
Week-3

Genetic drift, mutation and studying the shared evolutionary history of populations

***Aim:** Introduction to mutation and genetic drift. Studying the relationships (allele frequency correlations) between populations: Introduction to F-statistics.*

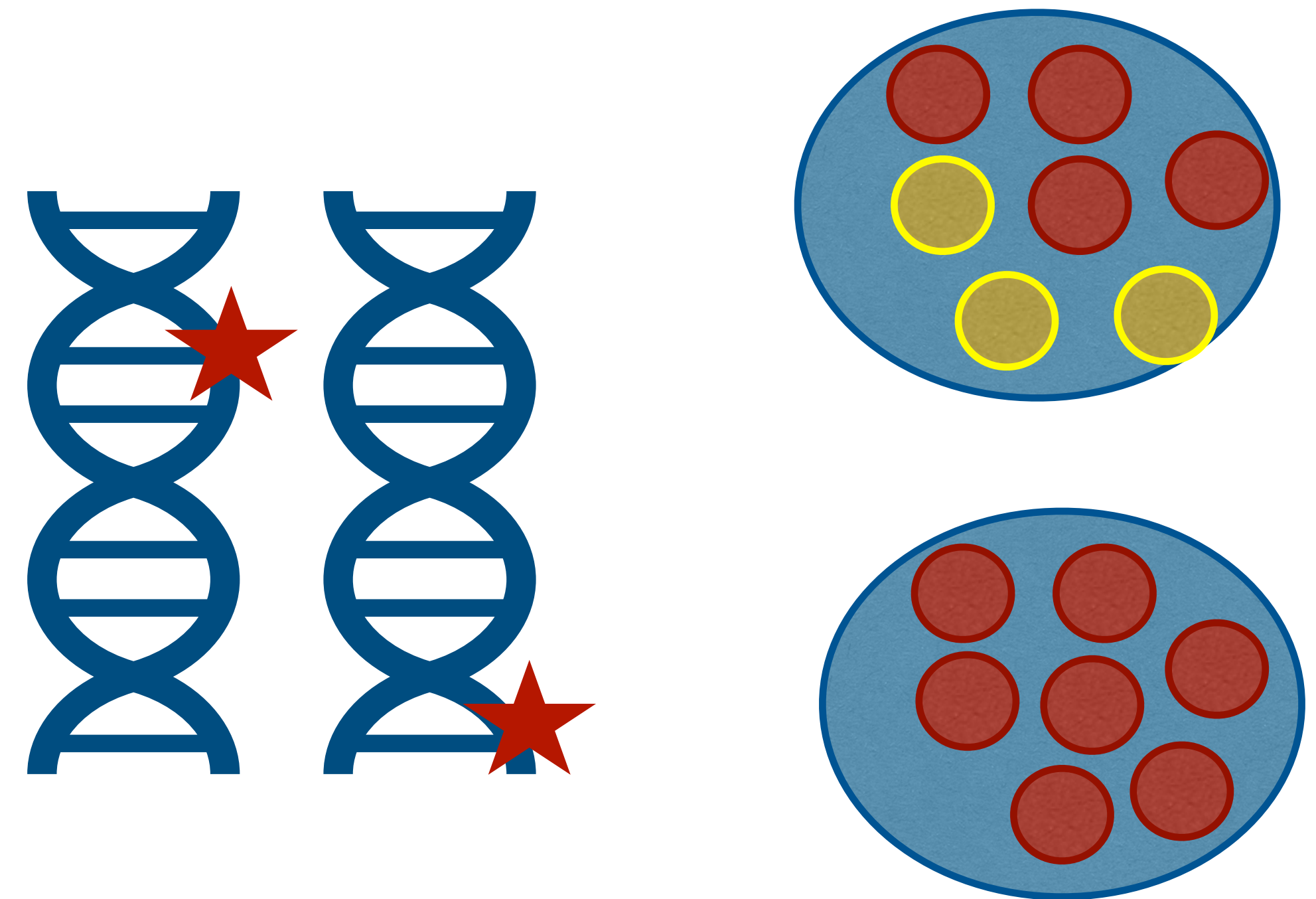
***Hands-on:** Becoming familiar with the AdmixTools, running F-statistics*

