATCACTCCACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |
| Phaco | AATTCCAGTCAACATGCGTATCACTCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG | |

Analysis of Nucleotide Variability

- Calculate statistics that define variability:
 - Number of total mutations
 - Average number of differences between 2 sequences
 - Frequency spectrum of the mutations ... between others...

Analysis of Nucleotide Variability

- Calculate statistics that define variability:
 - Number of total mutations
 - Average number of differences between 2 sequences
 - Frequency spectrum of the mutations ... between others...

| sel=0 | 281 | AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
|-------------|-----|--|
| AY5Brek | | AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYLW2 | | AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYP2 | | AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYSBlackJ | | AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| DQ3Kwb | | AACTCTAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| DQ0Kwb | | AACTCTAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB1riuk | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB2riuk | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB5riuk | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChYun | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChFu | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChHai | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWbdb6 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb26 | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFwb371 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb48 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb47 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb65 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EUCHiN | | AACTCTAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbViet | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| TypeIILanyu | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| TypeILanyu | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYMeish1 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYMeish2 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFXiang | | GATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYJNP13 | | AACTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| Phaco | | AATTCCAGTCAACATGCGTATCACCACCATAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |

Analysis of Nucleotide Variability

- Number of mutations:

Analysis of Nucleotide Variability

- Number of mutations:

n=26

| sel=0 | 281 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
|-------------|-----|--|
| AY5Brek | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYLW2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYP2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYSBlackJ | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| DQ3Kwb | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| DQ0Kwb | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB1riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB2riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AB5riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChYun | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChFu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbChHai | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWbdb6 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb26 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFwb371 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb48 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb47 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| ChWb65 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EUChiN | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFWbViet | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| TypeIILanyu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| TypeIILanyu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYMeish1 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYMeish2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| EFXiang | | GATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| AYJNP13 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |
| Phaco | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCAGGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG |

Analysis of Nucleotide Variability

- ## ● Number of mutations:

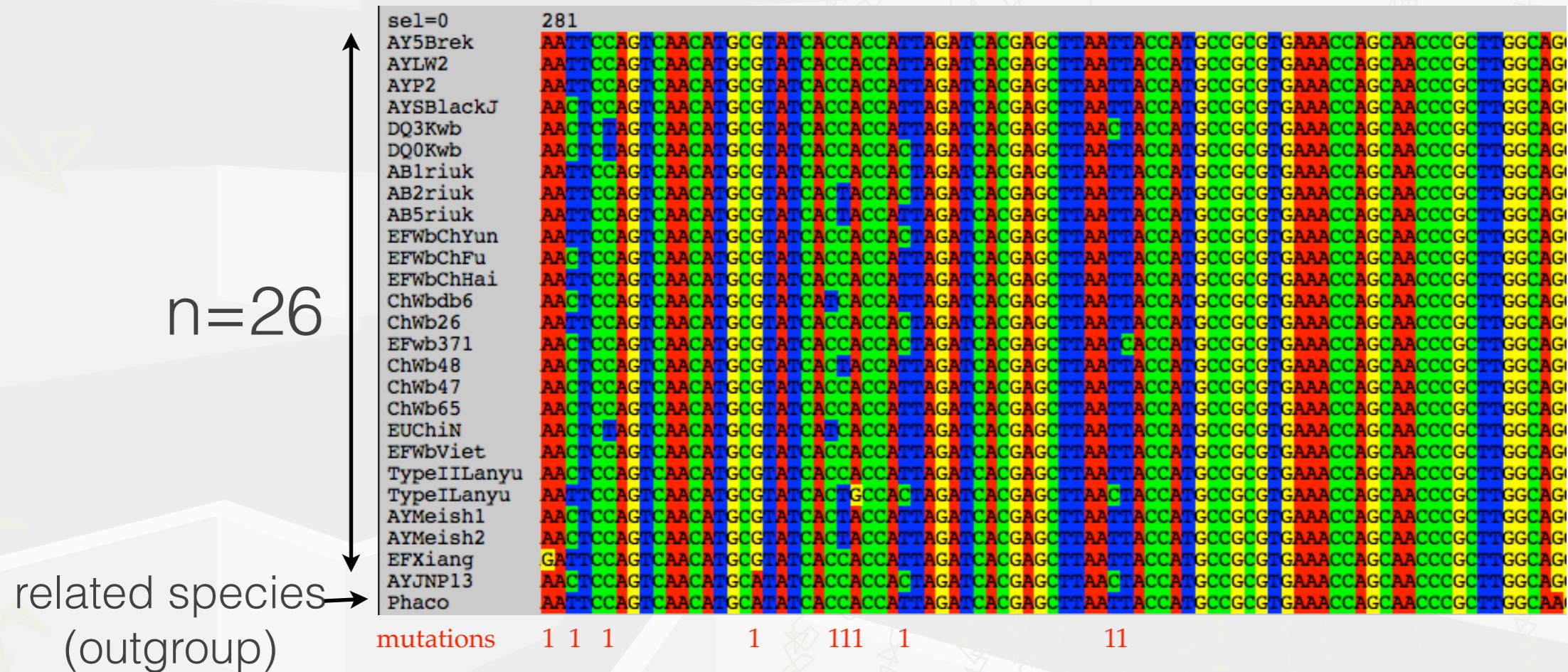
n=26

related species →
(outgroup)

| sel=0 | 281 |
|-------------|--|
| AY5Brek | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYLW2 | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYP2 | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYSBlackJ | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| DQ3Kwb | AACTCTAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| DQ0Kwb | AACTCTAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AB1riuk | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AB2riuk | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AB5riuk | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFWbChYun | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFWbChFu | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFWbChHai | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| ChWbdb6 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| ChWb26 | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFwb371 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| ChWb48 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| ChWb47 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| ChWb65 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EUCHiN | AACTCTAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFWbViet | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| TypeIILanyu | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| TypeIILanyu | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYMeish1 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYMeish2 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| EFXiang | GATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| AYJNP13 | AACTCCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |
| Phaco | AATTCAGTCAACATGGTATCACCACCATAGATCACGAGCTTAATTACCA |

Analysis of Nucleotide Variability

- Number of mutations:

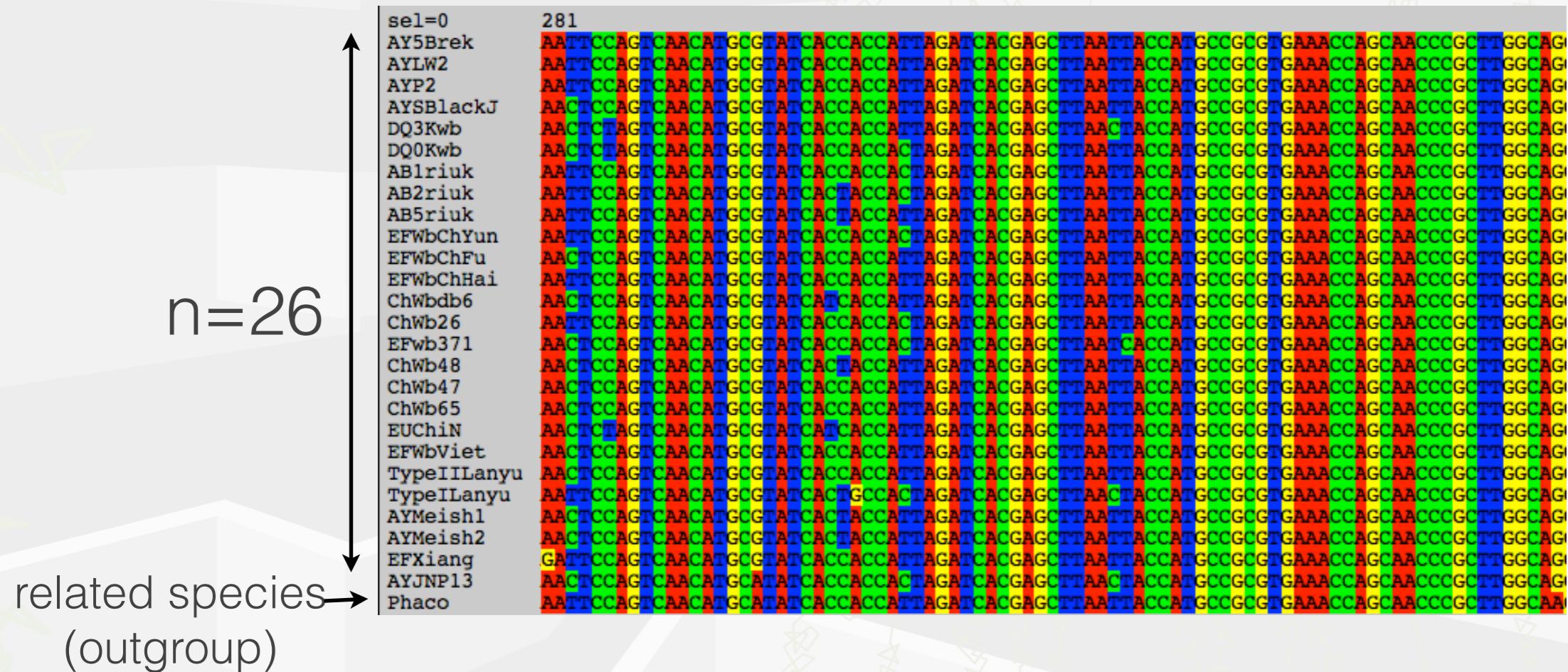


Analysis of Nucleotide Variability

- Average number of differences:

Analysis of Nucleotide Variability

- Average number of differences:



Analysis of Nucleotide Variability

- Average number of differences:

Number of combinations 2 samples in 26 total: $(26^*25)/2$

| sel=0 | 281 | Sequence Data |
|-------------|-----|---|
| AY5Brek | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYLW2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYP2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYSBlackJ | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| DQ3Kwb | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| DQ0Kwb | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB1riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB2riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AB5riuk | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChYun | | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChFu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbChHai | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWbdb6 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb26 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFwb371 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb48 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb47 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| ChWb65 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EUChiN | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFWbViet | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| TypeIILanyu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| TypeIILanyu | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYMeish1 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYMeish2 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| EFXiang | | GATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| AYJNP13 | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |
| Phaco | | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG |

Analysis of Nucleotide Variability

- Average number of differences:

Number of combinations 2 samples in 26 total: $(26^*25)/2$

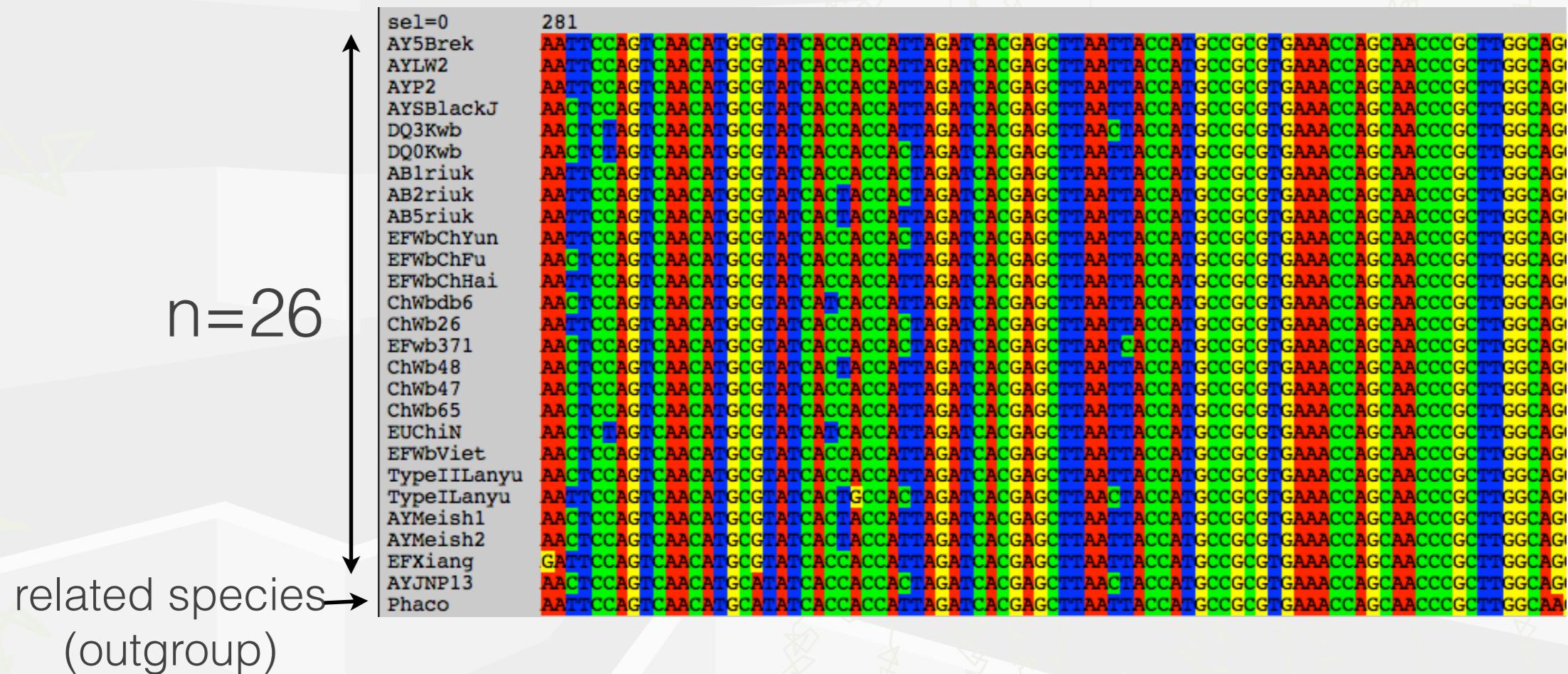
| sel=0 | 281 | |
|-------------|---|--|
| AY5Brek | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYLW2 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYP2 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYSBlackJ | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| DQ3Kwb | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| DQ0Kwb | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AB1riuk | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AB2riuk | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AB5riuk | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFWbChYun | AATTCCAGTCAACATGCGTATCACCACCAACTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFWbChFu | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFWbChHai | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| ChWbdb6 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| ChWb26 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFwb371 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| ChWb48 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| ChWb47 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| ChWb65 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EUChiN | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFWbViet | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| TypeIILanyu | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| TypeIILanyu | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYMeish1 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYMeish2 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| EFXiang | GATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| AYJNP13 | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |
| Phaco | AATTCCAGTCAACATGCGTATCACCACCAATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCGCTTGGCAG | |

Analysis of Nucleotide Variability

- Frequency spectrum of the mutations:

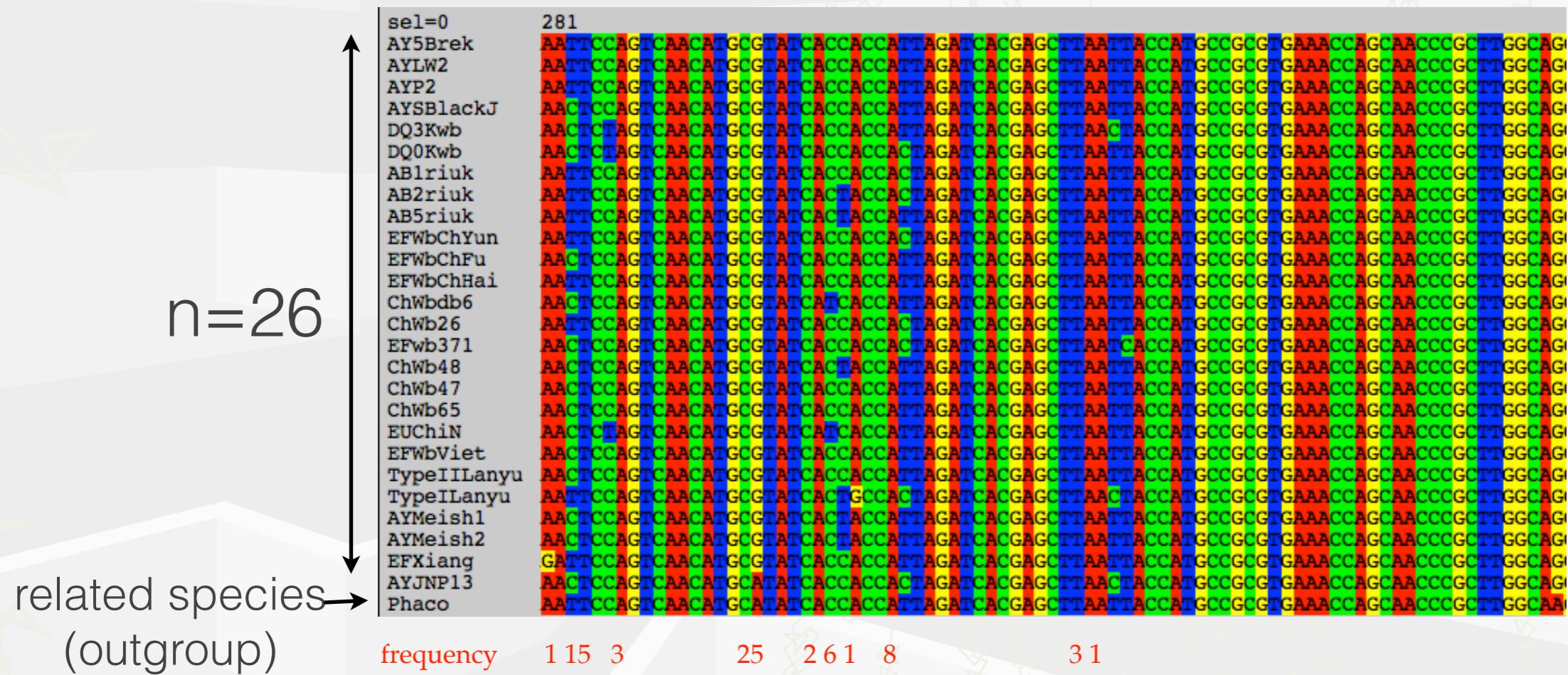
Analysis of Nucleotide Variability

- Frequency spectrum of the mutations:



Analysis of Nucleotide Variability

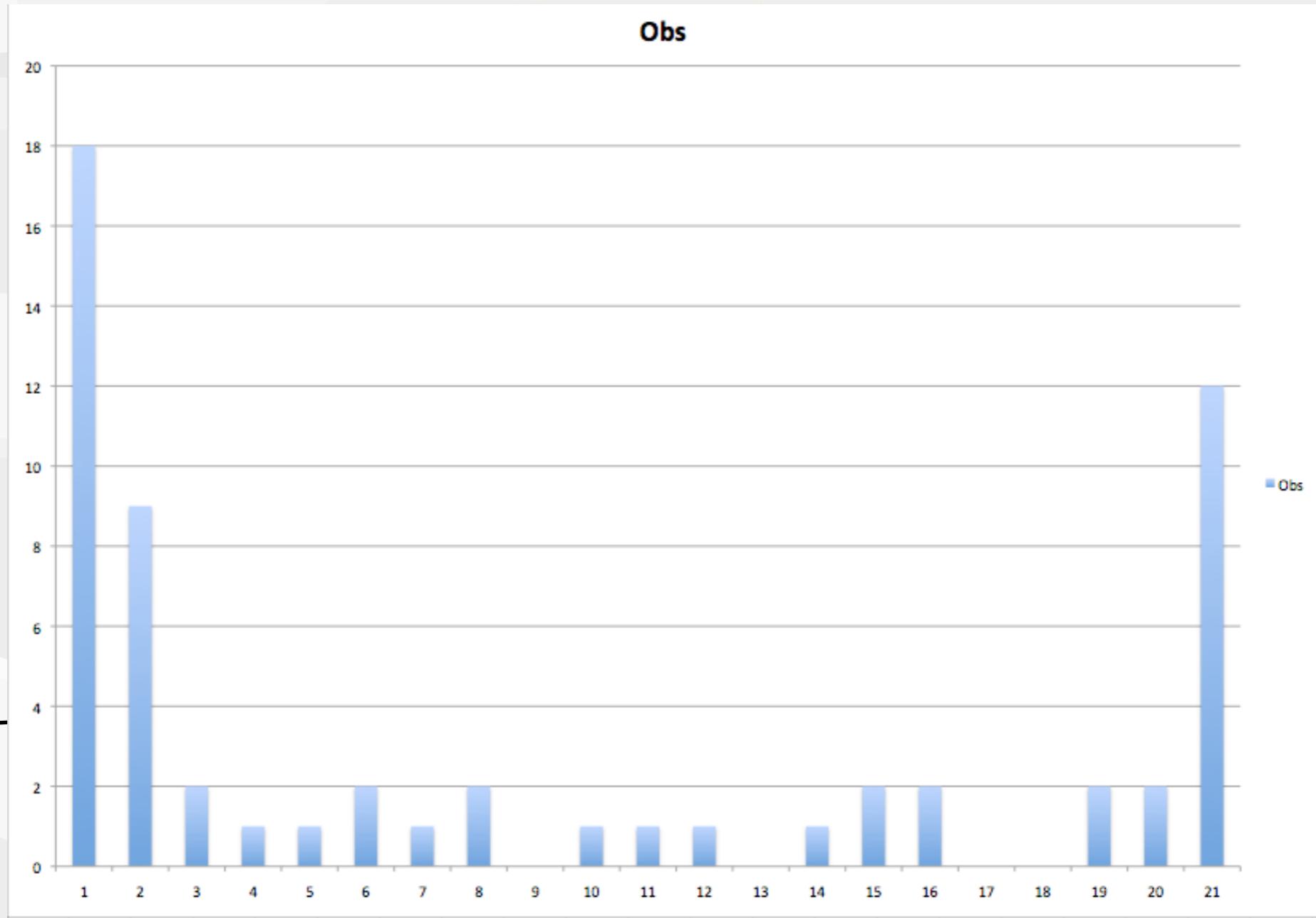
- Frequency spectrum of the mutations:



Analysis of Nucleotide Variability

- Frequency spectrum of the mutations:

n=26



Analysis of Nucleotide Variability

OK... We have calculated some statistics to estimate levels and patterns of variability.
... but What it means?

Levels and patterns of variability can be compared with
Evolutionary models to see **similarities and differences**

A population and its variability

Evolutionary models

Common Assumptions:

- Panmixia
- Discrete separated generations

Others:

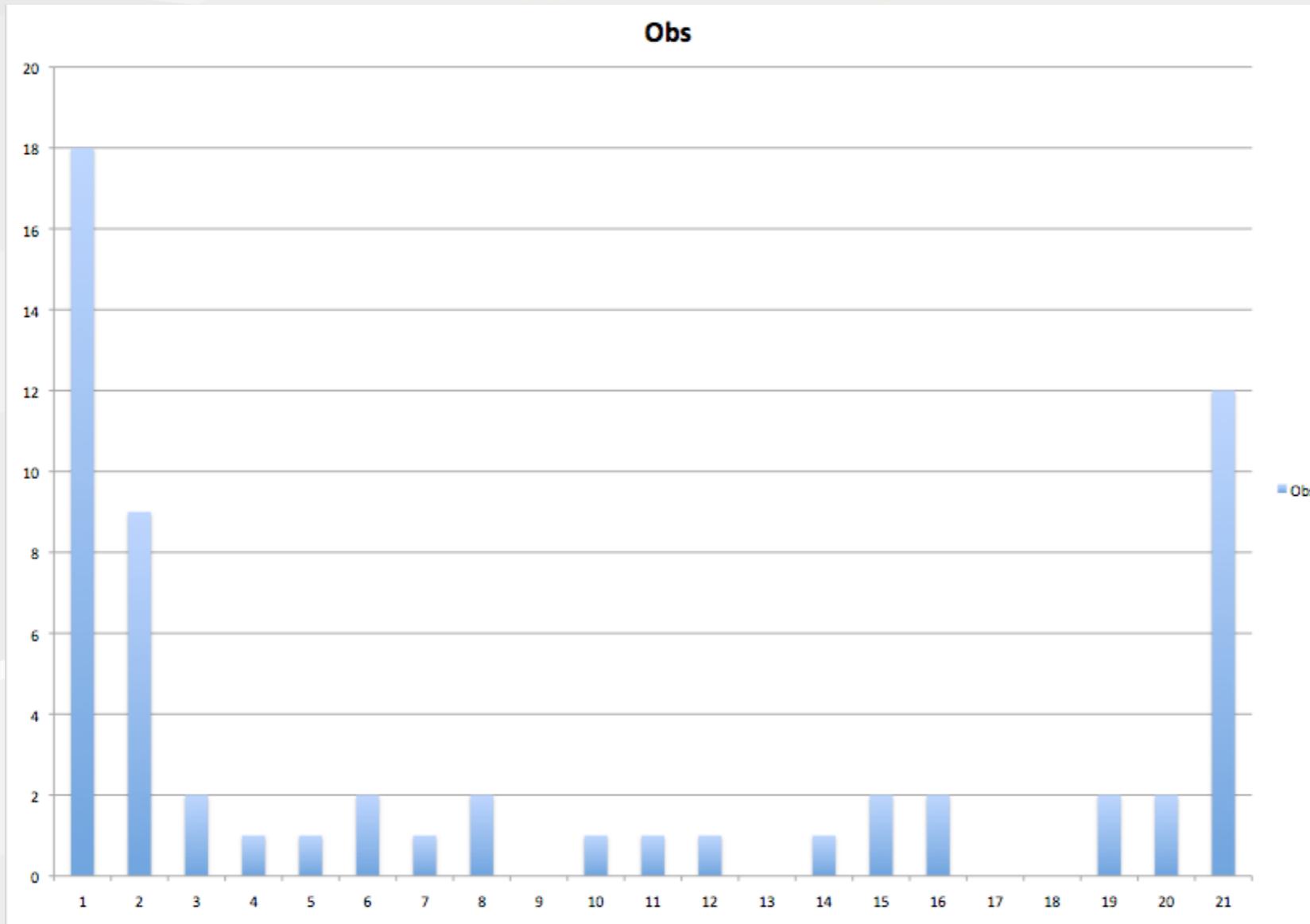
- Equal population size
- No migration
- No selection



from SVA National Veterinary Institute

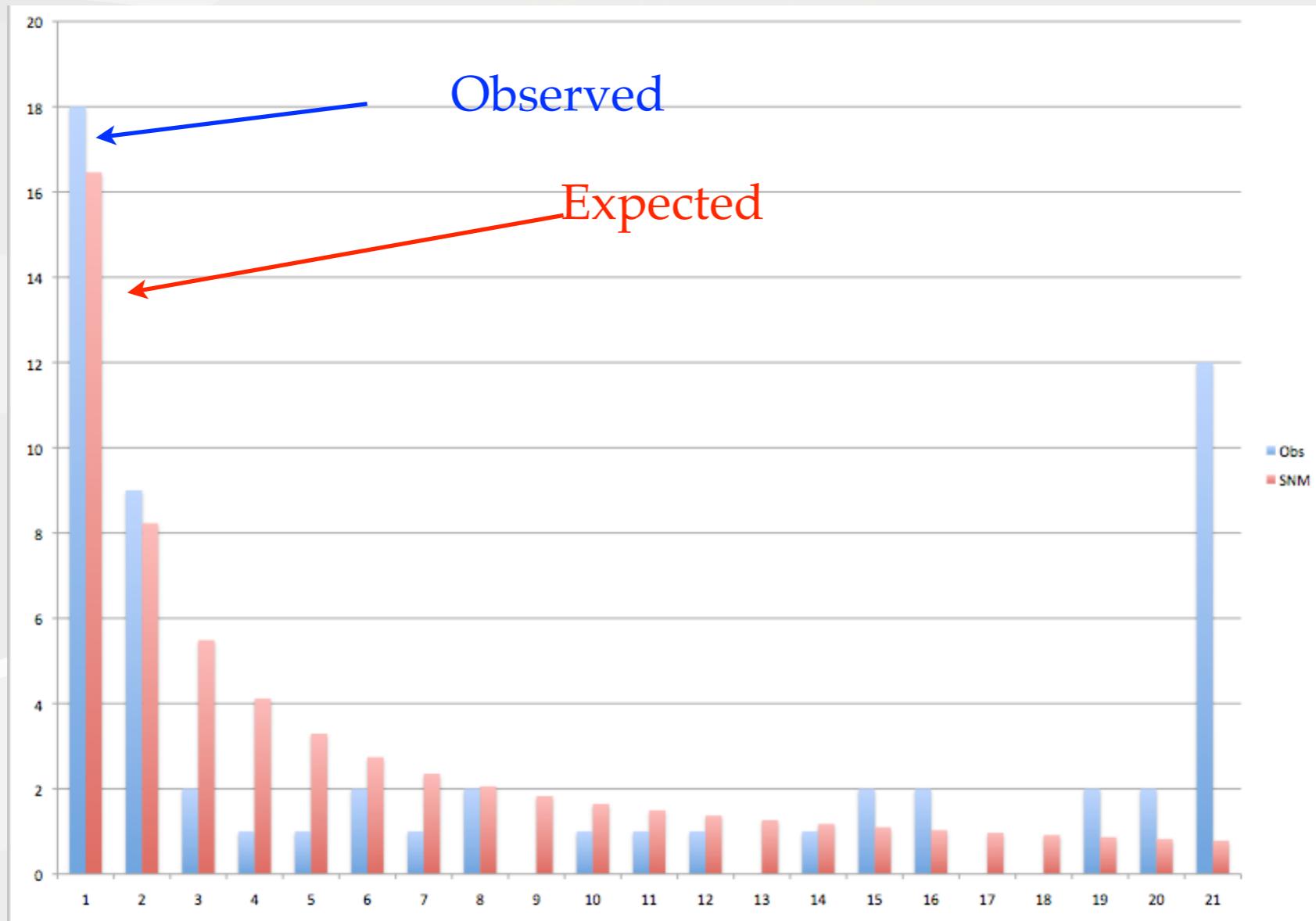
Analysis of Nucleotide Variability

For example, the frequency spectrum of mutations:



