



HAL
open science

Distinguishing shared ancestral polymorphism from recent introgression in genes with recombination

Miguel Navascués, Frantz Depaulis

► **To cite this version:**

Miguel Navascués, Frantz Depaulis. Distinguishing shared ancestral polymorphism from recent introgression in genes with recombination. Réunion 2008 du GDR 1928 Génomique des populations et génomique évolutive, GDR 1928 Génomique des populations et génomique évolutive., Oct 2008, Sète, France. hal-02814597

HAL Id: hal-02814597

<https://hal.inrae.fr/hal-02814597>

Submitted on 10 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

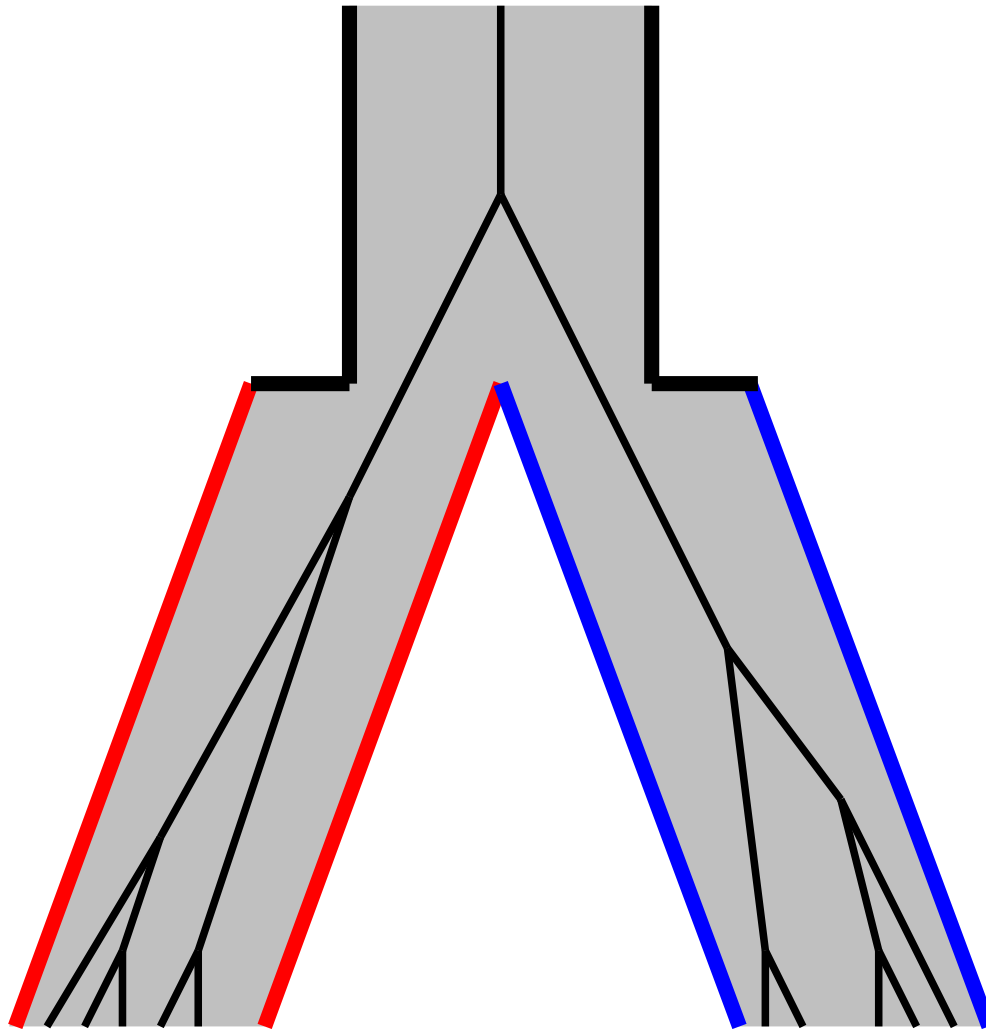
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distinguishing shared **ancestral polymorphism** from recent **introgression**

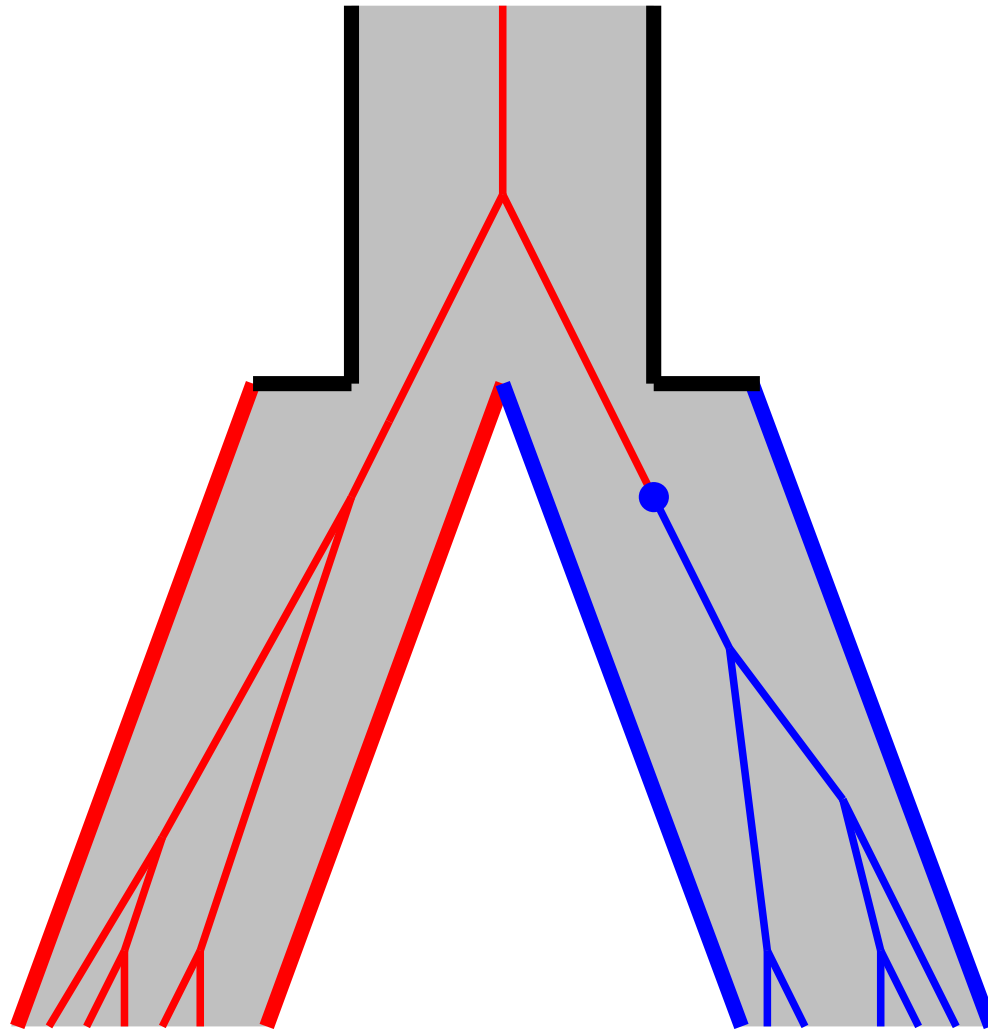
Miguel Navascués & Frantz Depaulis
CNRS UMR 7625 Écologie et Évolution (ENS/UPMC)