***Reading suggestions:** Nielsen and Slatkin, 2013, Chapter 2*

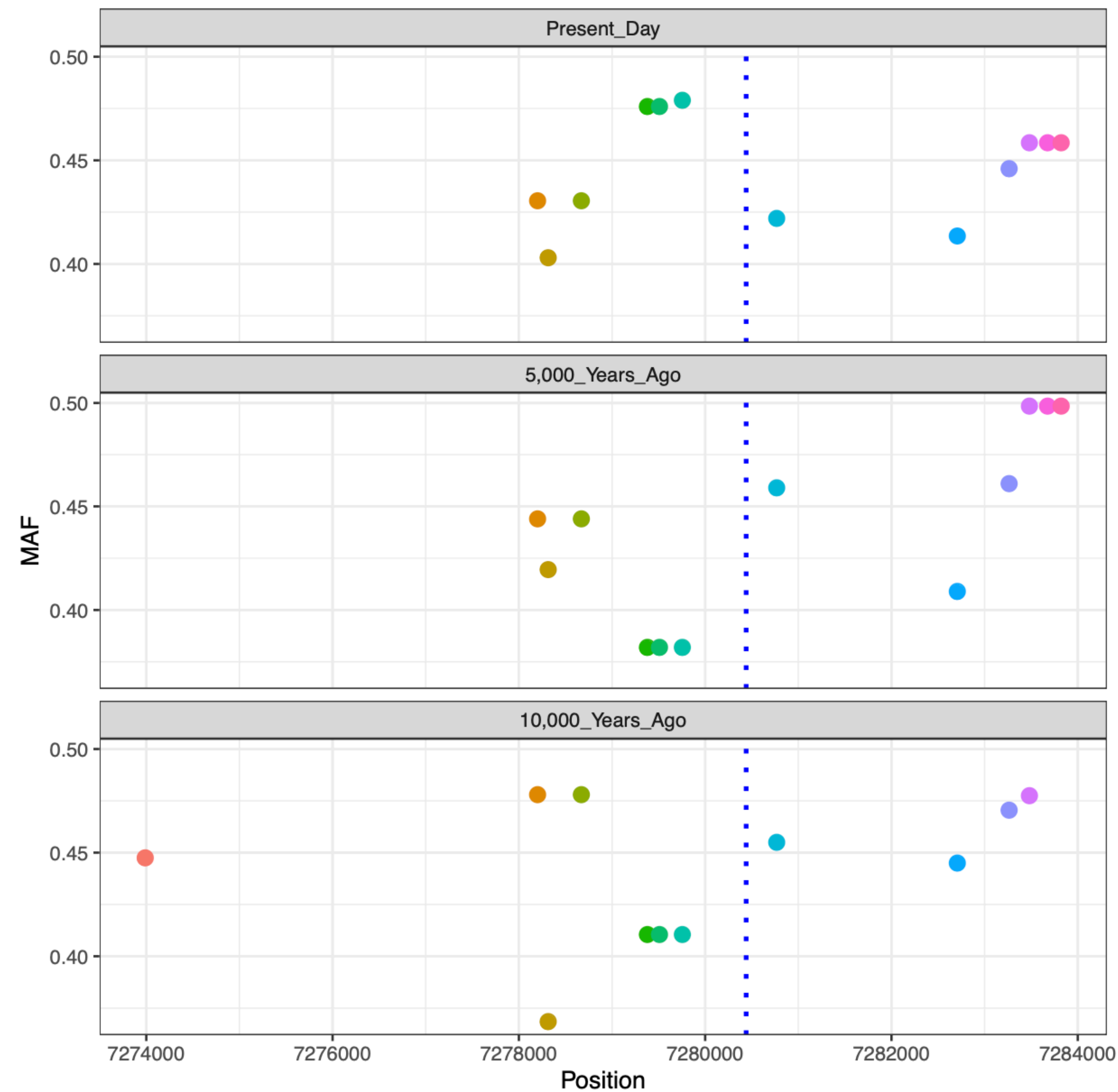
[Peter B, 2016, doi:10.1534/genetics.115.183913](https://doi.org/10.1534/genetics.115.183913)

[Patterson et al, 2012 doi: 10.1534/genetics.112.145037](https://doi.org/10.1534/genetics.112.145037)

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



Why allele frequencies change? Through which evolutionary mechanisms?



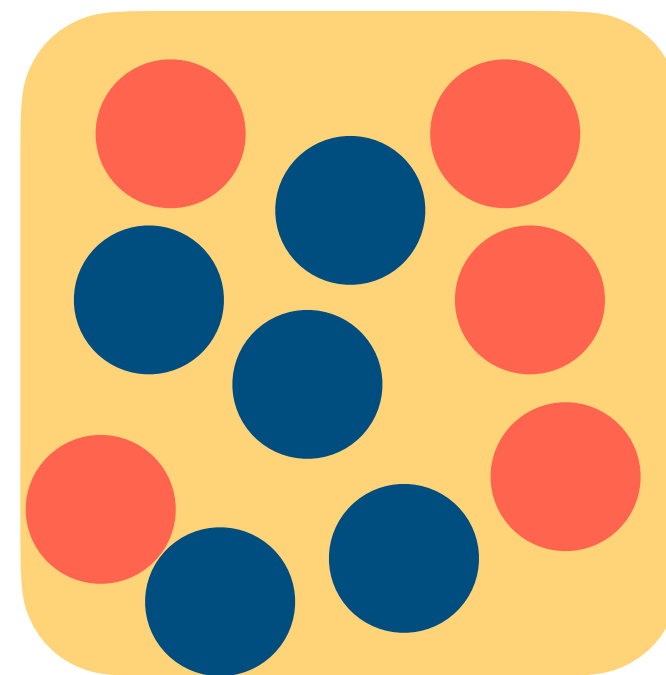
Why allele frequencies change? Through which evolutionary mechanisms?

