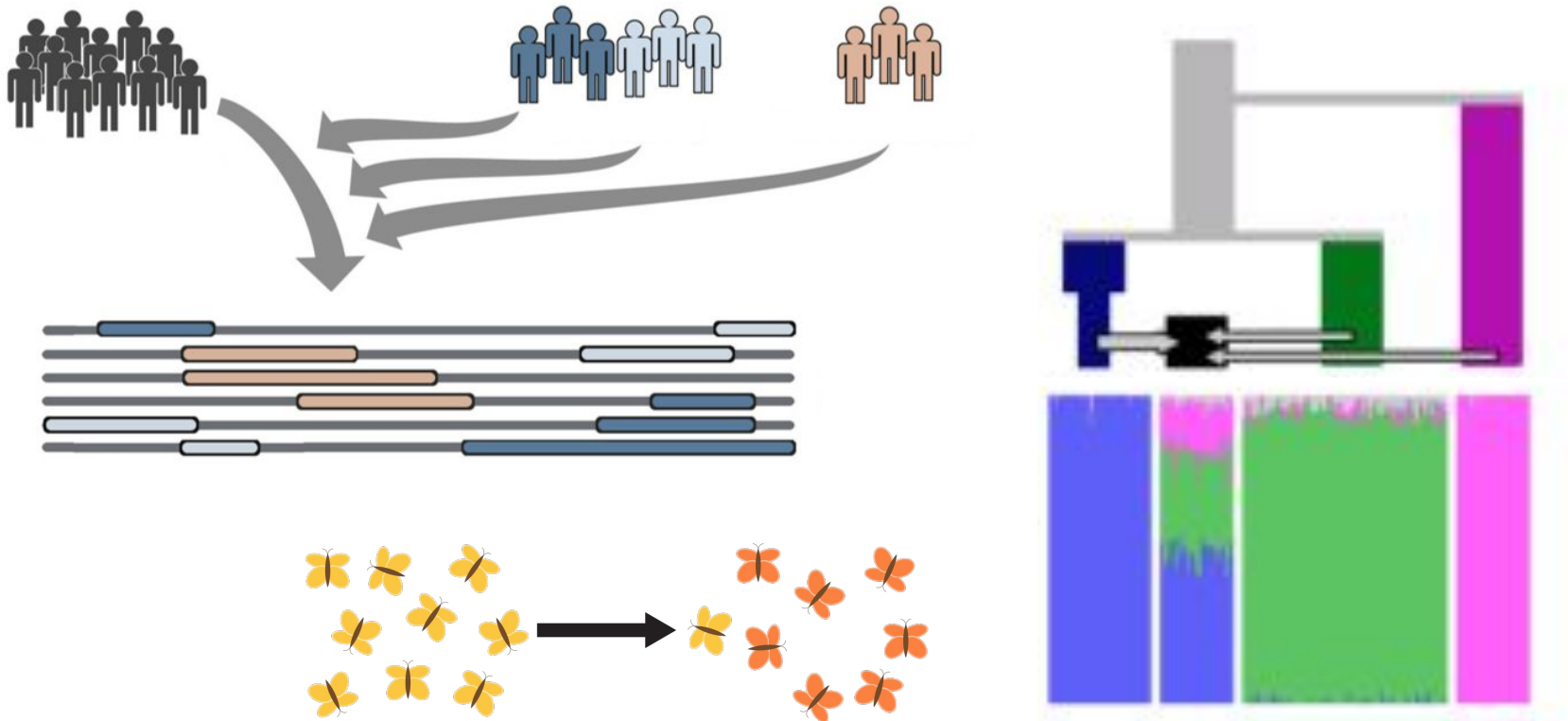


# Module 4: Population genomics of admixture and introgression



Shyamalika Gopalan, PhD

AGAR Workshop, July 28th 2022

## Part 2 objectives

- Understand the effects of gene flow/migration on patterns of neutral diversity and population differentiation
- Be able to explain the advantage of having multi-locus data for estimating gene flow
- Understand the principle behind methods for inferring gene flow
- Understand the difference between global and local ancestry

## Review of Exercise 2

How do **split time** and **effective population size** impact the process of **population differentiation**?

## How are populations related to each other?

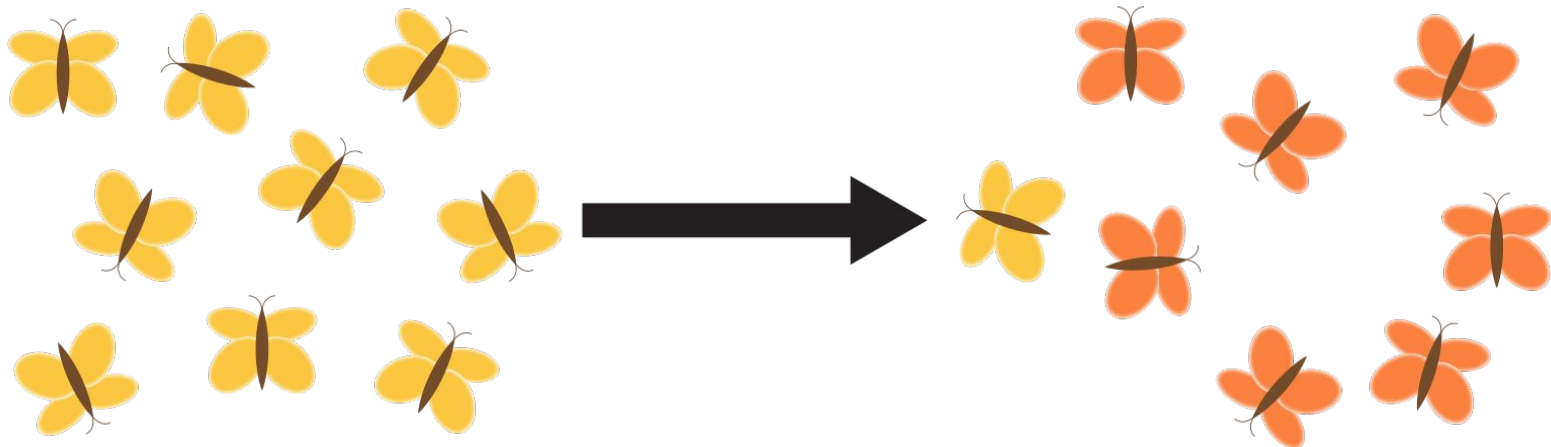
- When did populations begin evolving independently?
- Have they interacted since splitting?
- Have these dynamics changed over time, and if so, why?

# Populations evolve by four main mechanisms:

1. Mutation - generates new variation
2. Selection - directional change in frequency
3. Drift - random change in frequency
4. **Migration - introduces new variation**

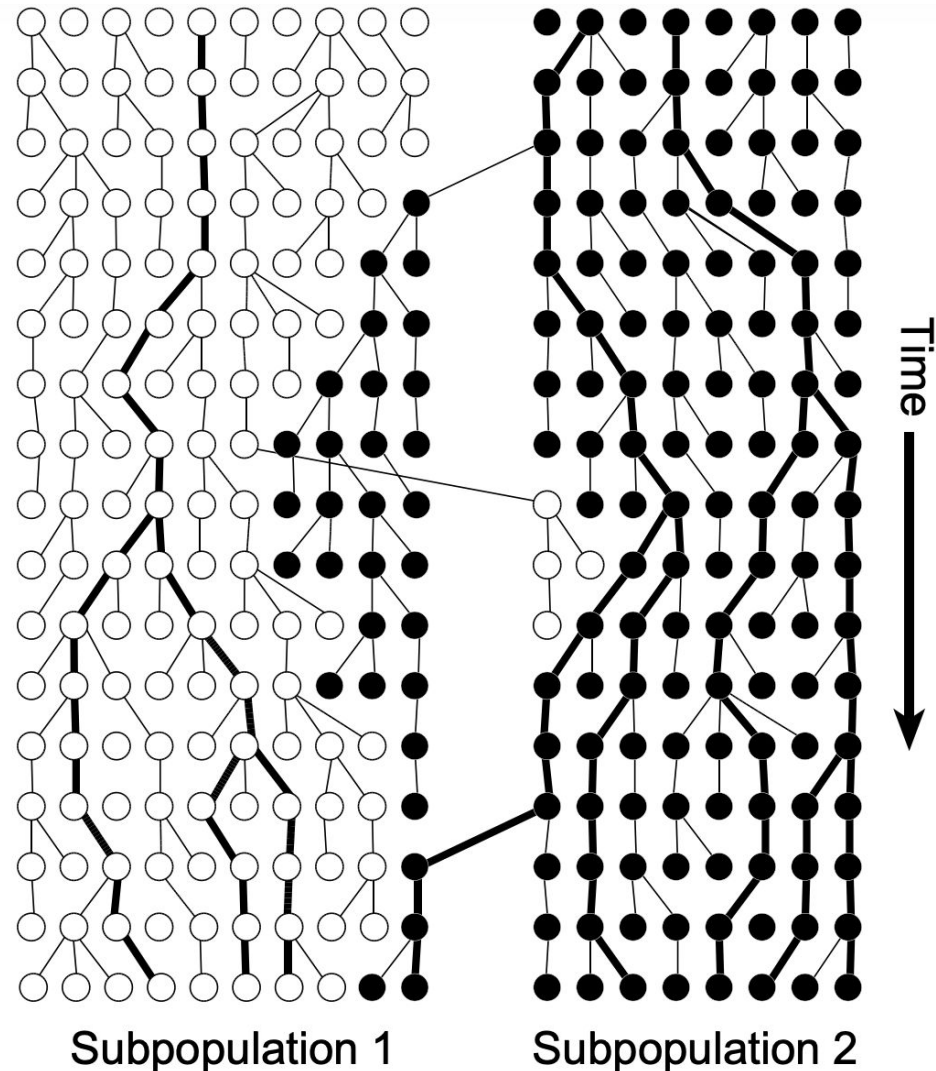
# Gene flow/migration

- The **movement of alleles** (or genetic lineages or individuals) from one population to another



# Gene flow/migration

- Parametrized as a **per generation rate**
- Describes the **proportion of the population** that is composed of migrant individuals (or lineages)



## Gene flow/migration

- The effect of gene flow **counters** that of genetic drift
- Neutrally evolving populations are expected to reach a **migration-drift equilibrium**

$$F_{ST} \sim \frac{1}{1 + 4N_e m}$$



# Exercise 3

**~20 min.**

[github.com/alanrogers/agar22/tree/main/introgression](https://github.com/alanrogers/agar22/tree/main/introgression)