Analysis of Nucleotide Variability

For example, the frequency spectrum of mutations:



Analysis of Nucleotide Variability

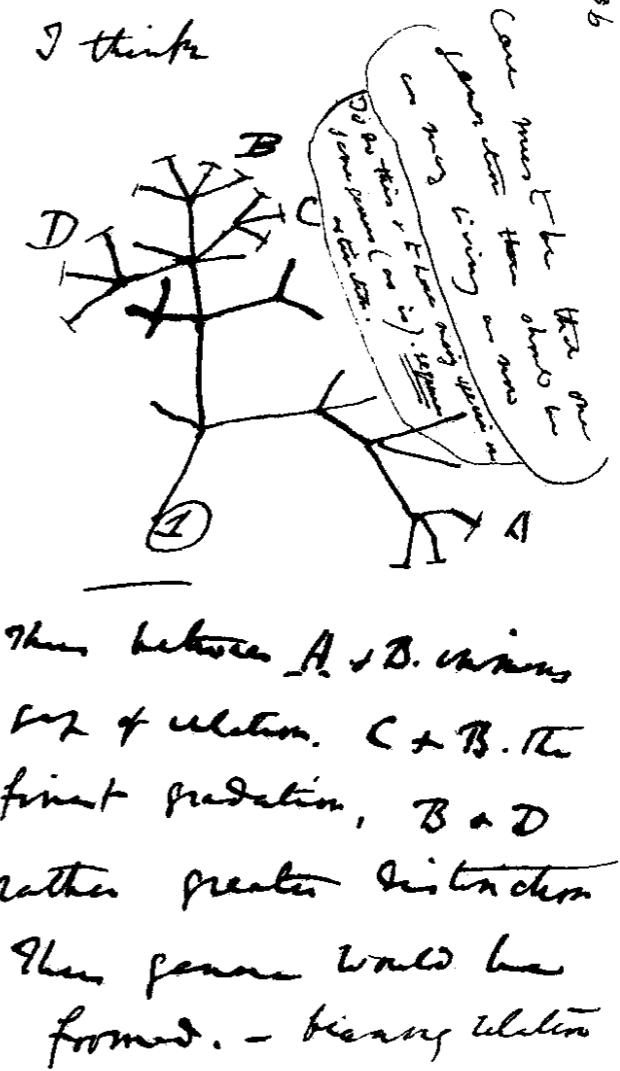
So, we are interested in knowing the **expected** levels and patterns of variability
(under some conditions)

Let's see then a bit of theory

A bit about Coalescent

What is the expected variability?

Imagine we collect a sample from one population.

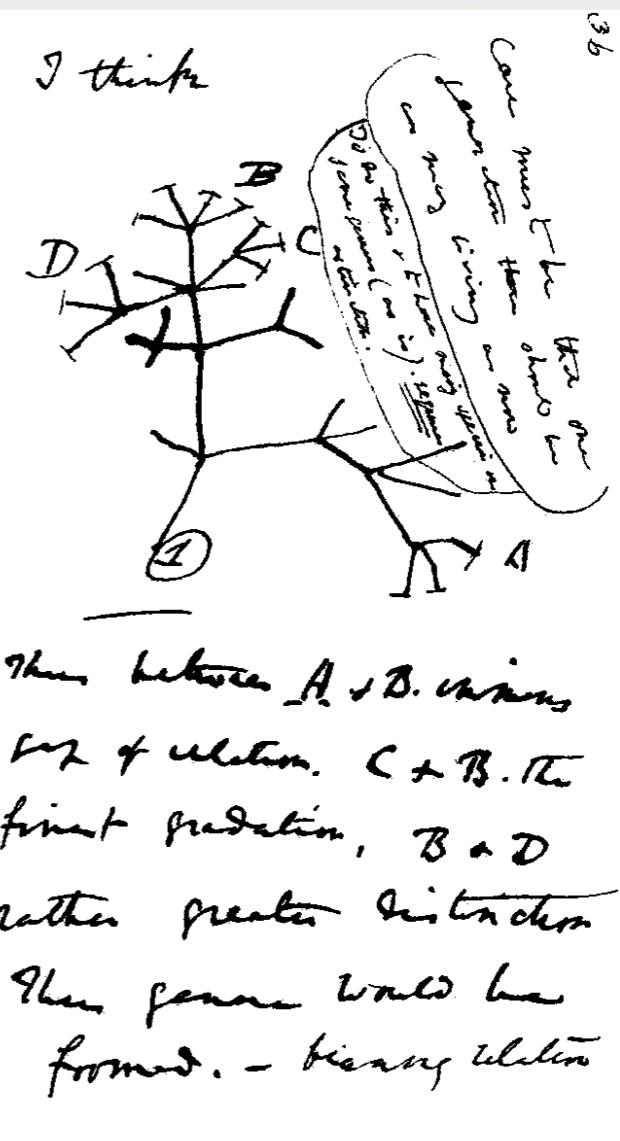


A bit about Coalescent

What is the expected variability?

Imagine we collect a sample from one population.

- We know all individuals have a common origin.

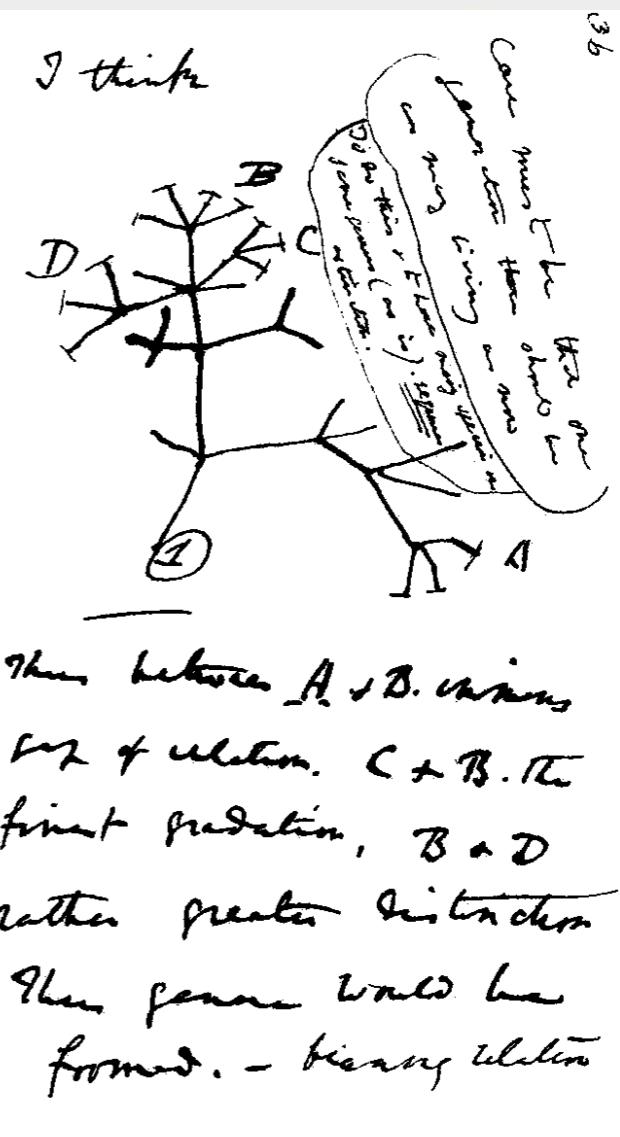


A bit about Coalescent

What is the expected variability?

Imagine we collect a sample from one population.

- We know all individuals have a common origin.
- Then, to understand the variation between individuals we must reconstruct the history until the ancestor.



A bit about Coalescent

What is the expected variability?

A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.

A bit about Coalescent

What is the expected variability?

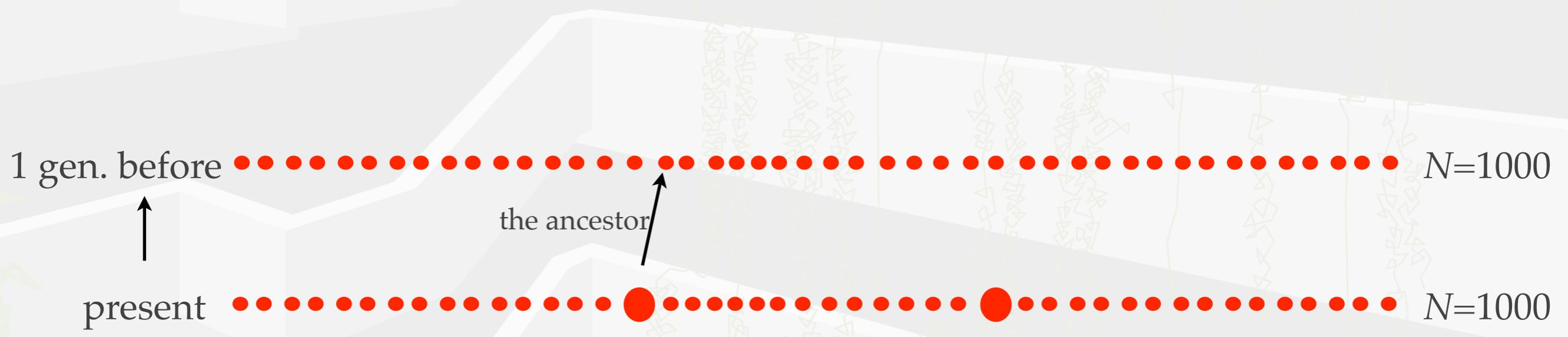
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

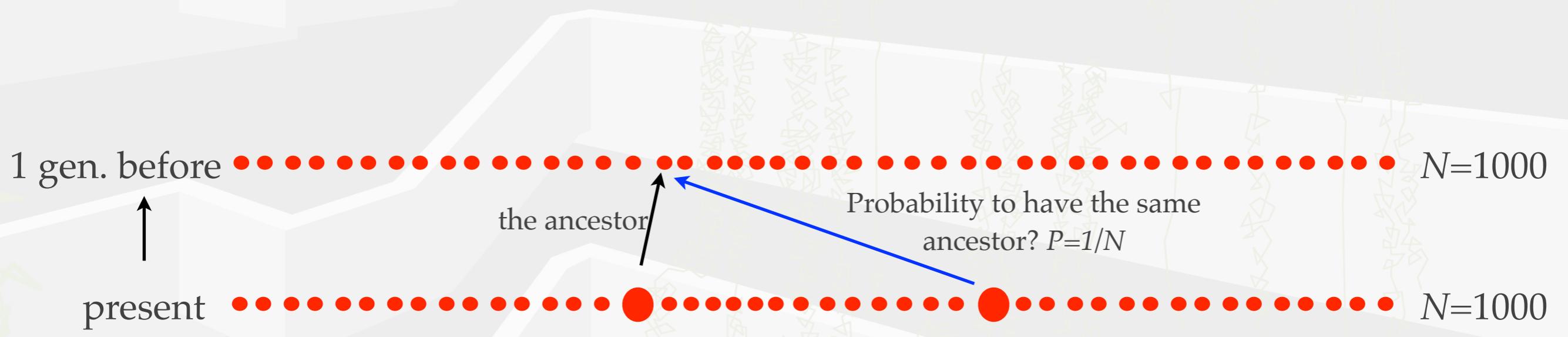
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

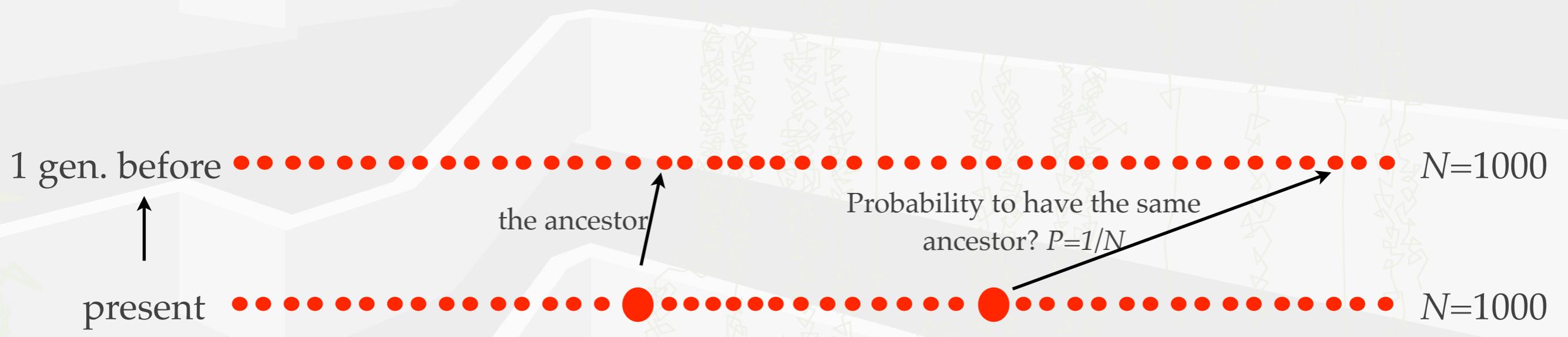
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

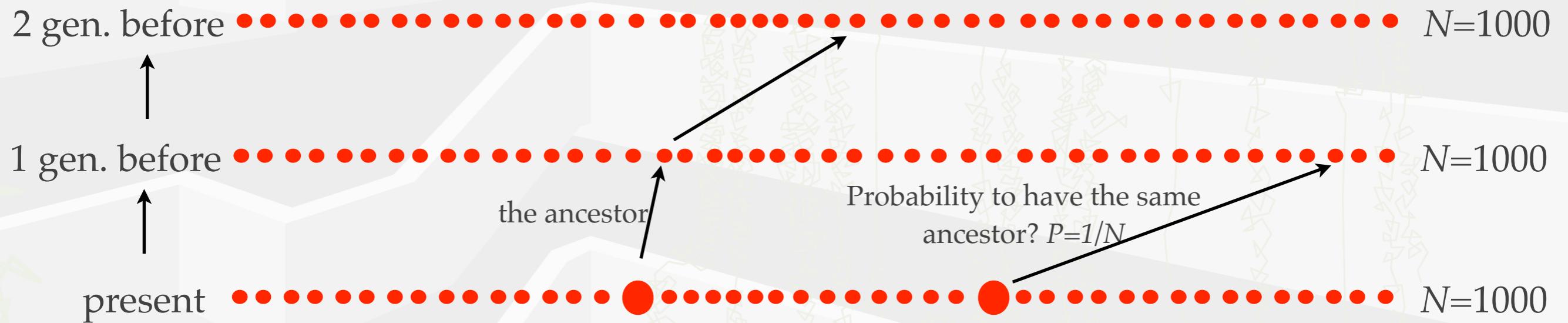
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

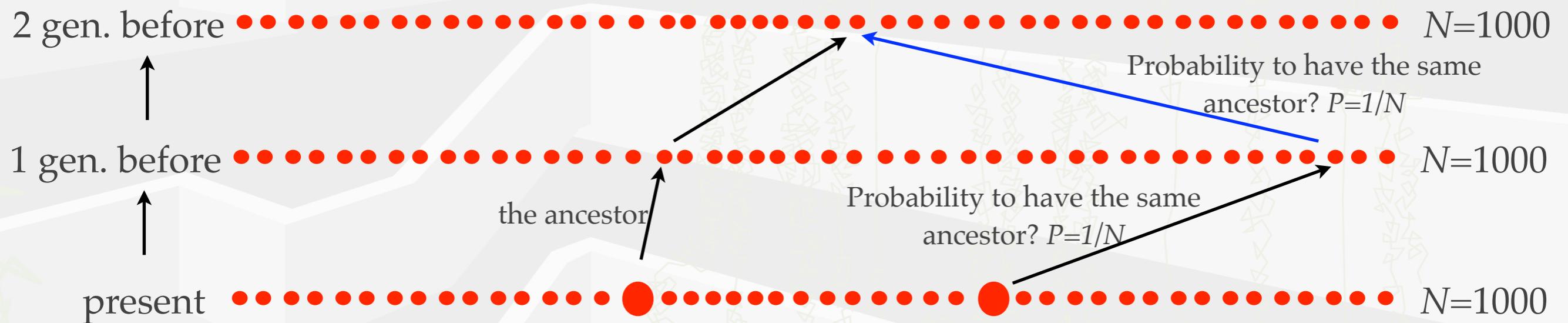
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

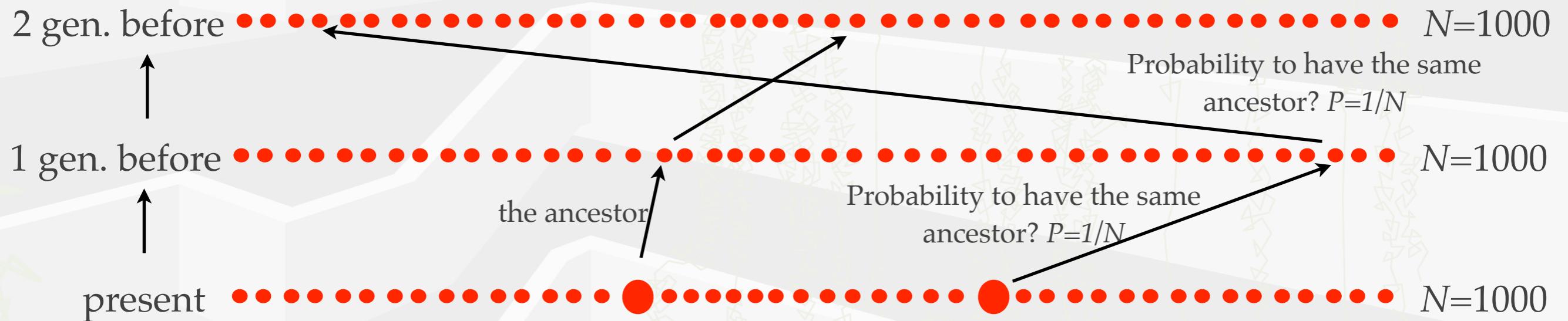
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

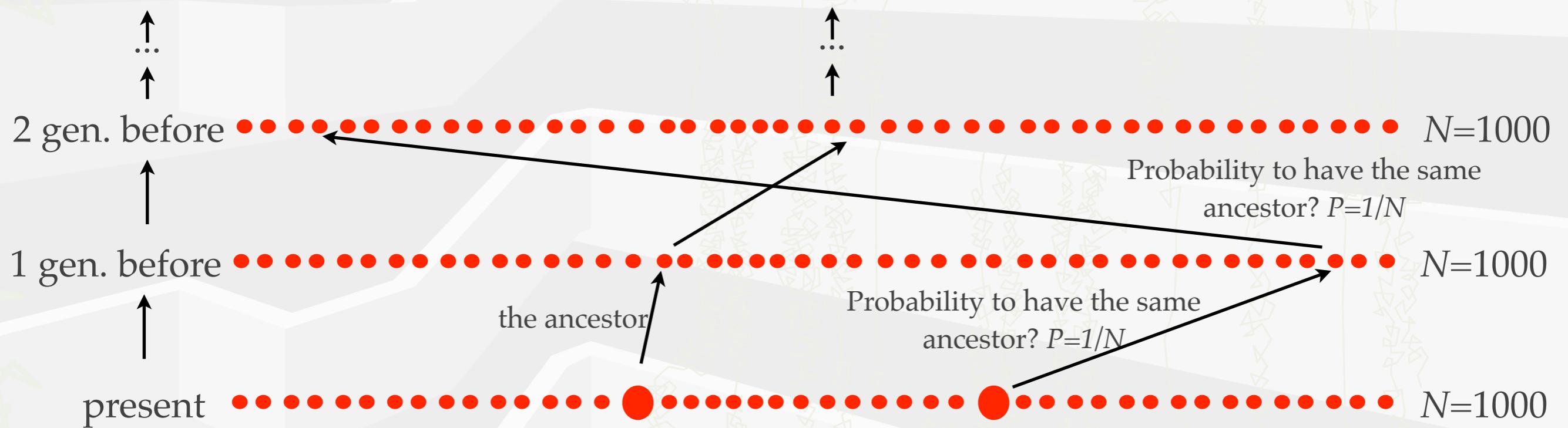
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.

A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of N=1000.
- Let's reconstruct the history and the number of expected mutations.

Probability of coalescence at generation t is:

$$P(t) = (1-P)^{(t-1)} \cdot P \approx P \cdot e^{-P \cdot t}$$

A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of N=1000.
- Let's reconstruct the history and the number of expected mutations.

Probability of coalescence at generation t is:

$$P(t) = (1-P)^{(t-1)} \cdot P \approx P \cdot e^{-P \cdot t}$$

... which is an exponential distribution when t is large
and P small

A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of N=1000.
- Let's reconstruct the history and the number of expected mutations.