Genetic drift

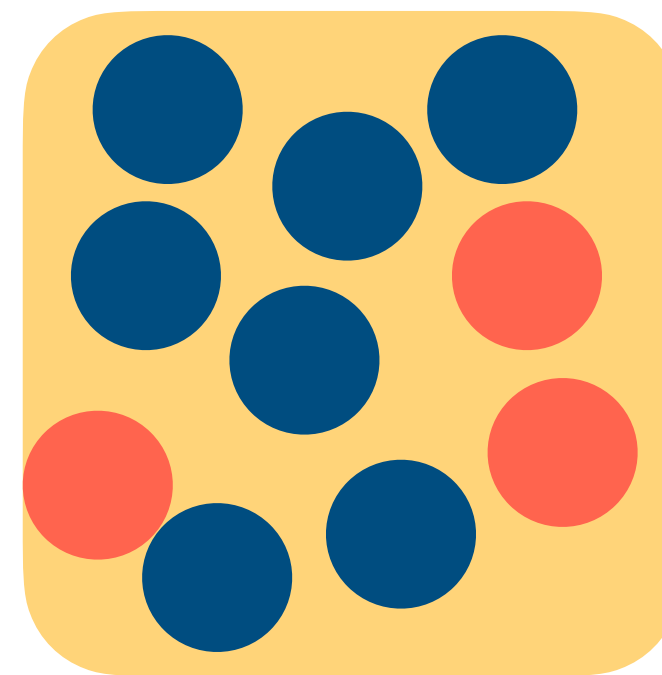
Selection

*Random change of allele frequencies
Finite population*

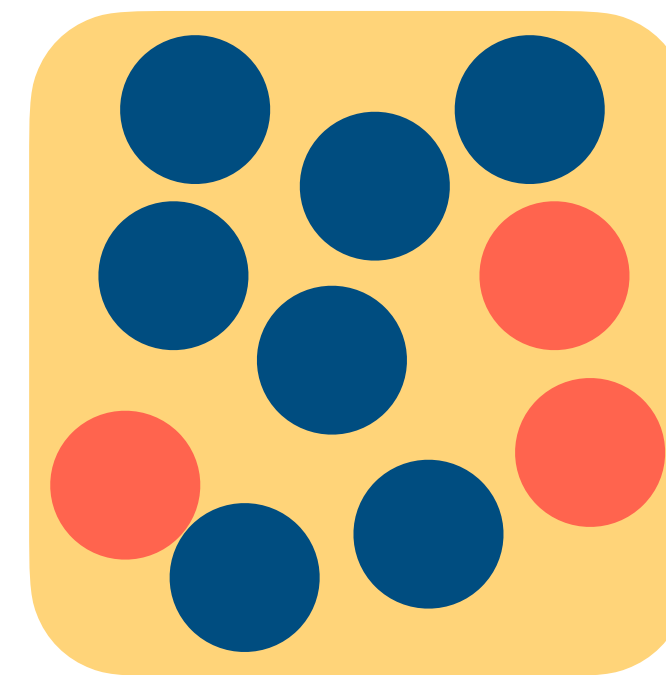
Due to other factors except selection: random, migration, death, etc...