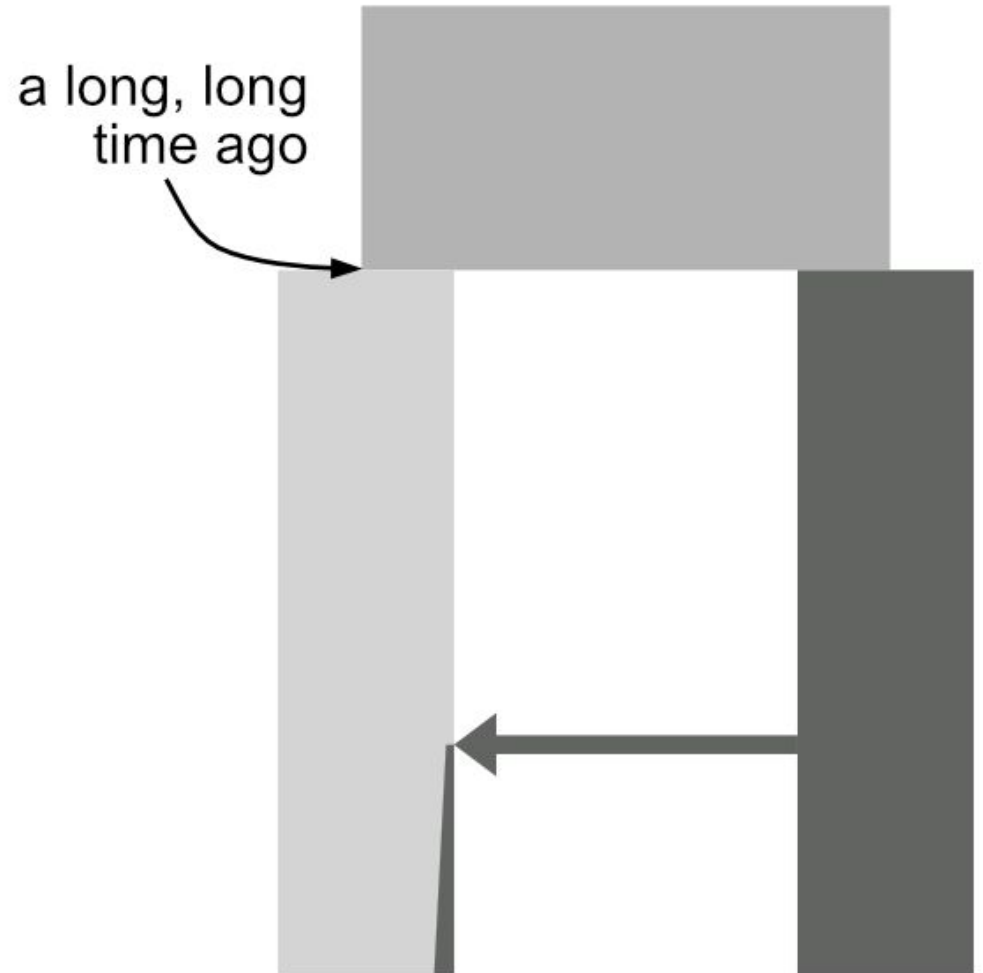
## Introgression and admixture

These terms involve the **same mechanism**, but tend to be used to describe **qualitatively different migration models**

# Introgression and admixture

## Introgression

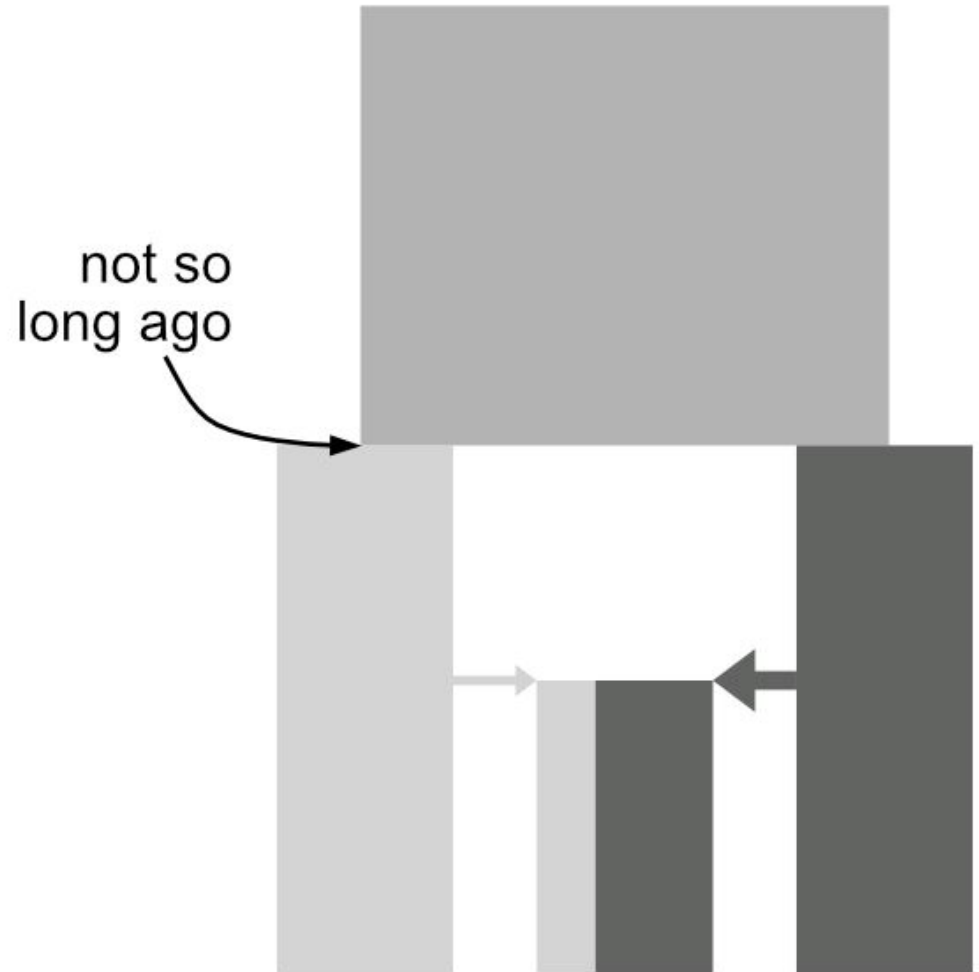
- Small proportions (~1-10%)
- Highly diverged populations (even different species)
- Pulse model



# Introgression and admixture

## Admixture

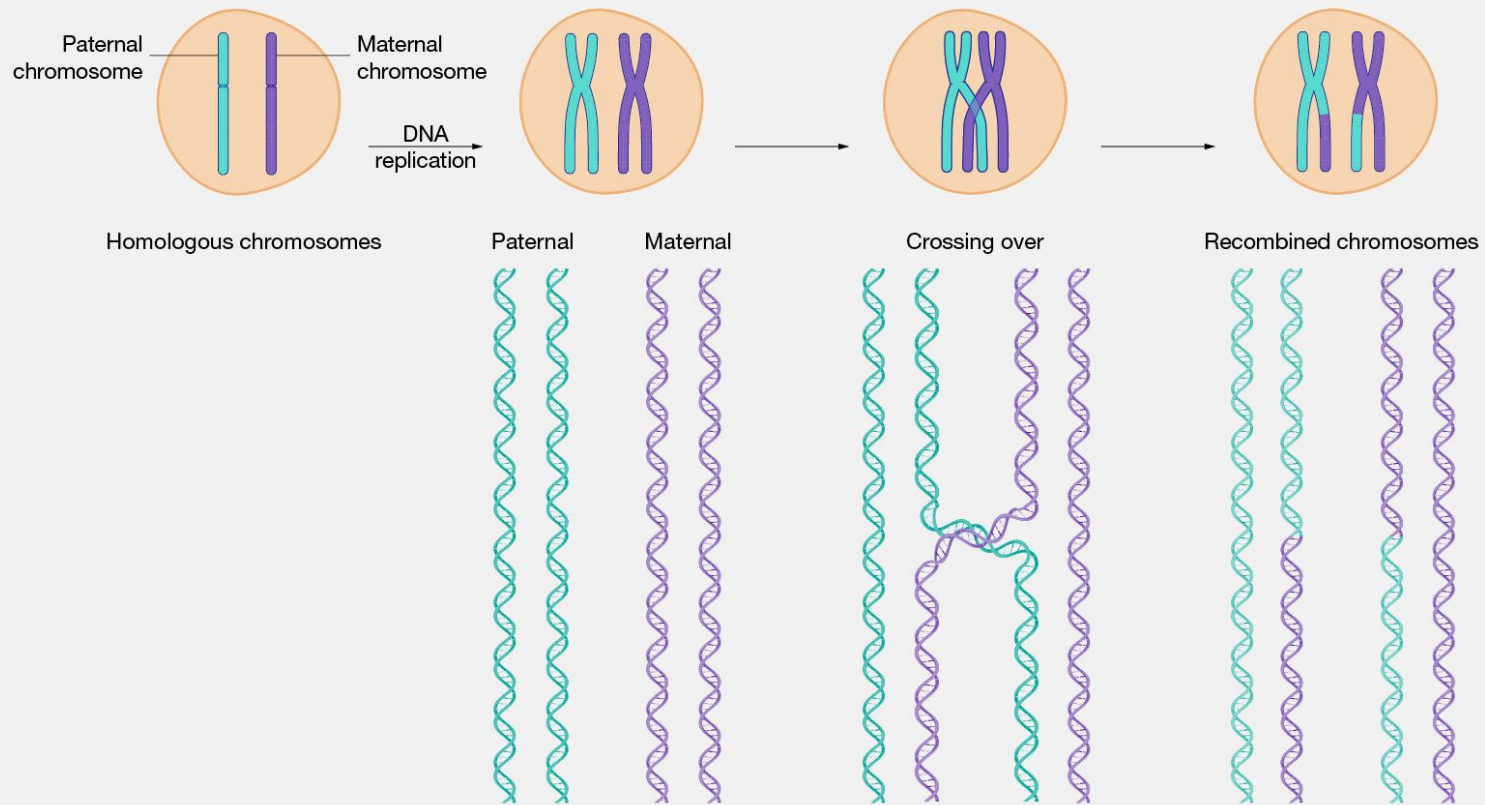
- Larger proportions (~10-90%)
- Moderately diverged populations
- Pulse model
- Often founds a new population