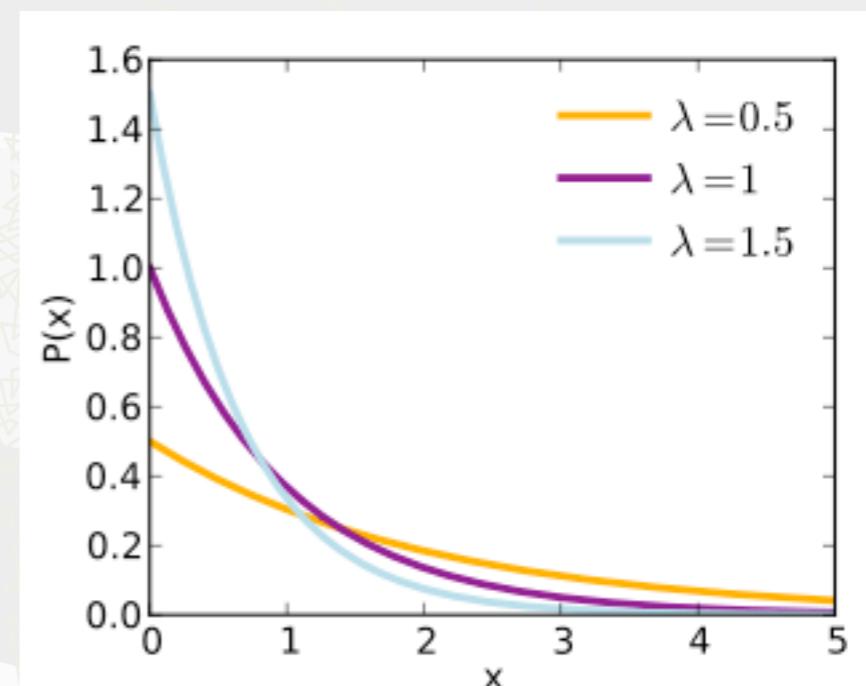
Probability of coalescence at generation t is:

$$P(t) = (1-P)^{(t-1)} \cdot P \approx P \cdot e^{-P \cdot t}$$

... which is an exponential distribution when t is large and P small

The mean is:

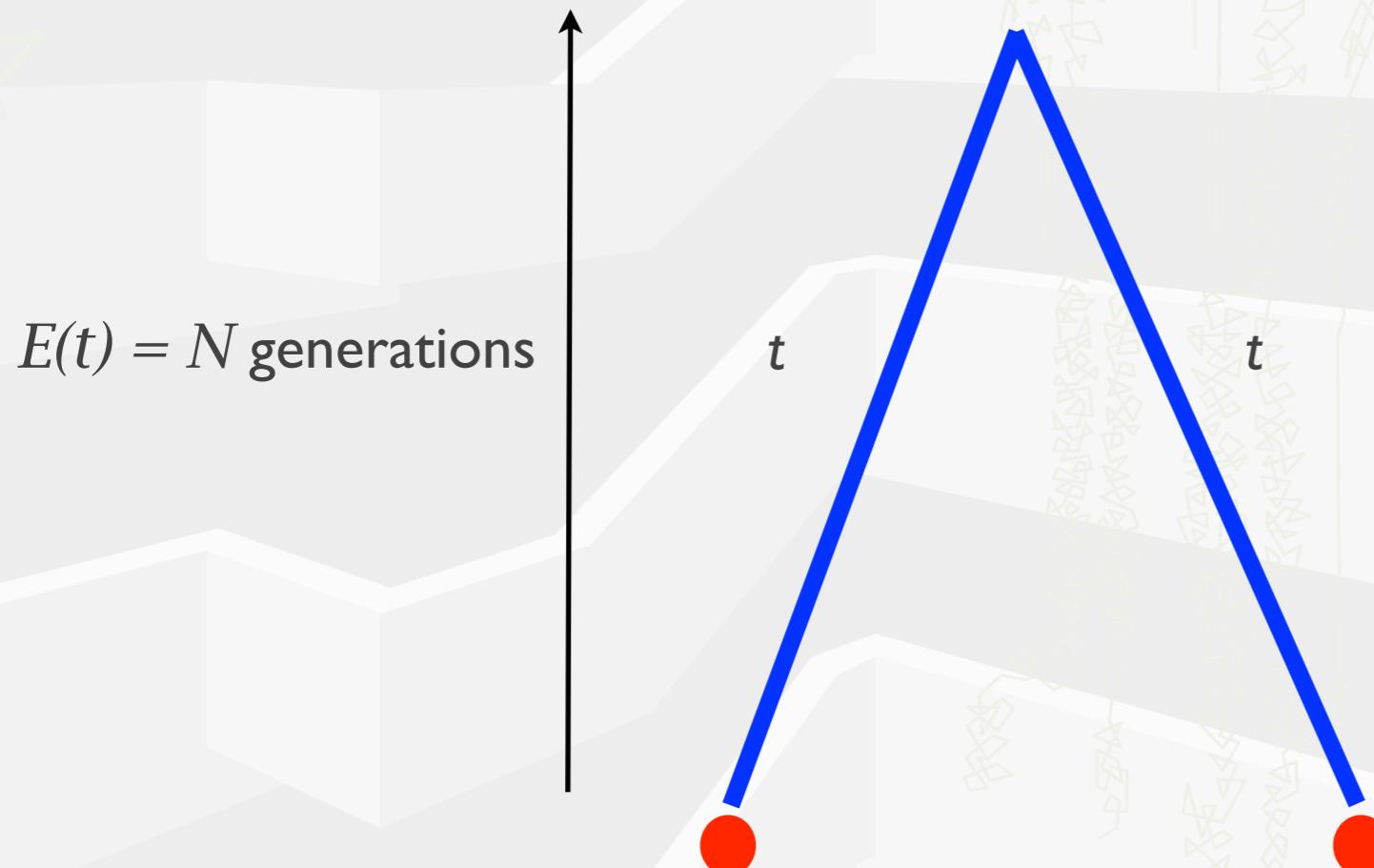
$$E(t) = 1/P = N$$



A bit about Coalescent

What is the expected variability?

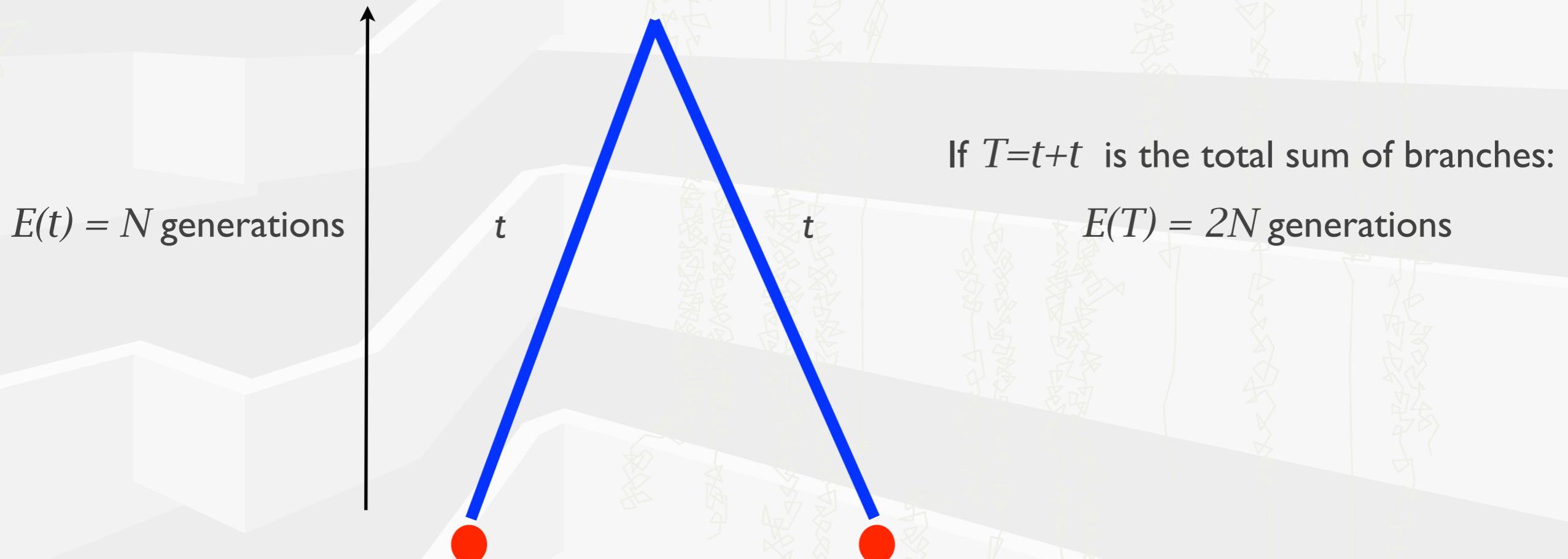
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.

A bit about Coalescent

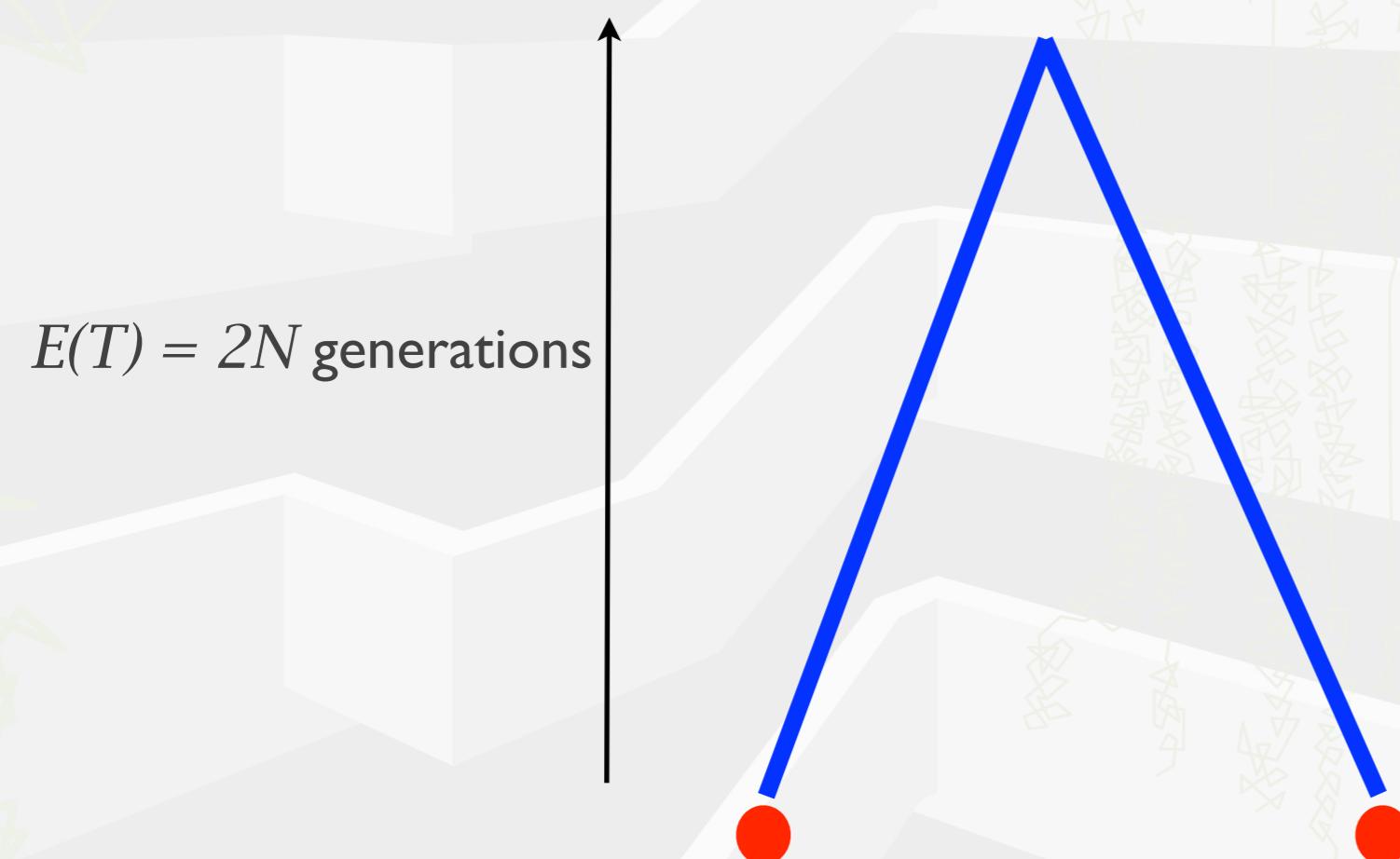
What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.
- We introduce μ (the mutation rate). The number of mutations per position and generation.
- μL is the number of mutations per gene and generation.
- The mutational process is random process and assume follows a Poisson distribution.

A bit about Coalescent

What is the expected variability?

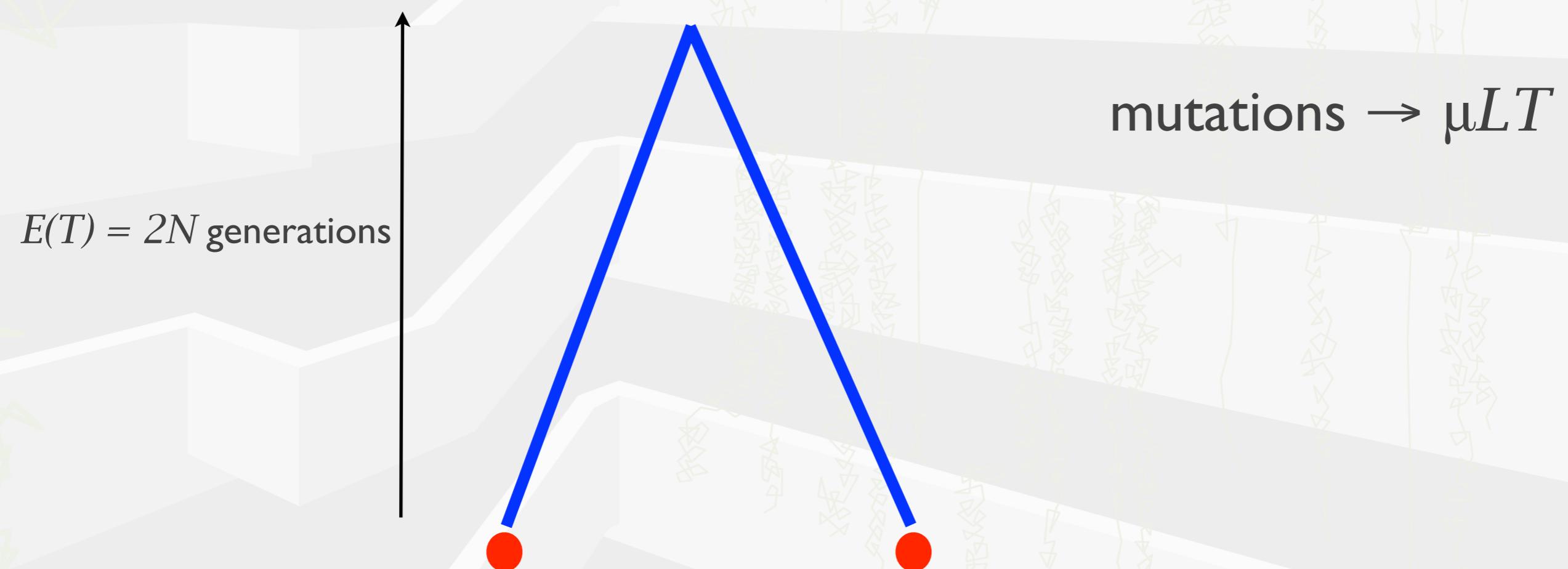
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

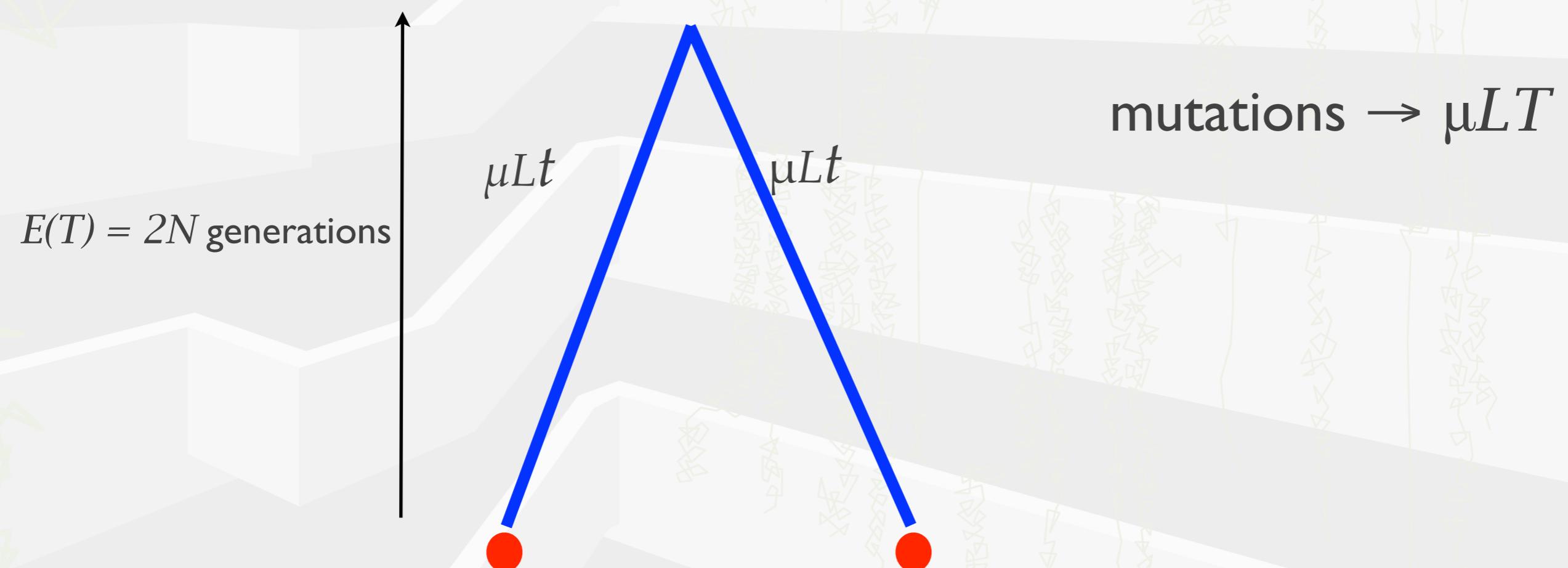
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

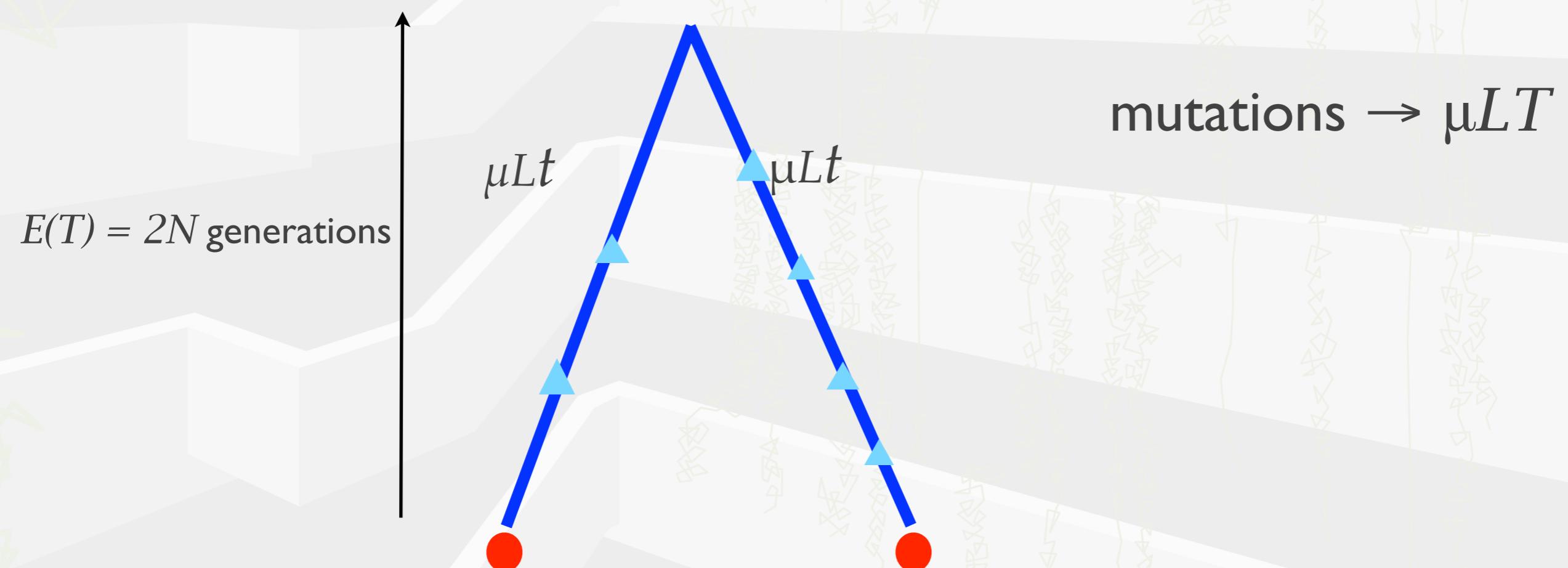
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

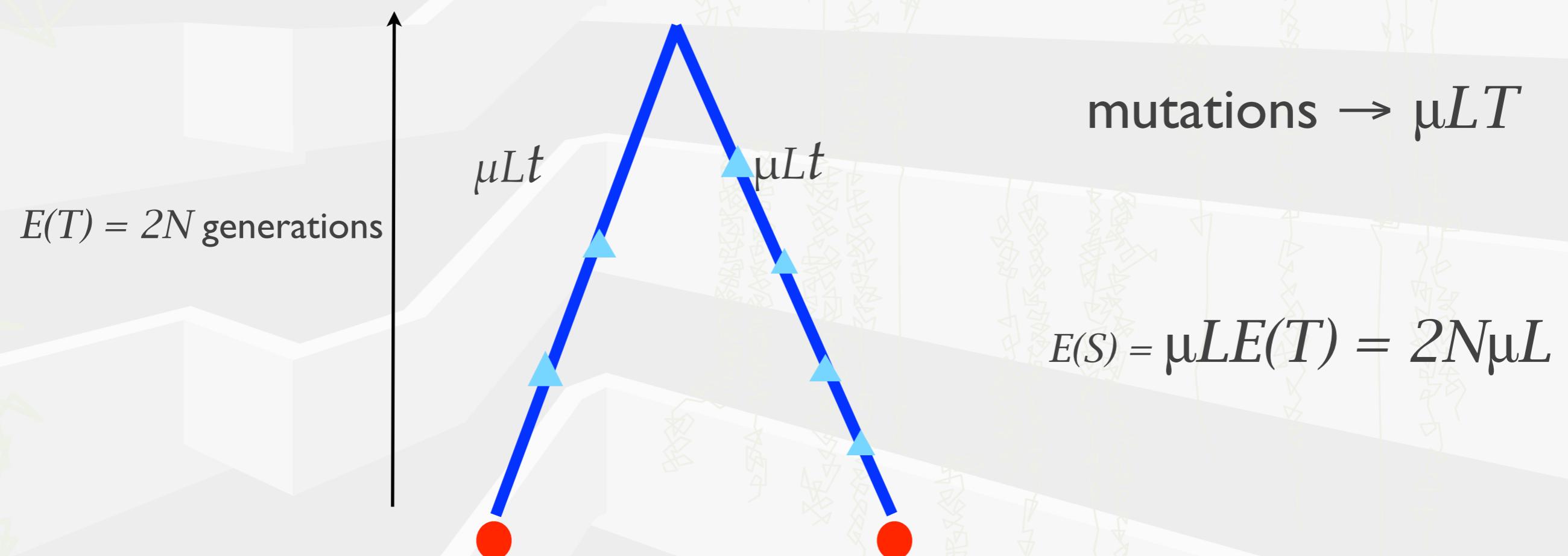
- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of two in a population of $N=1000$.
- Let's reconstruct the history and the number of expected mutations.



A bit about Coalescent

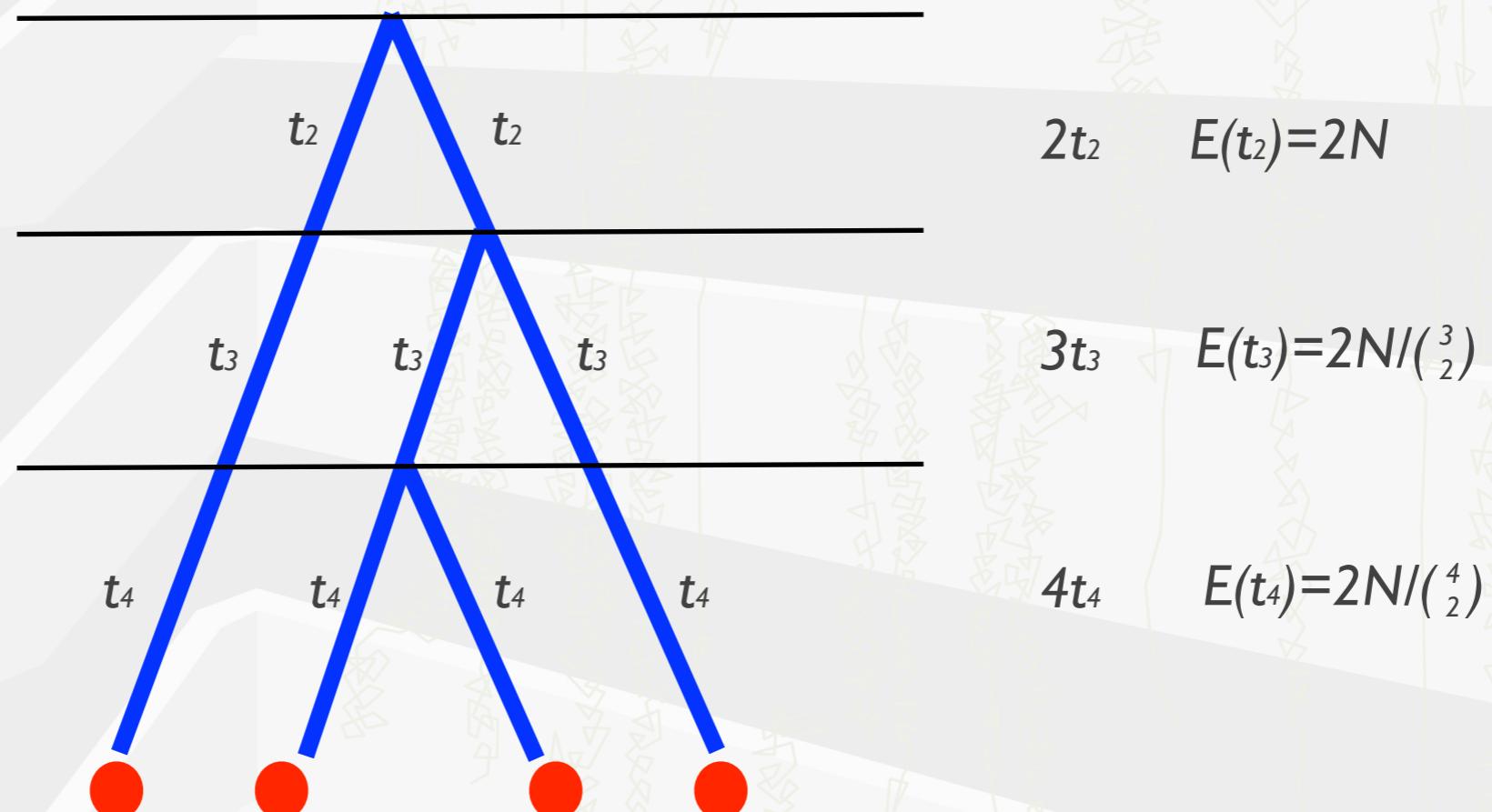
What is the expected variability?

