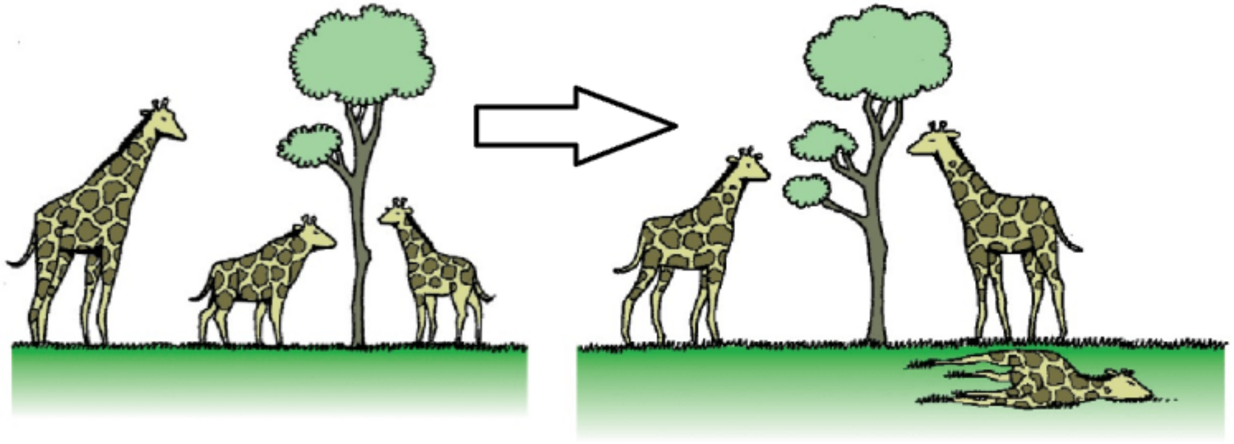
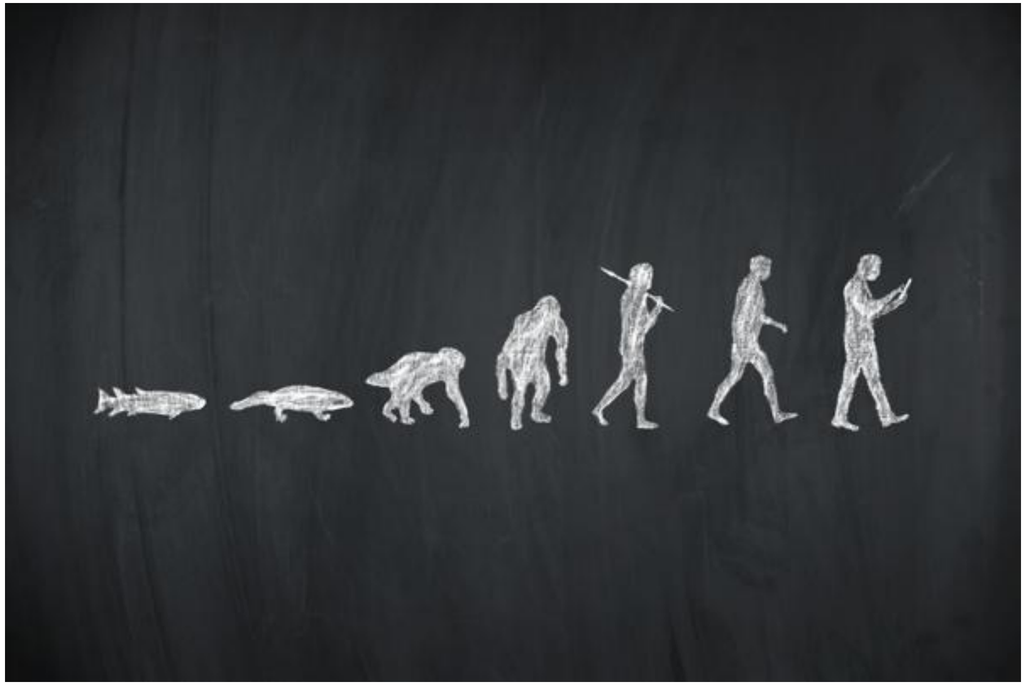
Population summary statistics & *Arlequin*