t_1

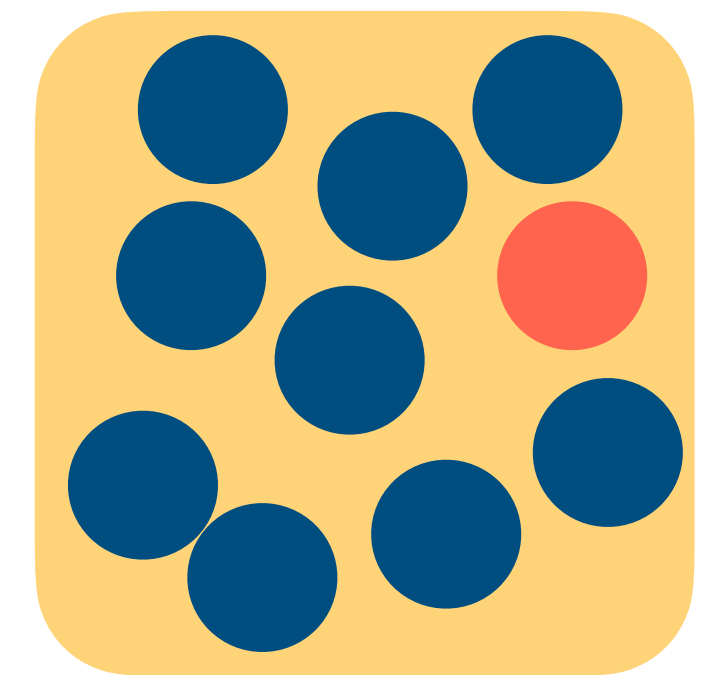


t_2



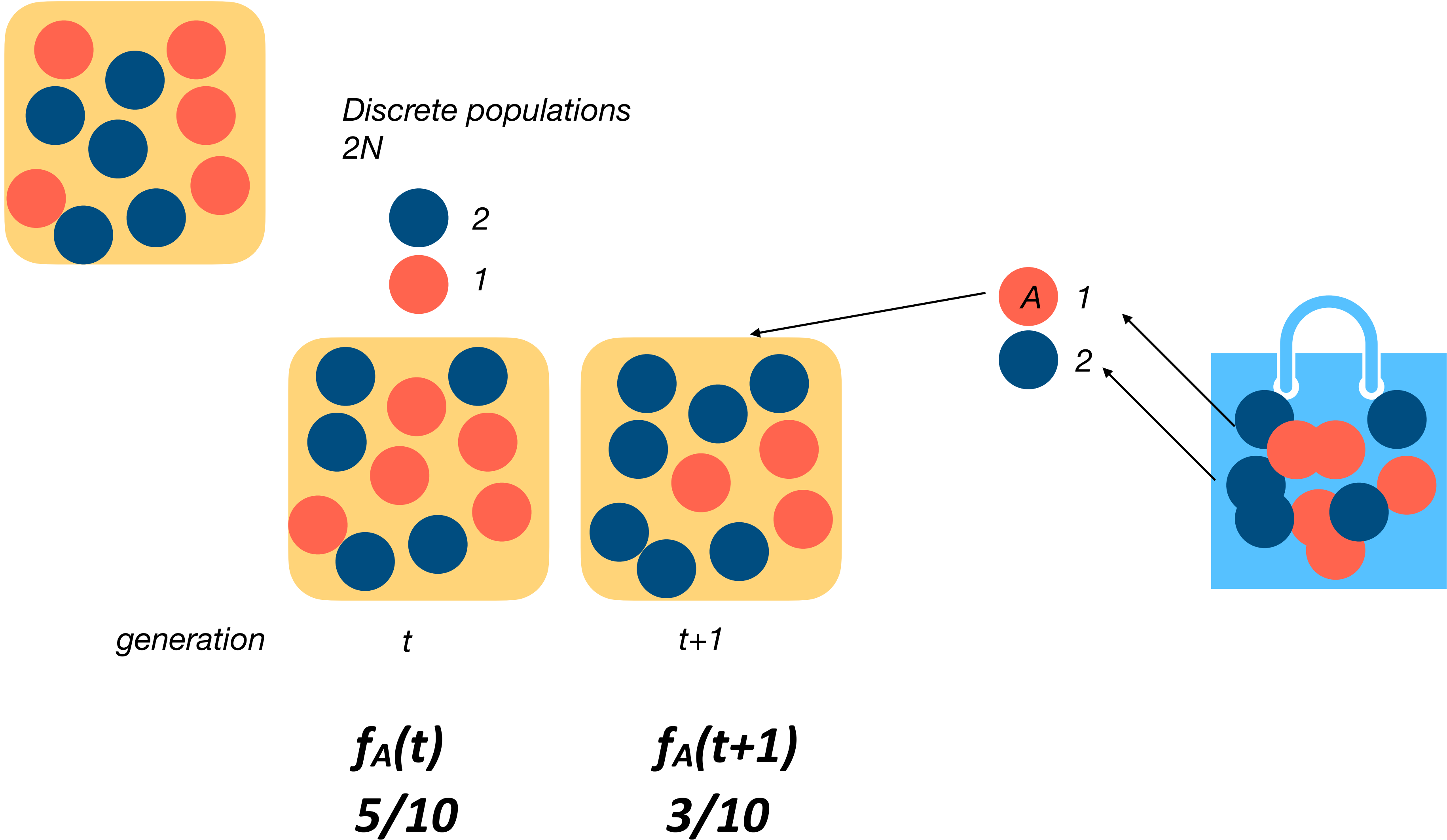
t_3

.....



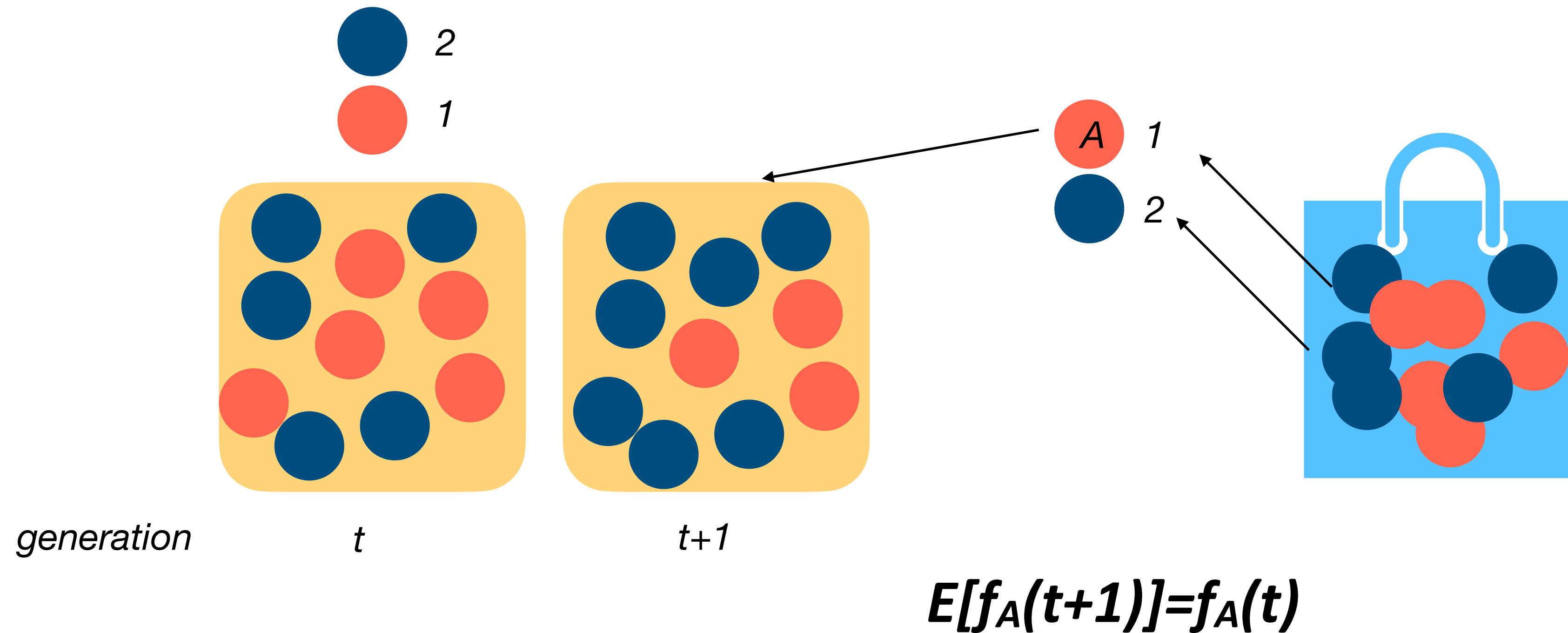
t_n

Wright-Fisher Model: A model to describe genetic drift



Using WF model - Characterize allele frequency change: *What is the probability of any allele at generation $t+1$*