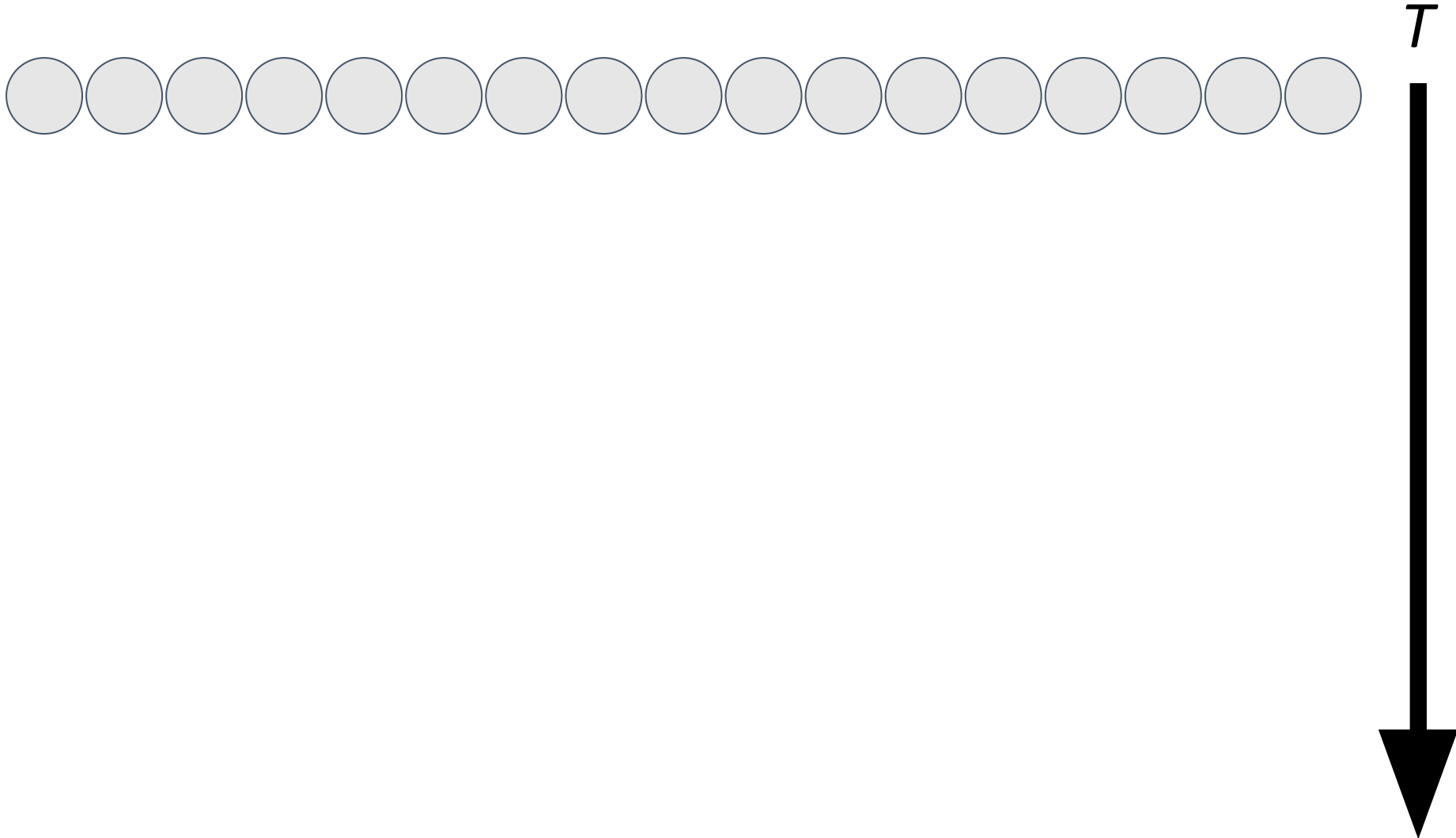
## Multi-locus genetic data

- A single genome from a **sexually reproducing** population is comprised of **multiple genetic lineages**
- **Recombination** breaks up contiguous stretches of genetic information (in humans, autosomes and X chromosome)