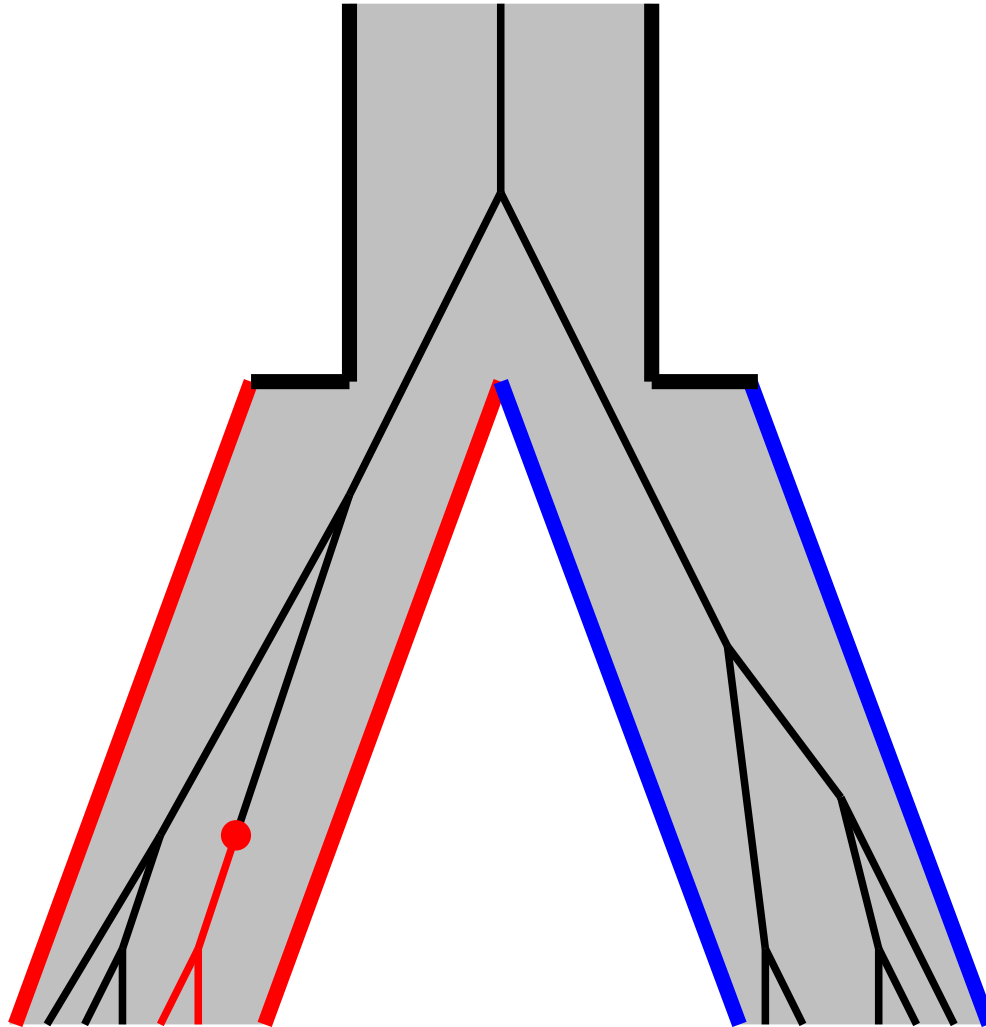
Polymorphisms between two species



Fixed polymorphism

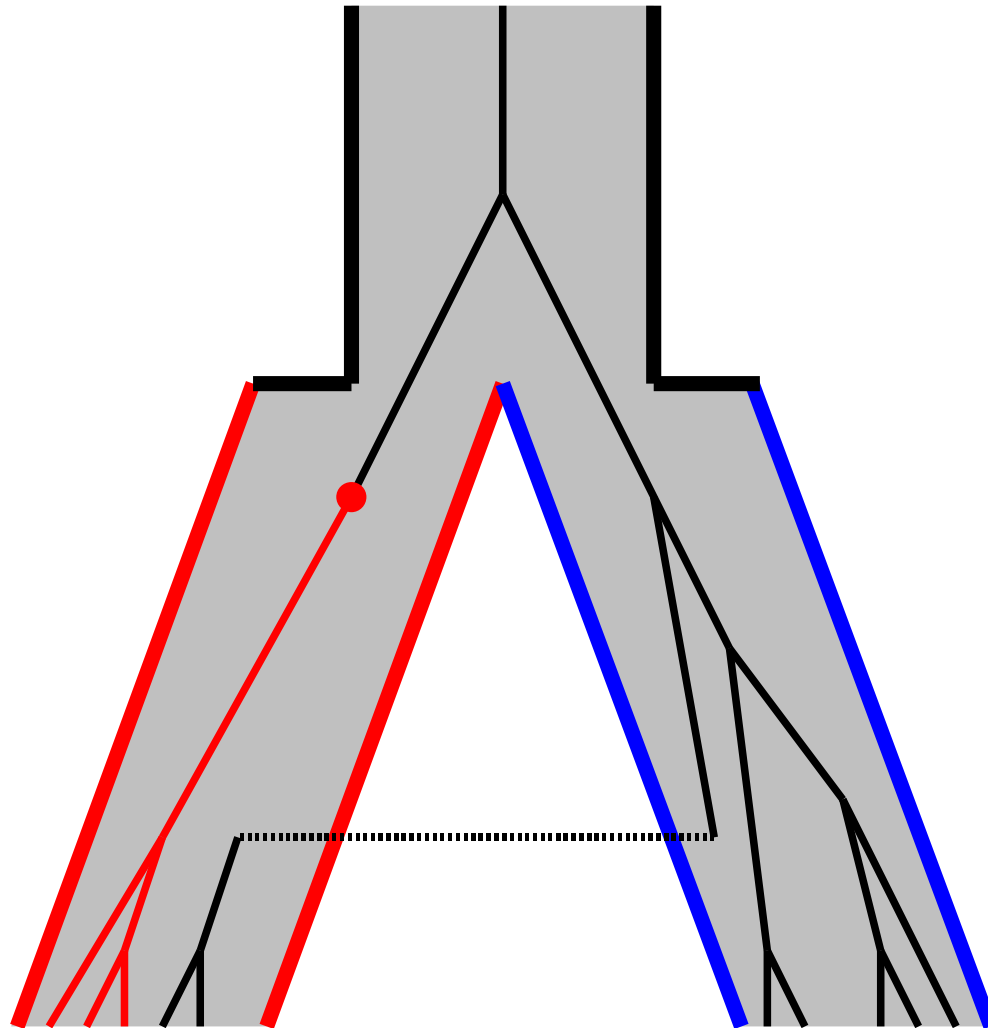


Exclusive polymorphism



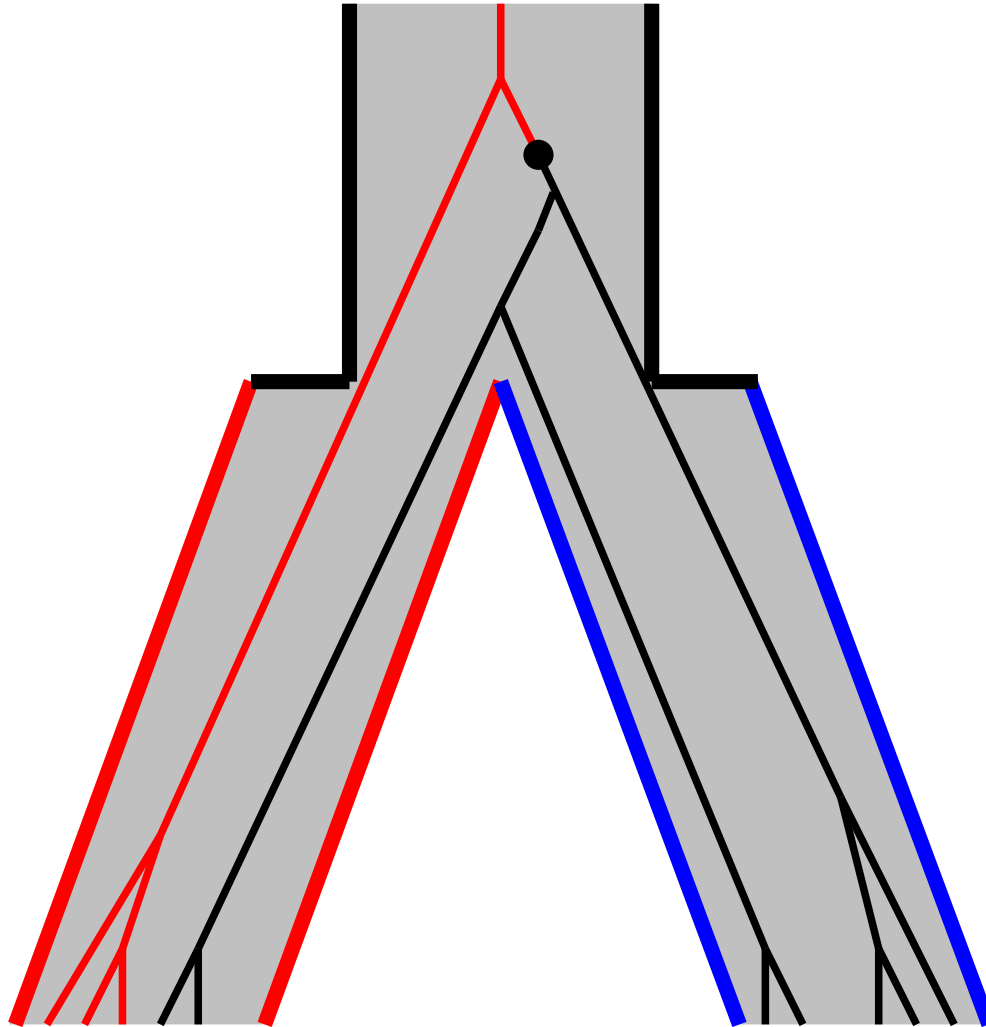
Exclusive polymorphism

Introgression



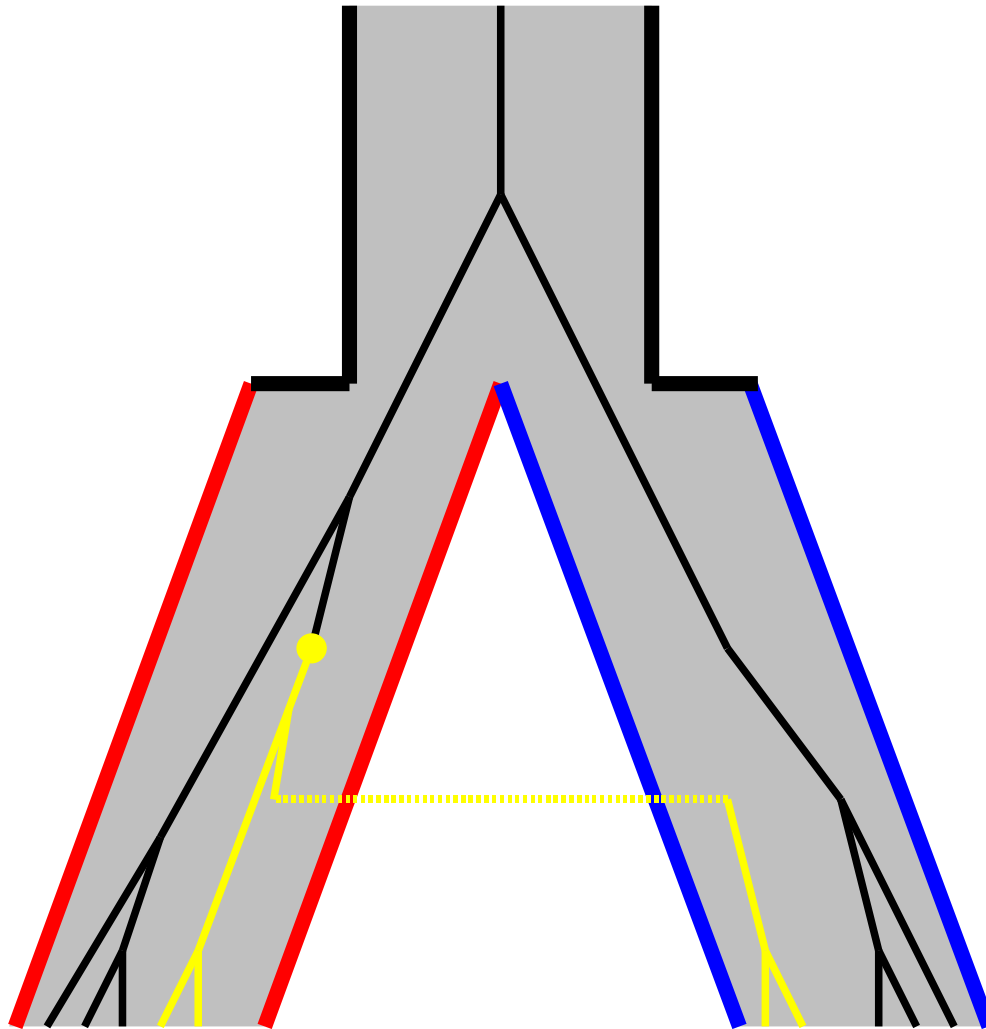
Exclusive polymorphism

Ancestral



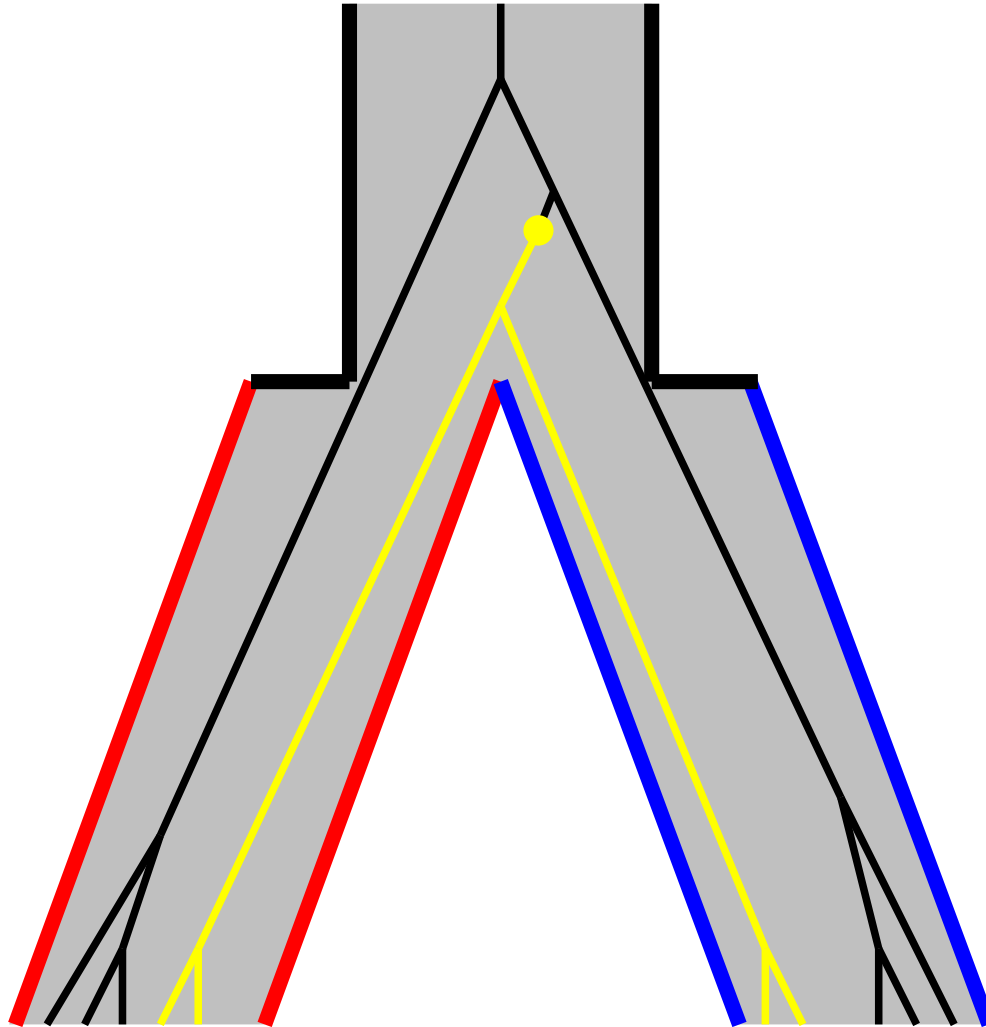
Shared polymorphism

Introgression



Shared polymorphism