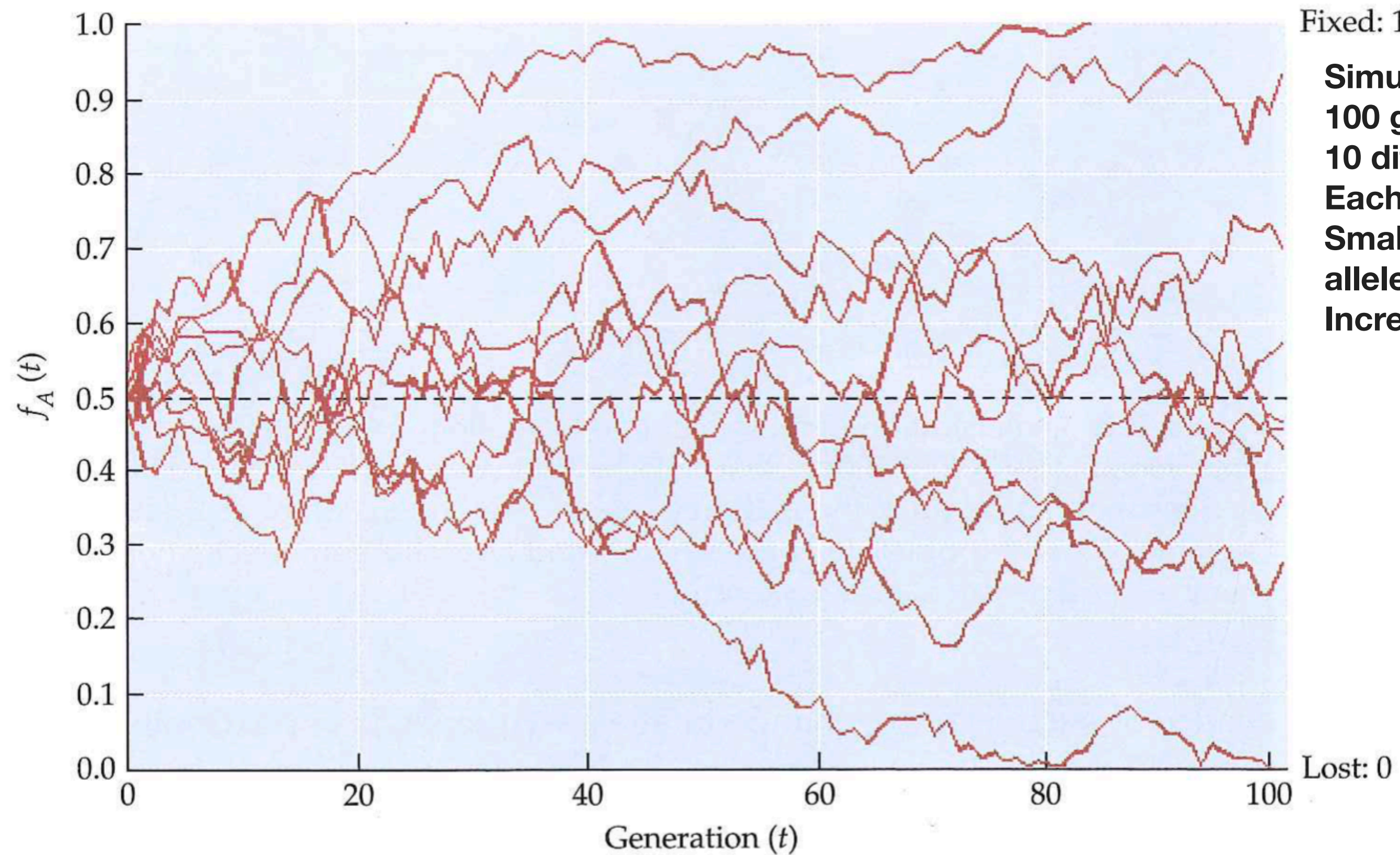
Random sampling \rightarrow chance of an allele at generation t to be A is the frequency of A in generation $t+1$



$f_A(t) = 1 \rightarrow$ fixed

$f_A(t) = 0 \rightarrow$ lost

If there is not any recurrent mutation \rightarrow allele gets fixed or lost



Fixed: 1

Simulation:

100 generations

10 different populations

Each evolves according to WF model

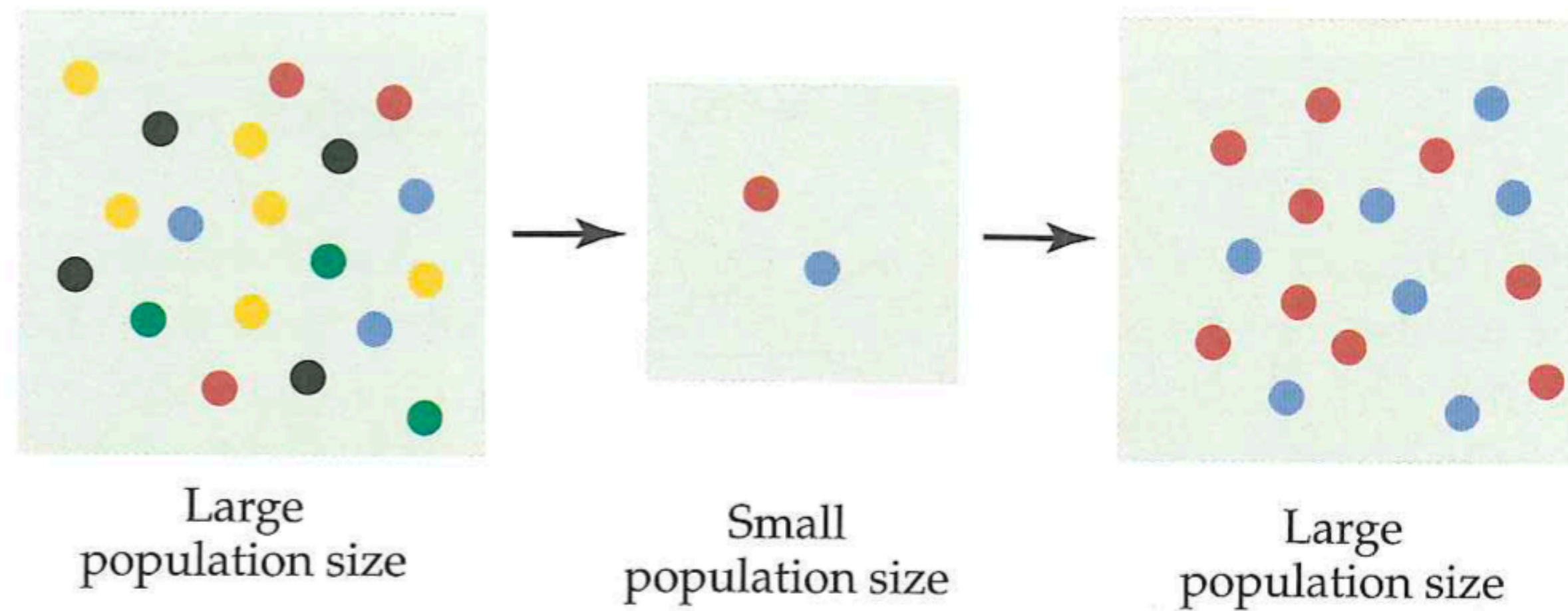
Small changes accumulate - after many generations, allele frequencies change importantly