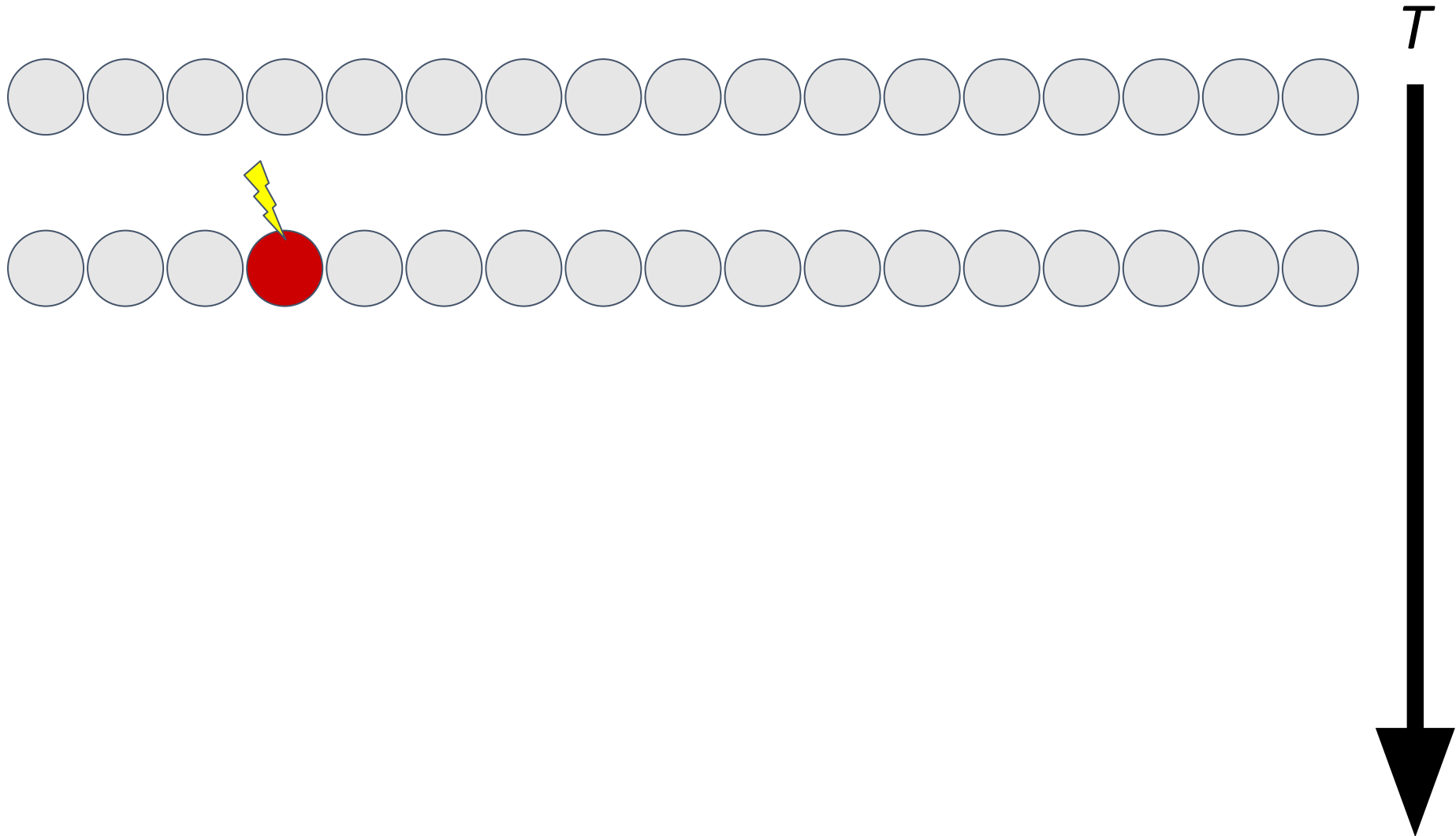
# Multi-locus genetic data



# Multi-locus genetic data

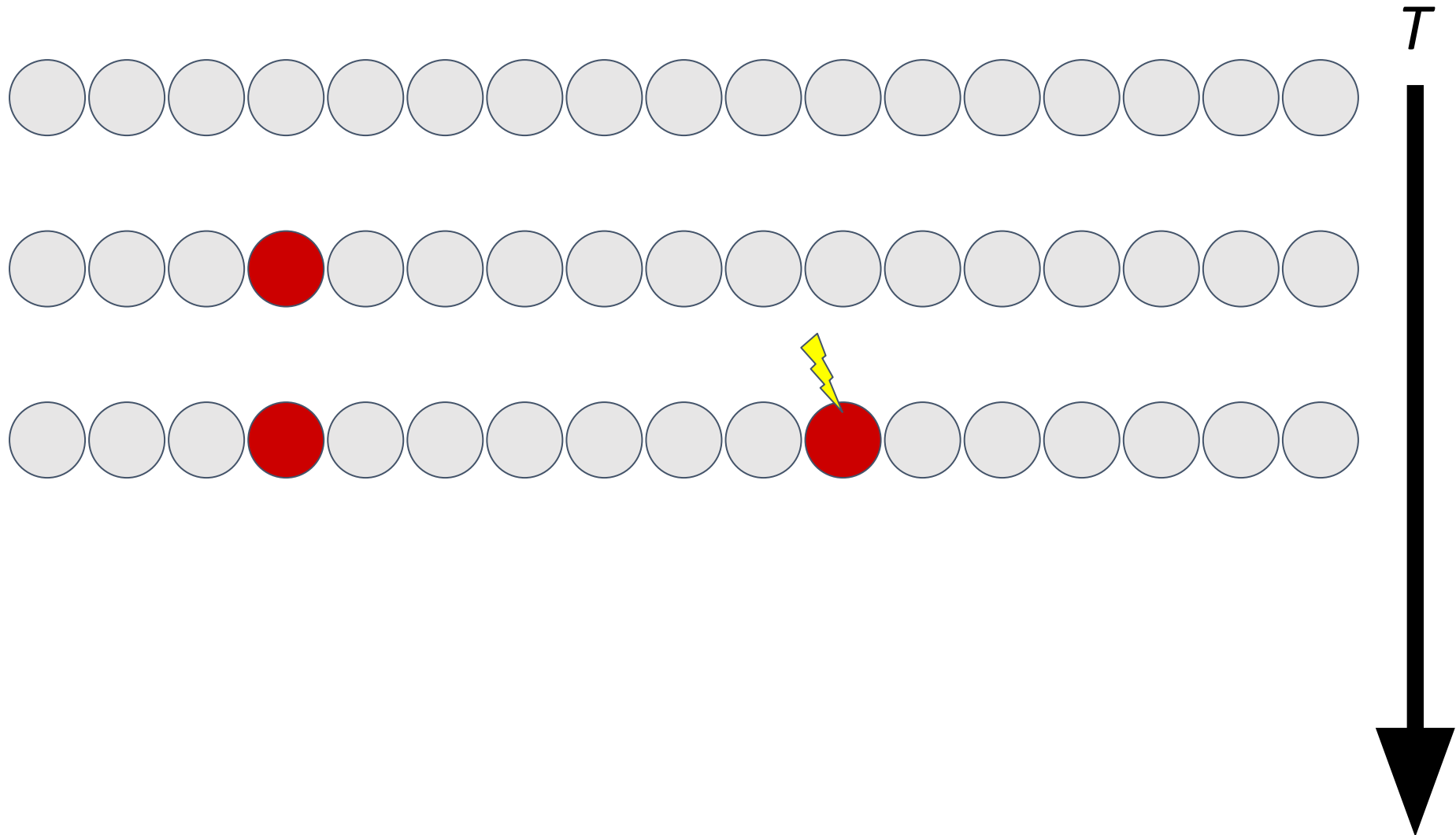


# Multi-locus genetic data

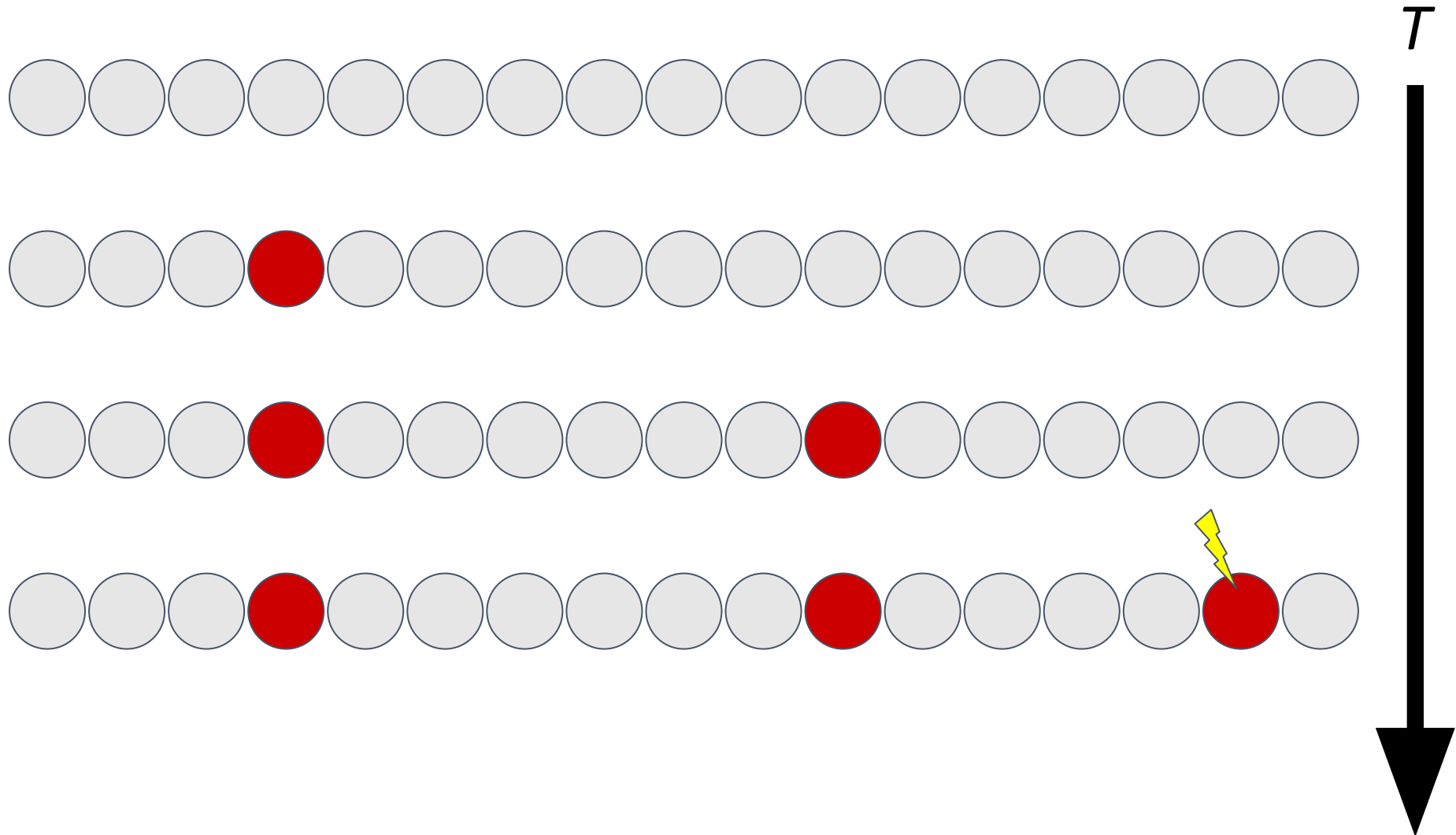




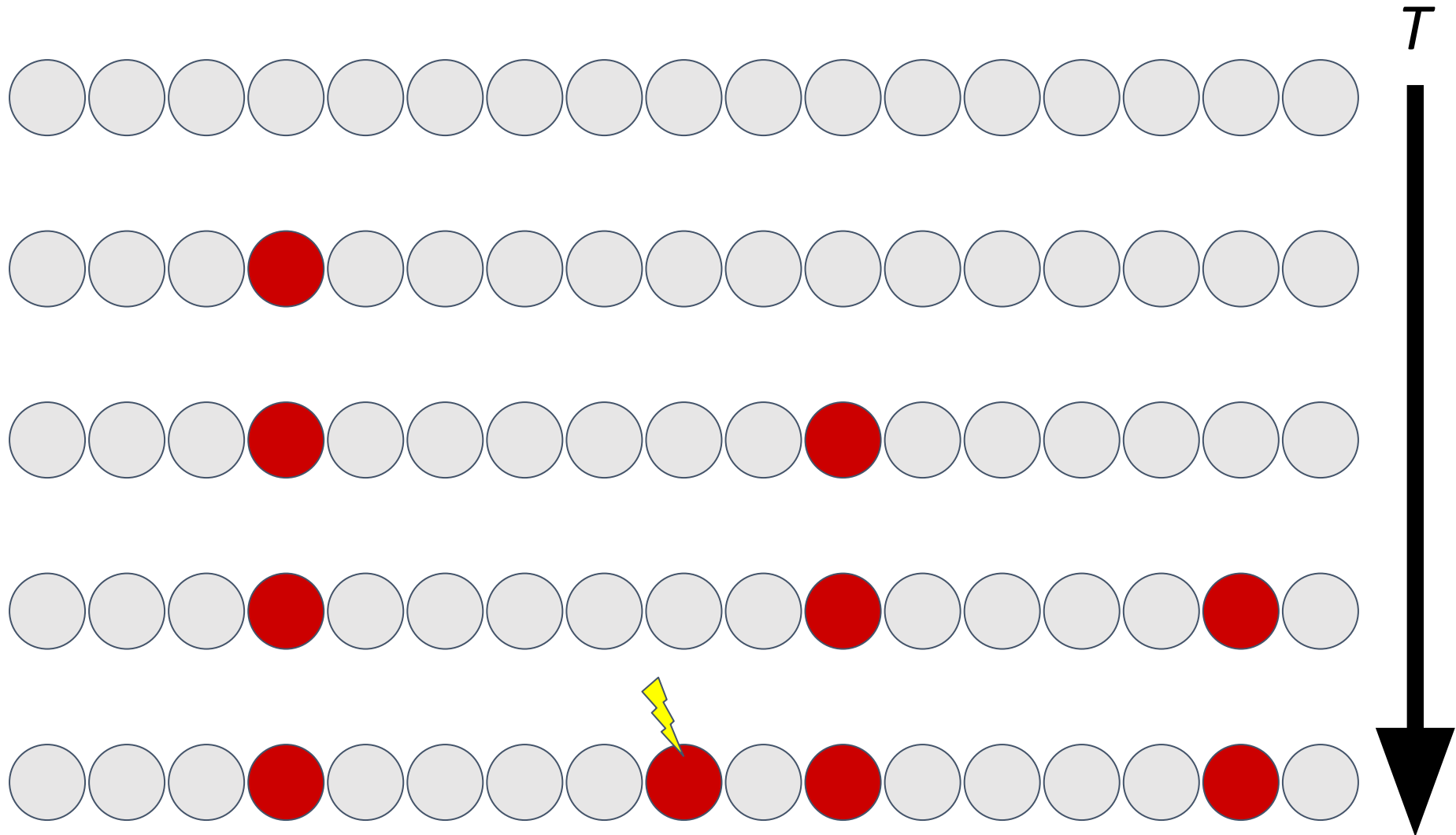
# Multi-locus genetic data



# Multi-locus genetic data



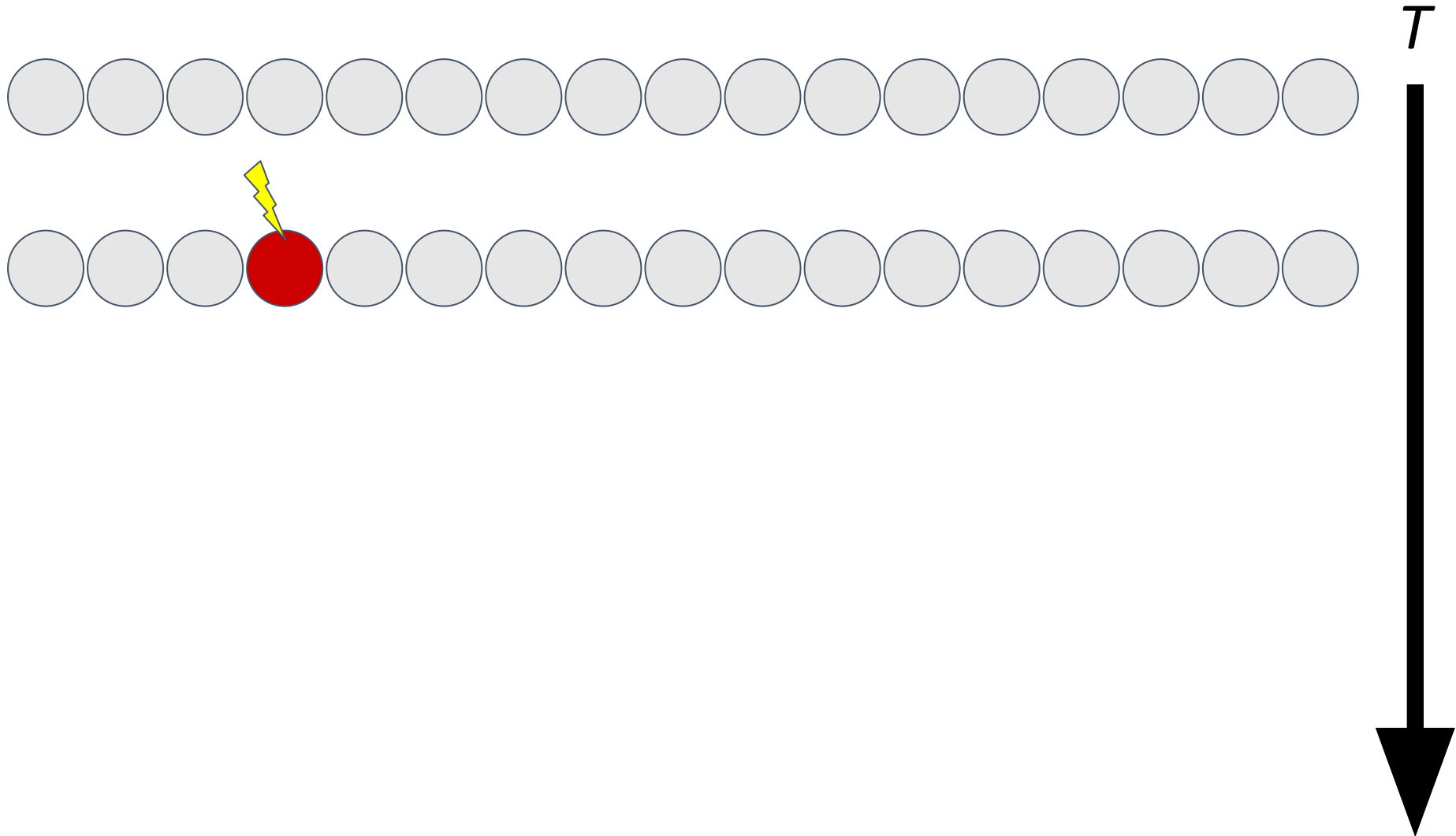