- In case a diploid population of $2N=1000$, $n=2$:
 - $E(T) = 4N$ generations
 - $E(S) = \mu L E(T) = 4N\mu L$

A bit about Coalescent

What is the expected variability?

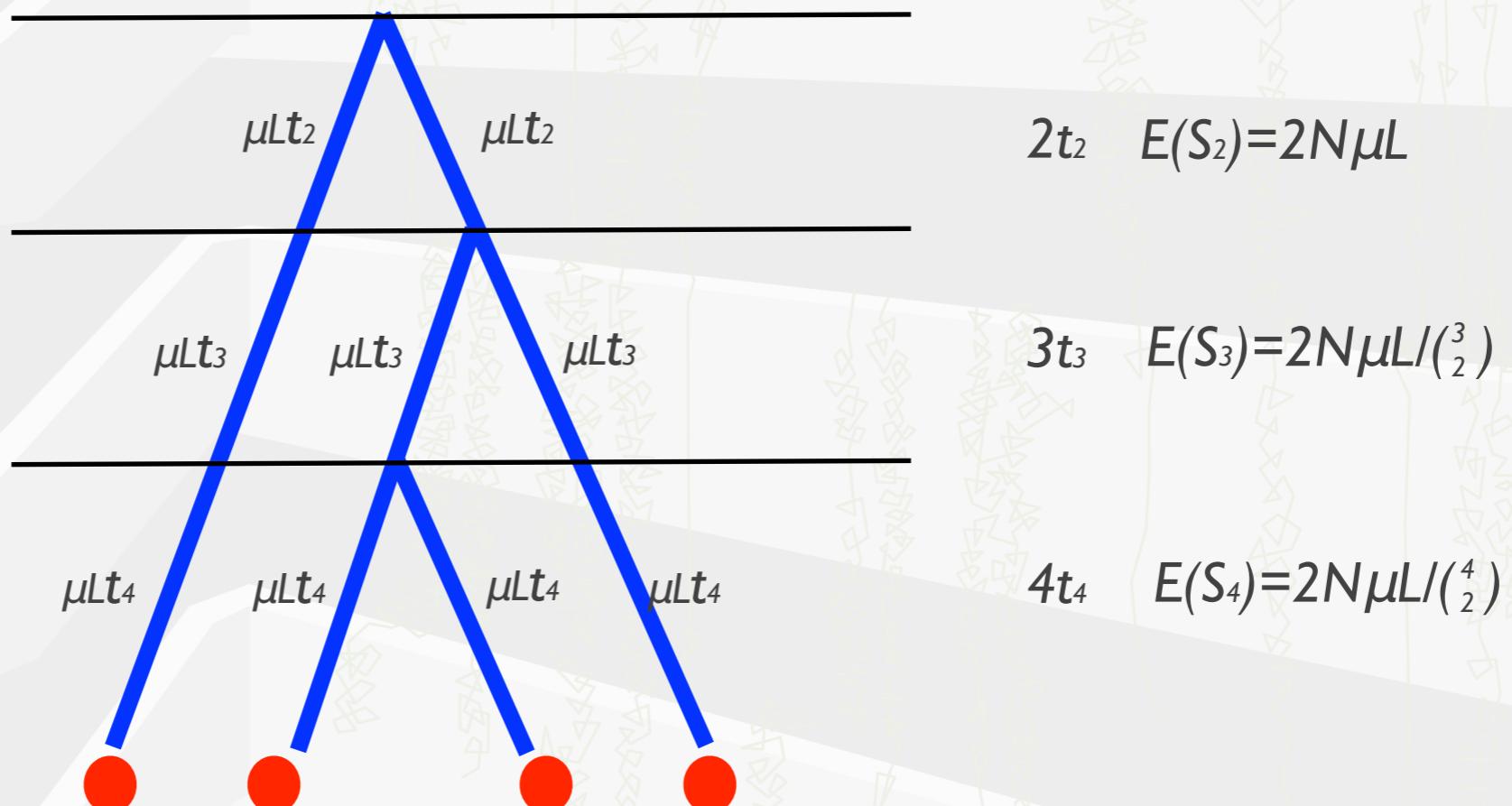
- Imagine a sample of “ n ” individuals in a population of $2N=1000$.
- Let’s reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

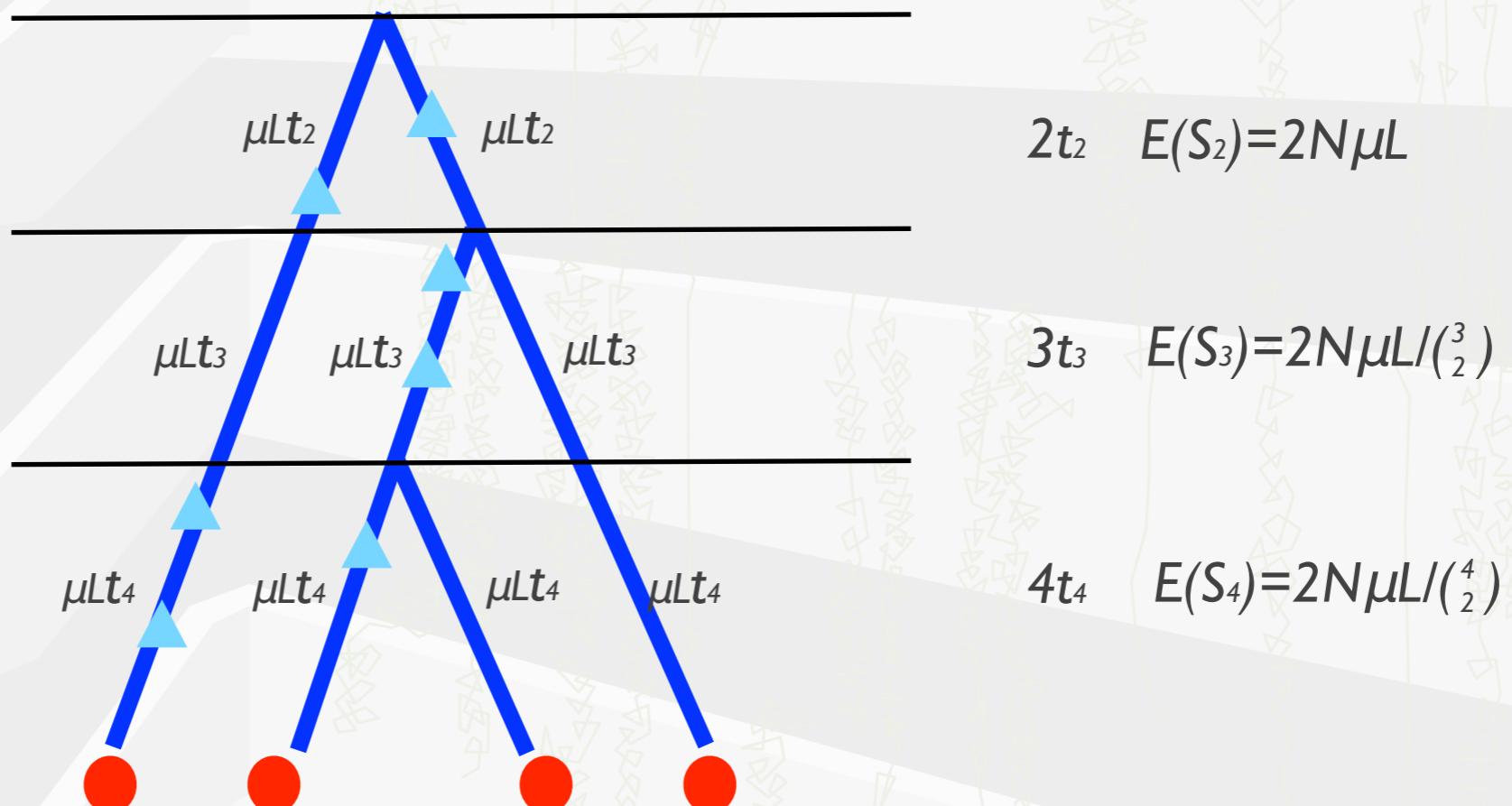
- Imagine a sample of “ n ” individuals in a population of $2N=1000$.
- Let’s reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- Imagine a sample of “ n ” individuals in a population of $2N=1000$.
- Let’s reconstruct the history and the number of expected mutations.



A bit about Coalescent

What is the expected variability?

- In case a diploid population and “n” samples:

$$\bullet E(T) = \sum_{i=2}^n E(T_{(i)}) = 4N \cdot a_n \text{ generations, } a_n = \sum_{i=1}^{n-1} 1/i$$
$$\bullet E(S) = \mu L E(T) = 4N\mu L \cdot a_n$$

Analysis of Nucleotide Variability

So, we are interested in knowing the **expected** levels and patterns of variability
(under some conditions)

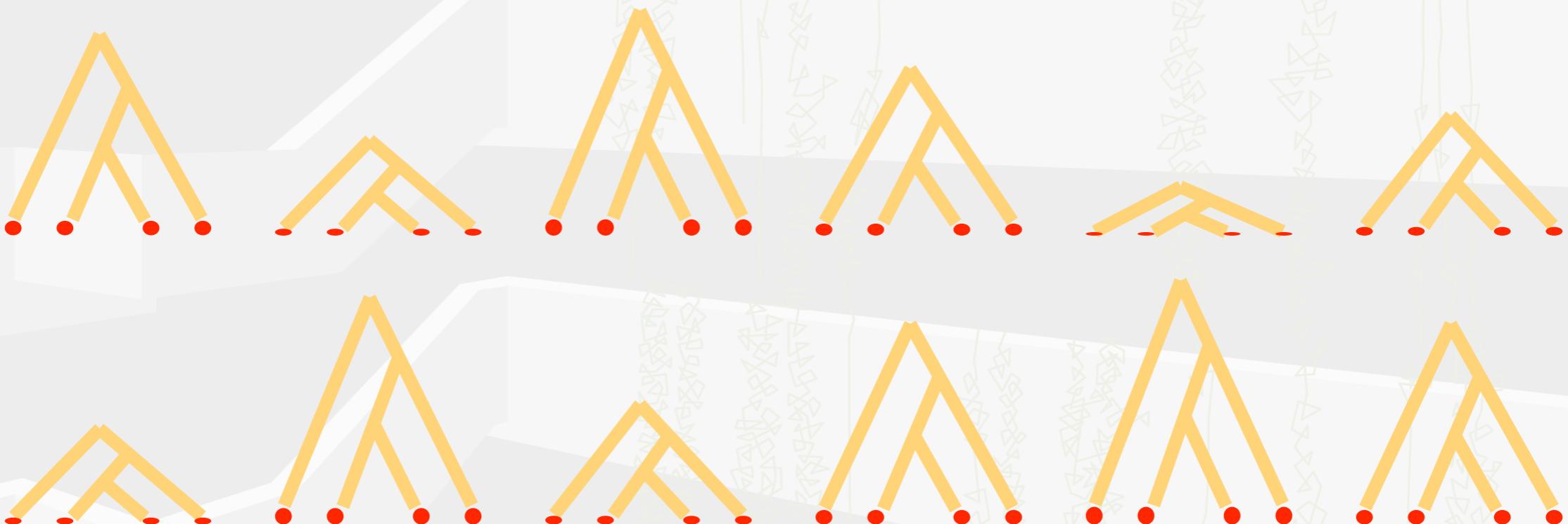
Analysis of Nucleotide Variability

So, we are interested in knowing the **expected** levels and patterns of variability (under some conditions)

We reconstruct many times our sample under a given model (that is, doing Monte Carlo simulations). Thus, we can contrast the observed results with the expected.

Monte Carlo Simulations

- ❖ Repeat many times the experiment under a given Model:



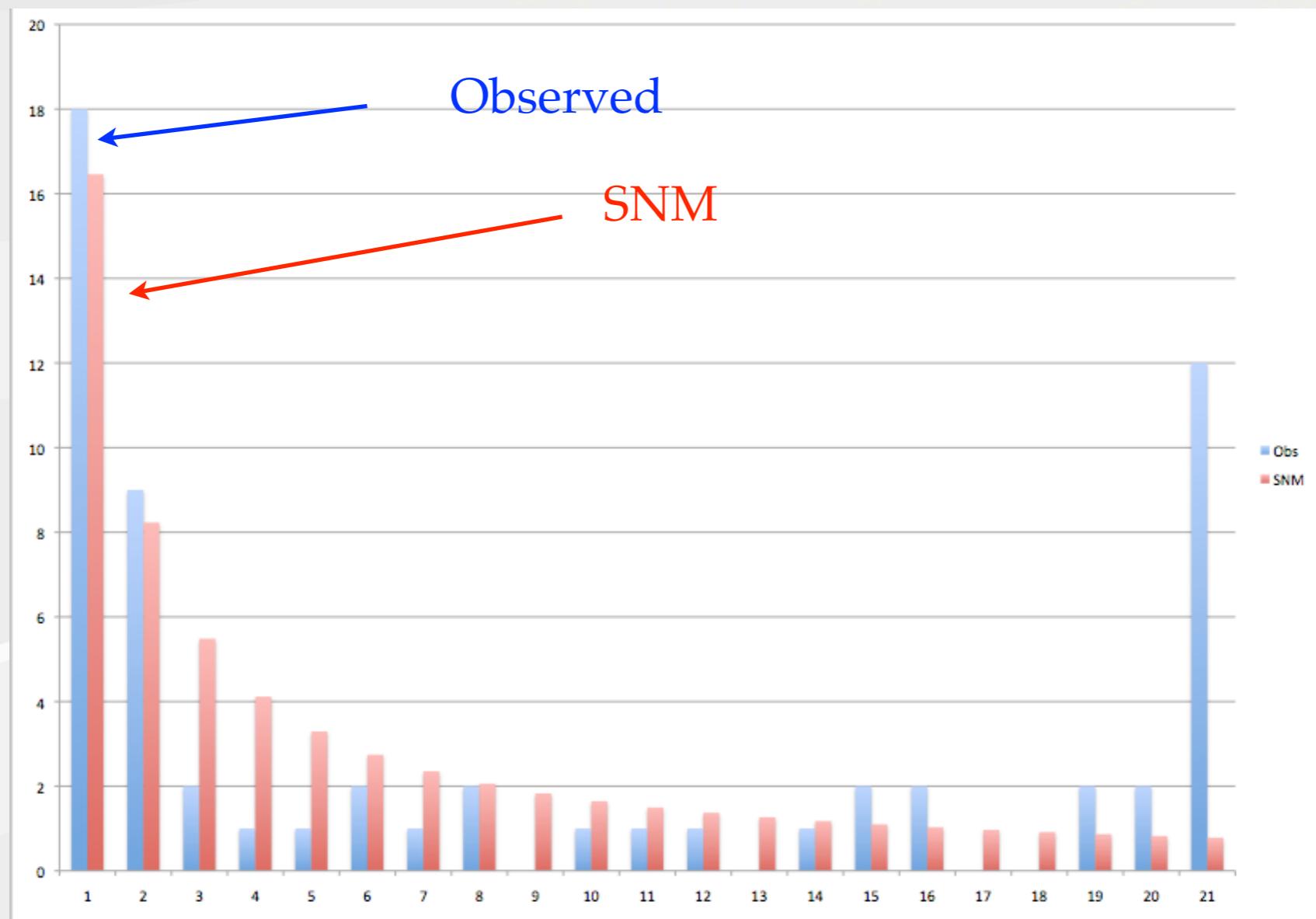
Monte Carlo Simulations

- ❖ Then calculate values (statistics) of interest and obtain the Expected distribution.



Analysis of Nucleotide Variability

Or the frequency spectrum of mutations:



Alternative Models

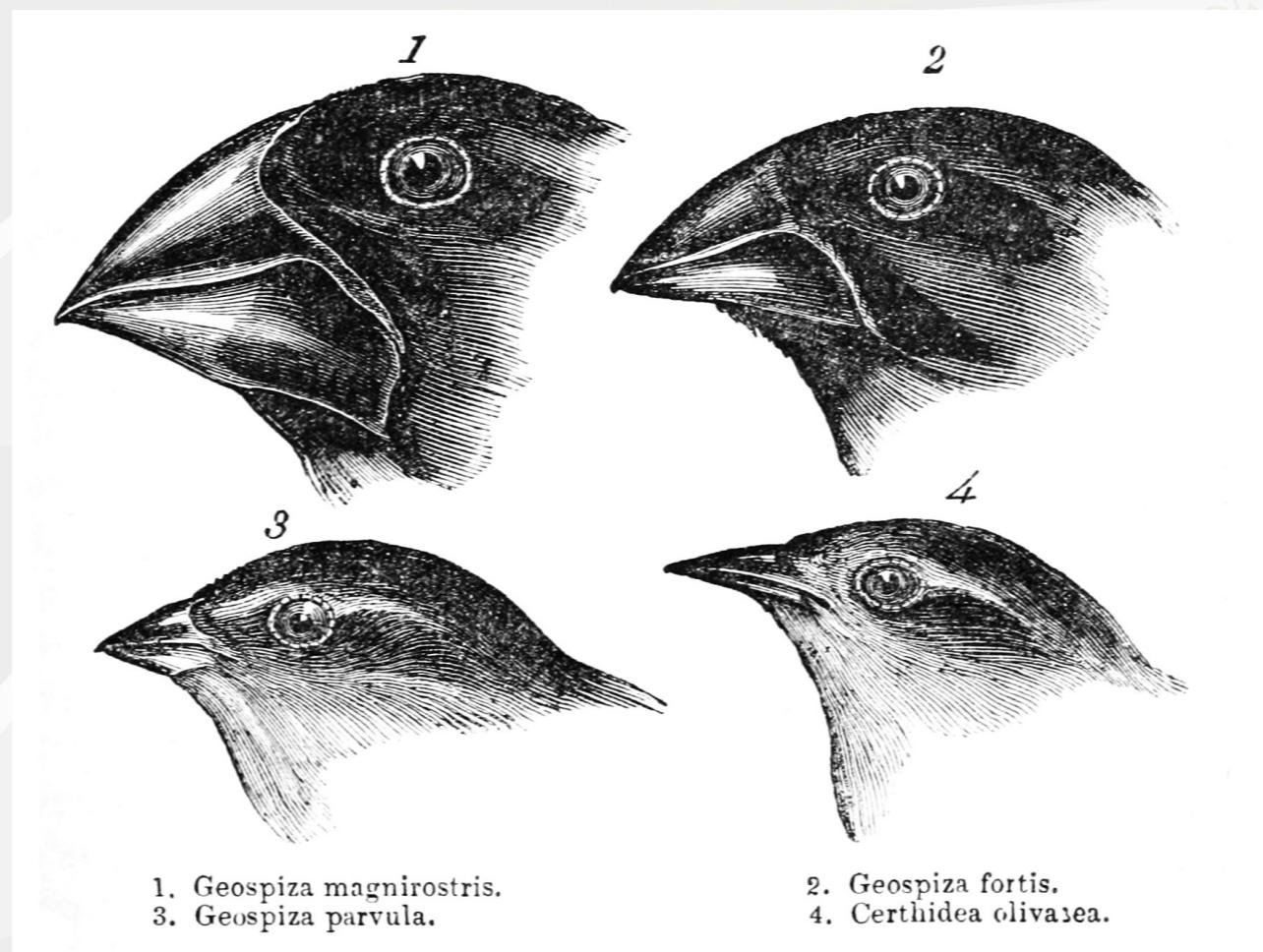
How to include other evolutionary processes?
(bottleneck, migration, selection...)

- Some are easy to include, others are complex (selection).
- Using coalescence, the key is to assume that only one (or none) event occur per generation. We simply look which of the processes occurs first by chance.

The Effect of Selection at Genome level

Adaptation to a given environment

- **Fitness** involves the ability of organisms to survive and reproduce in the environment in which they find themselves (Dobzhansky, 1959).



(Darwin 1845)

Adaptation to a given environment

- The differential ability of organisms to survive and reproduce (**Fitness**) may have a hereditary factor. Considering **Selection** (the process to chose the fittest individuals) and **Adaptation** (the hereditary adjustment of a species into an environment), different scenarios arise:
 - **Selection with Adaptation** (traits are totally or partially shaped by genetic effects). Individuals with higher fitness genetically change the frequencies in the next generation.
 - **Selection without Adaptation** (traits selected are entirely shaped by the environment). Individuals with higher fitness DO NOT genetically change the frequencies in the next generation.
 - **Adaptation without Selection** (e.g., phenotypic plasticity can modulate adaptation to a different environment and later convert these phenotypic changes irreversible at genetic level). NO Individuals had higher fitness in the process.

Adaptation to a given environment

- Individuals with better features will be able to contribute more to the next generation (**selection**).



Industrial melanism in peppered moth
(*Biston betularia*)

Adaptation to a given environment

- Individuals with better features will be able to contribute more to the next generation (**selection**).



Industrial melanism in peppered moth
(*Biston betularia*)

- If the traits selected have a genetic component, these individuals will have a **higher (relative) genetic fitness**.

Adaptation to a given environment

- Individuals with better features will be able to contribute more to the next generation (**selection**).



Industrial melanism in peppered moth
(*Biston betularia*)

- The **mean relative fitness** is the average contribution of the genotypes coming from these individuals to the global pool in the next generation.

| A₁A₁ | A₁A₂ | A₂A₂ |
|-----------------------------------|-----------------------------------|-----------------------------------|
| $p^2 \omega_{11}$ | $2pq \omega_{12}$ | $q^2 \omega_{22}$ |

$$\bar{\omega} = p^2 \omega_{11} + 2pq \omega_{12} + q^2 \omega_{22}$$

Adaptation to a given environment

- The fitness can be affected by few number of traits (**simple** environment) or alternatively by a large number of traits (a **complex** environment).
- Organisms living in simple environments are easier to achieve the theoretical optimum (maximum) fitness, while in the high multidimensional space of complex environments this optimum can be less accessible. **Larger populations give more chance to achieve optimum** (more mutations, more options).
- **Genetic load** is the relative difference between the optimal and the current fitness.

The Fitness Distribution

- The global effect of fitness and their dynamics in a population is a consequence of the the strength of the fitness and the number of positions involved, that is, of the **Distribution of Fitness Effects** (DFE).
- The fitness effect of genome positions, -positive or negative-, -strong or weak-, in a given environment can explain the capacity and the dynamics of the population to adaptation and the expected patterns of variability.

we may think that each position has assigned a specific s

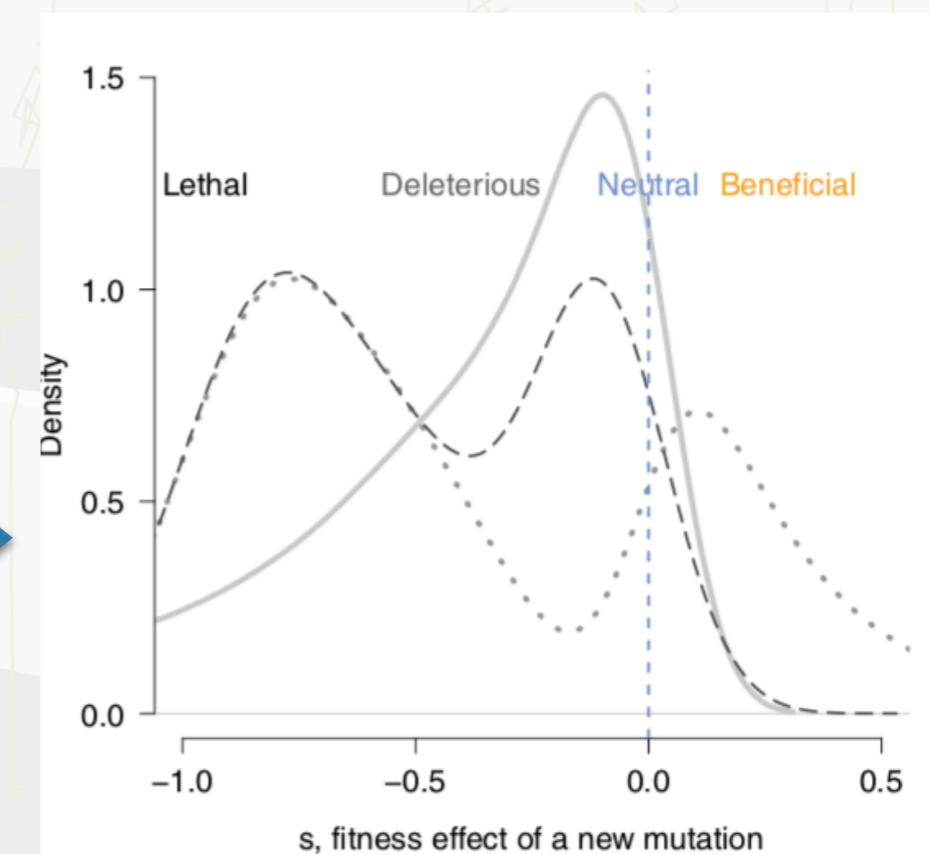
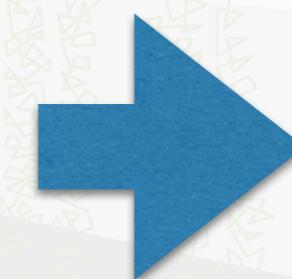
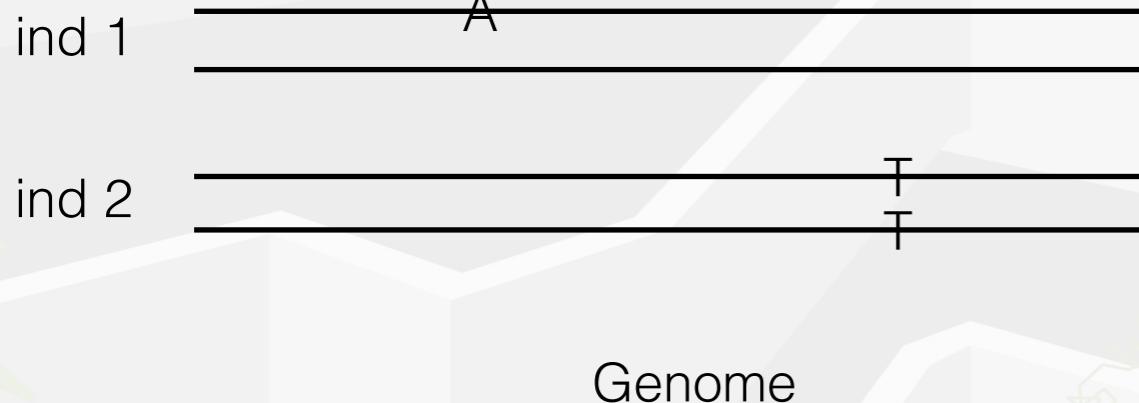


Figure 1. Hypothetical whole distributions of fitness effects.
(Bataillon & Bailey Ann NY Acad. Sc..2014)

The Fitness Distribution

- In case that the mean fitness of a population is **close to the optimum** (small genetic load), beneficial effects are rare, it is predicted:
 - an invariant and **L-shaped** (exponential) distribution of **beneficial** effects (Orr 2003).
 - **most mutations** are displacing from the optimum and thus **are deleterious**. A **gamma** distribution is predicted for deleterious effects (but see Cotto et al. 2023)...