Increase / Decrease - Random change, Not favoring A or a

Lost: 0

Figure 2.2 The Wright–Fisher model simulated for 10 populations, with $2N = 100$, over 100 generations (solid lines) for an initial allele frequency of 50%. Allele frequencies change randomly due to genetic drift. The expected (mean) allele frequency is shown by the dashed line.

Effect of population size



Nielsen and Slatkin 2013 *An Introduction to Population Genetics* Chapter 2

If pop size is small, alleles can easily get fixed or lost

If there is not any recurrent mutation -> allele gets fixed or lost

If there is not selection/mutation -> $Pr(\text{Allele A fixed}) = f_A(t)$

Mutation: Produces genetic variation and genetic drift acts on the variation

WF model - no mutation -> alleles get lost or fixed - so population does not contain genetic variation

Real populations -> mutation

point mutations


insertions & deletions

inversions

translocations

$2N$

How does mutation change the allele frequency?

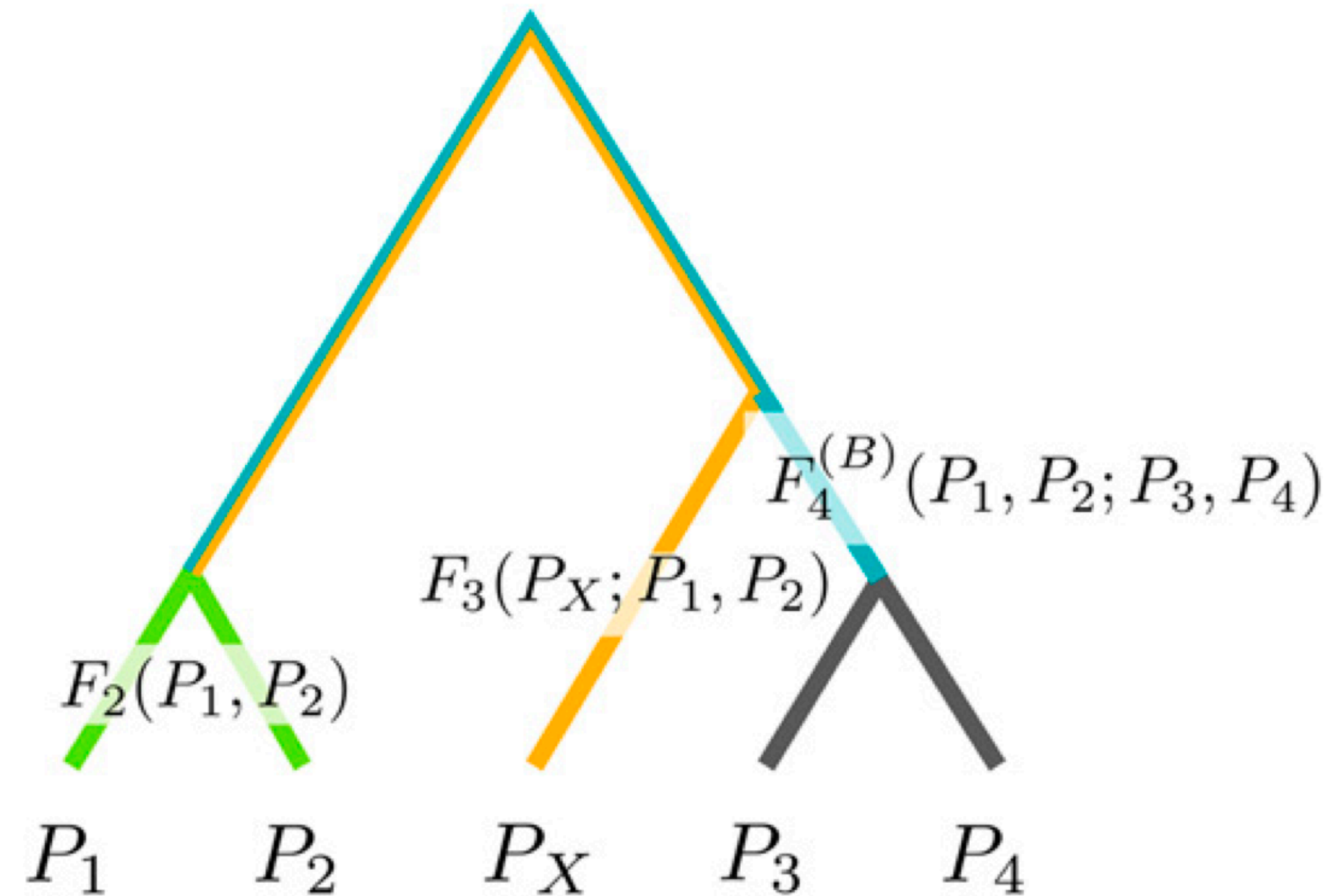
WF Model  **$E[f_A(t+1)] = f_A(t)$**

WF Model + a mutates to A with probability of u (mutation rate)


$$E[f_A(t+1)] = f_A(t) + ufa(t)$$

Allele frequency correlations between populations: *F*-statistics (Reich et al 2009)