Qiaoyun Ai
Jan 8, 2019

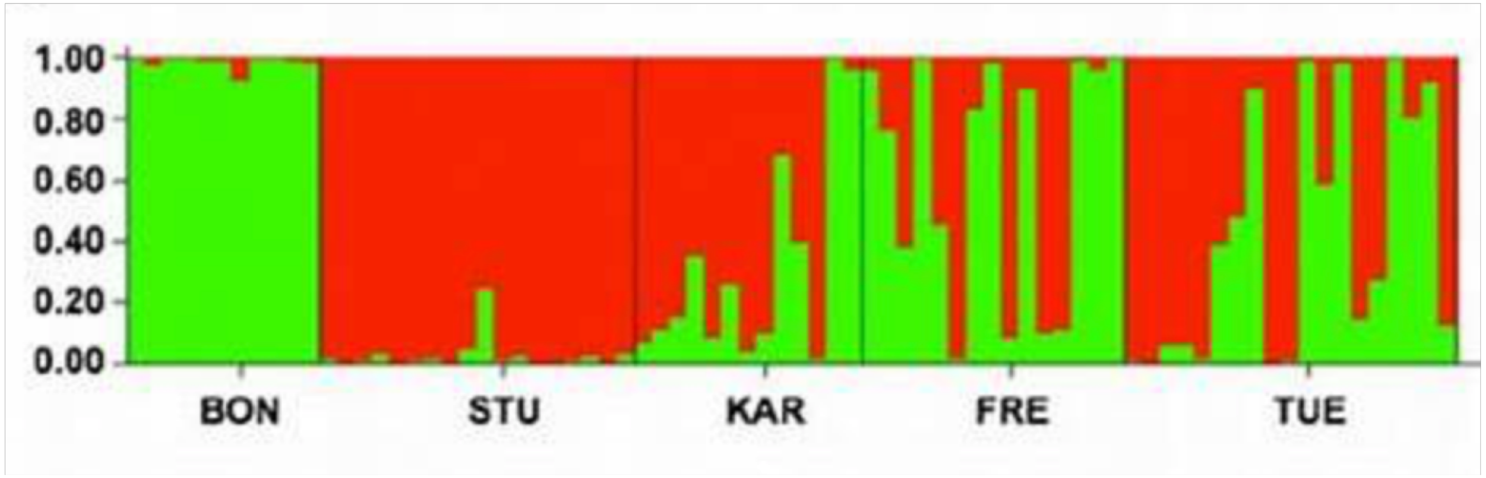
Population genetics



Adaptation



Speciation

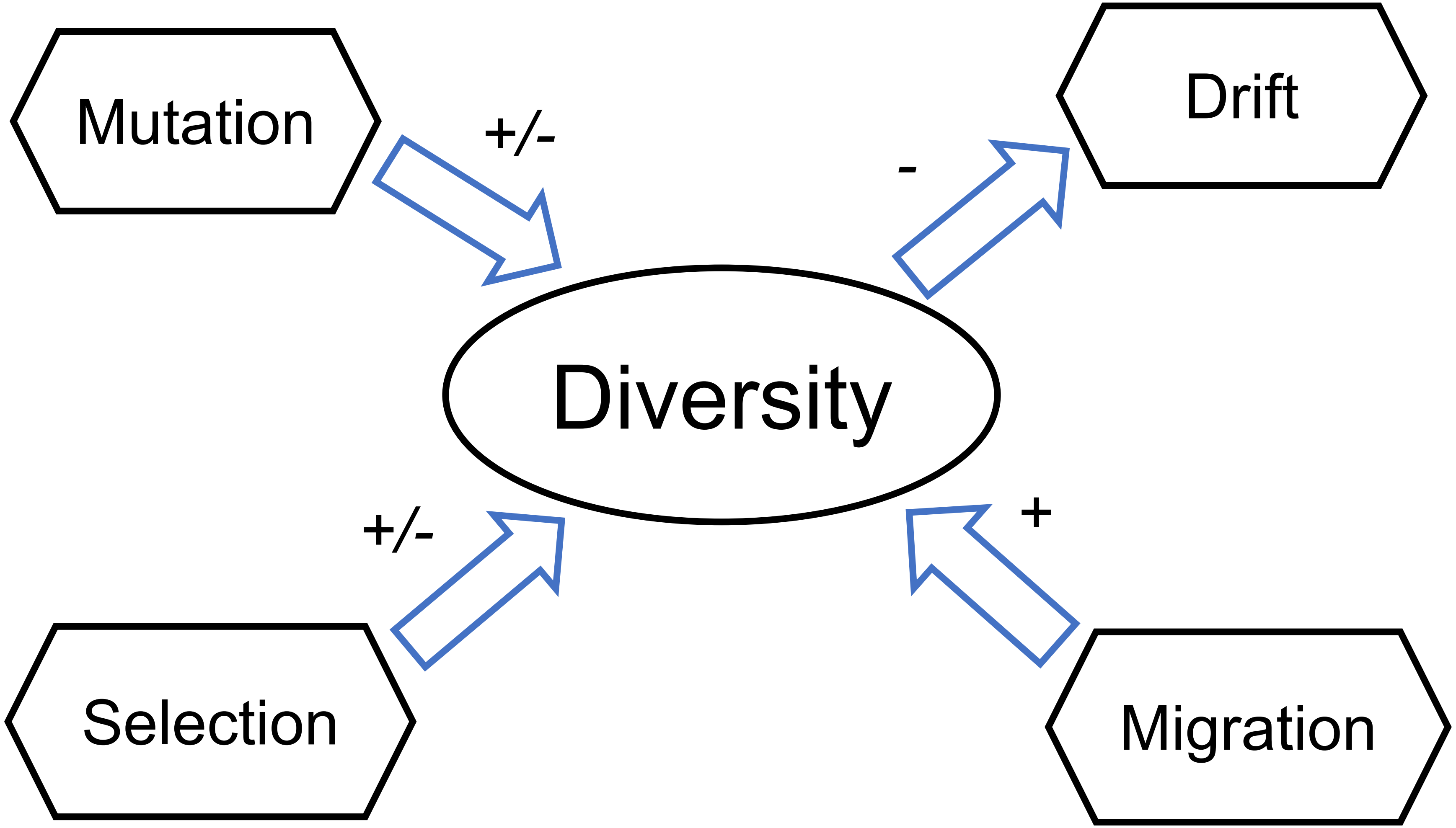


Population structure

Genetics diversity

Genetics diversity study for selective breeding