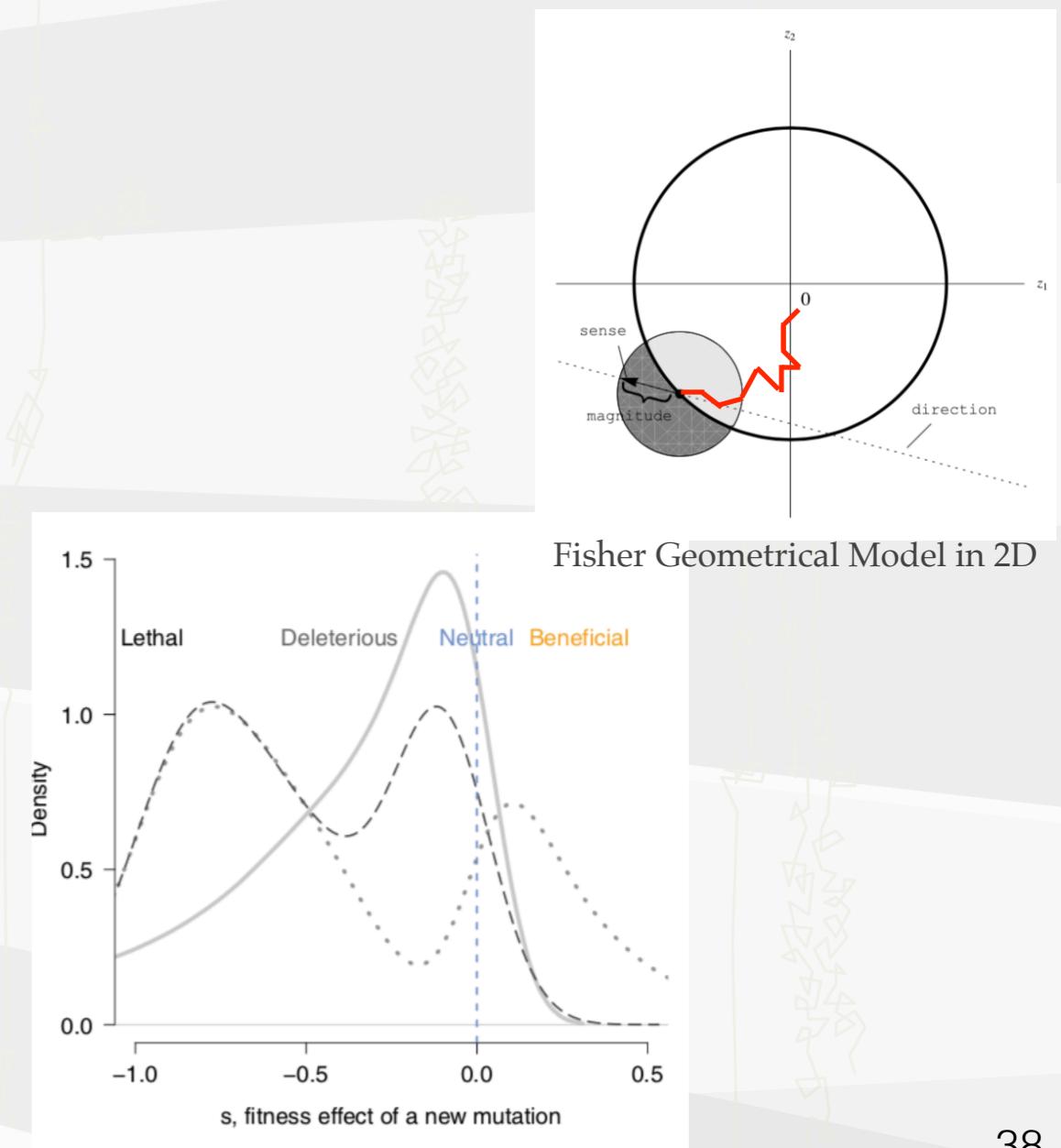


Figure 1. Hypothetical whole distributions of fitness effects.

(Bataillon & Bailey Ann NY Acad. Sc..2014)

Methods to measure Fitness

Methods to measure Fitness

- How many mutations affect the fitness and what importance each mutation has? What is the Genome Architecture in relation to Fitness? What is the role of pleiotropy, epistasis, linkage, redundancy...? What is the role of spatial and temporal variation in fitness optimal?
 - 1. Experimental studies measuring fitness components in contemporary populations (phenotype).**
 - 2. Experimental Evolution in Real Time (consider genotype and phenotype).**
 - 3. Inference from Sequence Variability data (consider genotype).**

Methods to measure Fitness

- How many mutations affect the fitness and what importance each mutation has? What is the Genome Architecture in relation to Fitness? What is the role of pleiotropy, epistasis, linkage, redundancy...? What is the role of spatial and temporal variation in fitness optimal?

1. Experimental studies measuring fitness components in contemporary populations (considers phenotype).

- Observation of viability, fecundity at different stages and during all lifetime organisms.

2. Experimental Evolution in Real Time (considers genotype and phenotype).

- Evolve populations in controlled (divergent common) environments. From virus and microbial to yeast, Drosophila or Daphnia and others..
- Estimate the effect of spontaneous mutations. / Mutagenesis: random or site-directed mutagenesis. / Mutation-accumulation experiments. / Fitness trajectories in adapting populations.

3. Inference from Sequence Variability data (considers genotype).

- Use polymorphisms and divergence data from functional and non-functional regions to infer the Distribution of Fitness Effects.

Methods to measure Fitness

- How many mutations affect the fitness and what importance each mutation has? What is the Genome Architecture in relation to Fitness? What is the role of pleiotropy, epistasis, linkage, redundancy...? What is the role of spatial and temporal variation in fitness optimal?

1. Experimental studies measuring fitness components in contemporary populations (considers phenotype).

- Give important information about the fitness effects at different stages at phenotypic level.

2. Experimental Evolution in Real Time (considers genotype and phenotype).

- From mutagenesis experiments, infer the distribution of the new positive and negative mutations.
- Evolution in action experiments are quite powerful because consider genotype and phenotype information. Many different experimental designs are used.

3. Inference from Sequence Variability data (considers genotype).

- Powerful and elegant, although phenotype is not considered. Many questions related to genomic architecture could not be answered.

Inference from Variability data

Inference from Variability data

- How many mutations affect the fitness and what importance each mutation has?

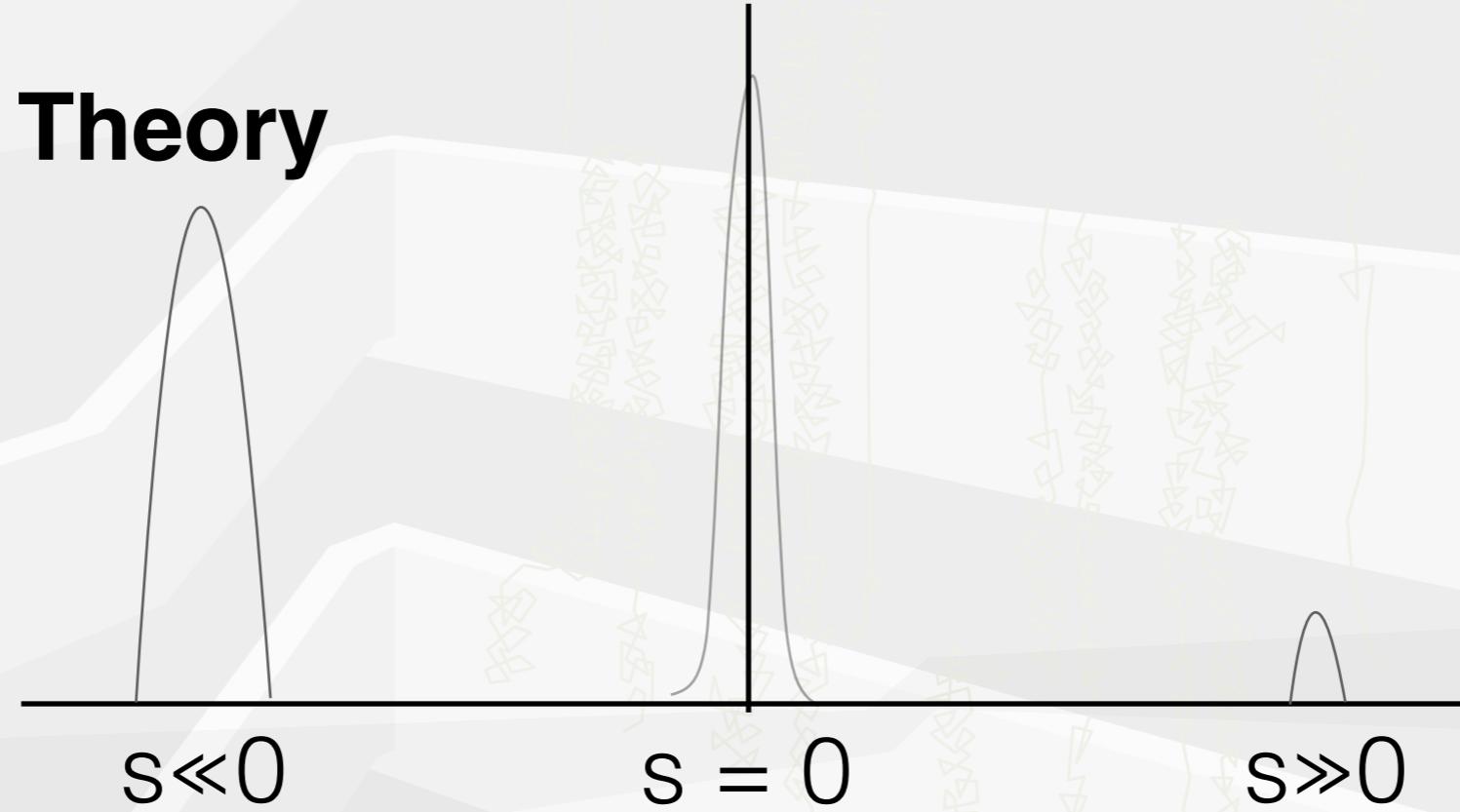
Inference from Variability data

- How many mutations affect the fitness and what importance each mutation has?
- What proportion of mutations are beneficial?

Inference from Variability data

- How many mutations affect the fitness and what importance each mutation has?
- What proportion of mutations are beneficial?

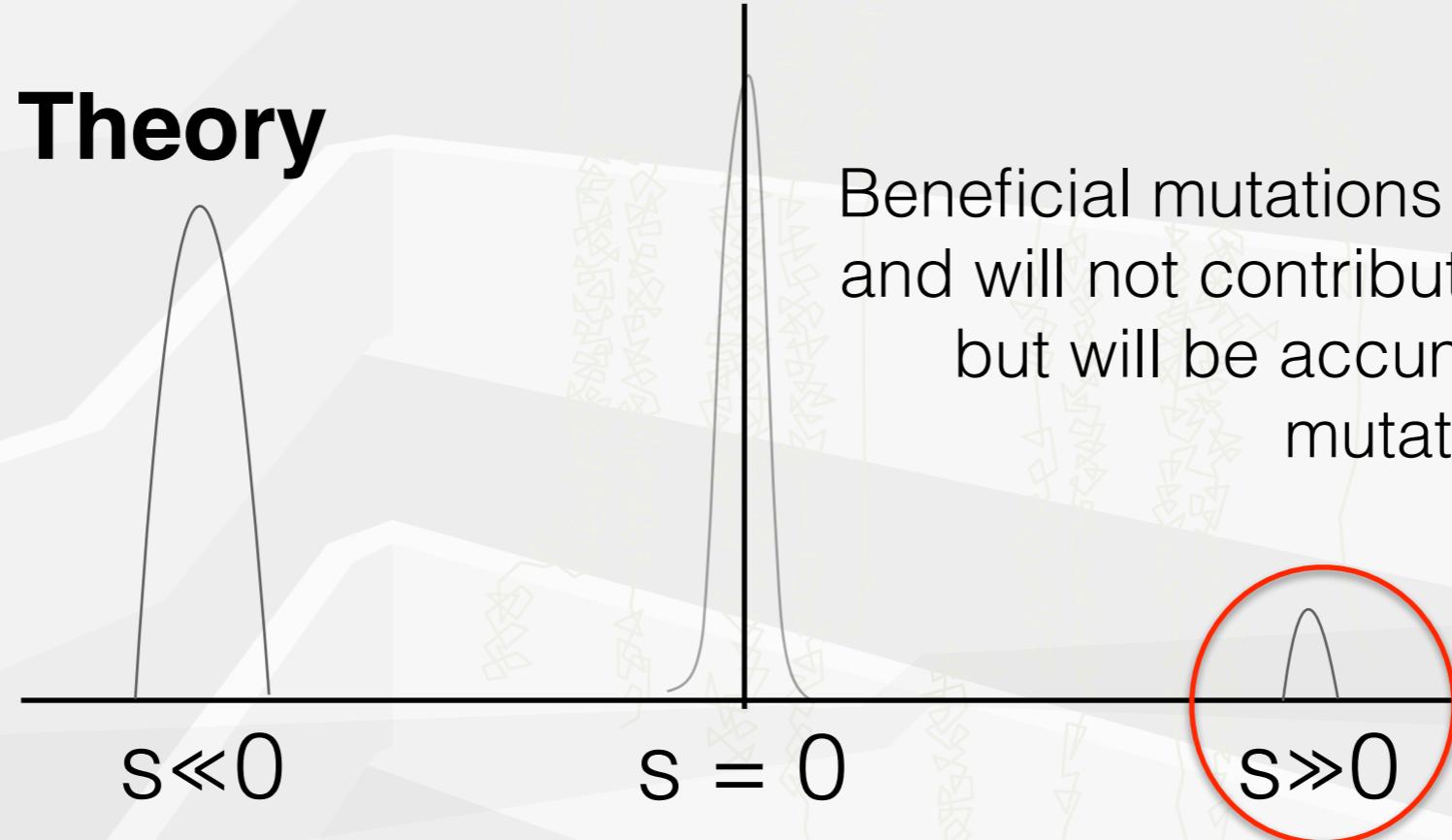
Neutral Theory



Inference from Variability data

- How many mutations affect the fitness and what importance each mutation has?
- What proportion of mutations are beneficial?

Neutral Theory



Beneficial mutations will be rapidly fixed and will not contribute to polymorphism, but will be accumulated as fixed mutations.

Inference from Variability data

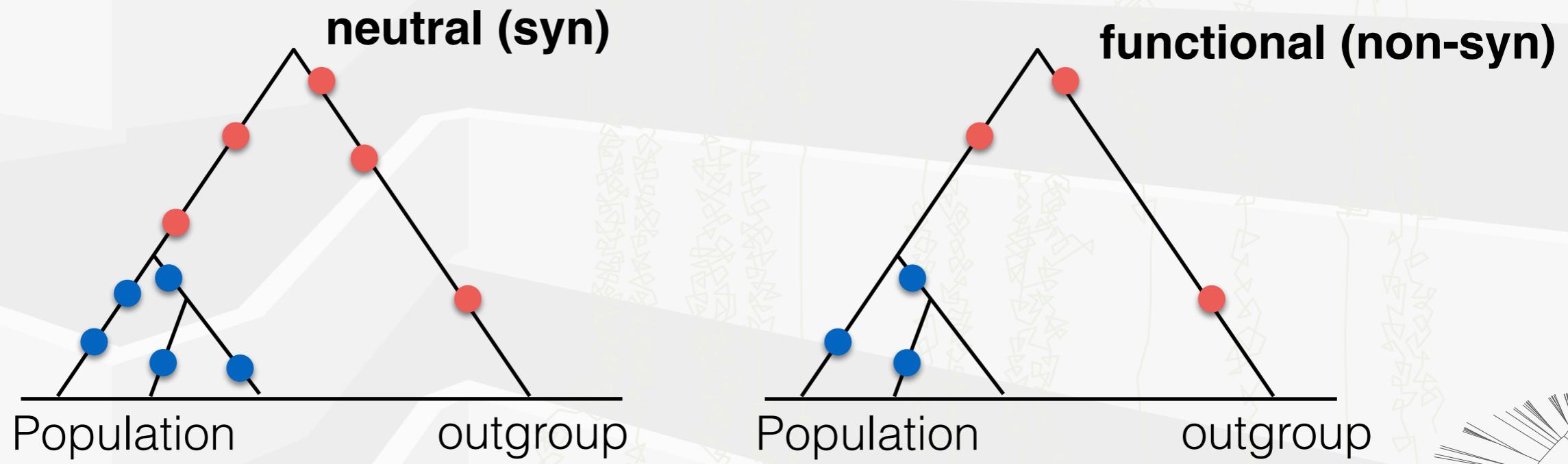
- How many mutations affect the fitness and what importance each mutation has?
- What proportion of mutations are beneficial?
- The MacDonald and Kreitman framework (1992)

Inference from Variability data

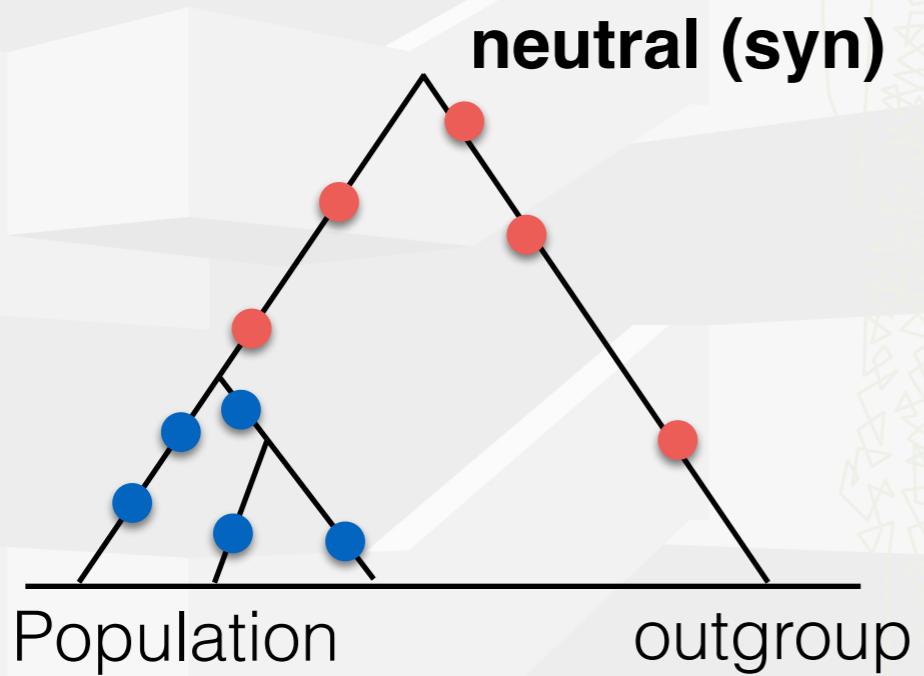
- The MacDonald and Kreitman framework (1992)
 - Separate the positions into neutral and others susceptible of selection (functional) Using coding regions: synonymous and non-synonymous.
 - Consider the polymorphism and the divergence with a close related species.

Inference from Variability data

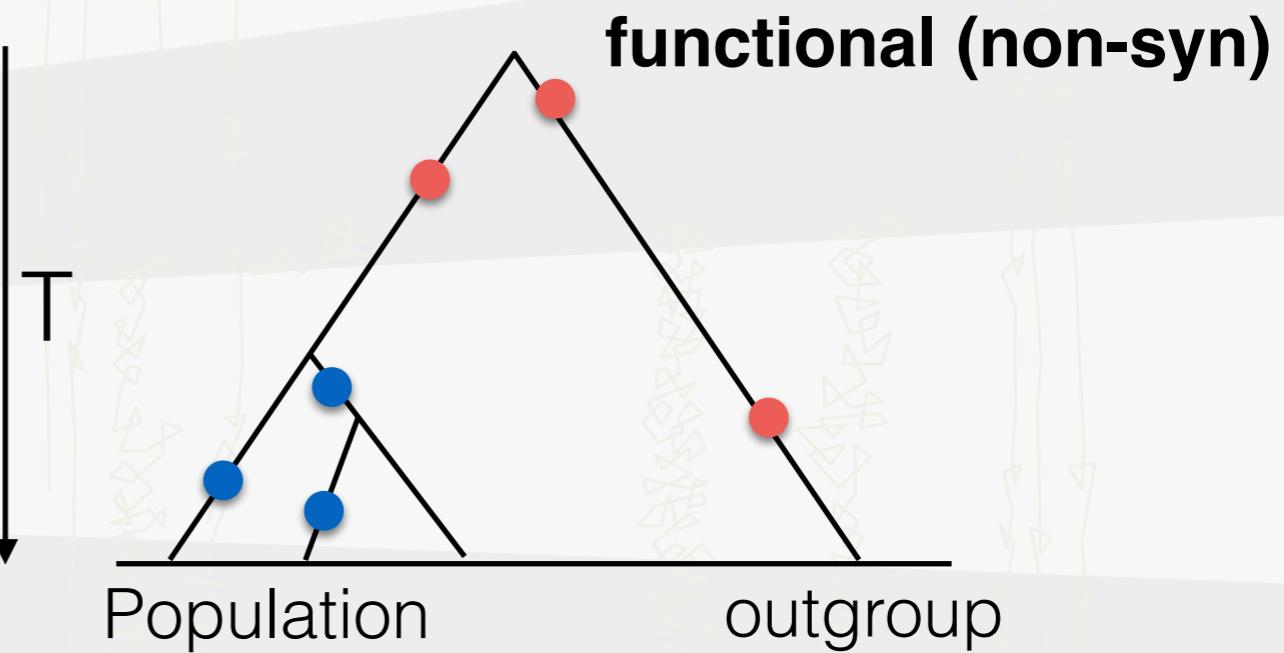
- The MacDonald and Kreitman framework (1992)
 - Separate the positions into neutral and others susceptible of selection (functional) Using coding regions: synonymous and non-synonymous.
 - Consider the polymorphism and the divergence with a close related species.