Shared genetic drift between populations -> Shared evolutionary history of populations



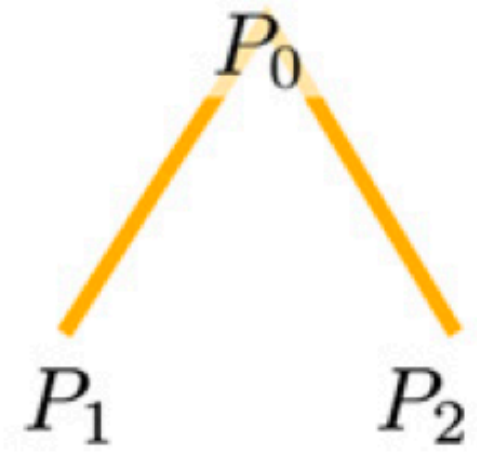
Population tree (population phylogeny)-> Shared evolutionary history of populations

Branch lengths -> amount of genetic drift

Branches bound by two pops -> shared drift between these pops

F_2 -statistics

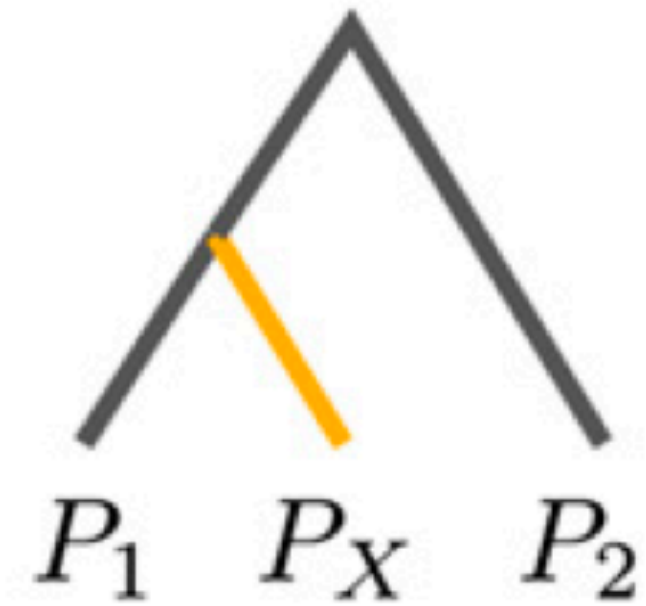
How much drift occurred between two populations? Genetic dissimilarity



$$F_2(P_1, P_2) = F_2(p_1, p_2) = \mathbb{E}(p_1 - p_2)^2$$

F_3 -statistics -> Three populations statistics

Aim: To test if P_X is admixed - admixture F_3 -statistics



Allele frequency differences between P_X to P_1 and P_2

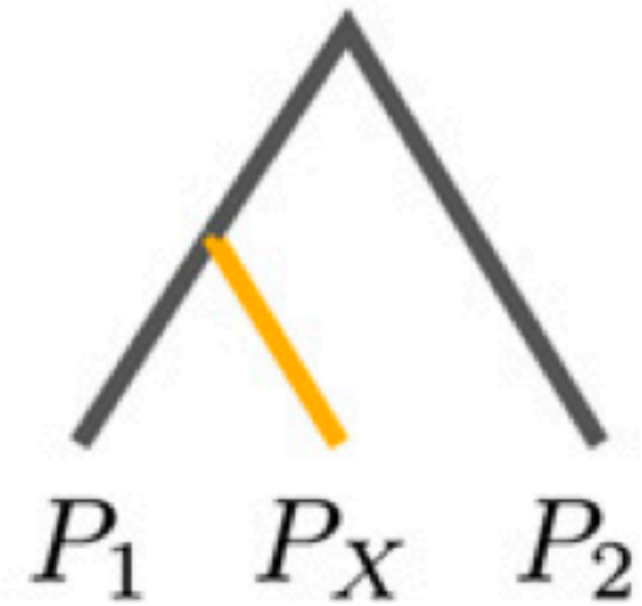
$$F_3(P_X; P_1, P_2) = F_3(p_X; p_1, p_2) = \mathbb{E}(p_X - p_1)(p_X - p_2)$$

P_X is related to P_1 and P_2 . Variance in allele frequency since P_X is split from P_1 and P_2

Could be negative if X descends from a mixture of P_1 and P_2

F_3 -statistics -> Three populations statistics

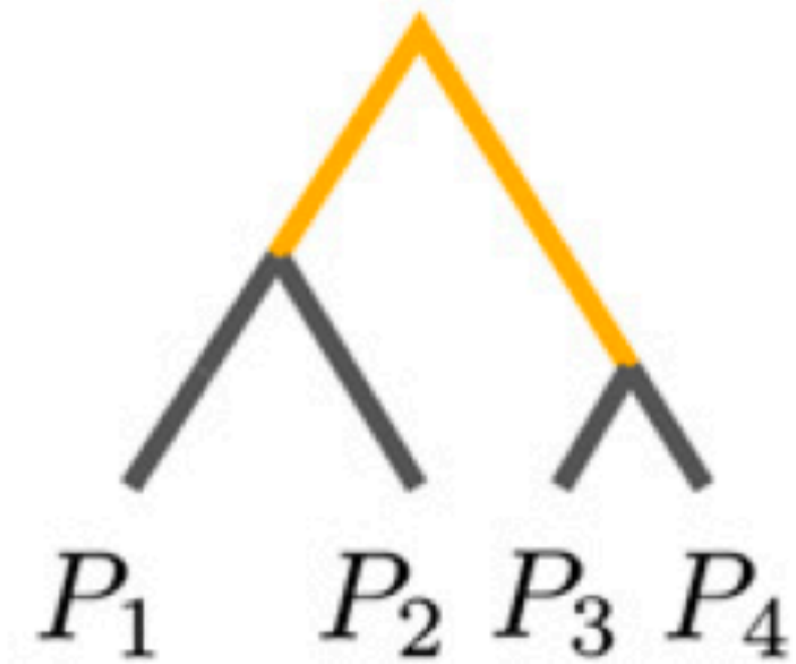
If P_X is not admixed -< outgroup F_3 -statistics