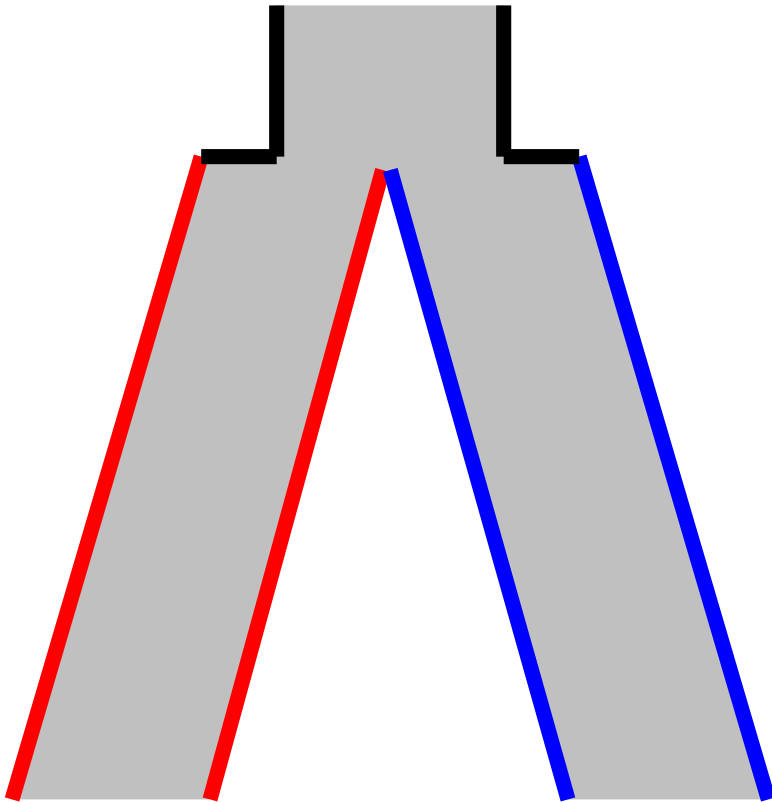
Ancestral



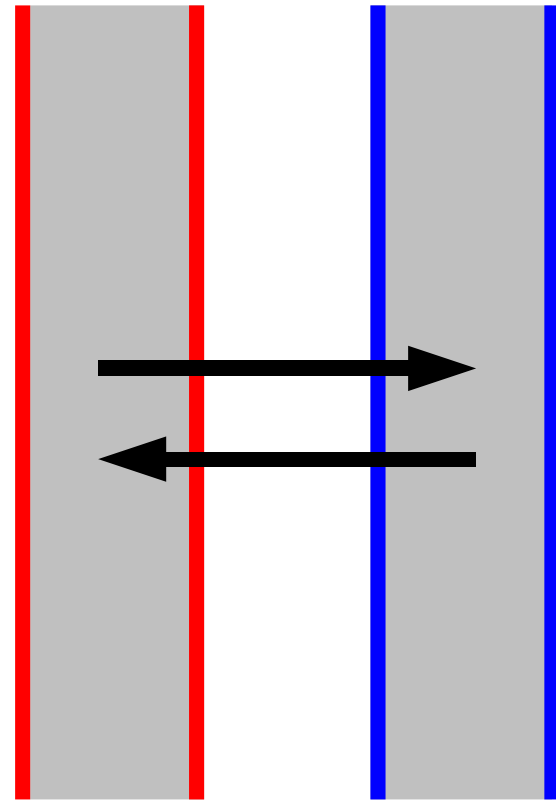
Interpretation of F_{ST}

Isolation vs. Migration

$$F_{ST} \approx 1 - e^{-t/2N}$$

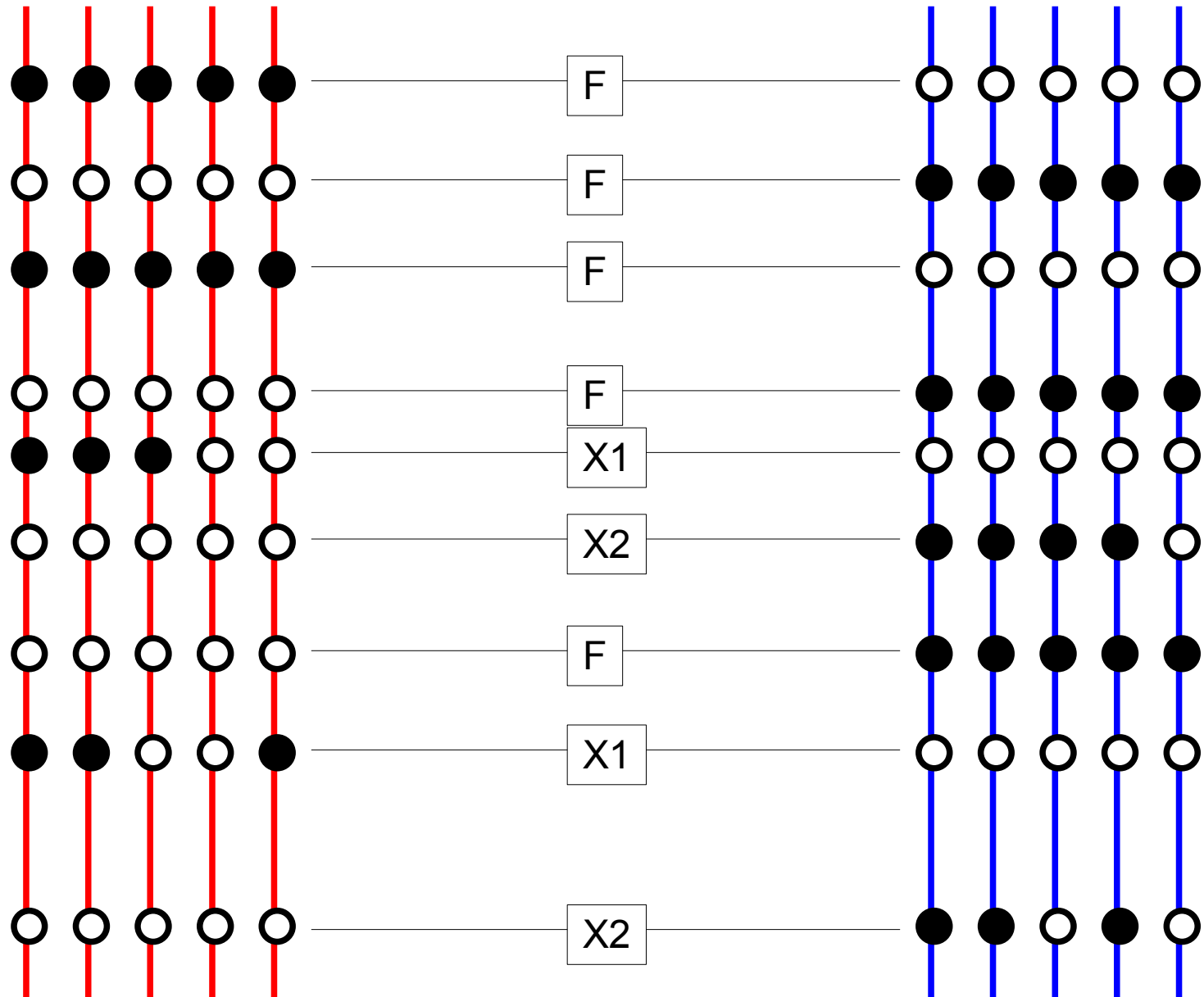


$$F_{ST} \approx \frac{1}{1 + 4Nm}$$



POPULATION 1

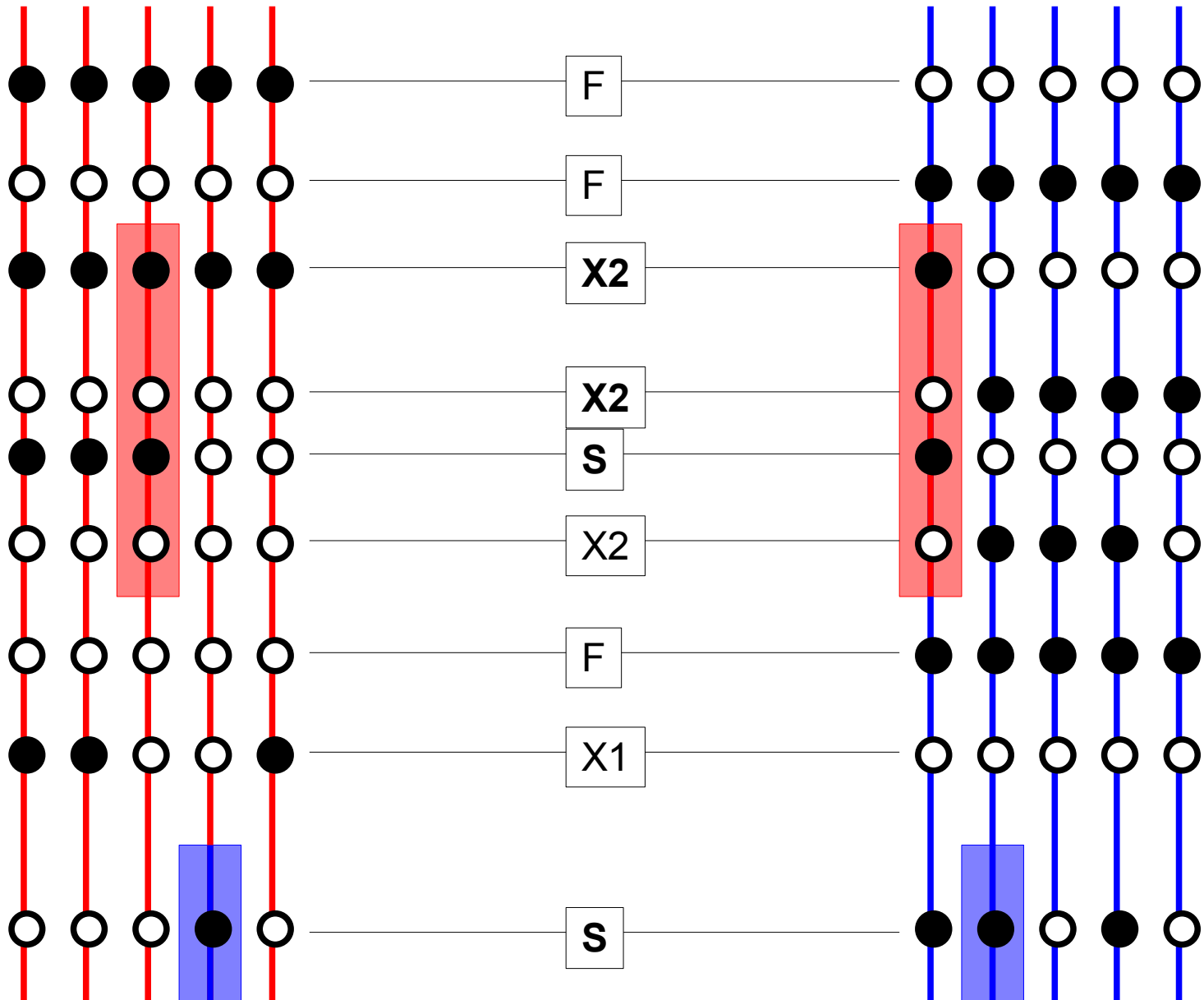
POPULATION 2



● derived state ○ ancestral state

POPULATION 1

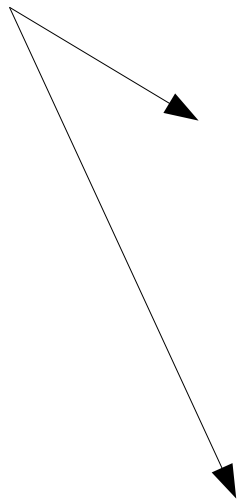