Factors control genetic diversity within population



Mutation

+/-

Drift

-

Diversity

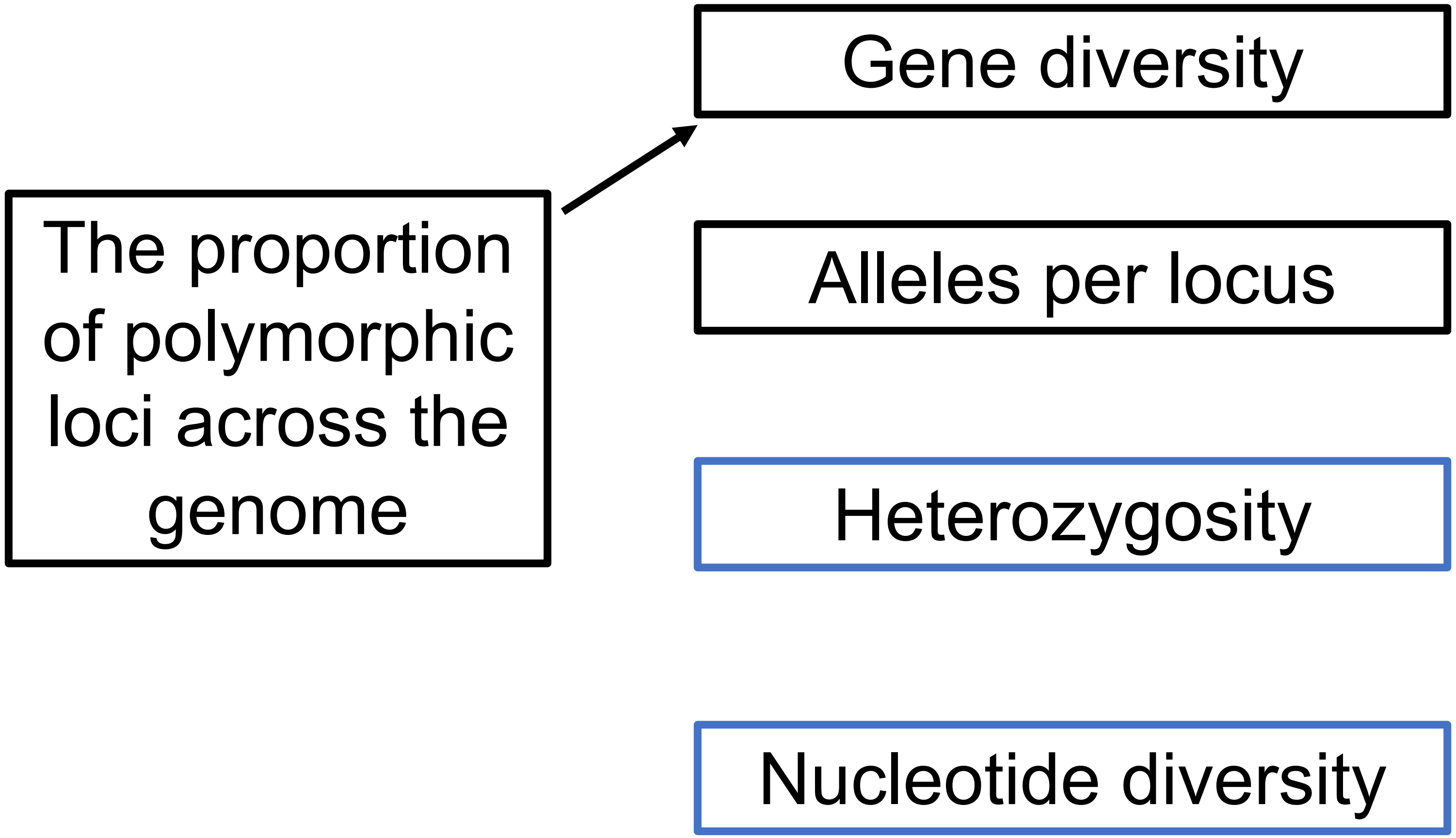
+/-

+

Selection

Migration

Genetic diversity



Gene diversity

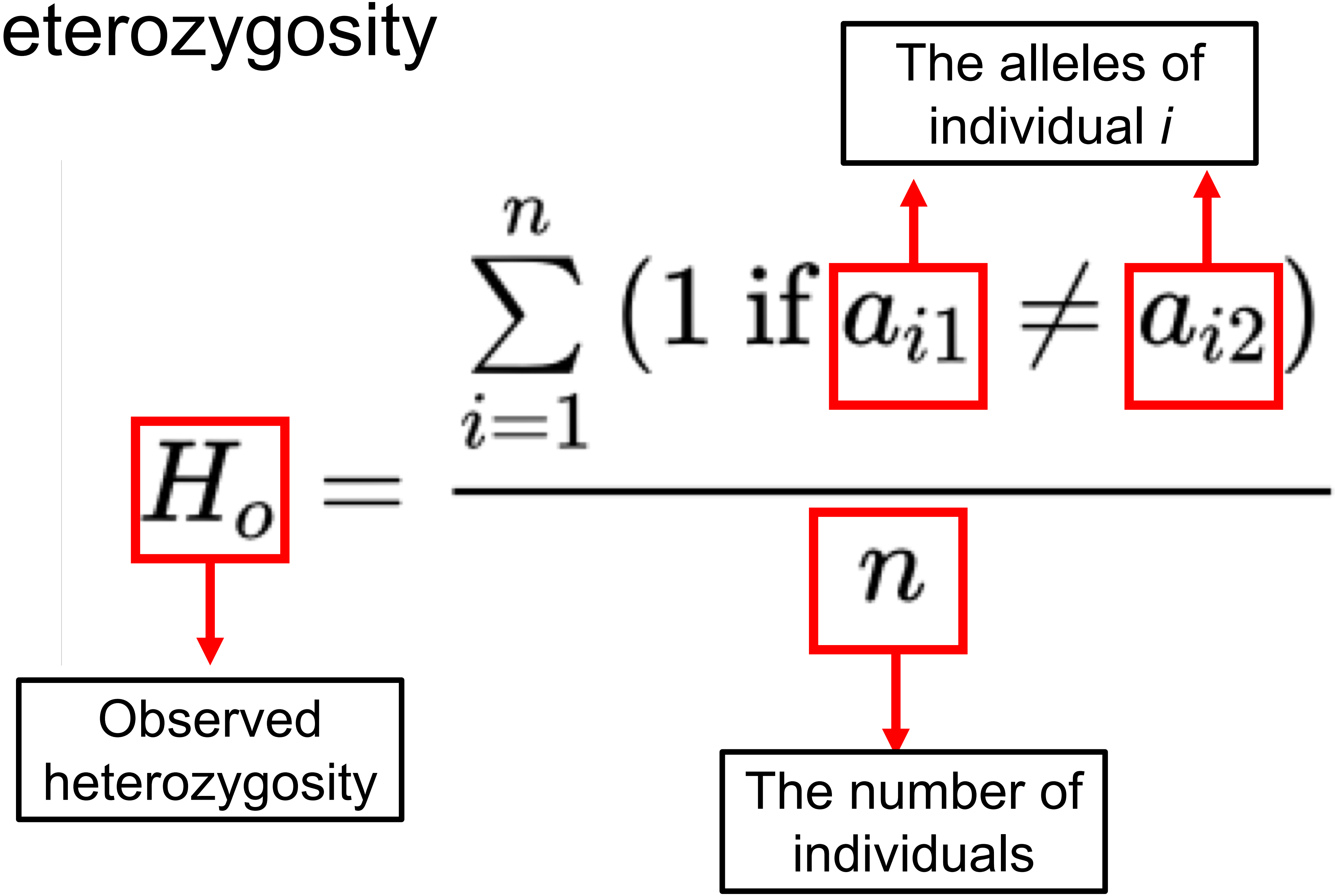
The proportion of polymorphic loci across the genome