$$F_3(P_X; P_1, P_2) = F_3(p_X; p_1, p_2) = \mathbb{E}(p_X - p_1)(p_X - p_2)$$

When P_X is unadmixed -> more positive the result -> more shared drift between P_1 and P_2

F_4 -statistics -> Four populations test

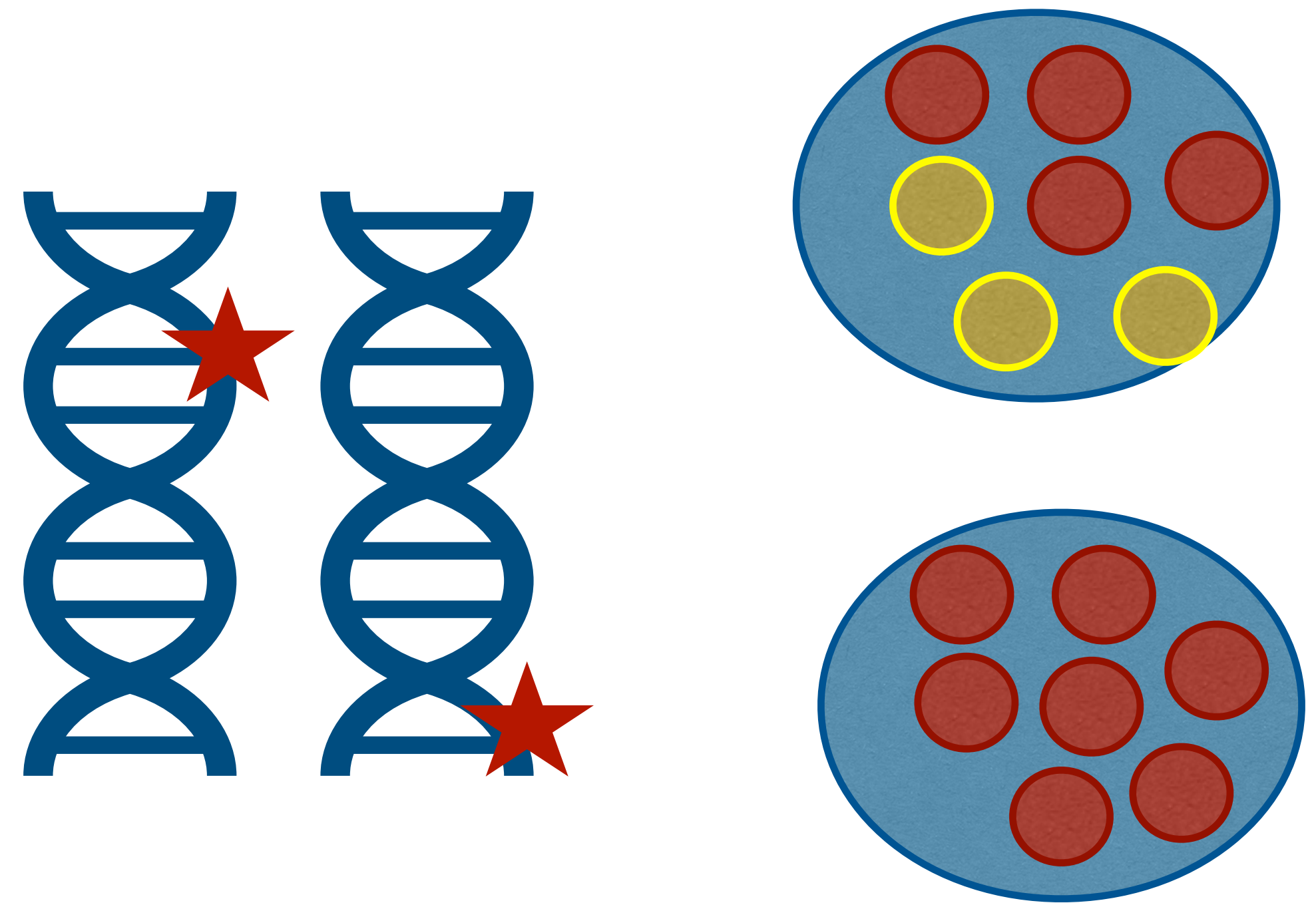


$$F_4(P_1, P_2; P_3, P_4) = F_4(p_1, p_2; p_3, p_4) = \mathbb{E}[(p_1 - p_2)(p_3 - p_4)]$$

Week-3

Hands-on

*Aim: Becoming familiar with the
AdmixTools, running F-statistics*



<https://github.com/DReichLab/AdmixTools>

or

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

or

<https://github.com/pontusssk/popstats>

Our server: /usr/local/sw/AdmixTools

PLAN

- 1- Run outgroup *f3*-statistics
- 2- Run admixture *f3*-statistics
- 3- Run *f4*-statistics
- 4- Plot the results in R
- 5- Interpret the results