Inference from Variability data



$$4N\mu_{\text{syn}}$$

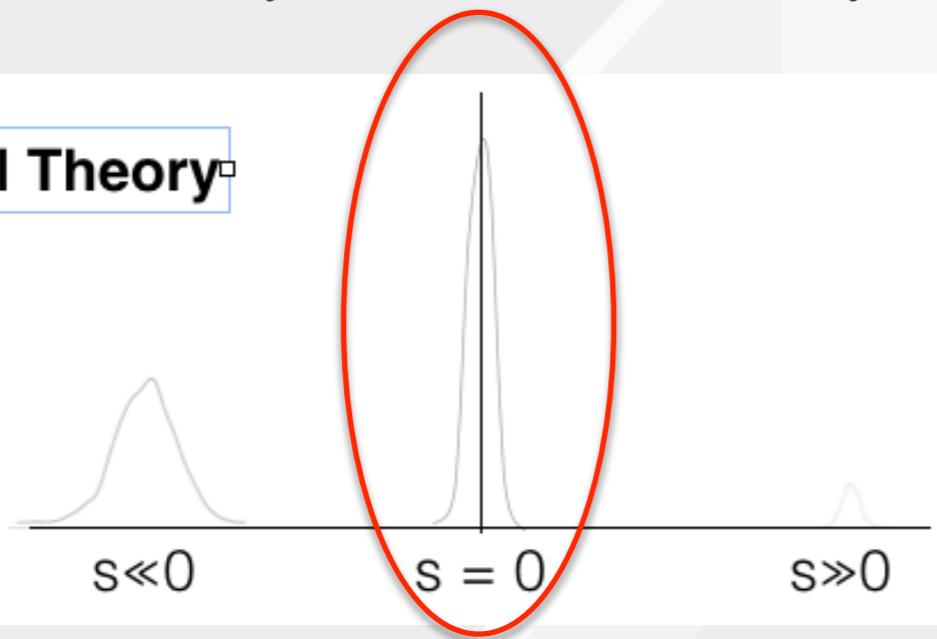


$$2\mu_{\text{syn}}T$$

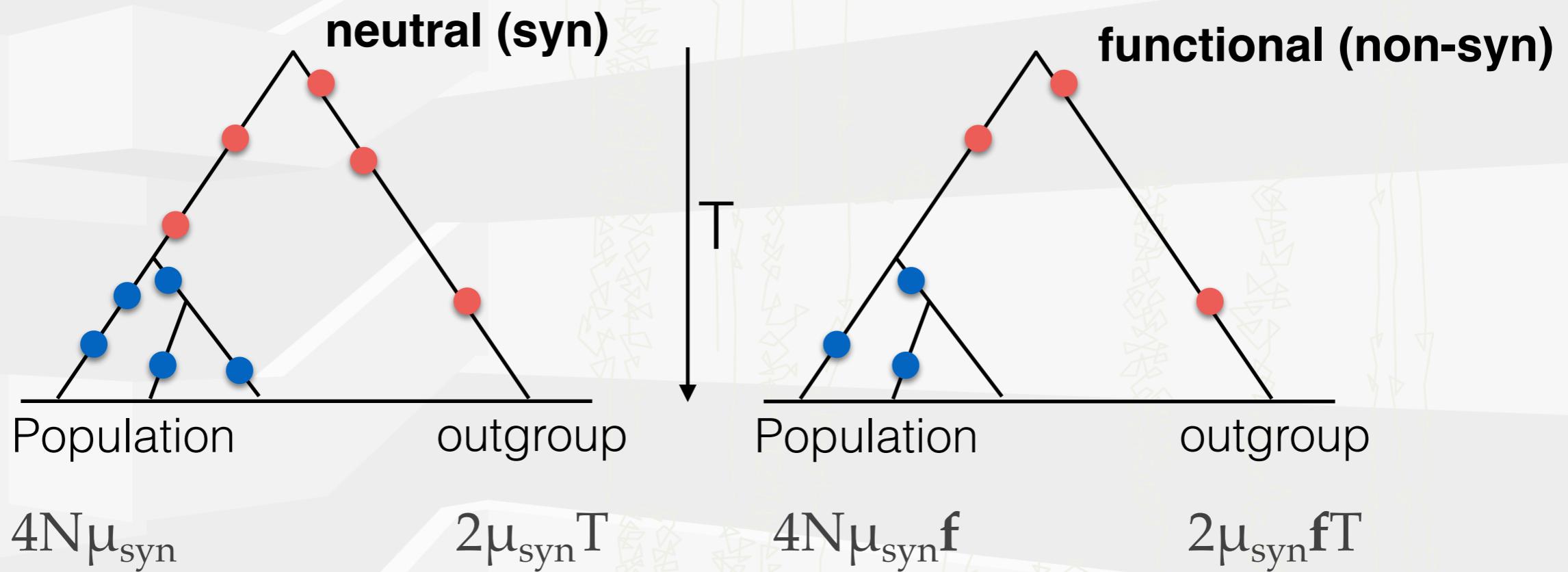
$$4N\mu_{\text{syn}}f$$

$$2\mu_{\text{syn}}fT$$

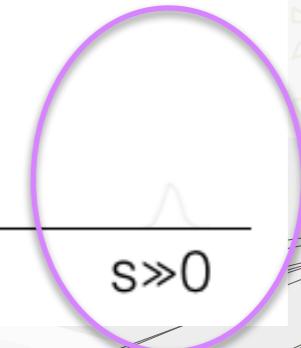
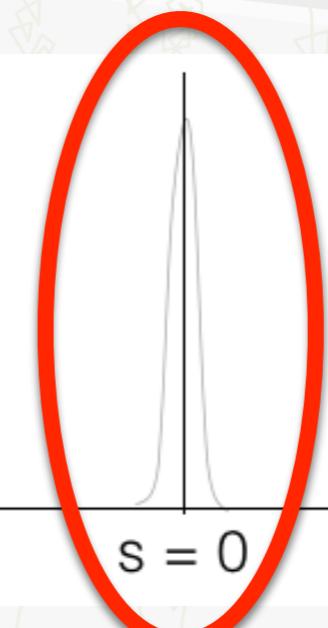
Neutral Theory



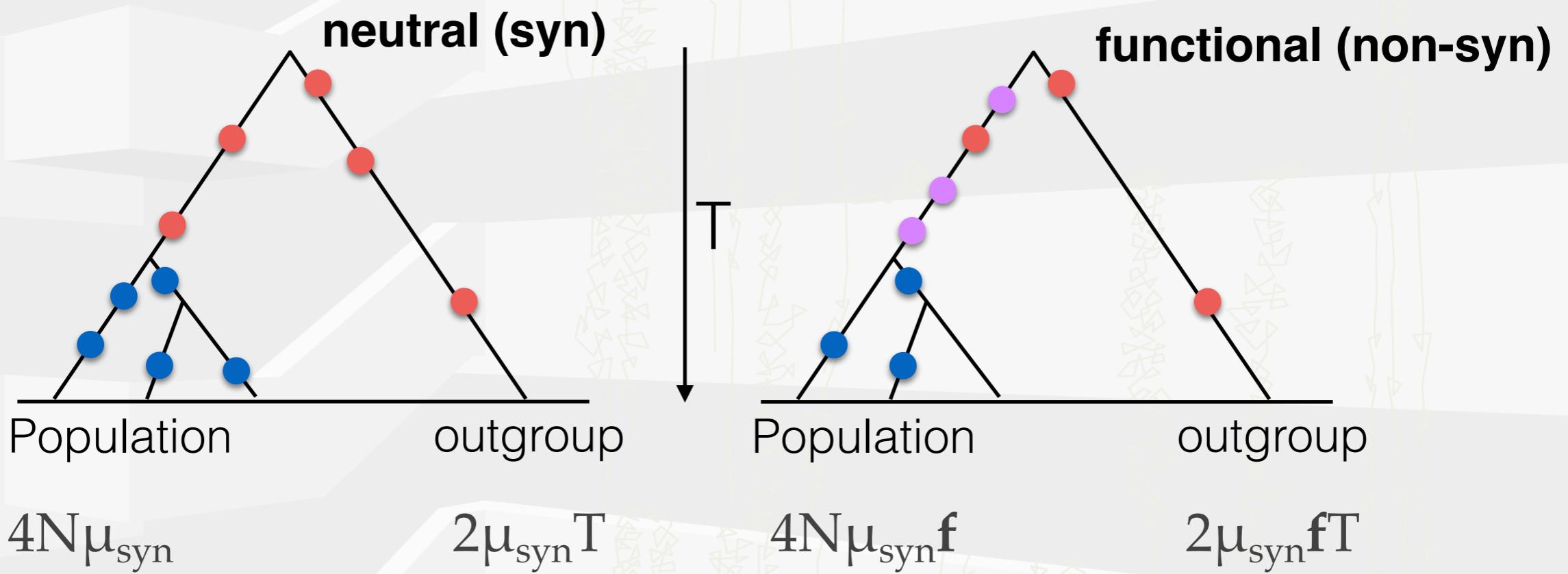
Inference from Variability data



Neutral Theory

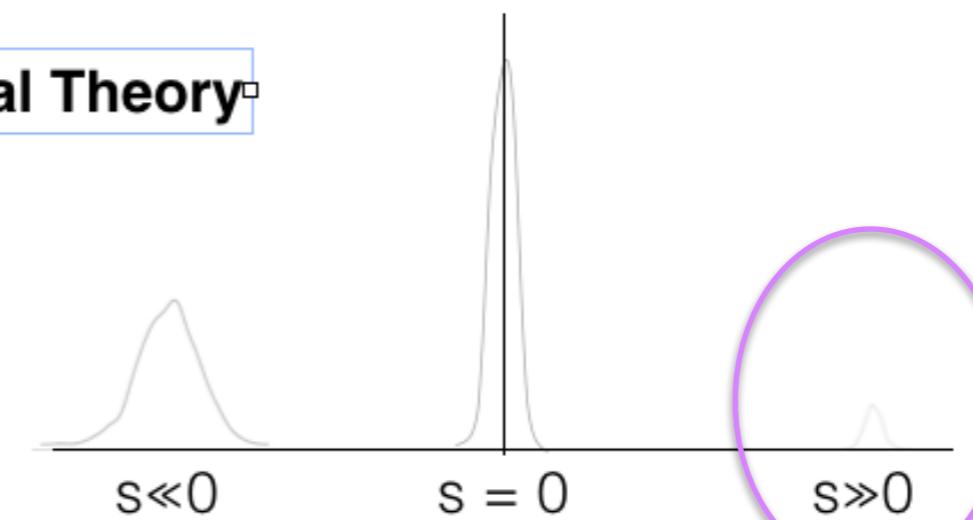


Inference from Variability data



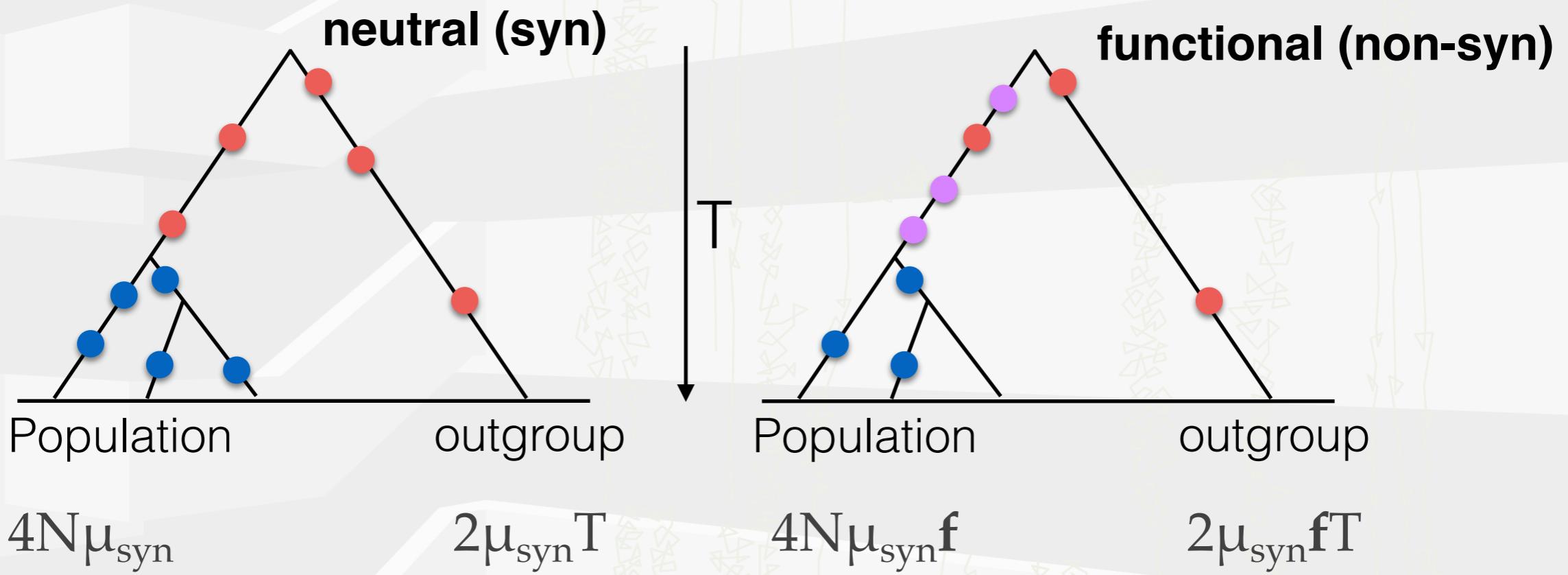
Beneficial mutations contribute only to divergence!

Neutral Theory



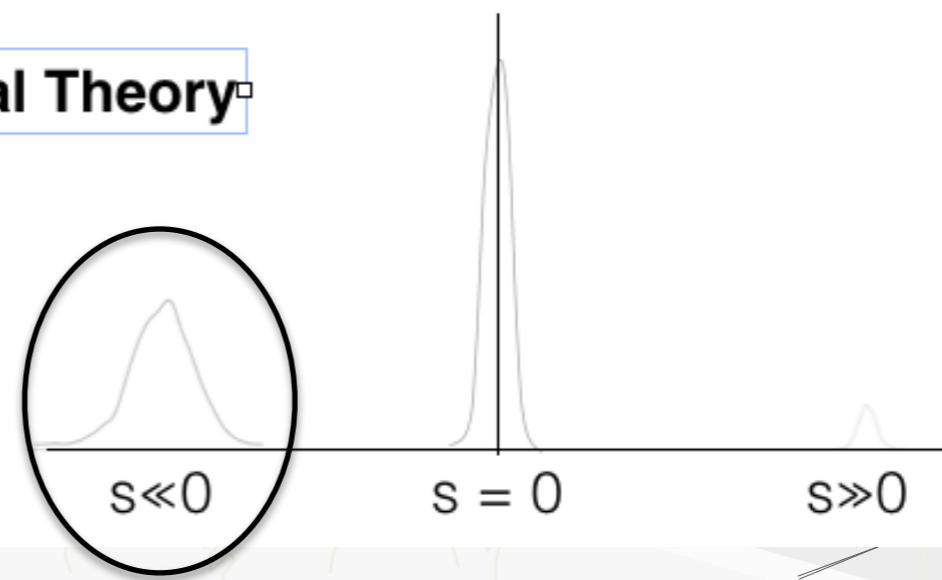
53

Inference from Variability data

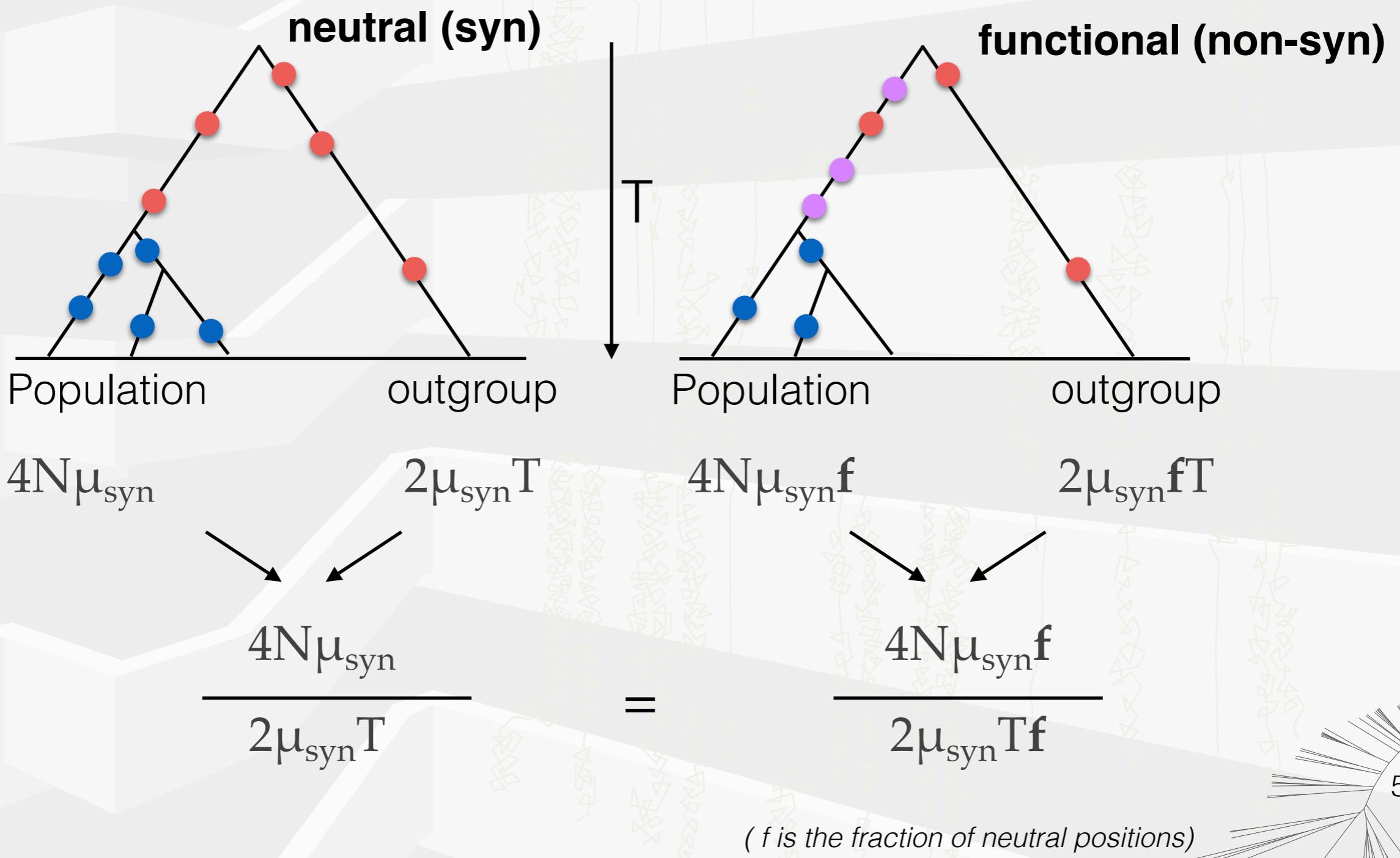


Deleterious mutations are lost directly. Do not contribute (just like reducing mutation in a factor f).

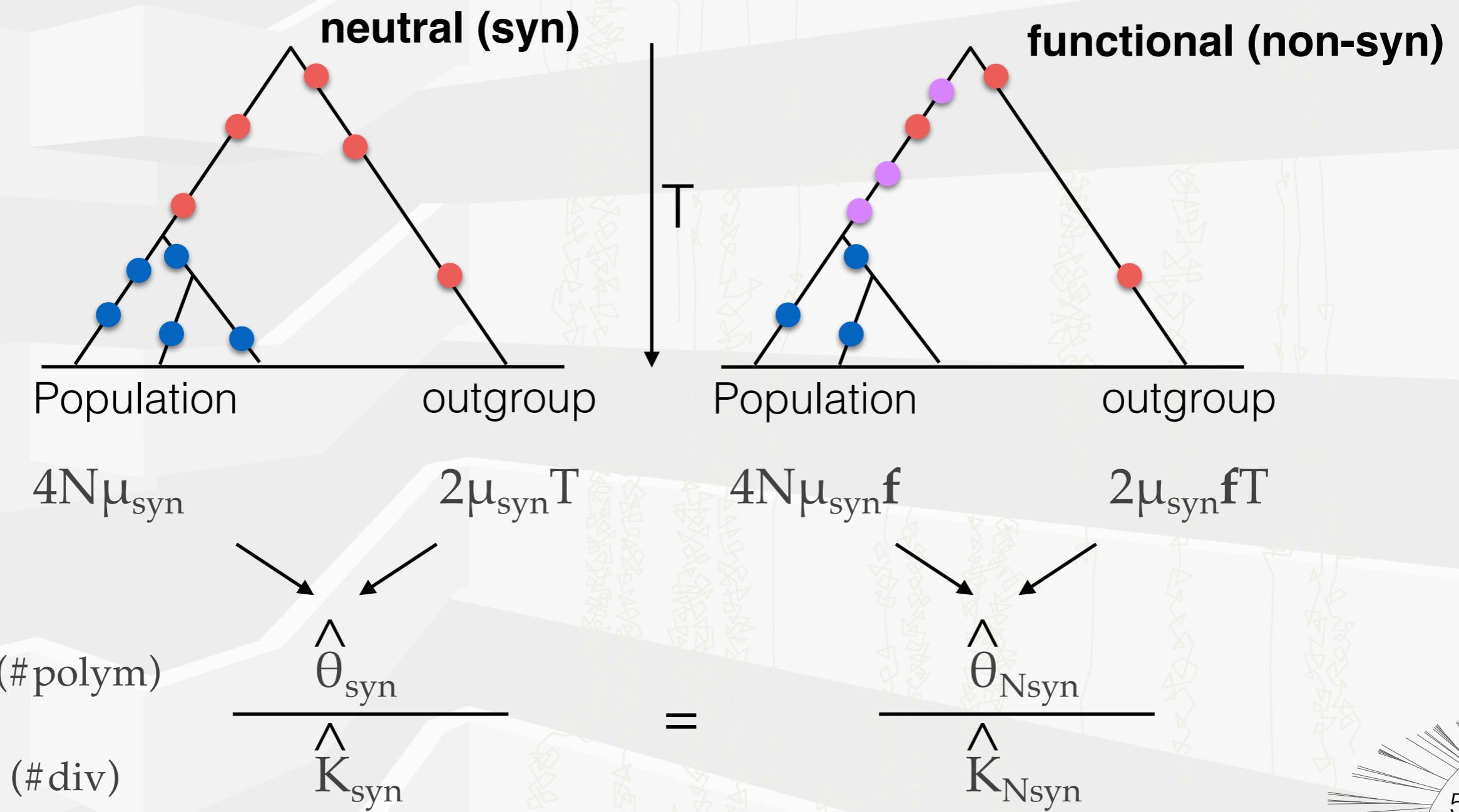
Neutral Theory



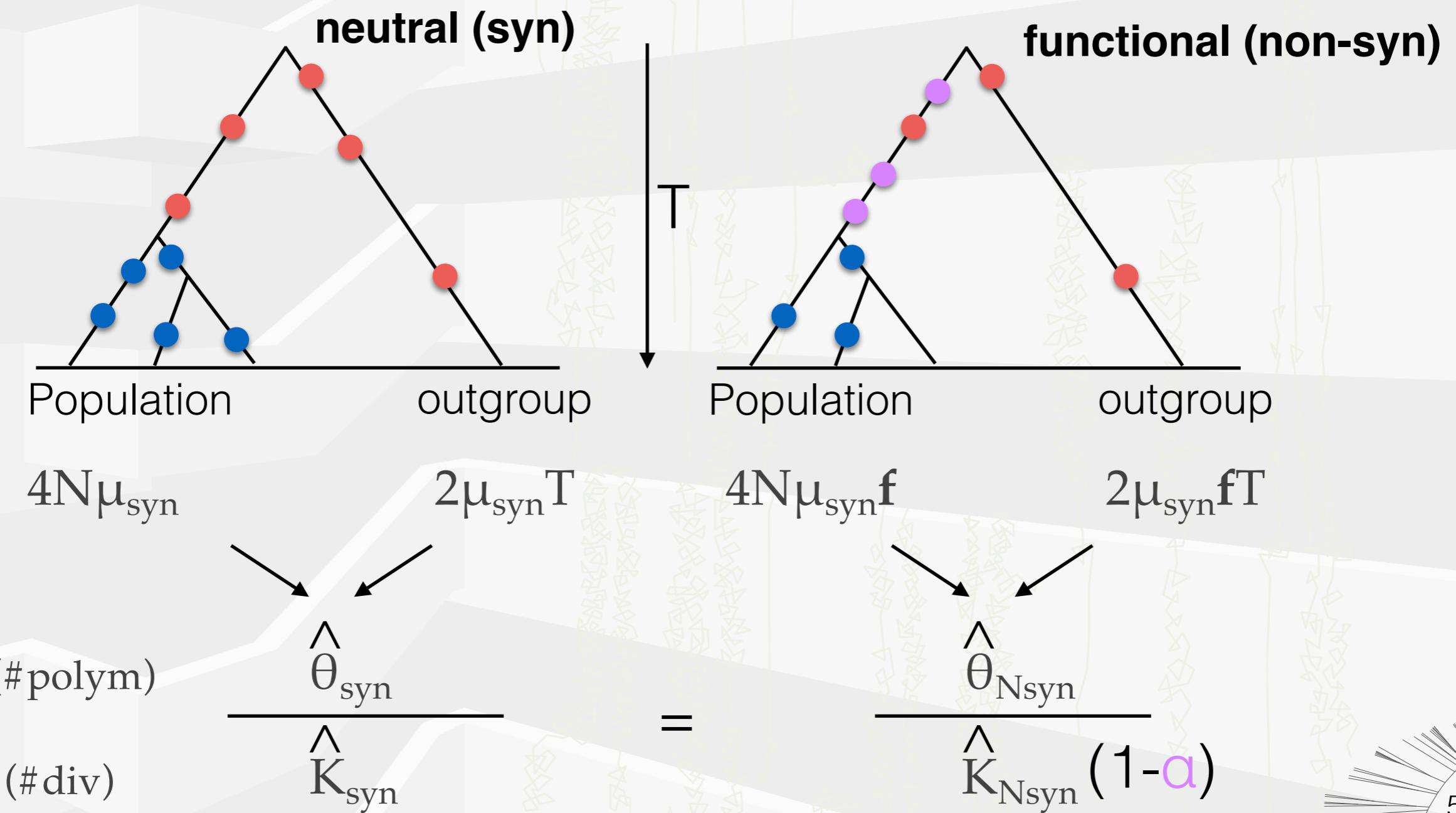
Inference from Variability data



Inference from Variability data



Inference from Variability data



Considers that beneficial mutations affect divergence but not the polymorphism:

Inference from Variability data

$$\frac{\hat{\theta}_{\text{syn}}}{\hat{K}_{\text{syn}}} = \frac{\hat{\theta}_{\text{Nsyn}}}{\hat{K}_{\text{Nsyn}}(1-a)}$$

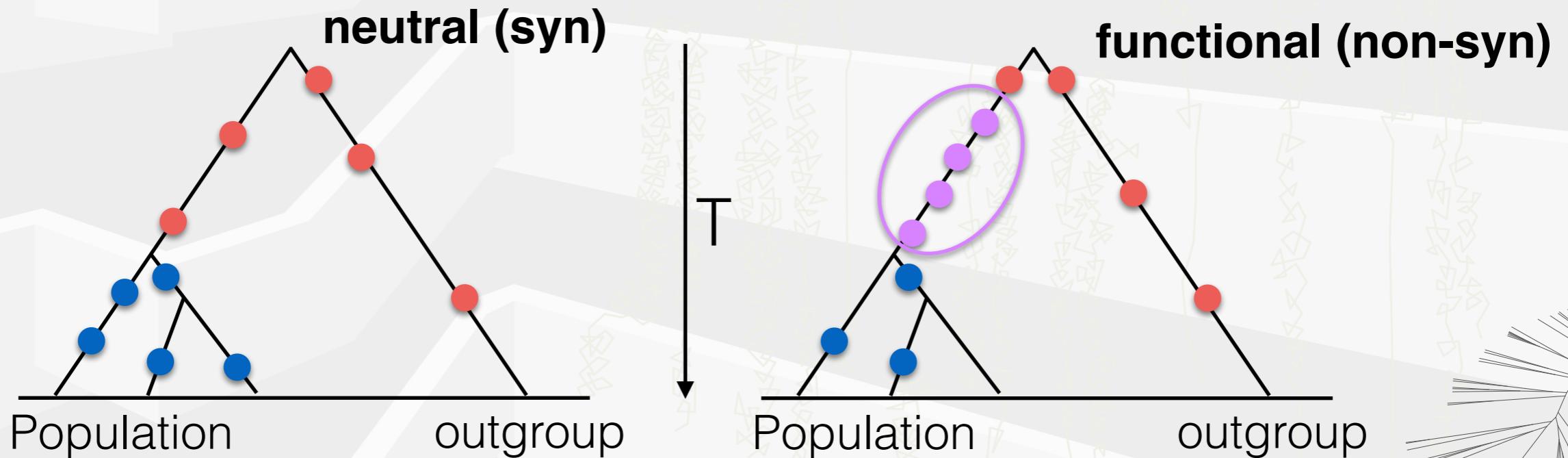
a is the proportion of beneficial (adaptive) substitutions.