POPULATION 2



● derived state ○ ancestral state

Statistics: number of runs

$n^+ = 8, n^- = 7$



****--*-****-*---**
1 2 345 678

NR=8

random

*******-----**
1 2

NR=2

clustered

Statistics: length of runs

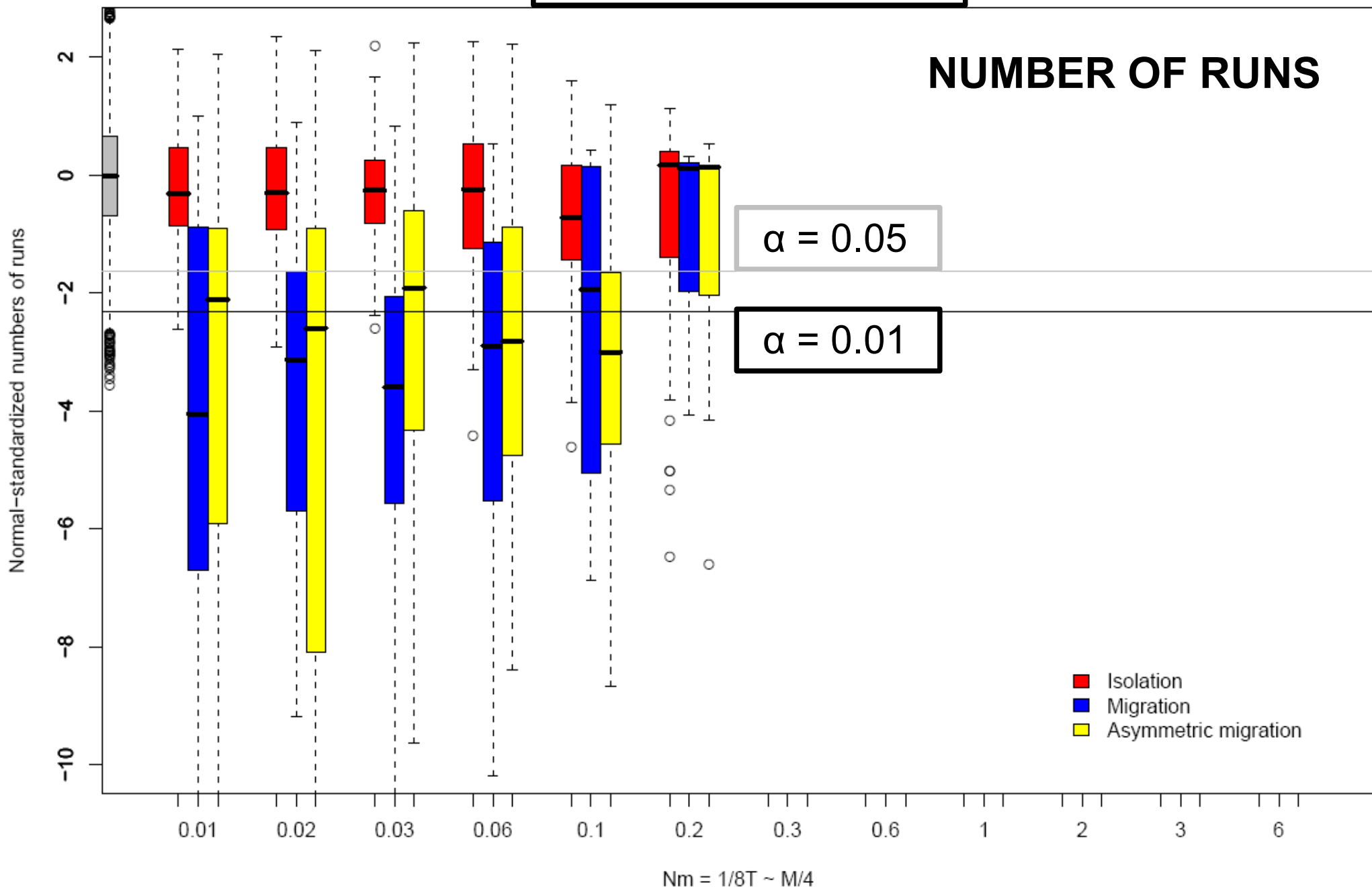
$$n^+ = 7, n^- = 8$$

$$\begin{array}{l} \text{"-----"} \\ - * - ** - - - * - - ** - * \end{array} + 7 \text{ random cuts } (*) = \text{Lengths } (-): 1, 1, 0, 3, 2, 0, 1, 0$$