# Multi-locus genetic data



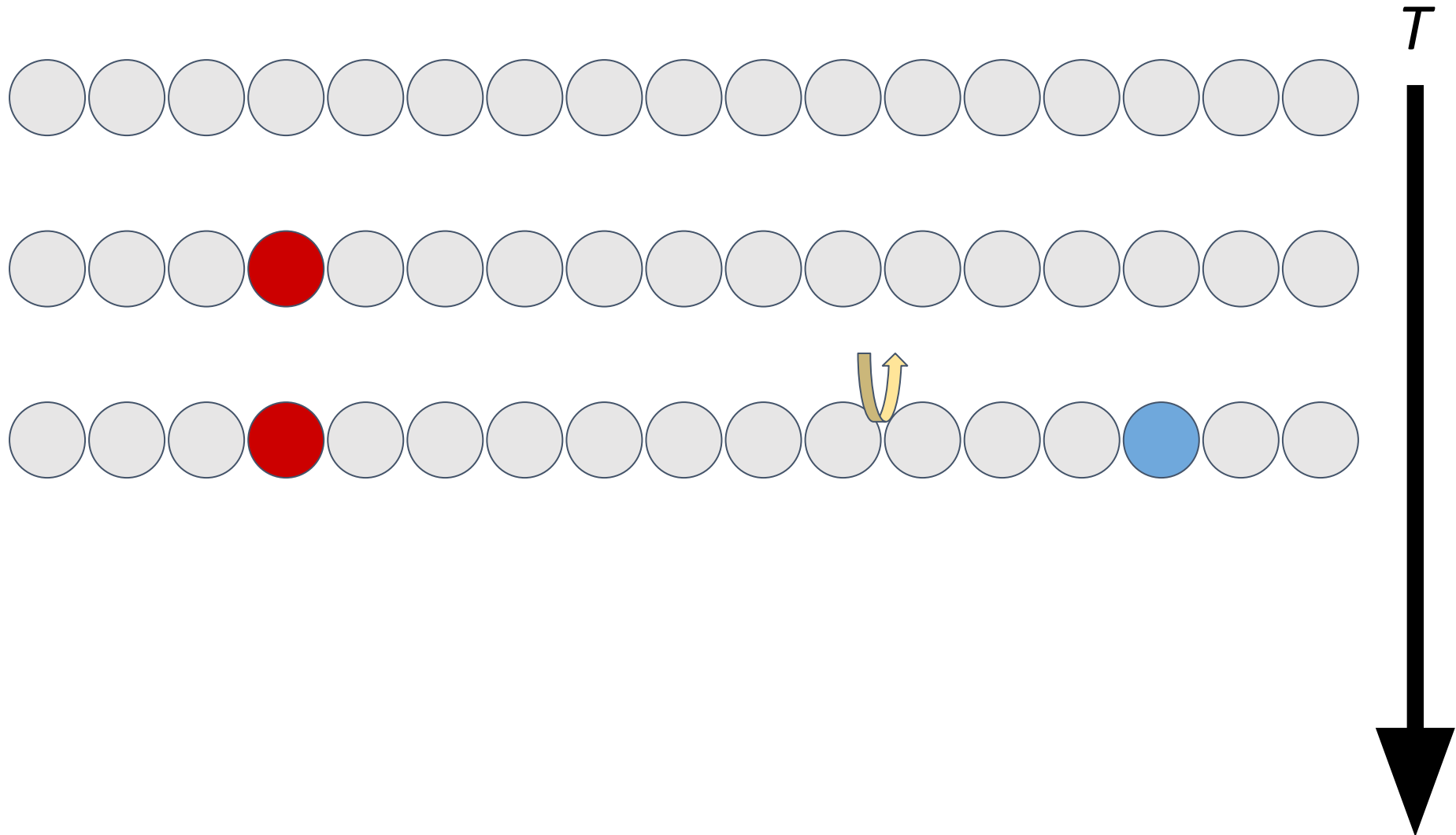
# Multi-locus genetic data



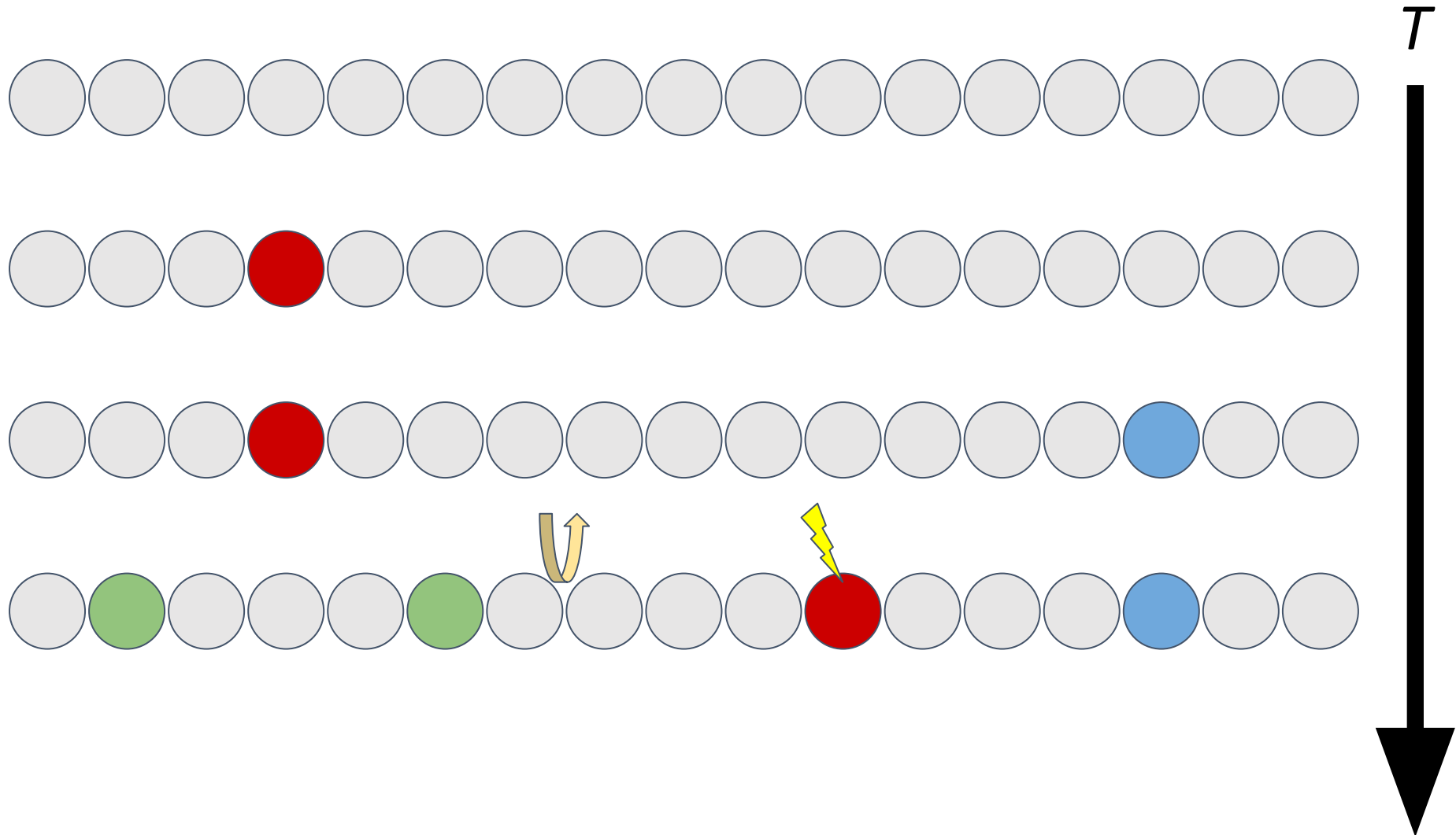
# Multi-locus genetic data



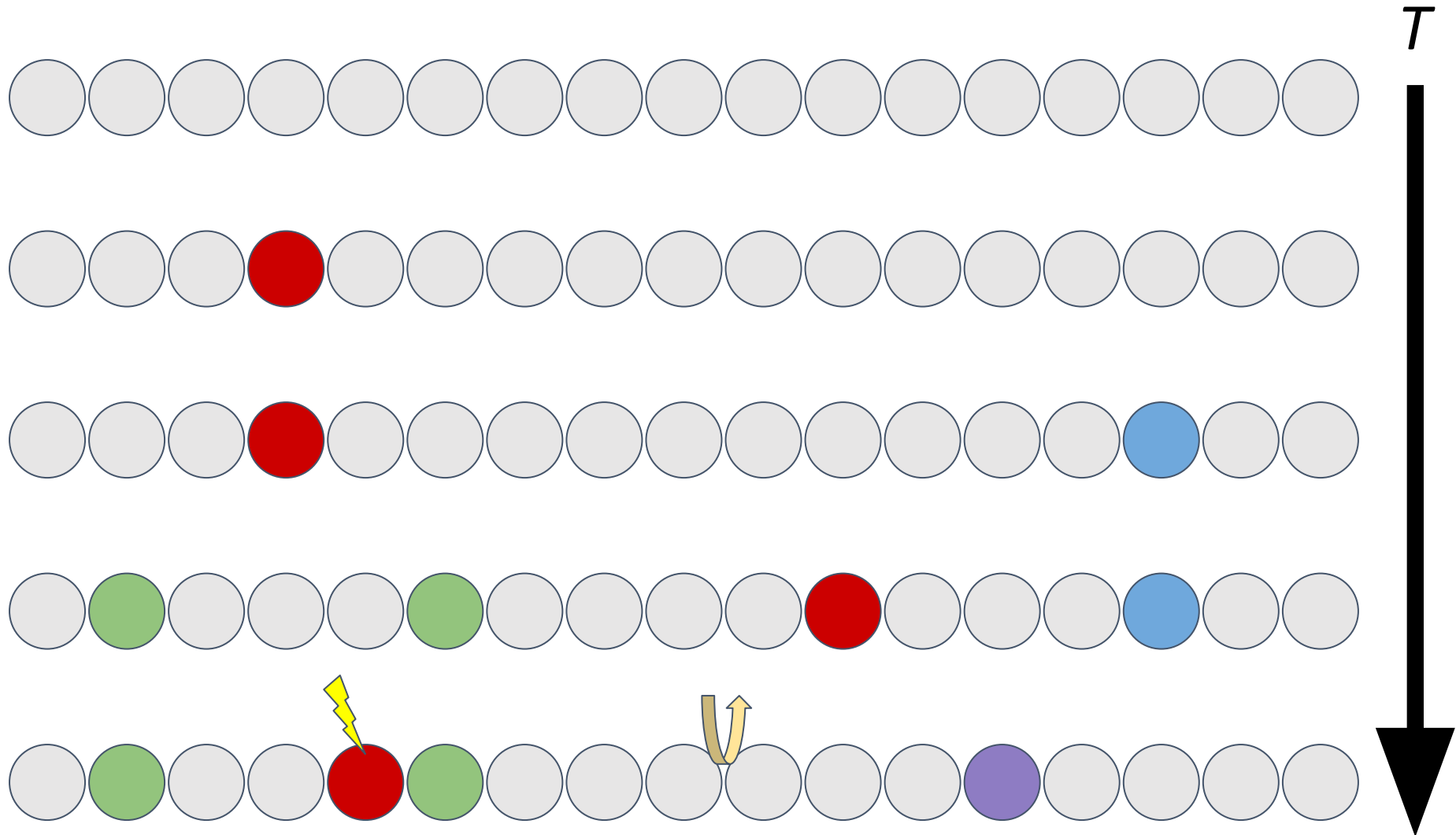
# Multi-locus genetic data



# Multi-locus genetic data



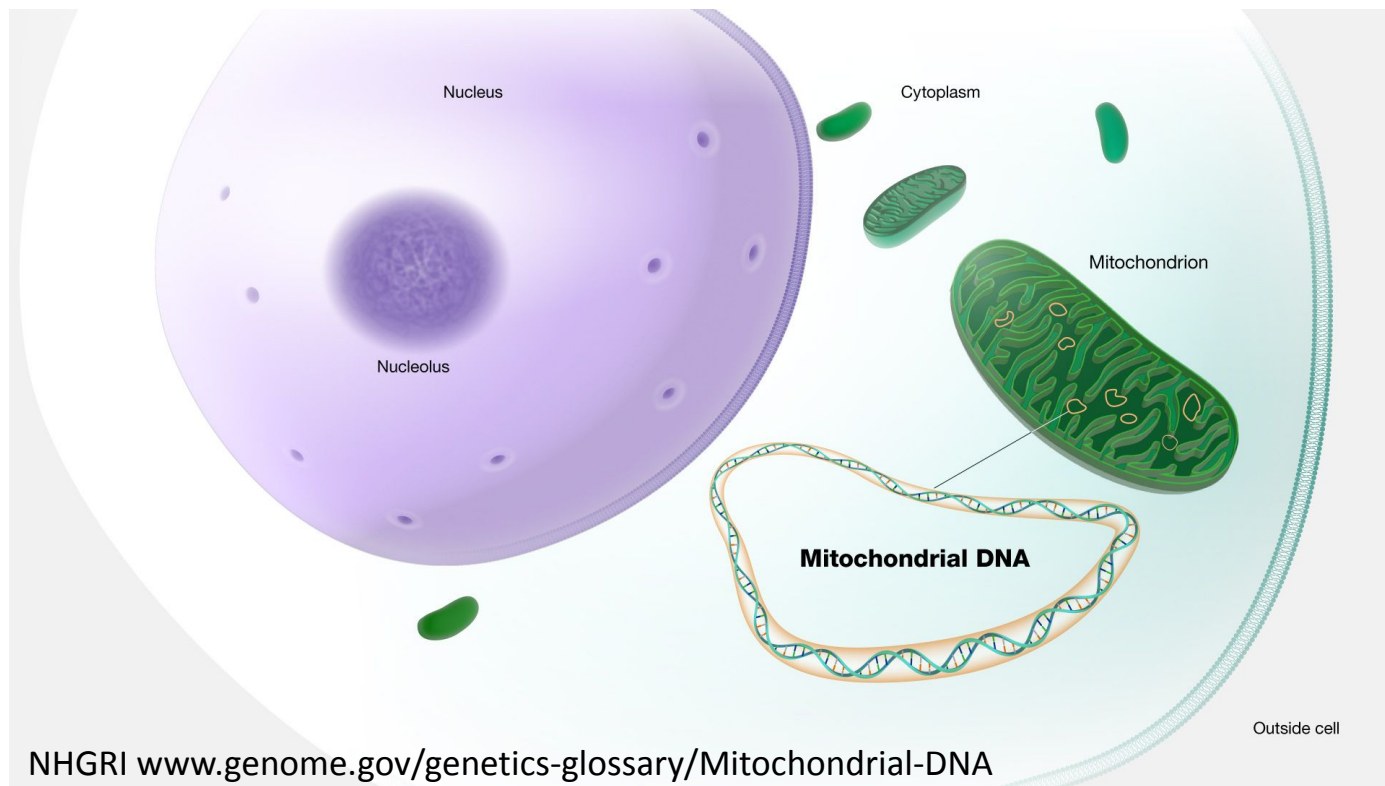
# Multi-locus genetic data





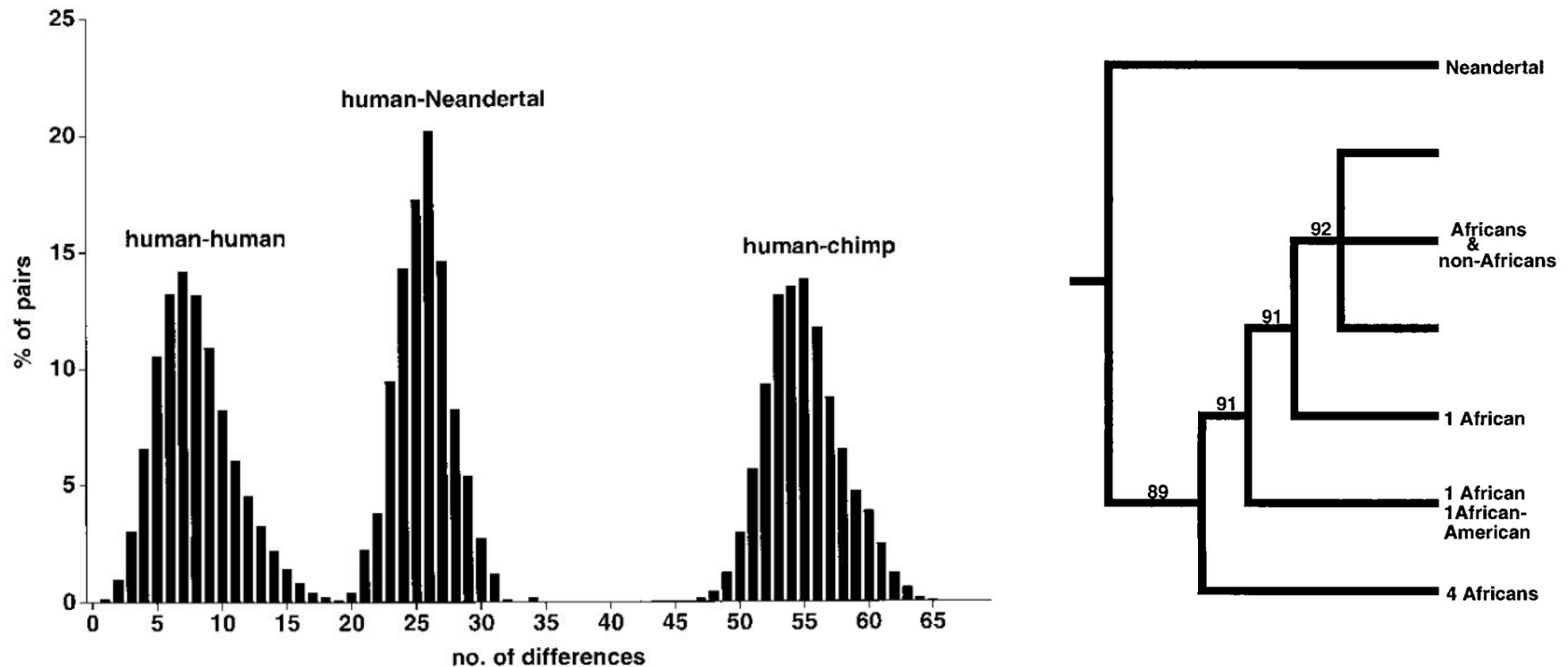