Alleles per locus

Heterozygosity

Nucleotide diversity

Heterozygosity



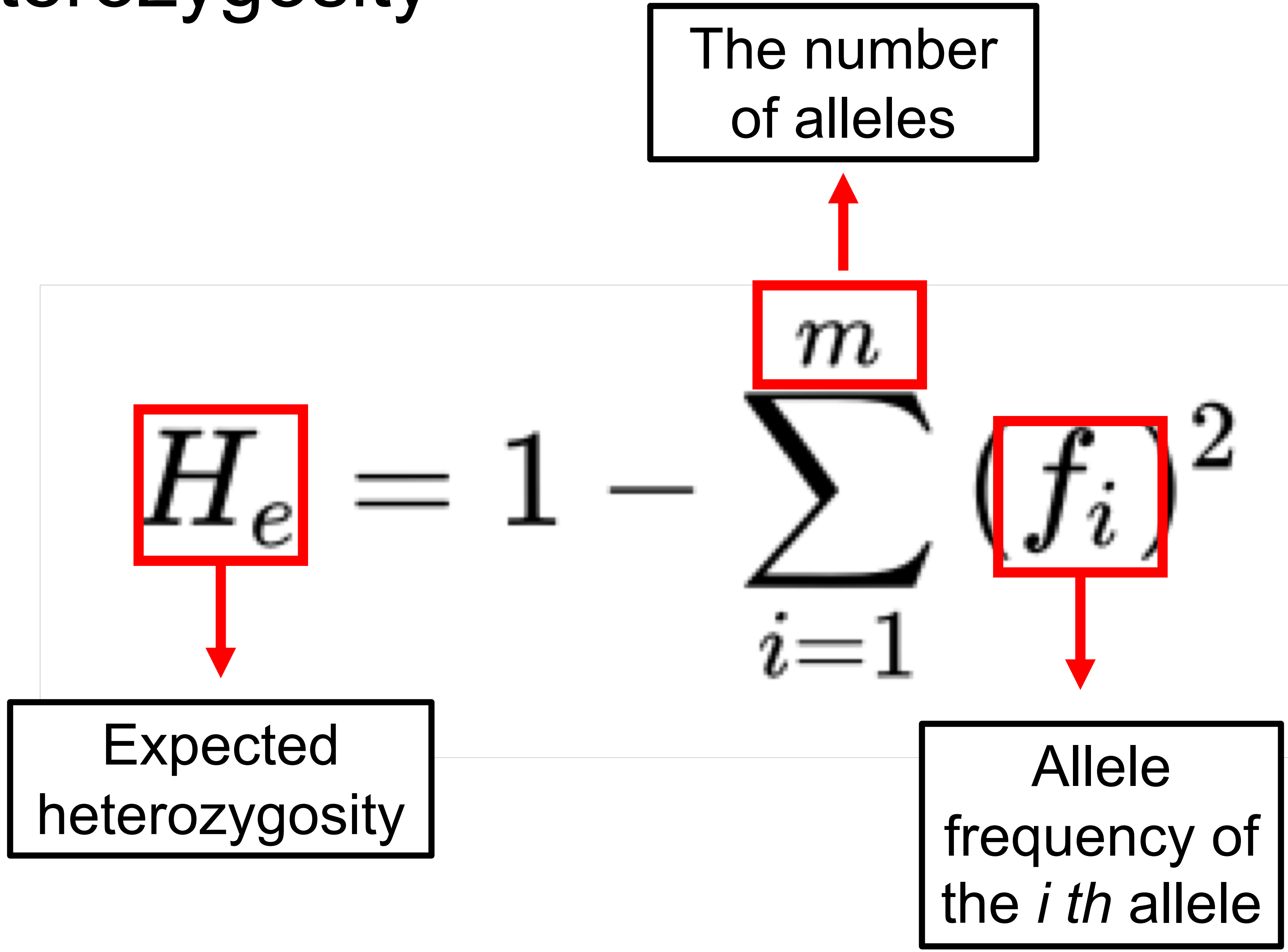
$$\boxed{H_o} = \frac{\sum_{i=1}^{n} (1 \text{ if } \boxed{a_{i1}} \neq \boxed{a_{i2}})}{\boxed{n}}$$

The alleles of individual *i*

Observed heterozygosity

The number of individuals

Heterozygosity



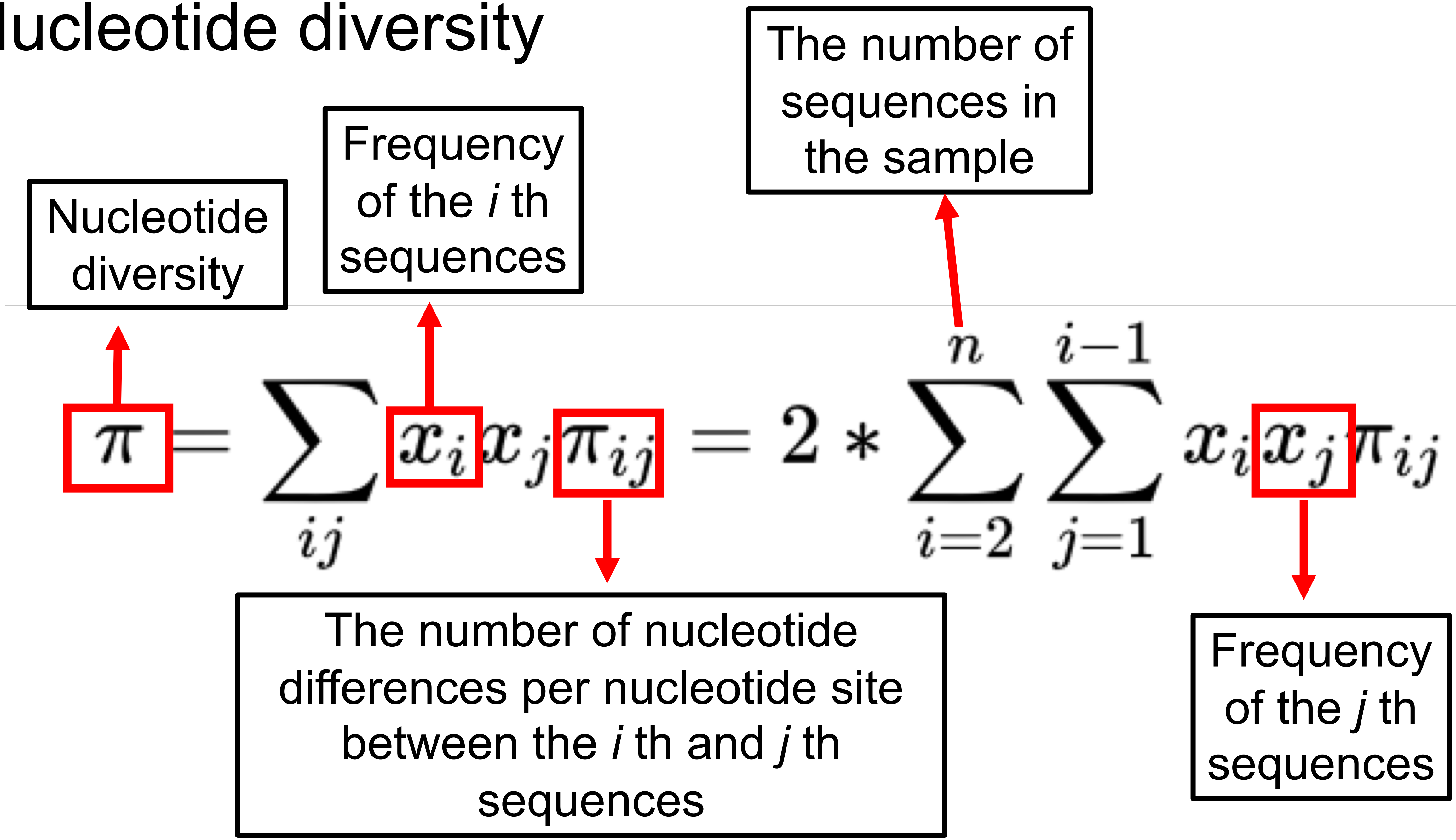
$$H_e = 1 - \sum_{i=1}^{m} (f_i)^2$$

The number of alleles

Expected heterozygosity

Allele frequency of the *i th* allele

Nucleotide diversity



Nucleotide diversity

Frequency of the *i* th sequences

The number of sequences in the sample

$$\boxed{\pi} = \sum_{ij} \boxed{x_i} x_j \boxed{\pi_{ij}} = 2 * \sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i \boxed{x_j} \pi_{ij}$$