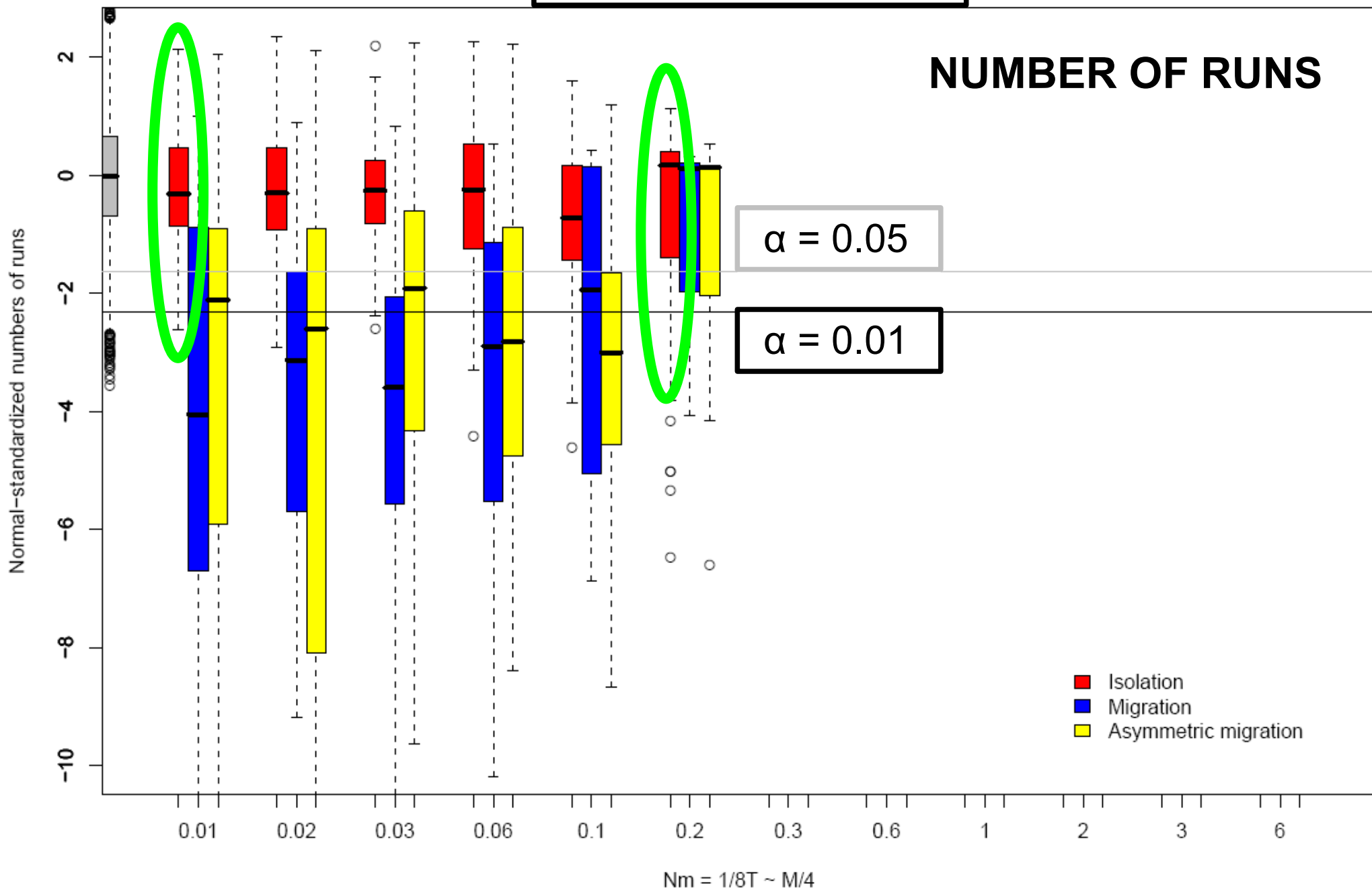
$$\begin{array}{l} \text{"-----"} \\ - - - - - * * * * * - - - - \end{array} + 7 \text{ clustered cuts } (*) = \text{Lengths } (-): 5, 0, 0, 0, 0, 0, 0, 3$$

F(*) vs. S+X(-)

NUMBER OF RUNS

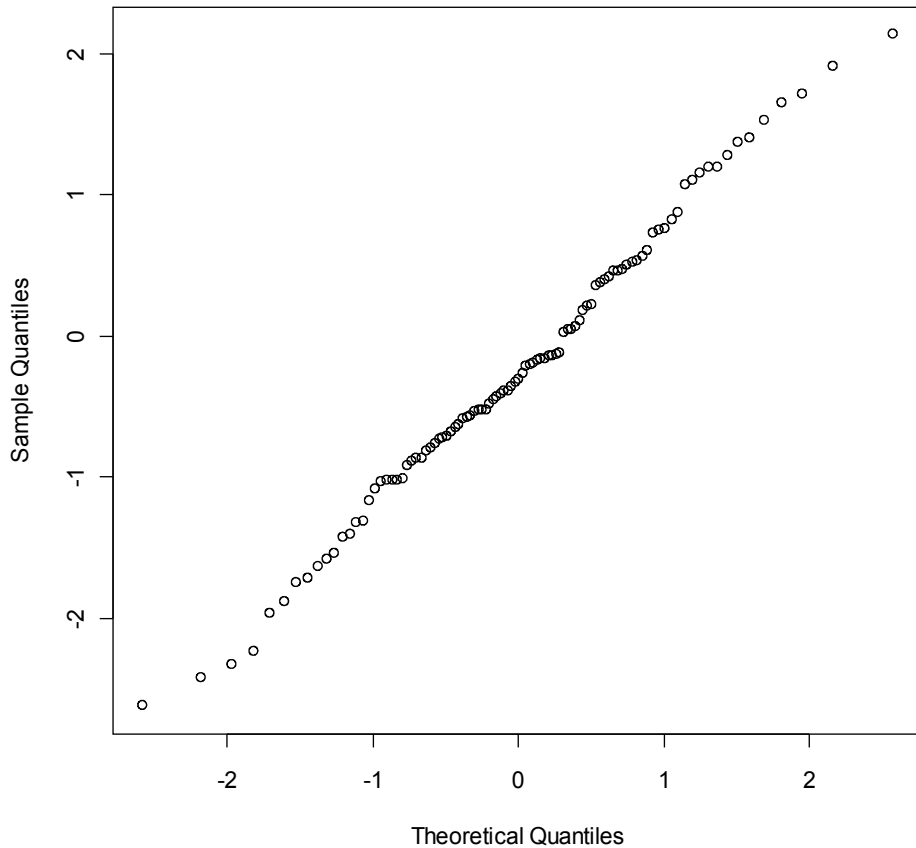


F(*) vs. S+X(-)