# Neanderthal introgression

- Earliest Neanderthal DNA came from the **mitochondrial genome**, a non-recombining genetic locus



# Neanderthal introgression

- The Neanderthal mtDNA sequence was **highly differentiated** from modern human mtDNA

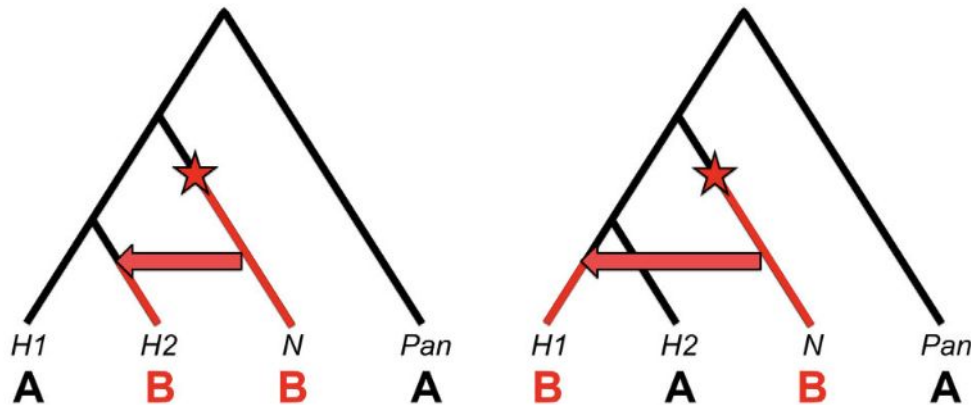


## Neanderthal introgression

Given the **mtDNA evidence**, how likely was it that anatomically modern humans and Neanderthals interbred?