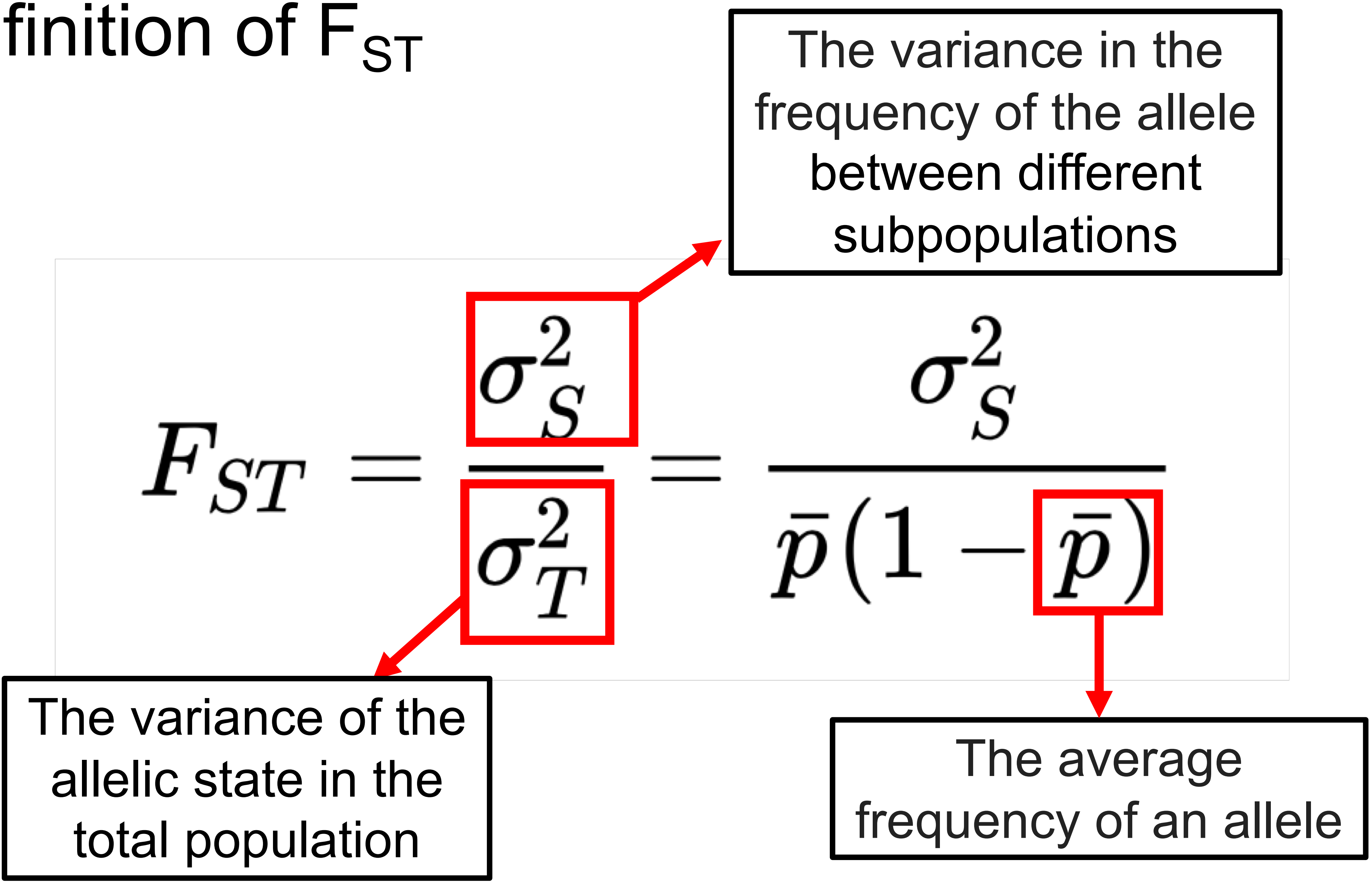
The number of nucleotide differences per nucleotide site between the *i* th and *j* th sequences

Frequency of the *j* th sequences

F_{ST} (fixation index)

F_{ST} is a measure of population differentiation due to genetic structure. It is frequently estimated from genetic polymorphism data, such as SNP or microsatellites.

Definition of F_{ST}



$$F_{ST} = \frac{\boxed{\sigma_S^2}}{\boxed{\sigma_T^2}} = \frac{\sigma_S^2}{\bar{p}(1-\boxed{\bar{p}})}$$

The variance in the frequency of the allele between different subpopulations

The variance of the allelic state in the total population

The average frequency of an allele

Definition of F_{ST}

$$F_{ST} = \frac{\bar{p}(1-\bar{p}) - \sum \overset{\text{The allele frequency in the } i \text{ th population}}{\boxed{c_i}}\underset{\text{The relative size of the } i \text{ th population}}{\boxed{p_i}}(1-p_i)}{\bar{p}(1-\bar{p})}$$

$$= \frac{\bar{p}(1-\bar{p}) - \overline{p(1-p)}}{\bar{p}(1-\bar{p})}$$

Estimation of F_{ST}

The average number of pairwise between two individuals sampled from different sub-populations

$$F_{ST} = \frac{\boxed{\pi_{\text{Between}}} - \boxed{\pi_{\text{Within}}}}{\pi_{\text{Between}}}$$

The average number of pairwise between two individuals sampled from same sub-populations

ARLEQUIN

Ver 3.5

Function introduction

Intra-population

Standard indices
(number of polymorphic sites, gene diversity)

Molecular diversity
(nucleotide diversity)