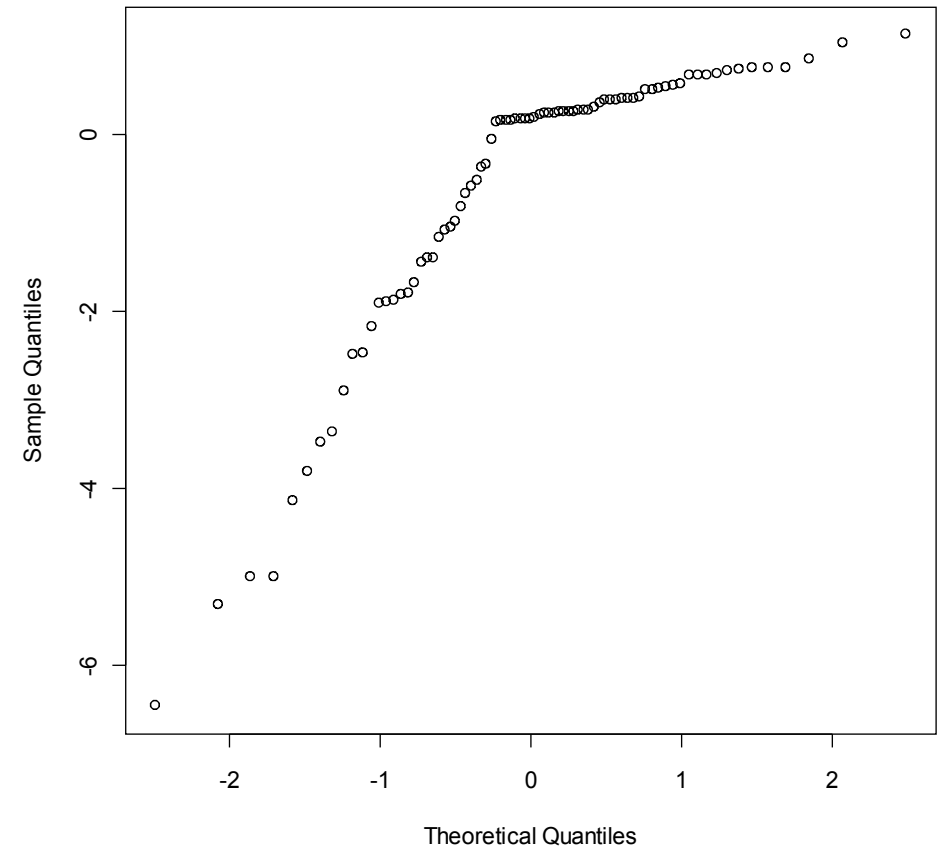
Isolation as null hypothesis

Normal Q-Q Plot



Isolation: $T=12.5$

Normal Q-Q Plot

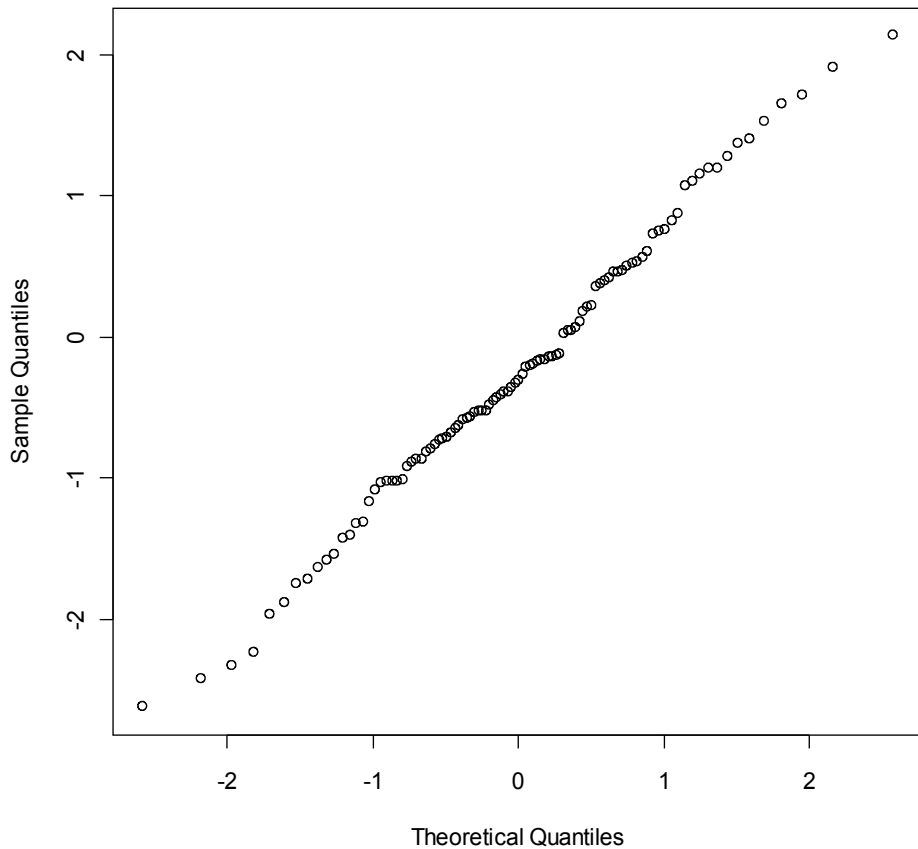


Isolation: $T=0.625$

(T: time scaled to population size)

Test with several loci

Normal Q-Q Plot



Kolmogorov–Smirnov test (on 10 loci)