# Neanderthal introgression

- In 2010, a draft sequence of the Neanderthal **nuclear genome** was published
- Authors developed a new statistic to **test for introgression**



$$D = \frac{\sum_{i=1}^n [C_{ABBA}(i) - C_{BABA}(i)]}{\sum_{i=1}^n [C_{ABBA}(i) + C_{BABA}(i)]}$$

# Neanderthal introgression

- Based on D-statistic values, authors estimated **2-3%** of the modern human genome derived from **Neanderthal introgression**

REVIEW ARTICLE

**Evolutionary Anthropology** WILEY  
ISSUES, NEWS, AND REVIEWS

## Inferring archaic introgression from hominin genetic data

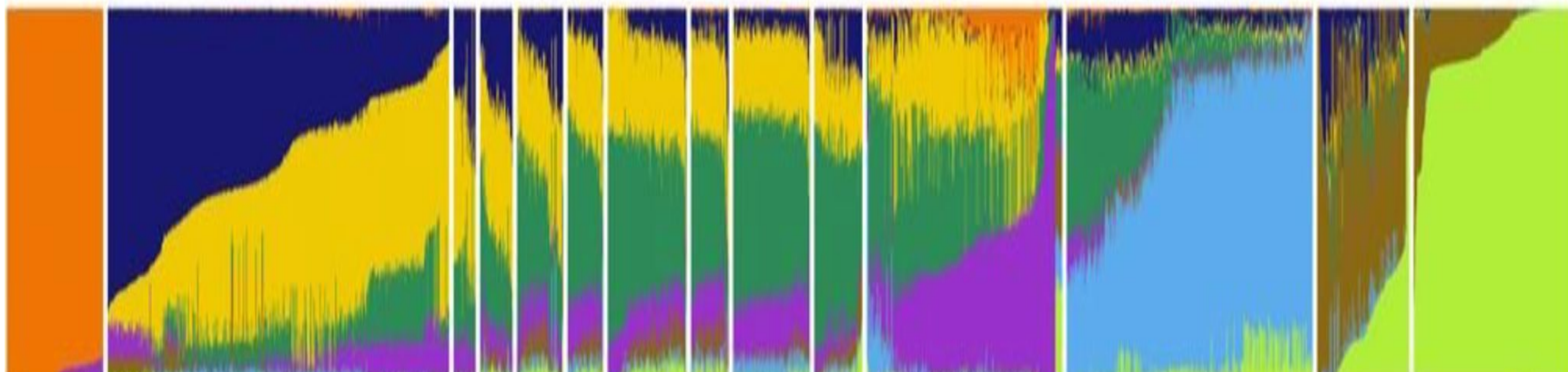
Shyamalika Gopalan<sup>1,2</sup>  | Elizabeth G. Atkinson<sup>1,3</sup> | Laura T. Buck<sup>4</sup> |  
Timothy D. Weaver<sup>5</sup> | Brenna M. Henn<sup>1,5,6</sup>

# Admixture

- After splitting, individuals from different populations can **merge**, founding a **new population** whose members carry genomic information from all the source populations

# Admixture

- Population level **multi-locus data** can give us rich insight into **population relatedness** and processes of **population mixture**
- If the parent populations are **sufficiently distinct**, it may be possible to parse these **different genetic components** statistically



# Admixture

- **ADMIXTURE** is a popular algorithm that estimates **genetic ancestry proportions** from **genotype data alone** (i.e. unsupervised)
- Models the sample of genomes as a combination of **K ‘ancestry clusters’**
- Assumes that loci are **approximately independent** of each other

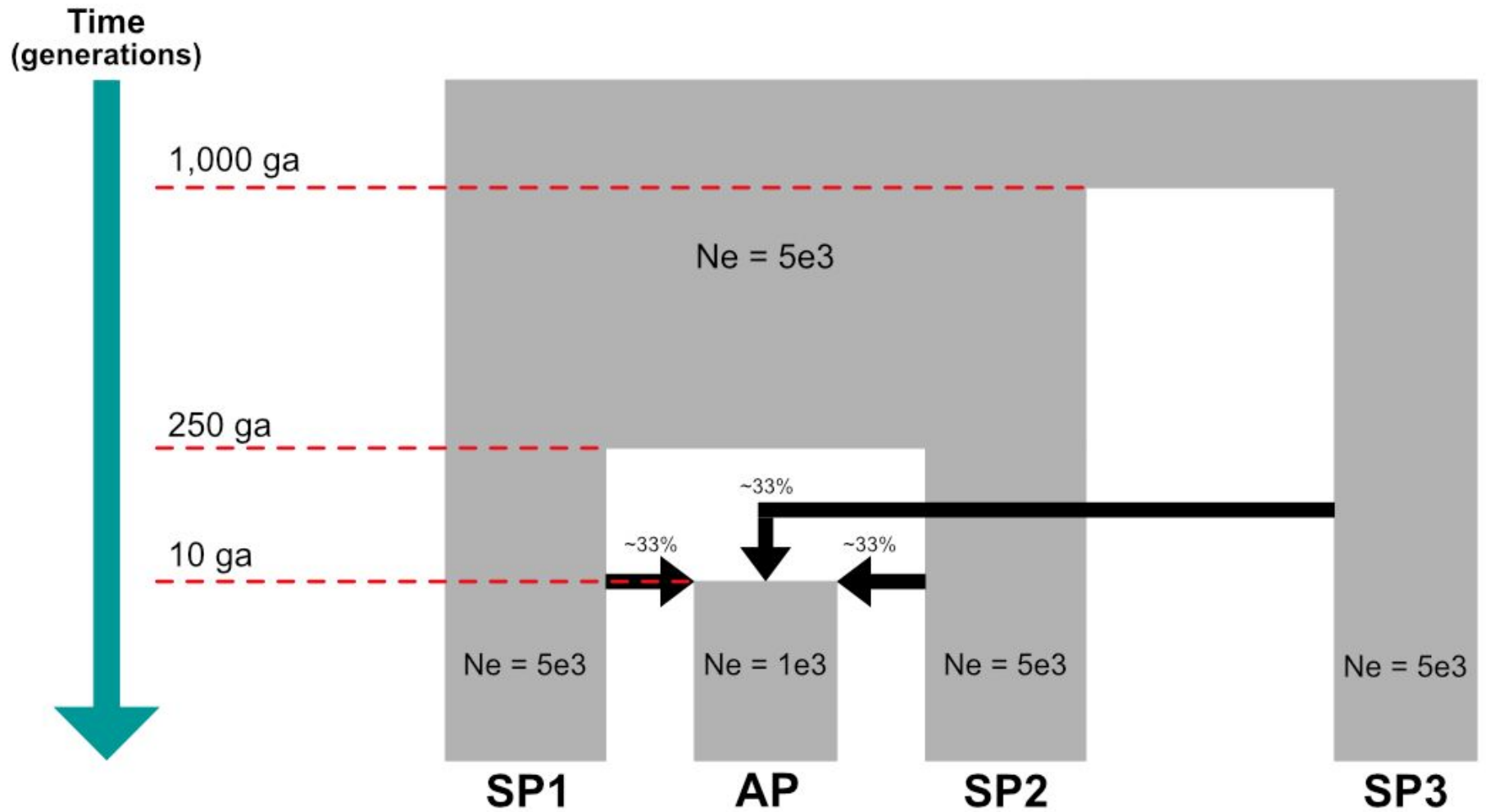


# Exercise 4

~25 min.

[github.com/alanrogers/agar22/tree/main/introgression](https://github.com/alanrogers/agar22/tree/main/introgression)

## Part 2 - Gene flow



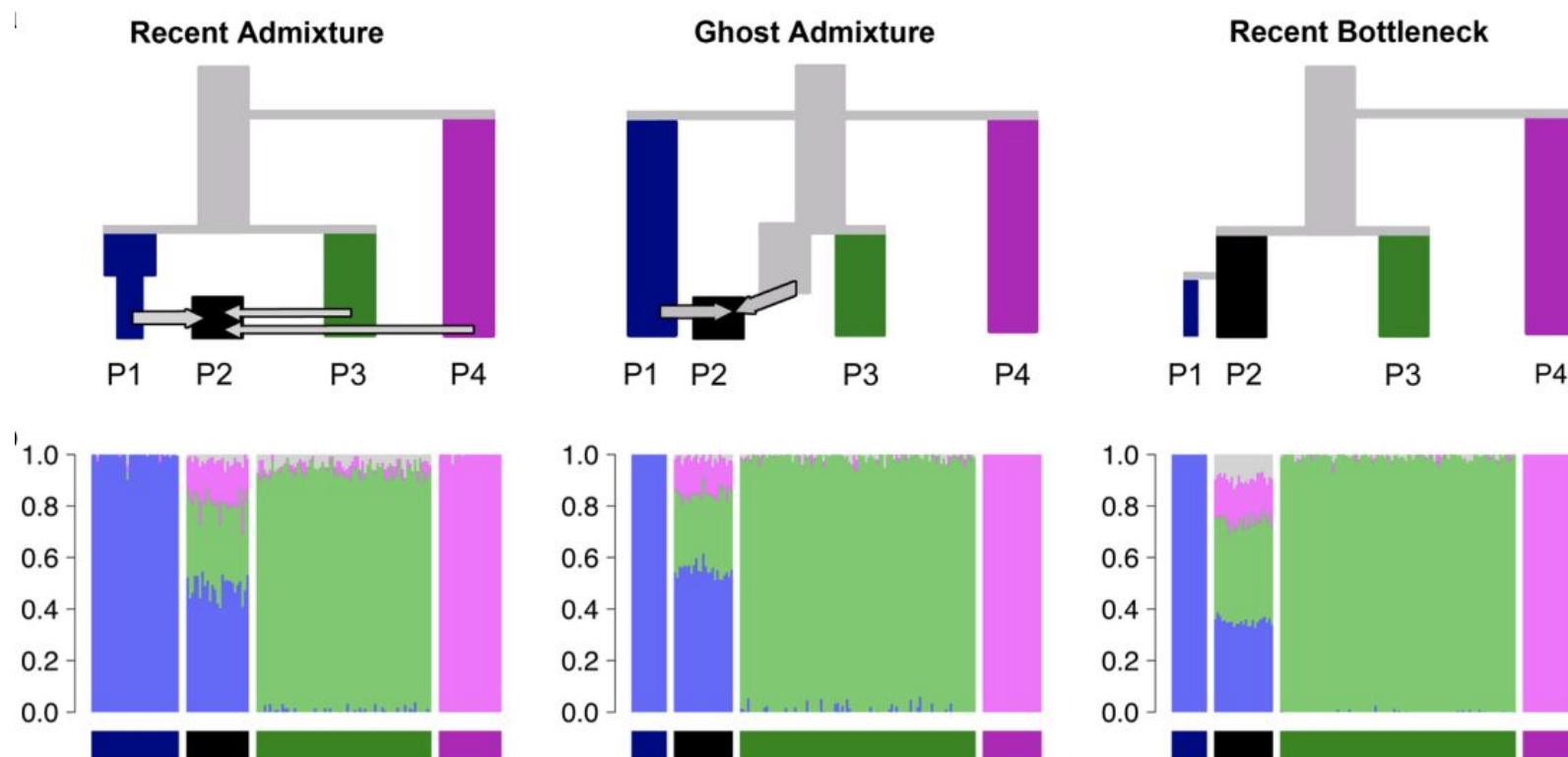
# Exercise 4

~25 min.