……

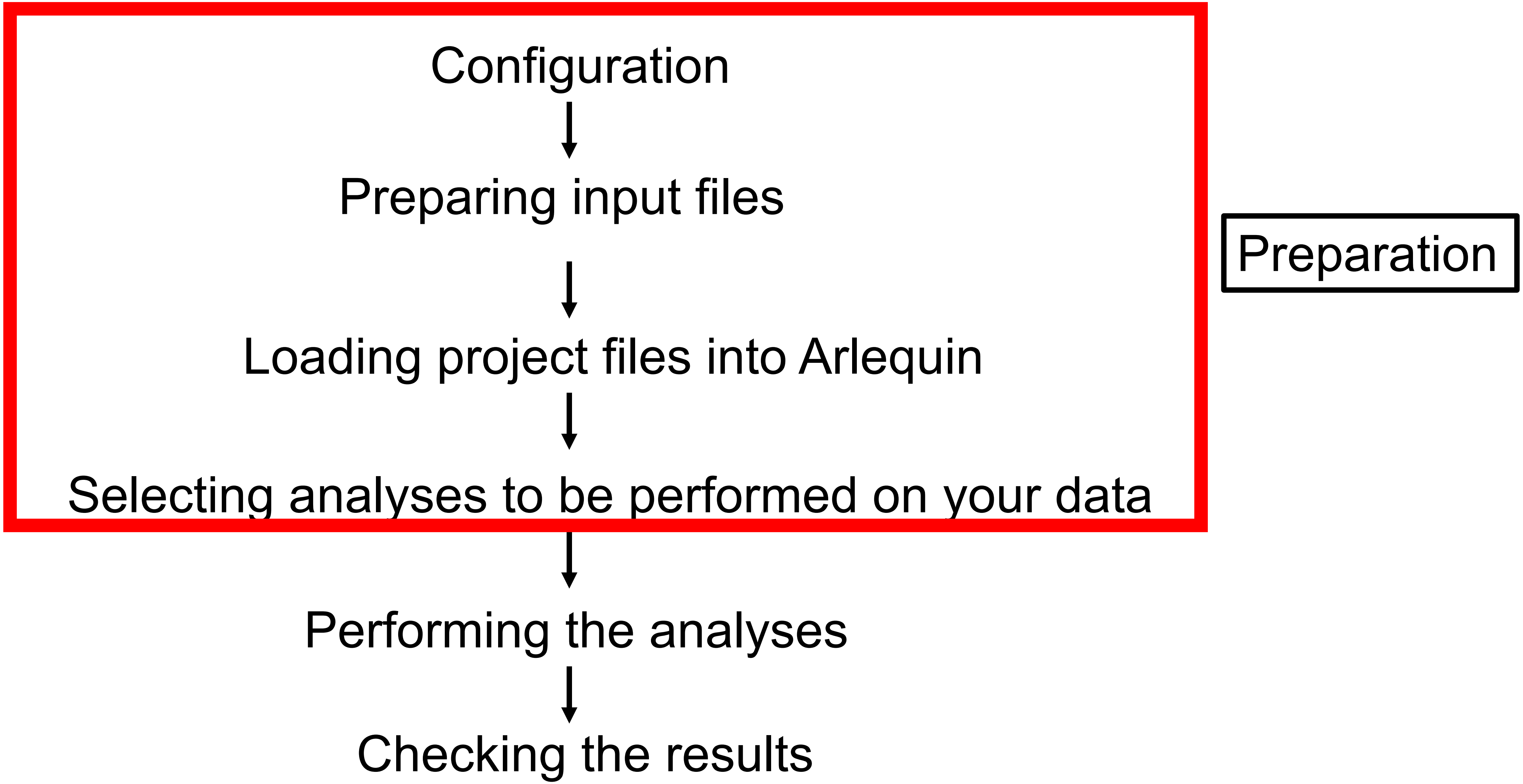
Inter-population

AMOVA
(Different hierarchical Analyses of Molecular Variance to evaluate the amount of population genetic structure)

F_{ST}-Pairwise genetic distances
(F_{ST}-based genetic distances for short divergence time)