Isolation: 1/10

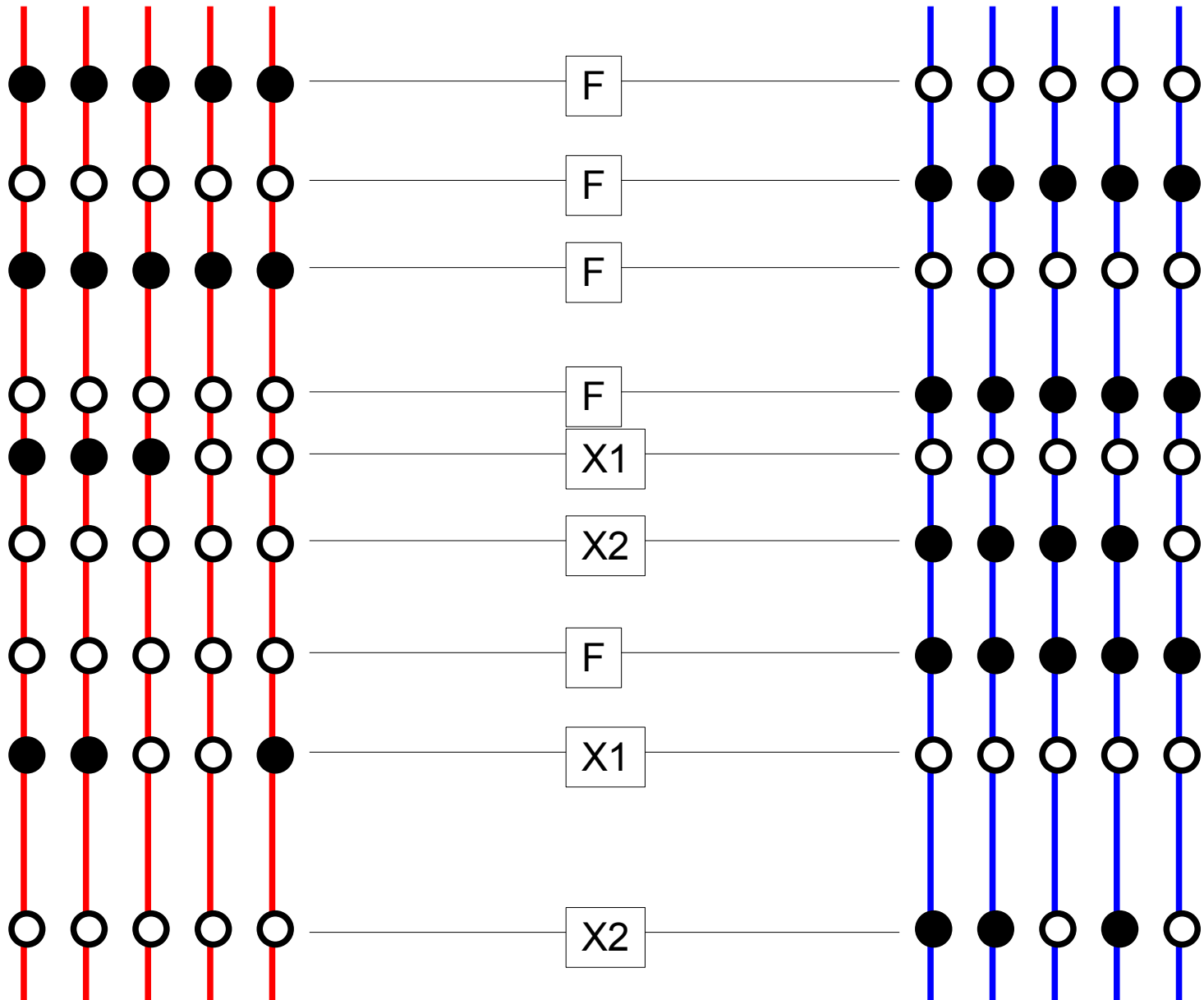
Migration: 9/10

Asymmetric mig.: 10/10

Asymmetric migration

POPULATION 1

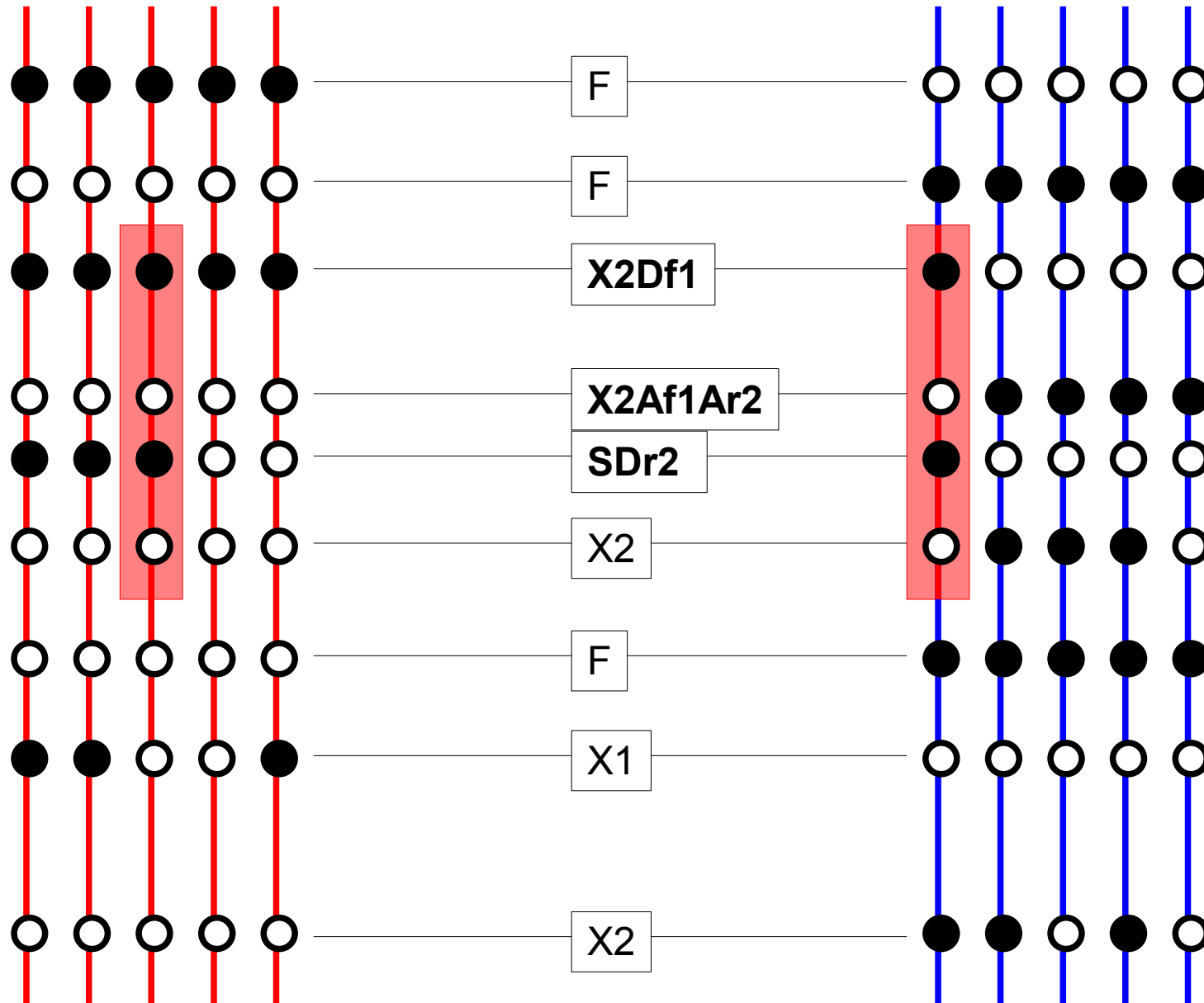
POPULATION 2



● Derived state ○ Ancestral state

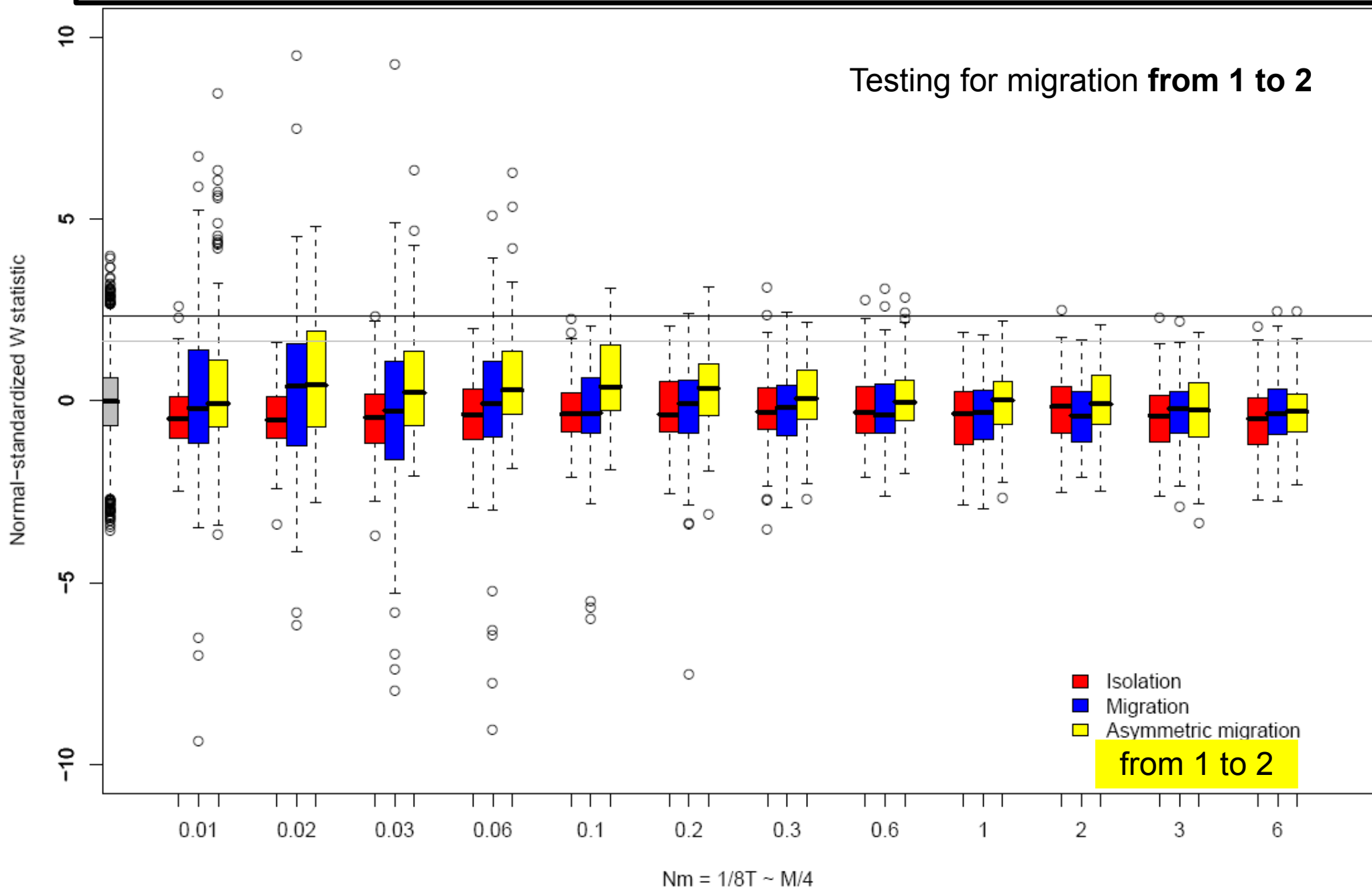
POPULATION 1

POPULATION 2



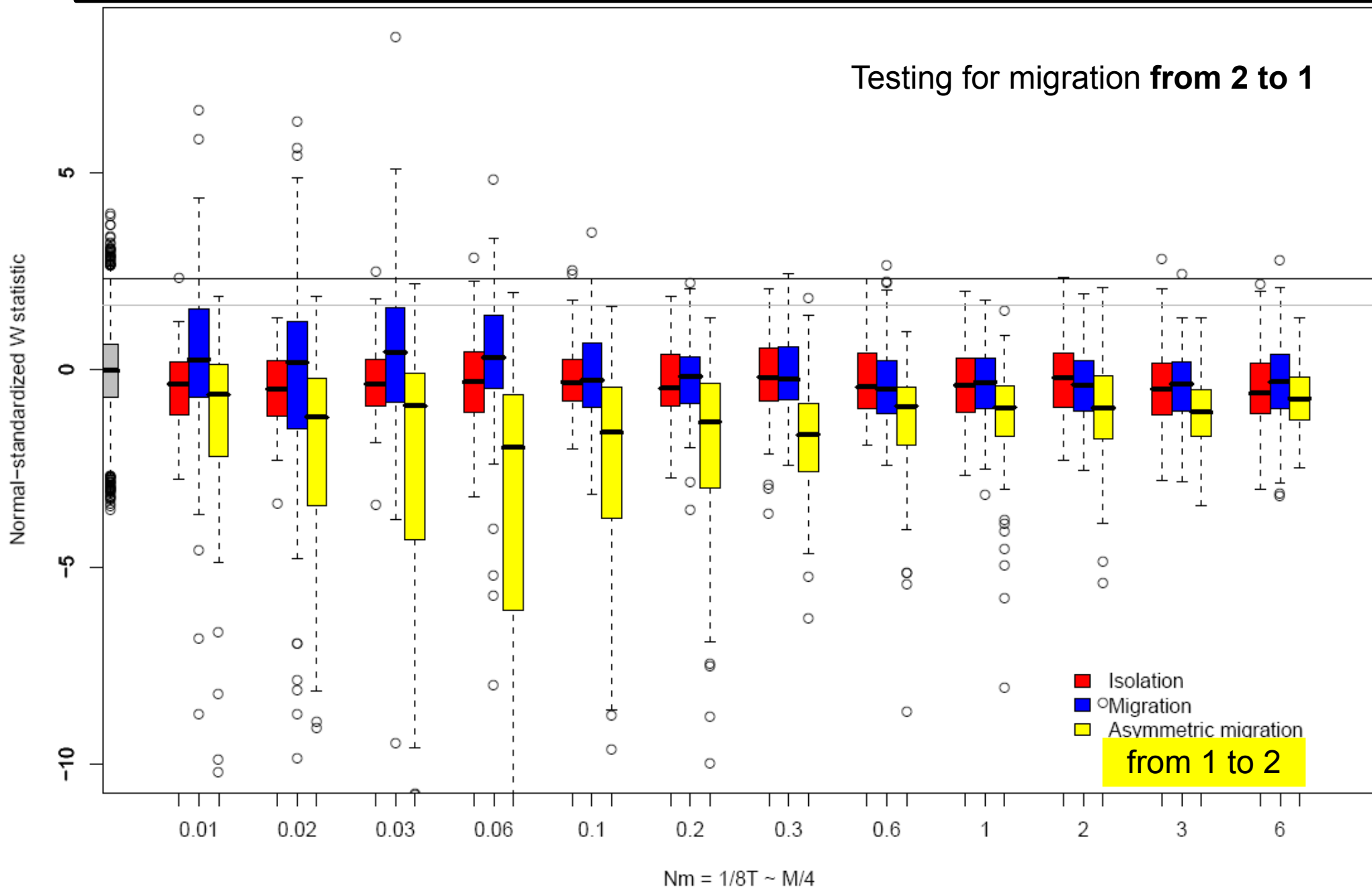
● Derived state ○ Ancestral state

SDr2+X2Df1+X2Af1Ar2 vs. SDr1+X1Df2+X1Af2Ar2



SDr1+X1Df2+X1Af2Ar1 vs. SDr2+X2Df1+X2Af1Ar1

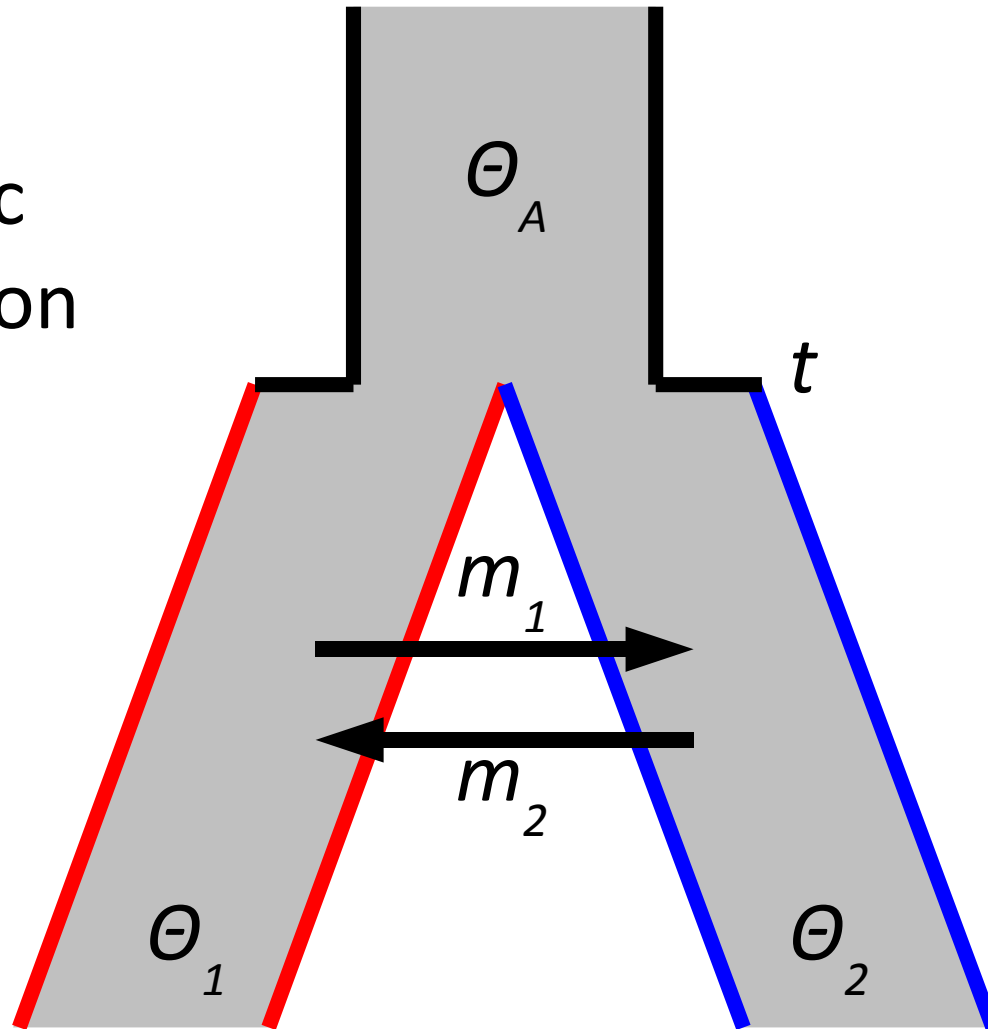
Testing for migration from 2 to 1



Likelihood Coalescent MCMC

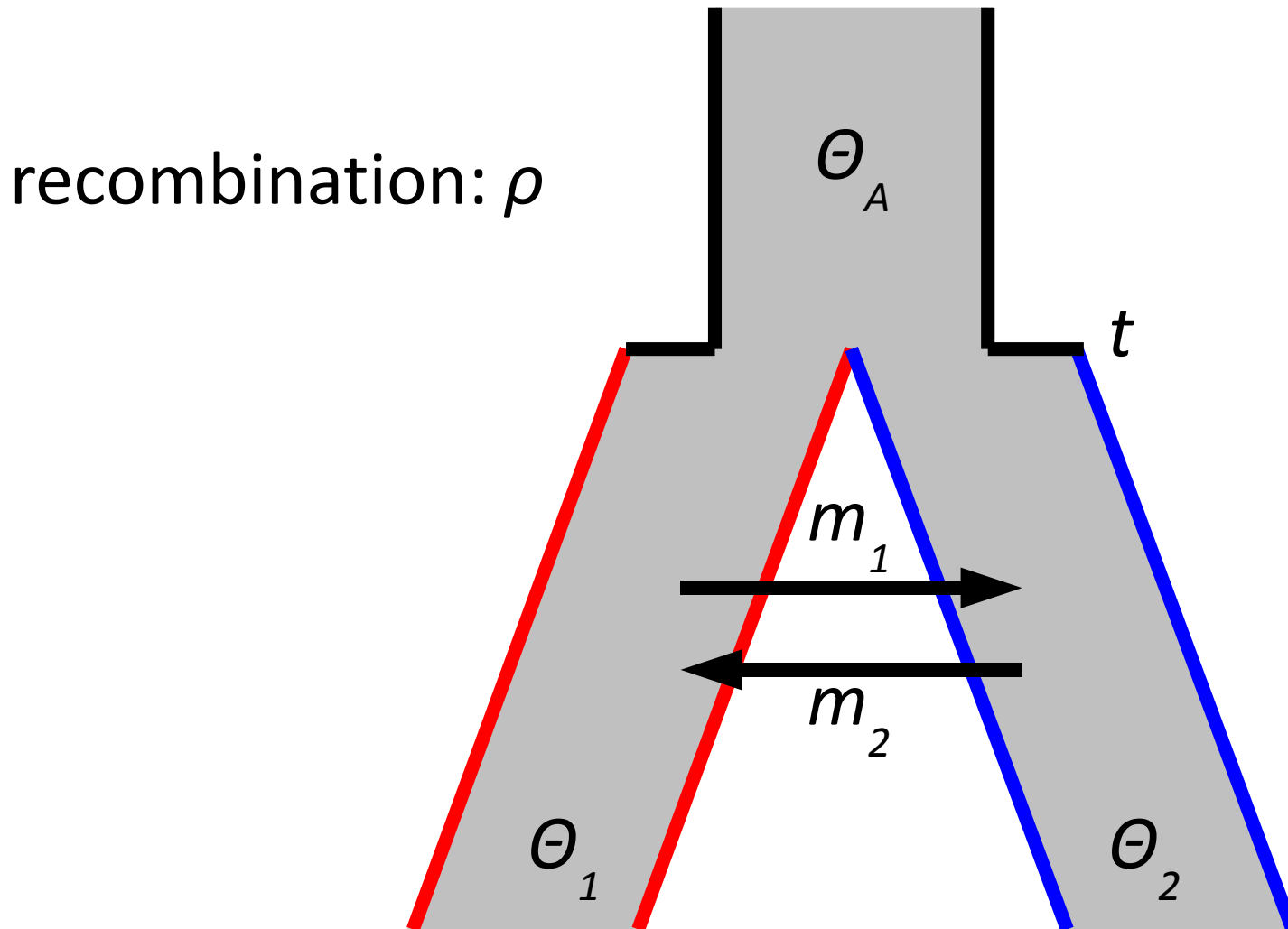