$$\hat{a} = 1 - \frac{\hat{\theta}_{\text{Nsyn}} \hat{K}_{\text{syn}}}{\hat{\theta}_{\text{syn}} \hat{K}_{\text{Nsyn}}}$$

Inference from Variability data

\hat{a} is the proportion of beneficial (adaptive) substitutions.

$$\hat{a} = 1 - \frac{\hat{\theta}_{\text{Nsyn}} \hat{K}_{\text{syn}}}{\hat{\theta}_{\text{syn}} \hat{K}_{\text{Nsyn}}}$$



The proportion of Adaptive Substitutions

$$\bar{\alpha} = 1 - \frac{\bar{D}_s}{\bar{D}_n} \left(\frac{P_n}{P_s + 1} \right)$$

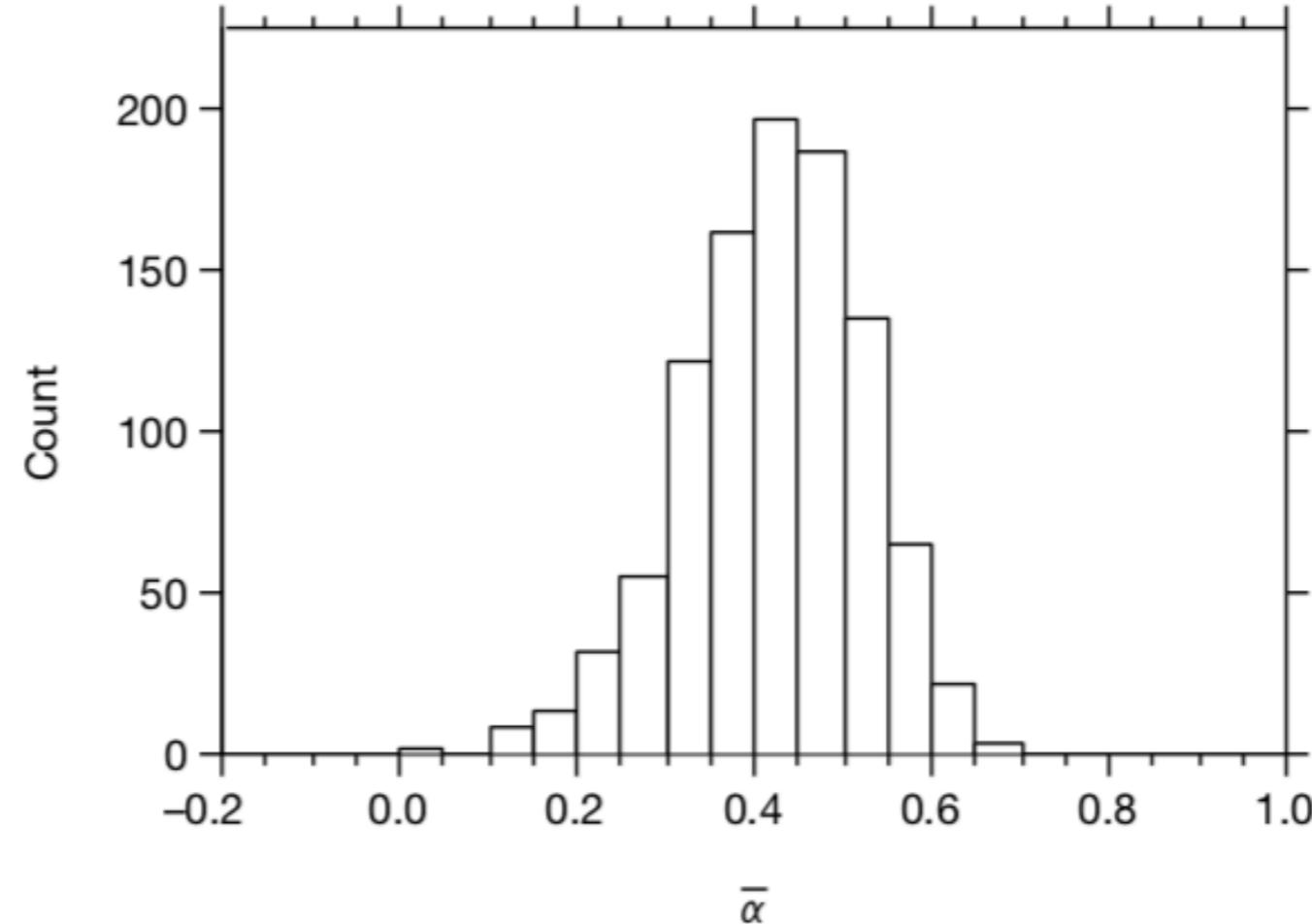


Figure 1 The distribution of 1,000 bootstrap values of $\bar{\alpha}$ for the divergence between *Drosophila simulans* and *D. yakuba* for genes in which $P_s > 5$. $\bar{\alpha}$ is the average proportion of amino-acid substitutions driven by positive selection.

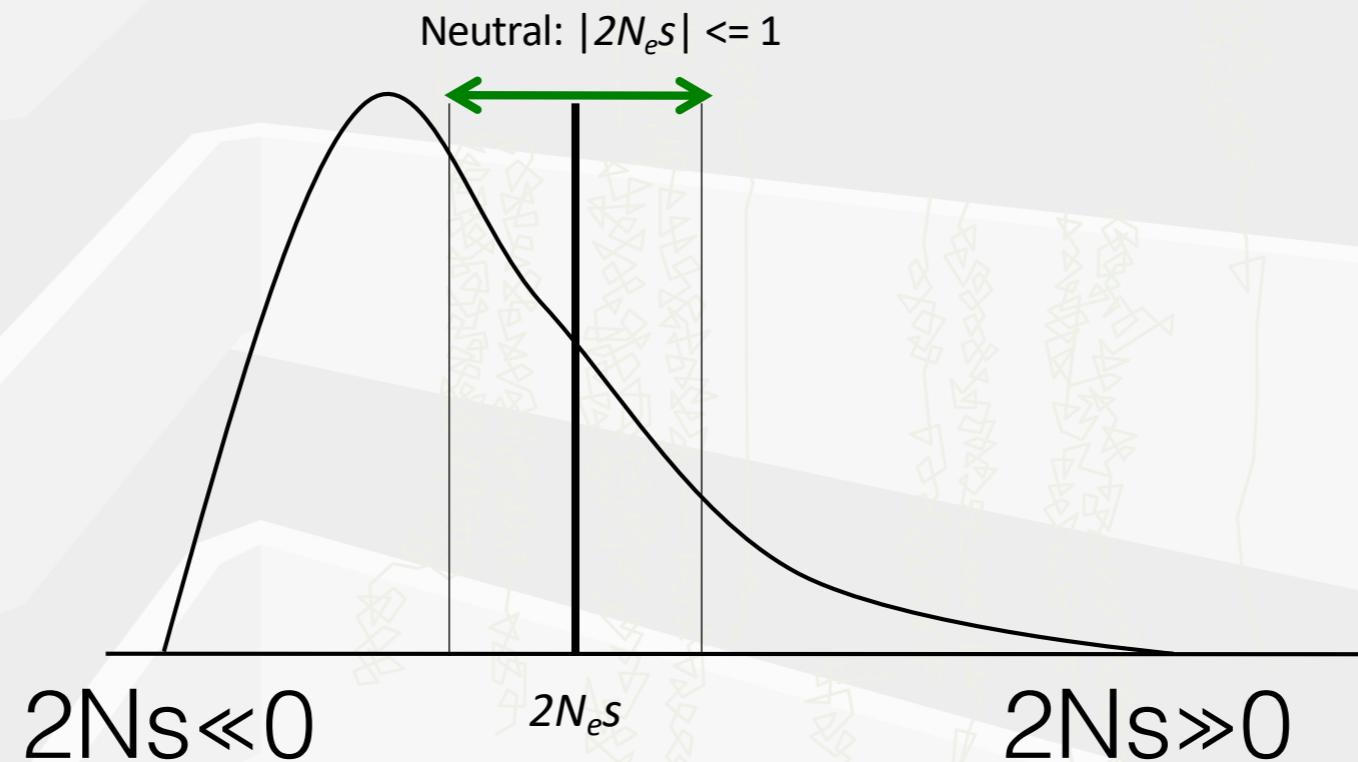
(Smith & eyre-Walker Nature 2002)

Expectations from other (more realistic) Scenarios

- **Not all polymorphisms are neutral.** The effect of segregating deleterious mutations has been observed across the variability of the genomes.
- Many mutations at functional positions are **segregating at low frequency** (weaker deleterious variants) but do not arrive to fixation in the proportions expected by SNM.

Expectations from Different Scenarios

- **Otha's nearly-neutral theory** define the selective effects in a more wide distribution. These selective effects are dependent on $N_e s$.



The Effect of Slightly Deleterious Mutations

- If consider deleterious mutations of different effect:

$$\frac{\theta_s}{\theta_N(1 - \beta_i)} = \frac{K_s}{K_N(1 - \alpha)}$$

The Effect of Slightly Deleterious Mutations

- If consider deleterious mutations of different effect:

$$\frac{\theta_s}{\theta_N(1 - \beta_i)} = \frac{K_s}{K_N(1 - \alpha)}$$

↑
eliminate deleterious present in polymorphisms ↑
eliminate beneficial that are fixed

The Effect of Slightly Deleterious Mutations

- If consider deleterious mutations of different effect:

$$\frac{\theta_s}{\theta_N(1 - \beta_i)} = \frac{K_s}{K_N(1 - \alpha)}$$

↑ ↑
eliminate eliminate
deleterious beneficial that
present in are fixed
polymorphisms

- Deleterious mutations are present at low frequency in the populations but are hardly fixed.
- Estimate of α considering only low or only high frequency polymorphisms are very different:

The Effect of Slightly Deleterious Mutations

- If consider deleterious mutations of different effect:

$$\frac{\theta_s}{\theta_N(1 - \beta_i)} = \frac{K_s}{K_N(1 - \alpha - \beta_d)}$$

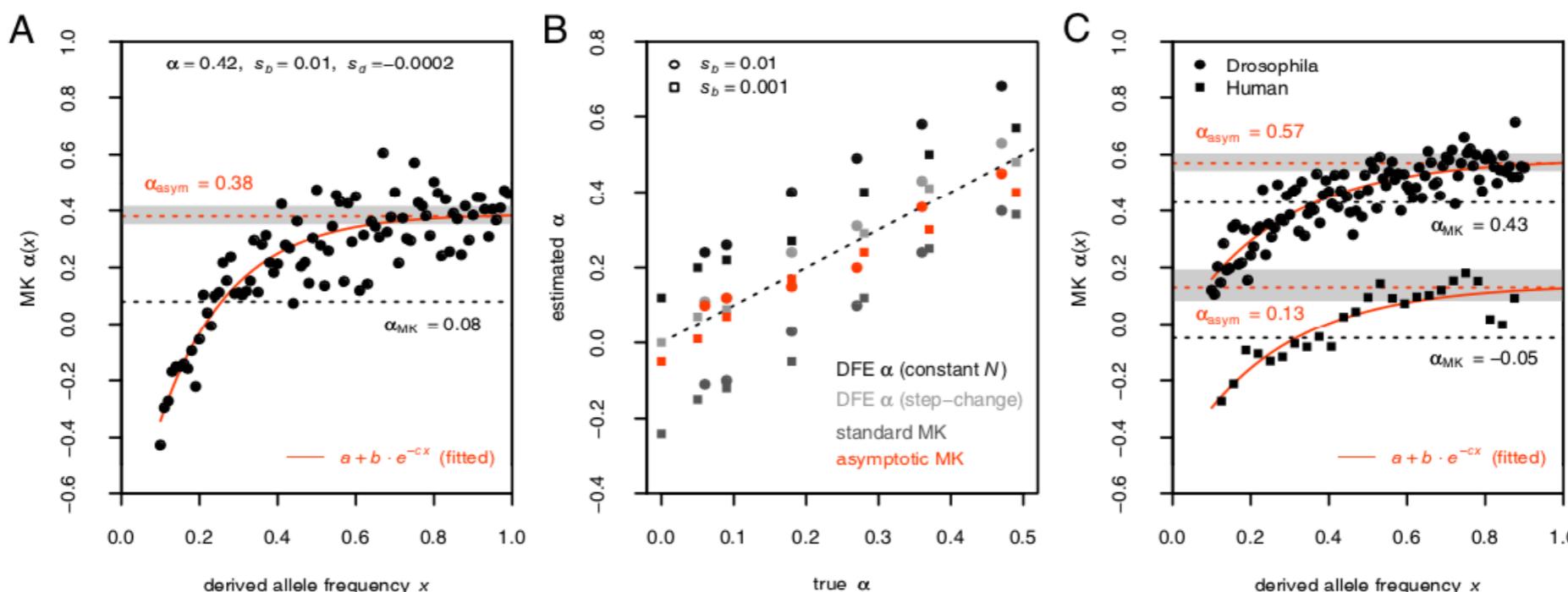
↑ ↑

eliminate deleterious present in polymorphisms

eliminate beneficial that are fixed

- Deleterious mutations are present at low frequency in the populations but are hardly fixed.
- Estimate of α considering only low or only high frequency polymorphisms are very different:

Messer and Petrov (PNAS 2013) estimate α using an asymptotic approach
(estimate α for each polymorphic frequency)

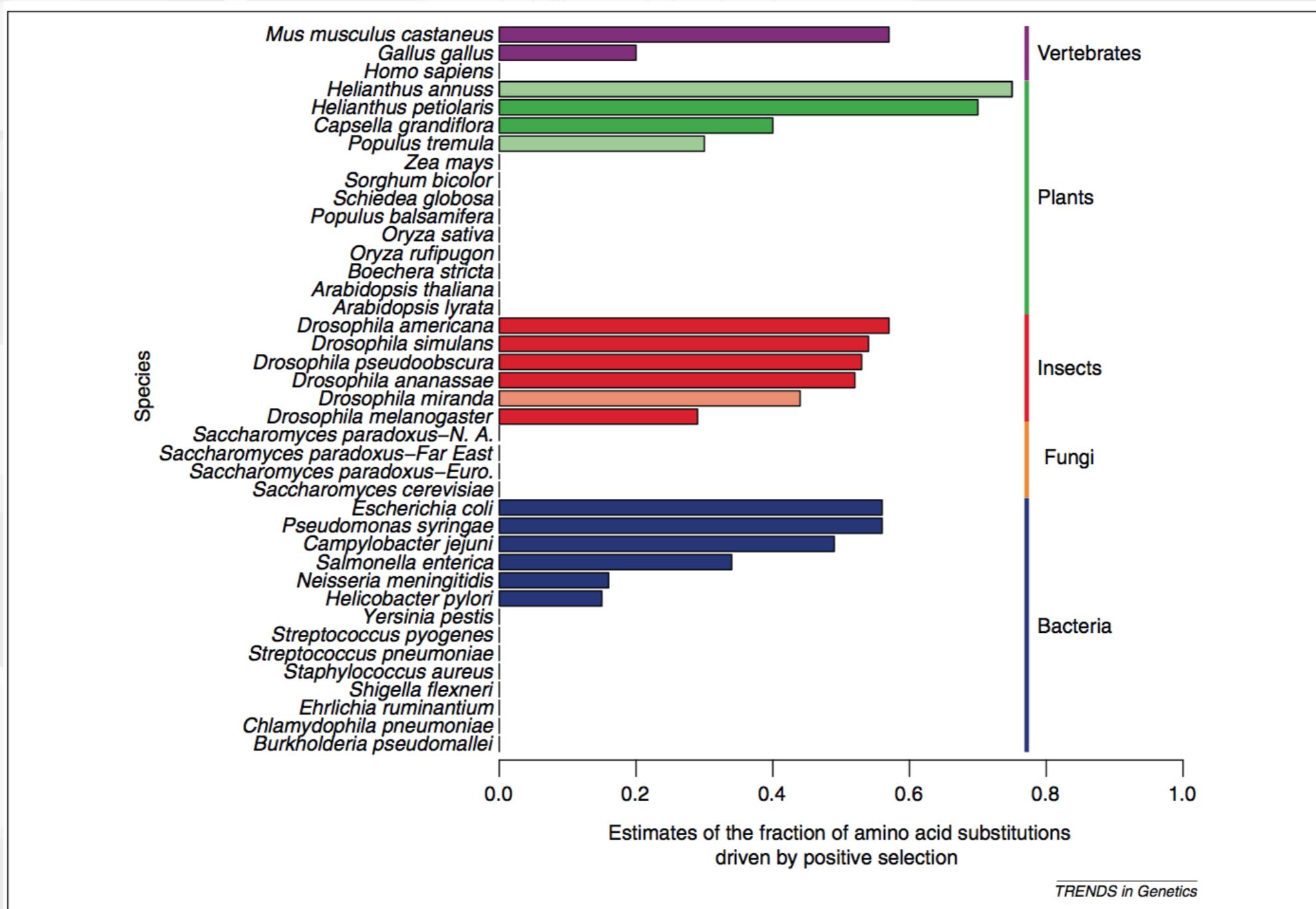


MKT modified framework

- Approaches mostly used to estimate a :
 - Exclude the low frequency variants to eliminate the effect of segregating deleterious mutations.
 - Estimate the asymptotic value of a at high frequencies.

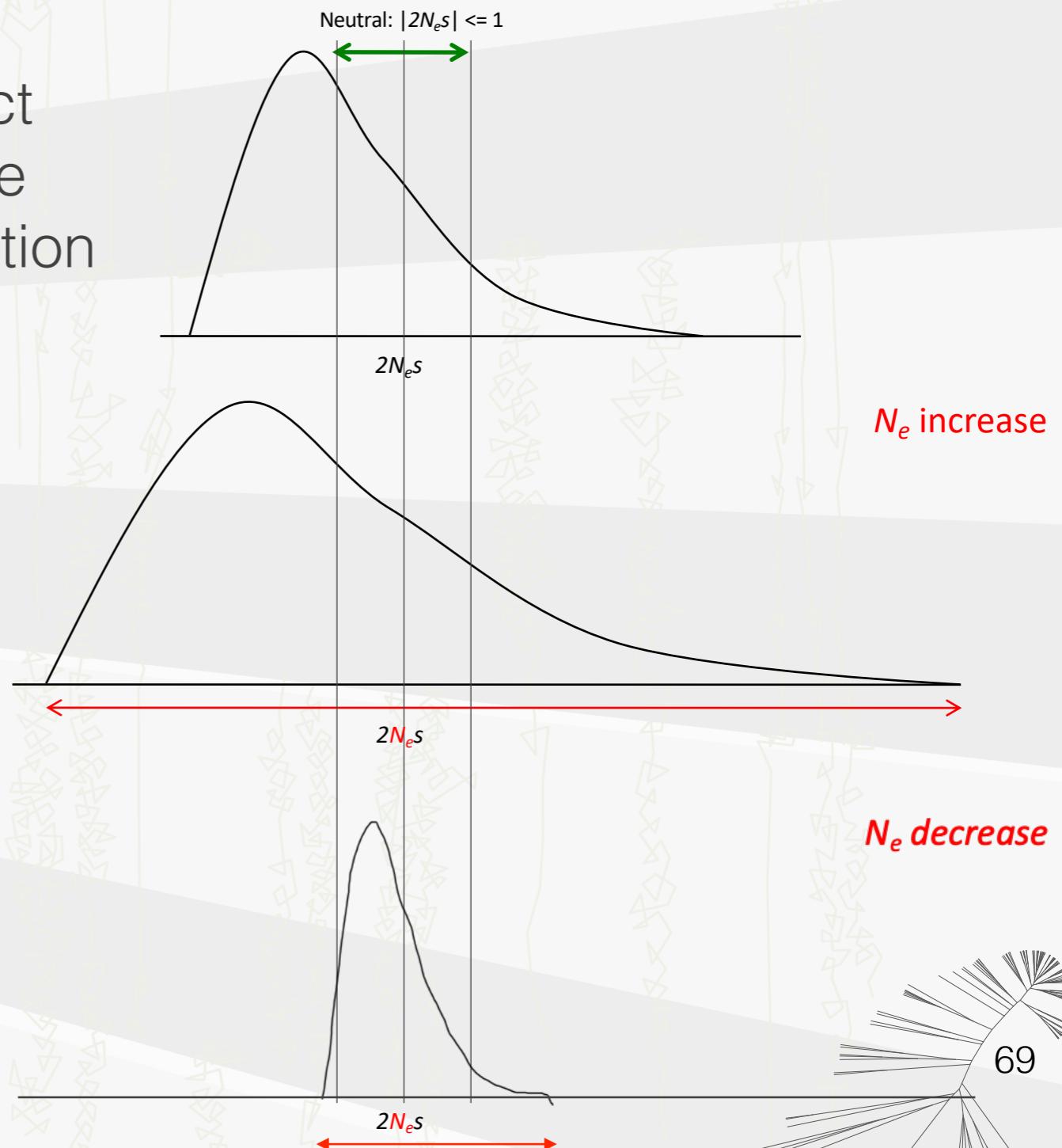
MKT modified framework

- Estimates of ω using a variety of methods:



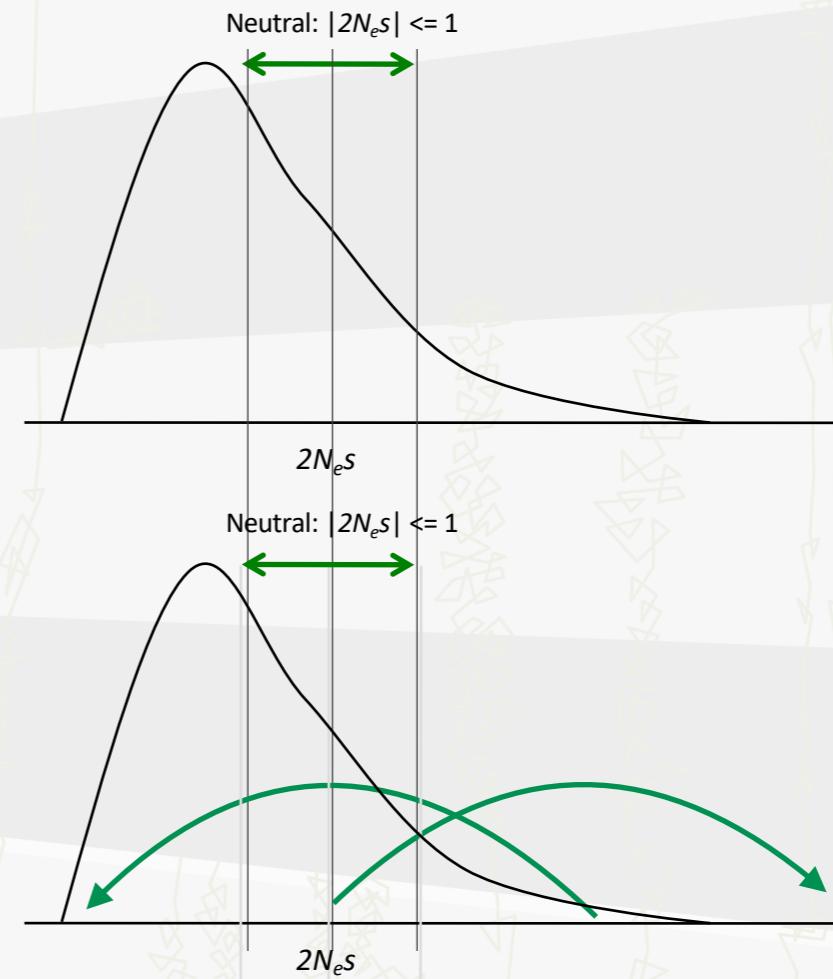
The Distribution of Fitness Effects (DFE)

- Demographic changes may affect the probability of fixation because they change the effective population size and thus the population selective effect $2N_eS$.



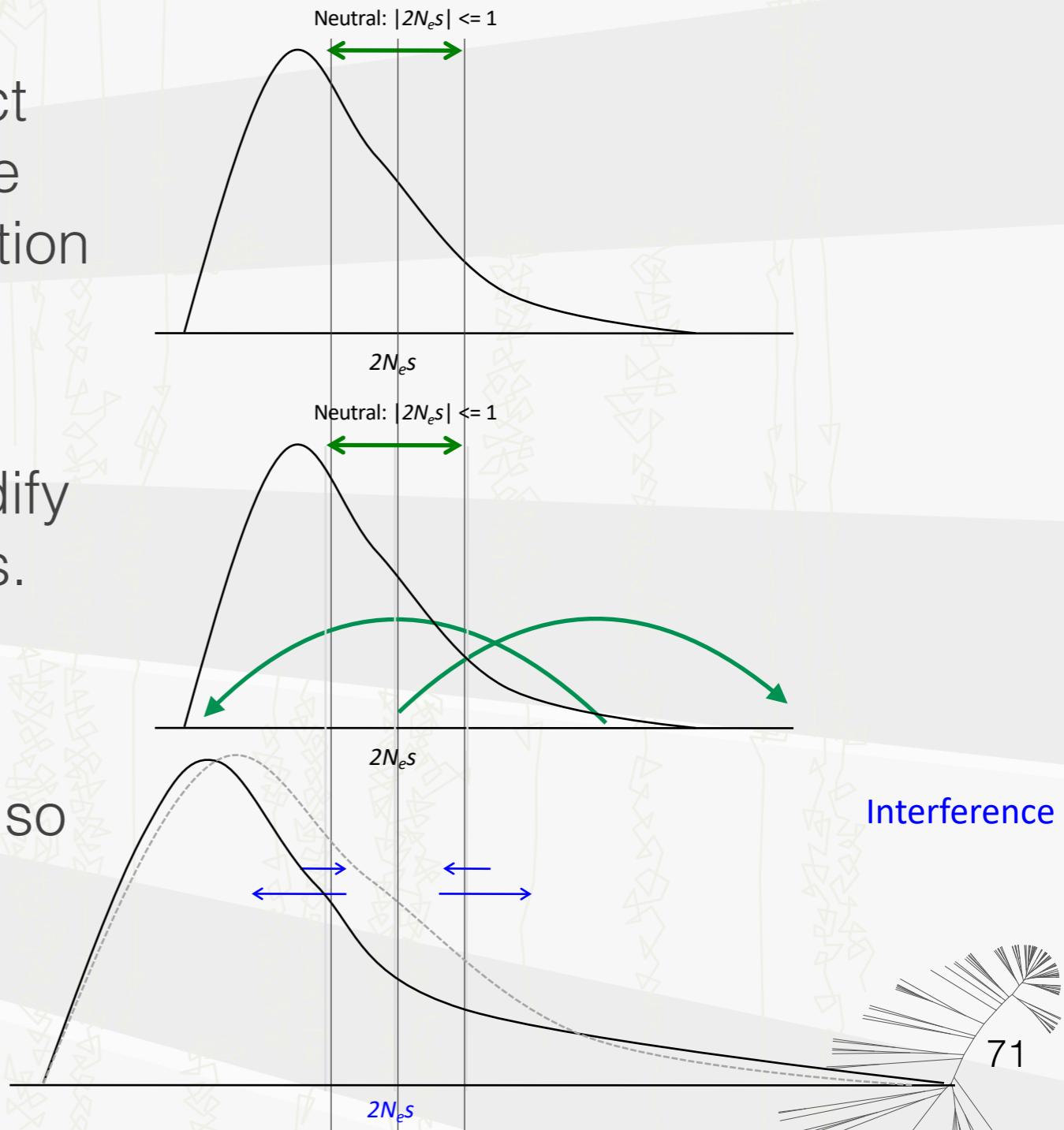
The Distribution of Fitness Effects (DFE)

- Demographic changes may affect the probability of fixation because they change the effective population size and thus the population selective effect $2N_eS$.
- Environmental changes may modify the selective effect of the variants.