……

Usage for graphical version



Configuration

↓

Preparing input files

↓

Loading project files into Arlequin

↓

Selecting analyses to be performed on your data

↓

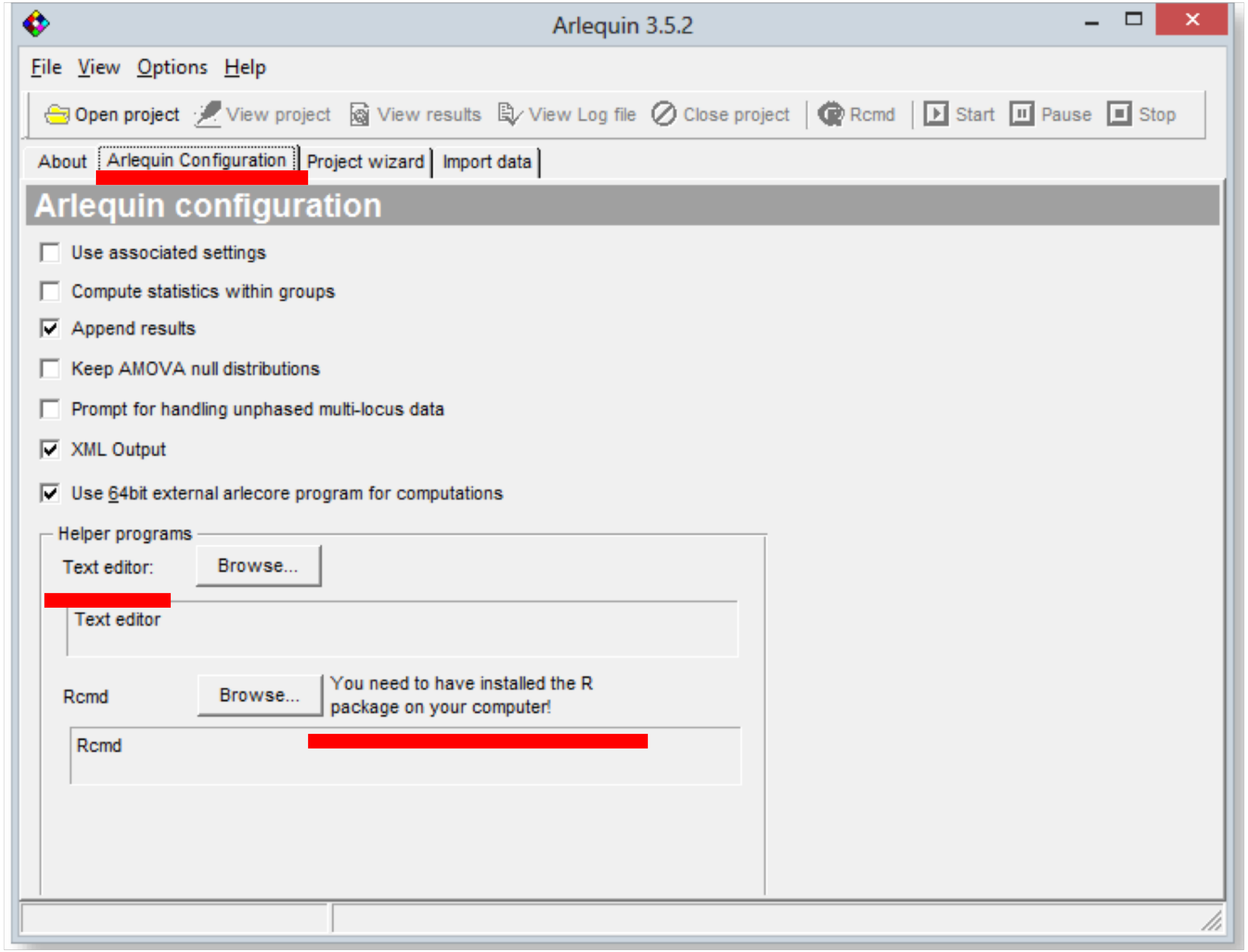
Performing the analyses

↓

Checking the results

Preparation

Configuration



Arlequin configuration

☐ Use associated settings

☐ Compute statistics within groups

☑ Append results

☐ Keep AMOVA null distributions

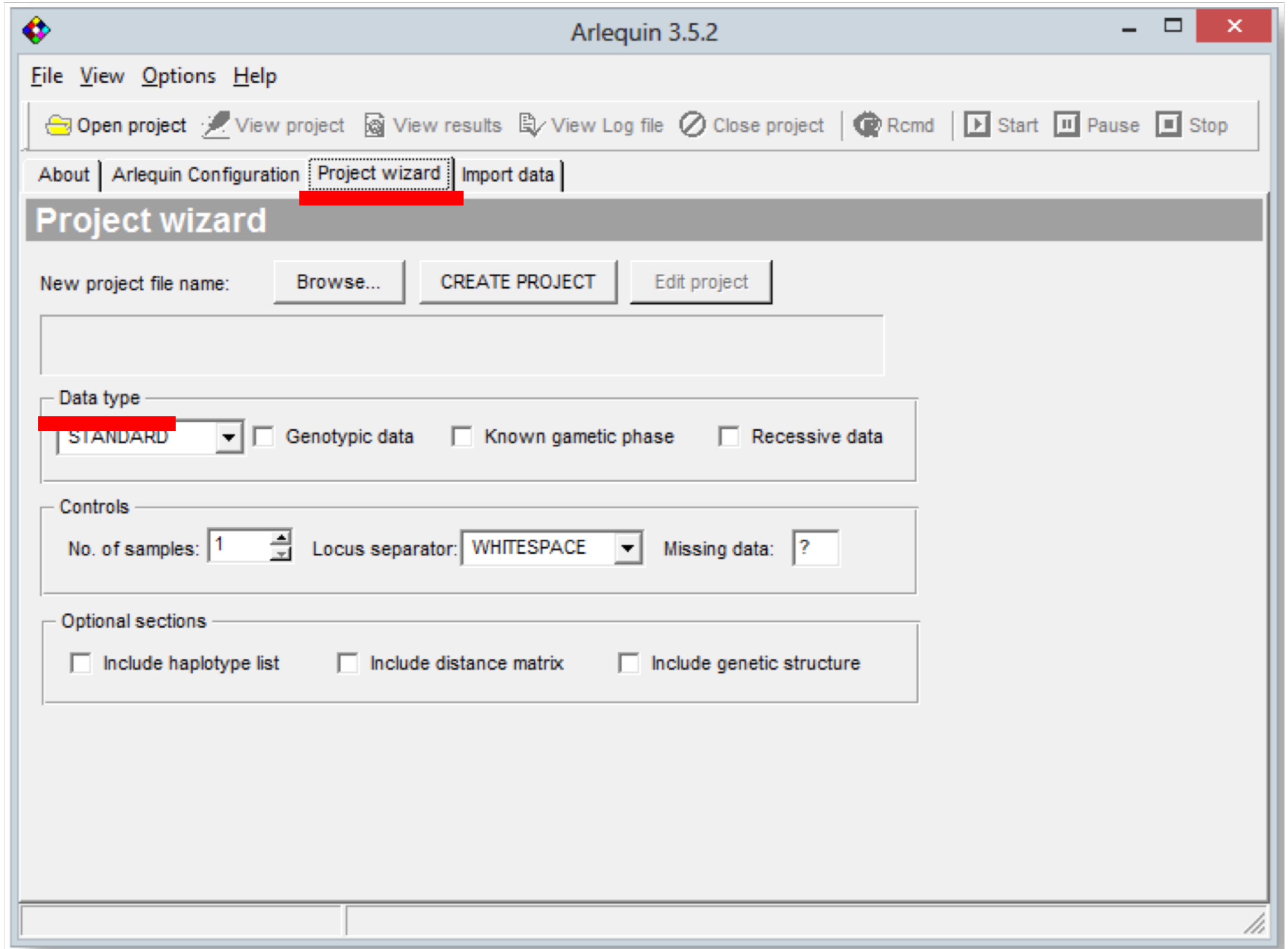
☐ Prompt for handling unphased multi-locus data

☑ XML Output

☑ Use 64bit external arlecore program for computations

You need to have installed the R package on your computer!

Preparing input files



New project file name:

Data type

STANDARD ☐ Genotypic data ☐ Known gametic phase ☐ Recessive data

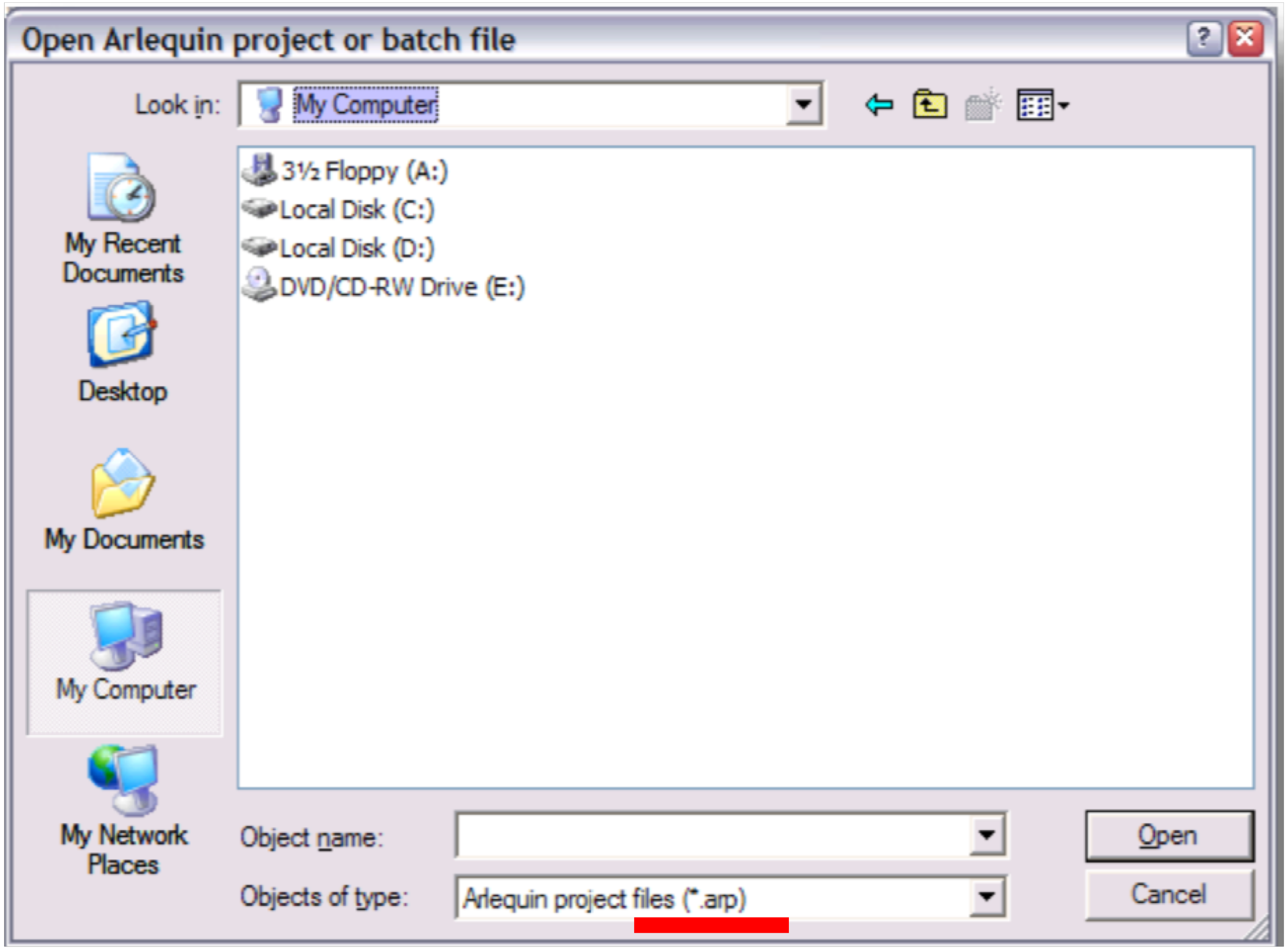
Controls

No. of samples: 1 Locus separator: WHITESPACE Missing data: ?