(Nielsen & Wakeley 2001, Hey & Nielsen 2004)

no intragenic recombination



Summary Statistics Coalescent MCMC

(Becquet & Przeworski 2007)

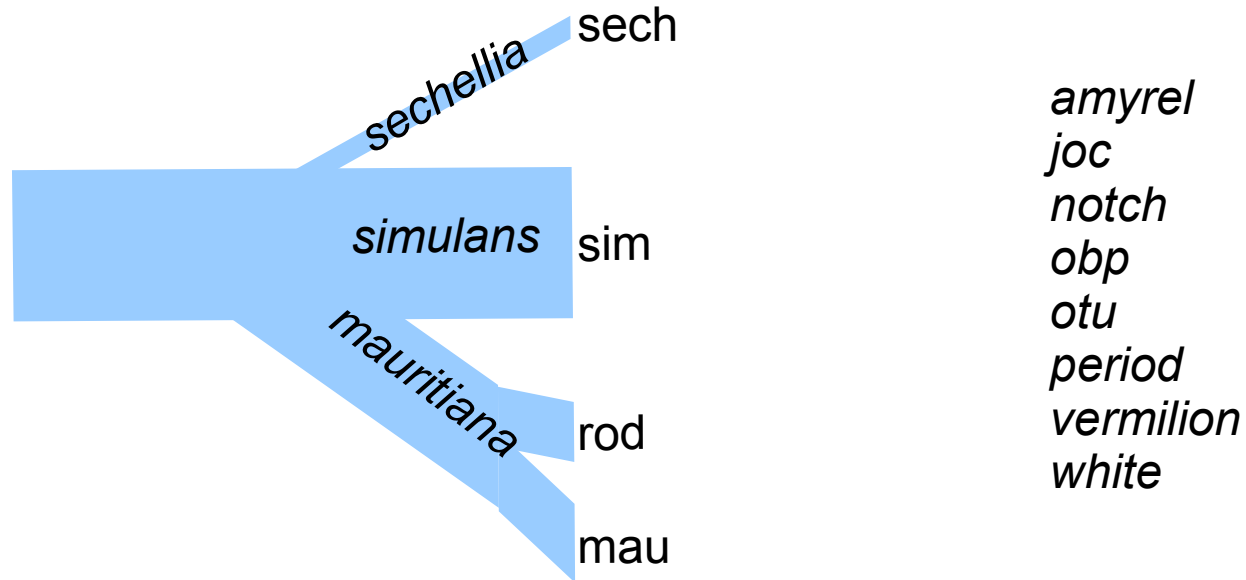


Summary

- Clustering of certain types of polymorphism
- Non-parametric statistics: based on runs
- Statistics sensitive to migration
- Statistics sensitive to direction of migration
- Summary statistics replacing likelihood (e.g. ABC)

Drosophila simulans, sechellia, mauritiana

Marie-Louise Cariou and team



Biodiversité: TRANSBIODIV

Arabidopsis: Xavier Vekemans and team

Millet: Thierry Robert and team

Oaks: Antoine Kremer and team

Mice: Pierre Boursot and team