[github.com/alanrogers/agar22/tree/main/introgression](https://github.com/alanrogers/agar22/tree/main/introgression)

# Admixture

- **Beware of overinterpreting!** Different scenarios can produce qualitatively similar results



# Admixture

- **Beware of overinterpreting!** Different scenarios can produce qualitatively similar results



ARTICLE

DOI: 10.1038/s41467-018-05257-7

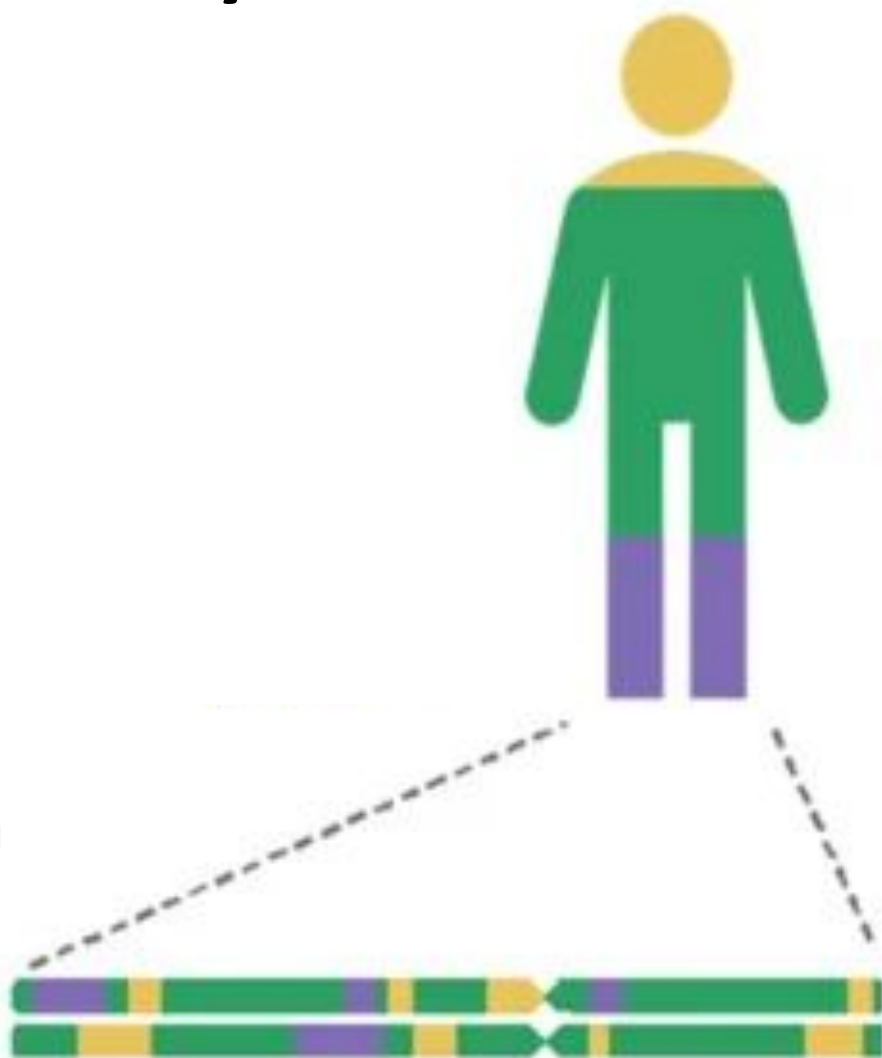
OPEN

## A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots

Daniel J. Lawson <sup>1</sup>, Lucy van Dorp<sup>2,3</sup> & Daniel Falush<sup>4</sup>

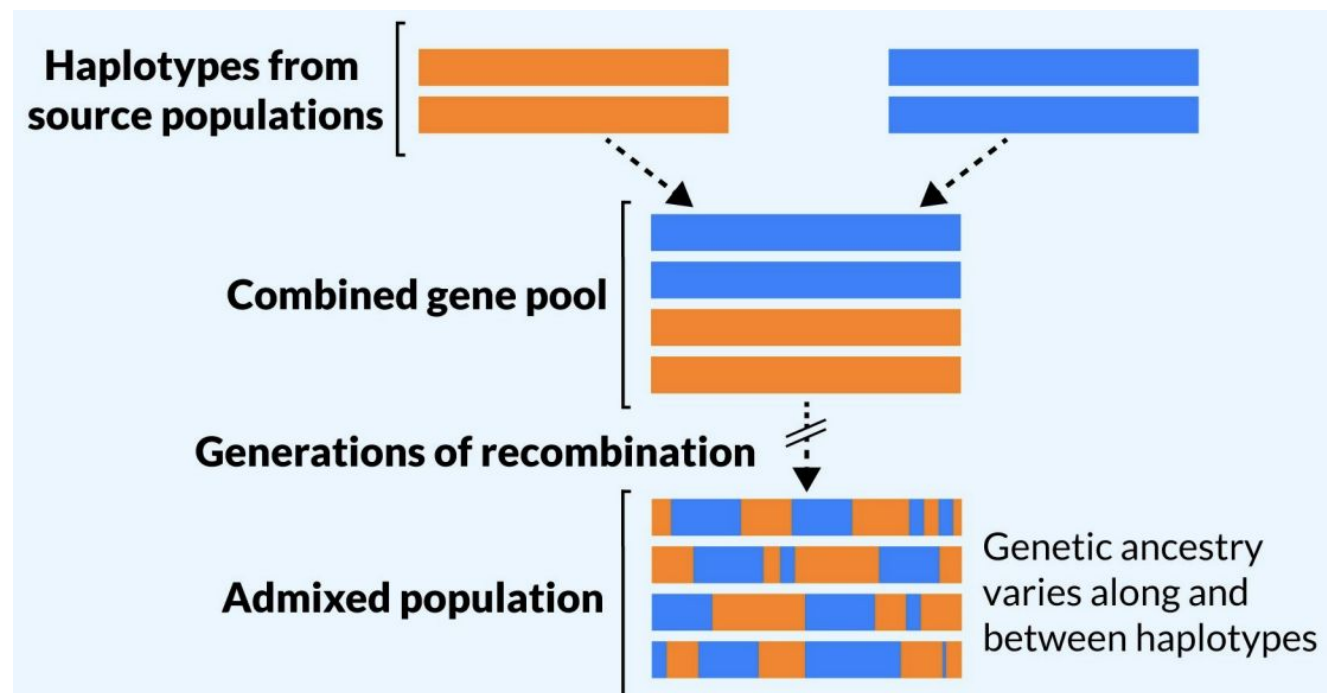
# Global versus local ancestry

- ADMIXTURE software calculates estimated ancestry proportions from the **entire genome**, or **global ancestry**
- In some cases, we might be interested in more **fine-scale** information



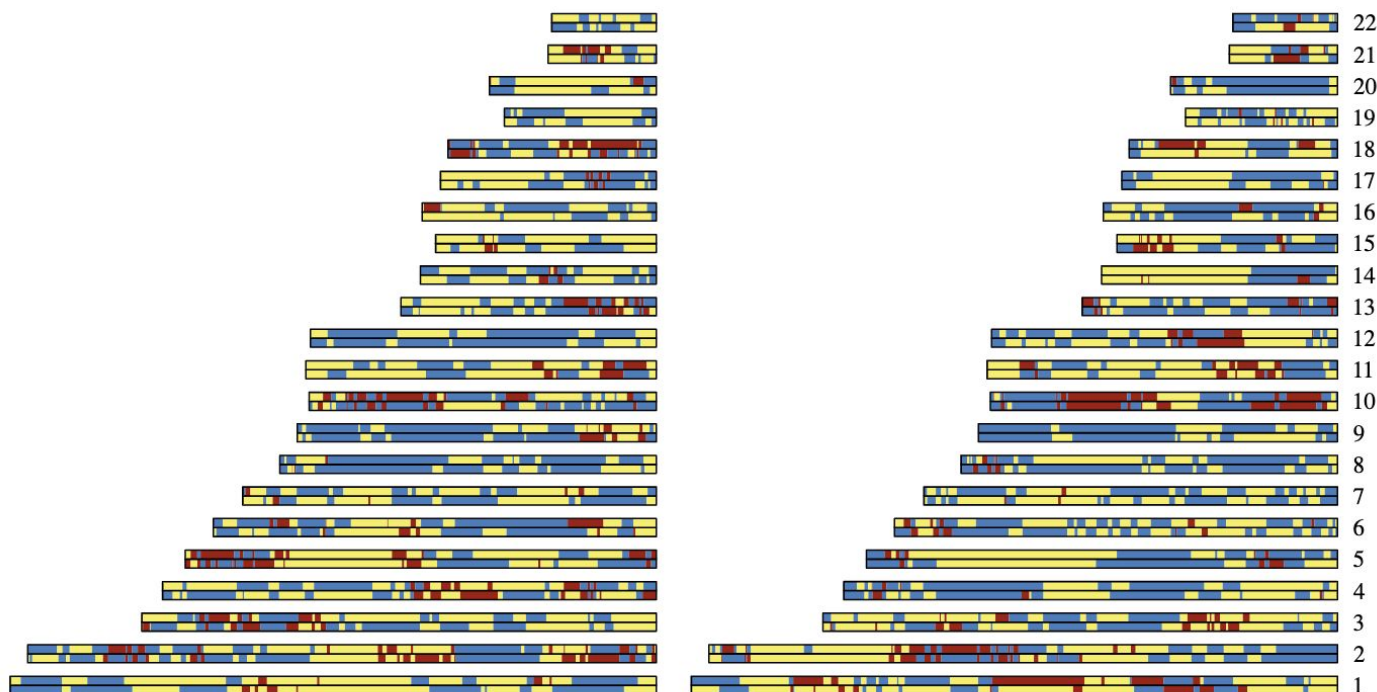
# Global versus local ancestry

- At each generation post-admixture, recombination breaks up **ancestry tracts** in individual genomes



# Global versus local ancestry

- Individuals with **nearly identical** global ancestry proportions can **differ significantly** in their local ancestry distributions





# Global versus local ancestry

- This gives us an extra layer of information to understand the **demographic processes** of admixture, **post-admixture selection**, and more

## Human genetic admixture through the lens of population genomics

---

Shyamalika Gopalan<sup>1</sup>, Samuel Pattillo Smith<sup>2,3</sup>, Katharine Korunes<sup>1</sup>, Iman Hamid<sup>1</sup>, Sohini Ramachandran<sup>2,3,4,†</sup> and Amy Goldberg<sup>1,†</sup>

---

<sup>1</sup>Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA

<sup>2</sup>Center for Computational Molecular Biology, <sup>3</sup>Department of Ecology, Evolution and Organismal Biology, and

<sup>4</sup>Data Science Initiative, Brown University, Providence, RI 02912, USA

# Exercise 5

**~20 min.**

[github.com/alanrogers/agar22/tree/main/introgression](https://github.com/alanrogers/agar22/tree/main/introgression)