Optional sections

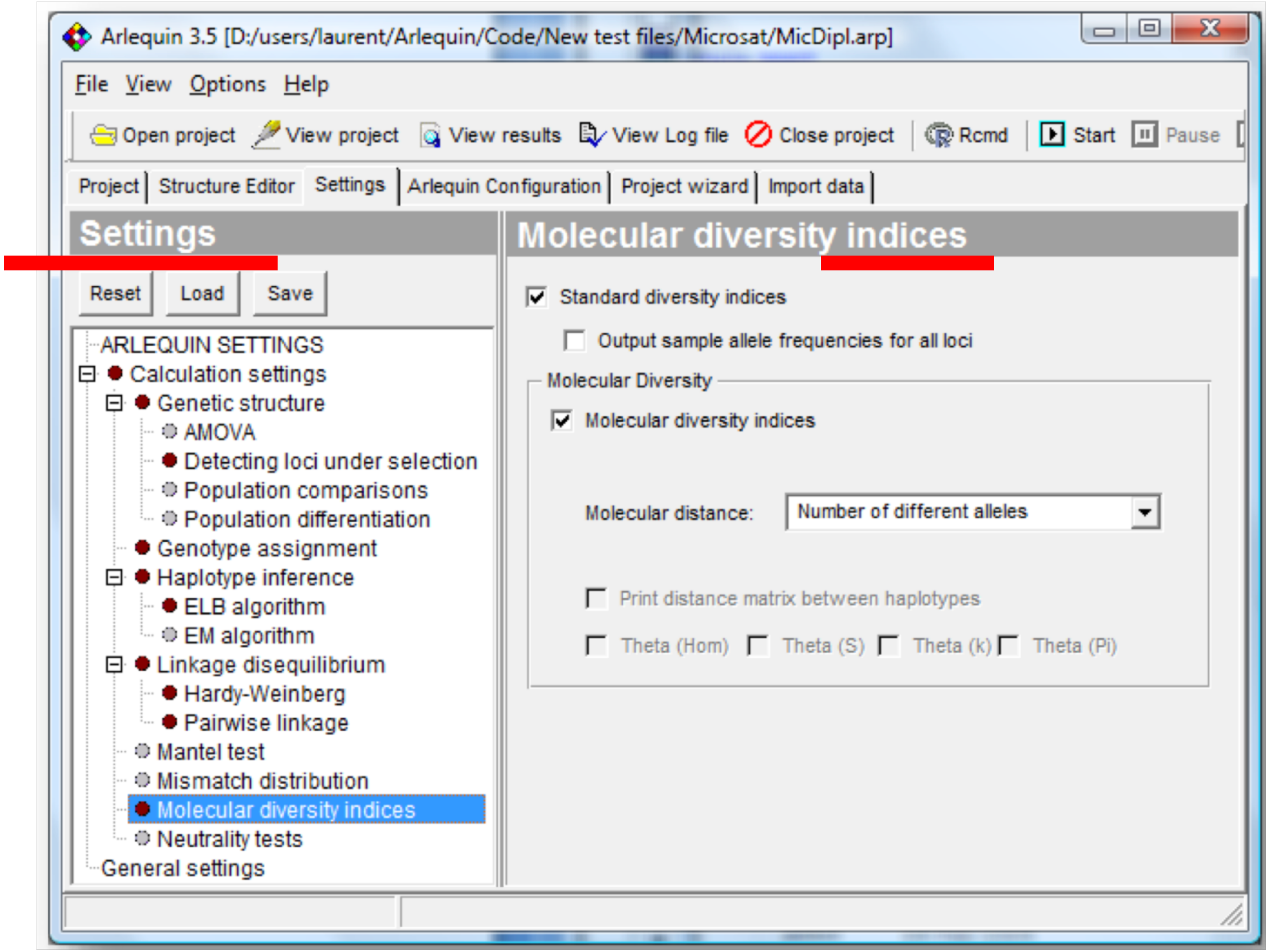
☐ Include haplotype list ☐ Include distance matrix ☐ Include genetic structure

Loading project files into Arlequin



Objects of type: Arlequin project files (*.arp)

Selecting analyses to be performed on your data



☑ Standard diversity indices

☐ Output sample allele frequencies for all loci

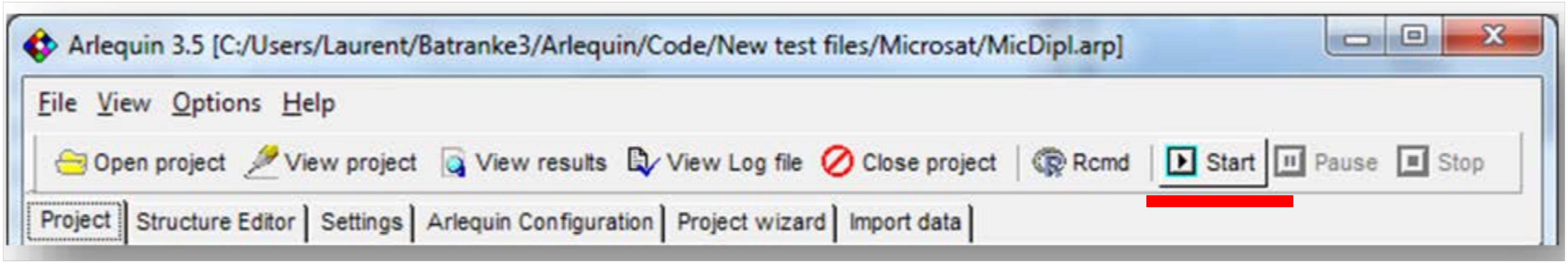
☑ Molecular diversity indices

Molecular distance: Number of different alleles

☐ Print distance matrix between haplotypes

☐ Theta (Hom) ☐ Theta (S) ☐ Theta (k) ☐ Theta (Pi)

Performing the analyses



Manual

http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf

Thank you for your attention!