The Distribution of Fitness Effects (DFE)

- Demographic changes may affect the probability of fixation because they change the effective population size and thus the population selective effect $2N_eS$.
- Environmental changes may modify the selective effect of the variants.
- The distribution of the selective variants on the genome and their relationship (e.g., linkage) may also modify the general effect of selection on the individuals.



Inference of DFE from Genomics Variability

- Usual Assumptions
 - Independence between positions. Use the **Site Frequency Spectrum** and the Divergence as Observations.
 - Few or no demographic processes (no migration, no population structure, few or no changes in N_e). But see the use of nuisance parameters.
 - Effect of Directional Selection on new mutations.
 - No correlation N_e and f

Inference of DFE from Genomics Variability

Table 3. Summary of studies inferring the distribution of scaled fitness effects, N_s , of nonsynonymous mutations

| Organism | Method/dem model | $-1 < N_s < 0$ | $-10 < N_s < -1$ | $-10 < N_s$ | Distribution(s) fitted | References |
|---------------------------------|--------------------------------|----------------|------------------|-------------|-----------------------------------|------------|
| Human | Diffusion + complex demography | 0.27 | 0.30 | 0.43 | Mix of normal exponential/neutral | 57 |
| Human | EWK2009 | 0.35 | 0.09 | 0.56 | Γ | 56 |
| <i>Mus musculus castaneus</i> | K&K | 0.19 | 0 | 0.81 | LN, Γ , β , Spikes | 59 |
| <i>Pan troglodytes</i> | EWK2009 | 0.09 | 0.06 | 0.74 | Γ | 76 |
| <i>D. melanogaster</i> | EWK2009 | 0.06 | 0.07 | 0.87 | Γ | 56 |
| <i>Saccharomyces cerevisiae</i> | | 0.25 | 0.25 | 0.5 | Γ | 63 |
| <i>S. paradoxus</i> | | 0.2 | 0.2 | 0.6 | | |
| Angiosperms | EWK2009 | 0.1–0.35 | 0.05–0.15 | 0.7–0.8 | Γ | 77 |
| <i>Medicago truncatula</i> | EWK2009 | 20–35 | 12–15 | 50–65 | Γ | 78 |

NOTE: EWK2009: diffusion based, simple demographic model fitted featuring a possible step change from population size N_1 to population size N_2 at some time t in the past (N_1 , N_2 , and t become “nuisance parameters” estimated alongside DFE and the fraction of favorable mutations).

K&K: discrete W–F matrix based, demographic model identical to EWK2009.

Dem, demographic; LN, log normal; spikes, spikes at different N_s class values.

Inference of DFE from Genomics Variability (e.g., polyDFE)

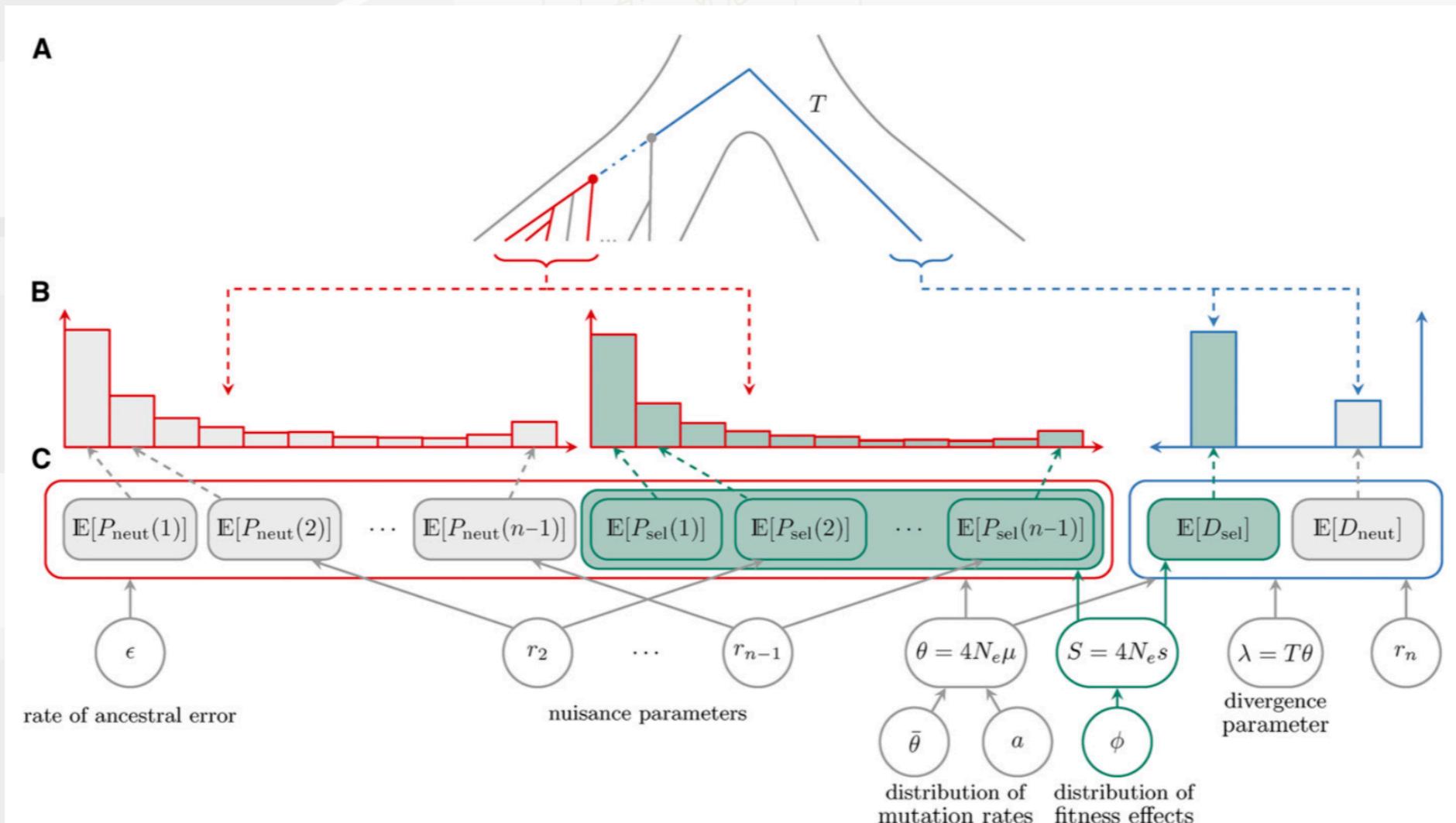


Figure 1 Schematic of data and the hierarchical model assumed in the inference method. Throughout the figure, gray and blue fill indicates sites that are assumed to be evolving neutrally or potentially under selection, respectively; while red and blue outlines indicates polymorphism and divergence data (expectations), respectively. (A) The history and coalescent tree of two populations: the ingroup (on the left side), for which polymorphism data are collected, and the outgroup (on the right side), for which divergence counts are obtained. A total of n sequences are sampled from the ingroup (marked in red), with the MRCA marked with a red circle. The MRCA of the whole ingroup population is marked with a gray circle. From the outgroup we typically have access to one sequence (marked in blue). The total evolutionary time between the MRCA of the sample (red circle) and the sampled outgroup sequence can be divided into the time from the MRCA of the sample to the MRCA of the whole ingroup population (blue dot-dash line) and T , the time from the ingroup MRCA to the sampled outgroup sequence (blue solid line). (B) Observed SFS and divergence counts $[p_z(i)]$ and d_z , with $z \in \{\text{neut}, \text{sel}\}$ and $1 \leq i < n$. (C) Expected counts $[\mathbb{E}[P_z(i)]$ and $\mathbb{E}[D_z]$, with $z \in \{\text{neut}, \text{sel}\}$ and $1 \leq i < n$, model parameters, and relations between parameters, expectations, and data. The dashed gray and blue ↓'s connect observed counts from (B) with matching expected counts from (C).

Inference of DFE from Genomics Variability (e.g., polyDFE)

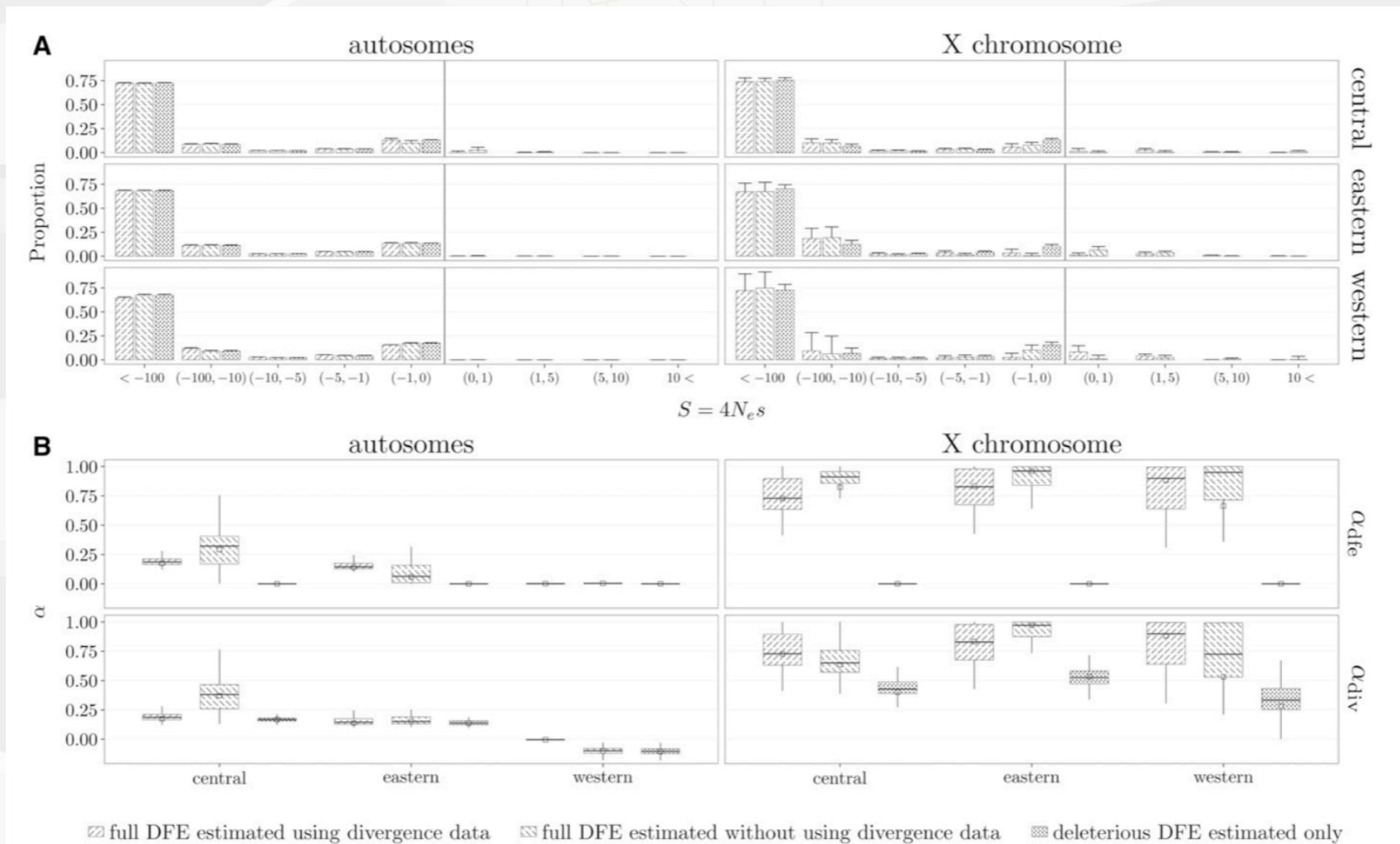


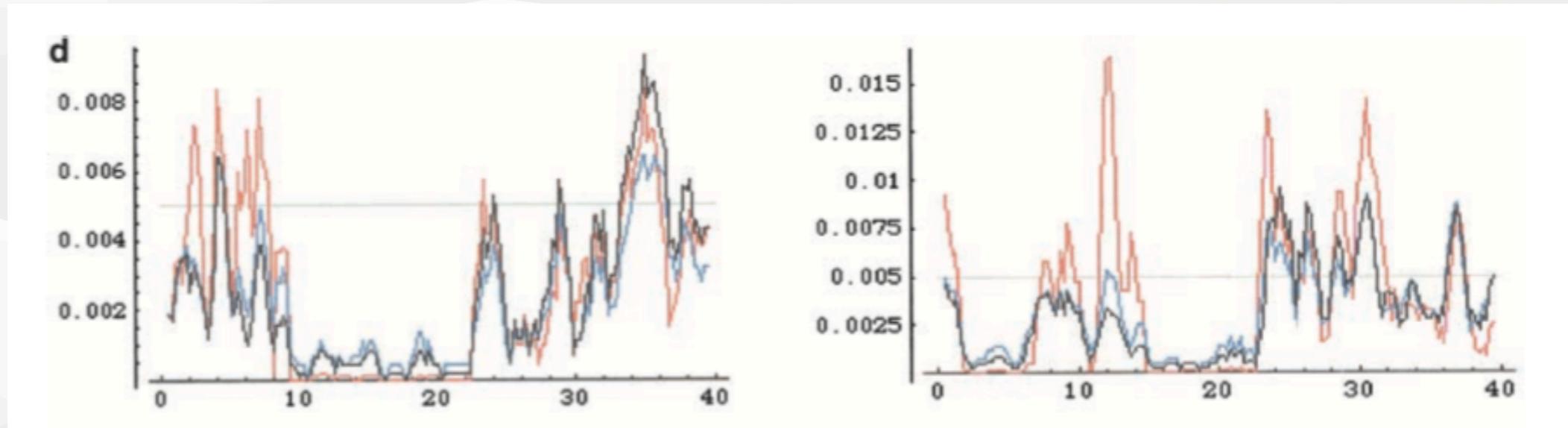
Figure 8 Inference of DFE and α (proportion of beneficial substitutions) on three chimpanzee subspecies. A full DFE was inferred from both polymorphism and divergence data, while also both a full DFE and deleterious DFE were inferred from polymorphism data only. (A) Inferred discretized DFE. The error bars indicate 1 SD obtained from the inferred discretized DFEs from 100 bootstrap data sets. (B) Box plot of inferred α_{dfe} and α_{div} from the 100 bootstrap data sets. The values of α inferred on the original data sets are given as empty squares. Note that when inferring a deleterious DFE only, α_{dfe} is zero.

Relationship between traits and fitness

- Traits can be uncorrelated with fitness (neutral traits) while others may be strongly associated.
- Traits can be determined by a single gene or alternatively by a large number of genes (Quantitative Trait Loci, QTL).
- Traits can change their association with fitness in function of the environment.
- Loci can affect single trait or alternatively affect a large number of traits (pleiotropic scenario).

Relationship between traits and fitness

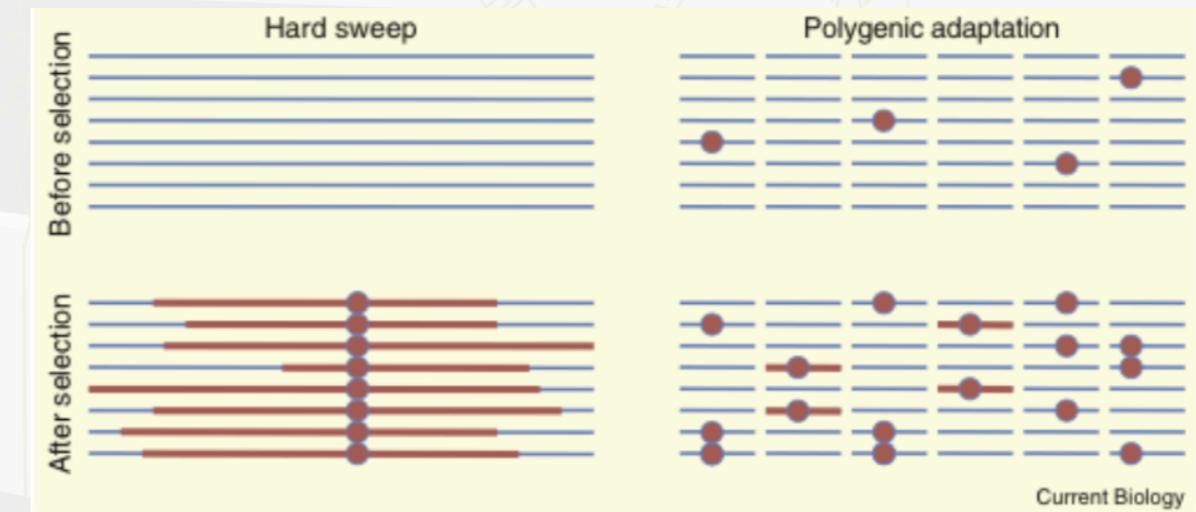
- The relation of the genes that control the trait and the relation of these traits with the fitness determine the pattern of variability at the genome.
 - Traits affecting severely the fitness that are determined by a single gene would show a selective sweep pattern.



(Kim and Stephan, Genetics 2002)

Relationship between traits and fitness

- The relation of the genes that control the trait and the relation of these traits with the fitness determine the pattern of variability at the genome.
- Instead, **a trait** affecting fitness that is determined by **many genes** may only show a slight increase in frequencies in all them (**infinitesimal effect**).
- In this case, it is very important to determine **what loci are affecting the trait** (for example, using GWAS methods).



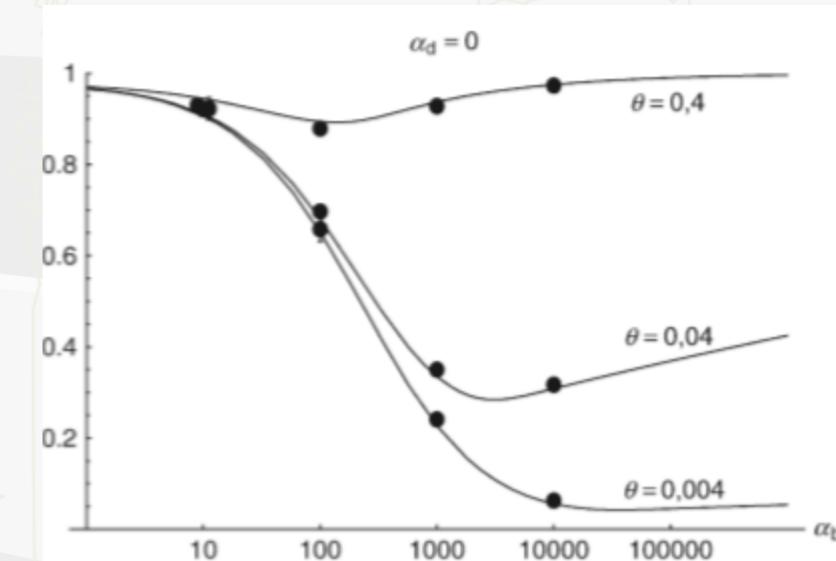
(Pritchard et al. Curr Biology 2010)

Relationship between traits and fitness

- The relation of the genes that control the trait and the relation of these traits with the fitness determine the pattern of variability at the genome.
- Perhaps **stabilizing selection** for a number of traits is maintaining the variability in many loci -instead of an equilibrium between mutation-directional selection-drift-
- Estimation of the effect of the loci implicated and their number may explain the observed patterns of association under this model. The role of genetic redundancy.

Relationship between traits and fitness

- The relation of the genes that control the trait and the relation of these traits with the fitness determine the pattern of variability at the genome.
- **Changes in the environment** can modify the phenotype - and consequently the variability in loci that are implicated in fitness - producing new genomic patterns.
Soft selective sweeps may be common.



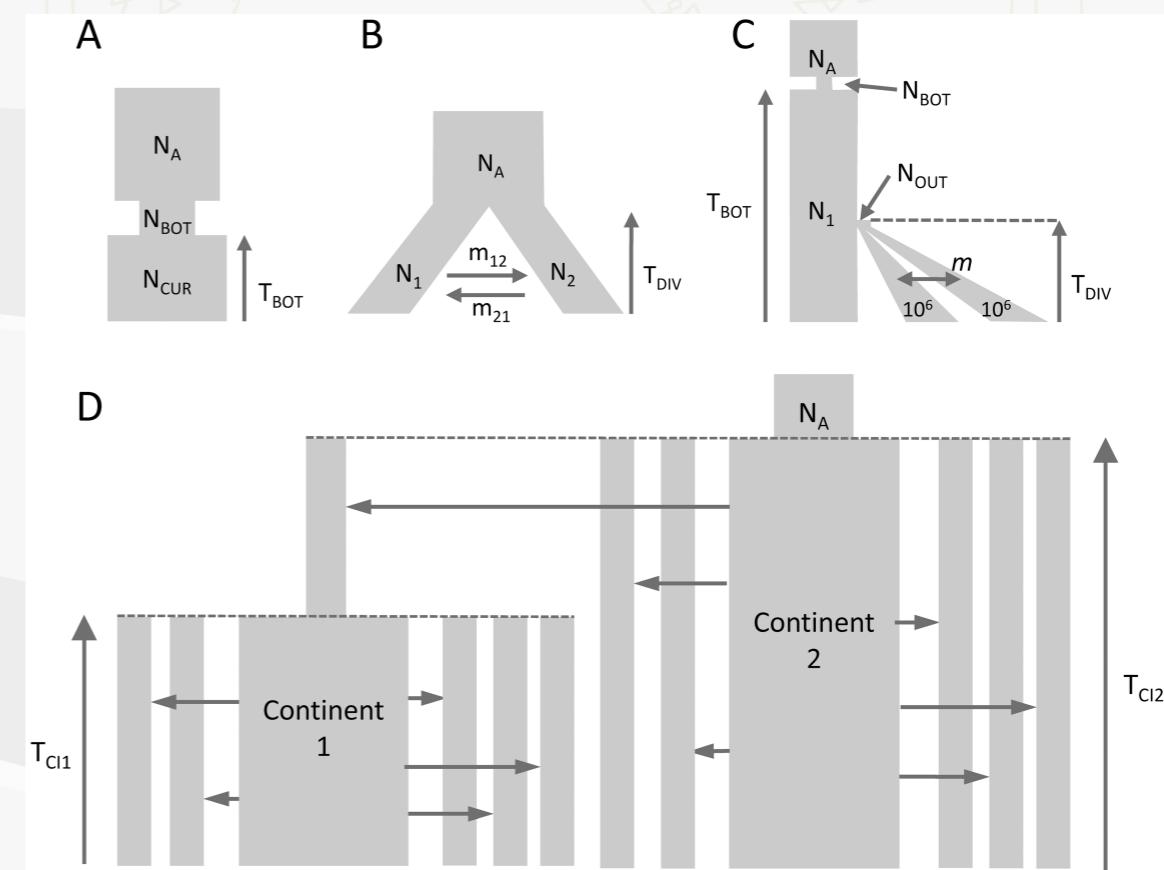
(Probability that an adaptive substitution is from standing genetic variation.

Hermission and Pennings. Genetics 2015)

α_d means deleterious/neutral selection before environmental change. α_b means beneficial after environmental change,

Relationship between traits and fitness

- The relation of the genes that control the trait and the relation of these traits with the fitness determine the pattern of variability at the genome.
- Demographic Changes** may play a role in the genetic architecture by the random modification of the number of implicated loci with different effect in fitness.



(Excoffier et al. PLoS Genet. 2013)

Review Lectures

- Fitness and its role in evolutionary genetics. H. Allen Orr. *Nature Reviews Genetics*. Volume 10, August 2009. pp 531-539.
- Effects of new mutations on fitness: insights from models and data. Thomas Bataillon and Susan F. Bailey. *Ann. N.Y. Acad. Sci.* 2014. pp. 76-92. ISSN 0077-8923.
- Molecular Population Genetics. Casillas S. and Barbadilla A. *Genetics*, Vol. 205, 1003–1035, 2017. doi: 10.1534/genetics.116.196493.
- A population genetic interpretation of GWAS findings for human quantitative traits. Simons YB, Bullaughey K, Hudson RR, Sella G. 2018. *PLoS Biol* 16(3): e2002985.
- The Importance of Genetic Redundancy in Evolution. Láruson, Yeaman and Lotterhos. *Trends in Ecology & Evolution*, 2020, Vol. 35, No. 9.
- A null model for the distribution of fitness effects of mutations. Cotto O, et al. *Proc Natl Acad Sci U S A*. 2023. PMID: 37252948.
- Phenotypic selection in natural populations: what have we learned in 40 years? Svensson E.I, *Evolution*, 2023, 77(7), 1493–1504. <https://doi.org/10.1093/evolut/qpad077>.