

The Neutral Model

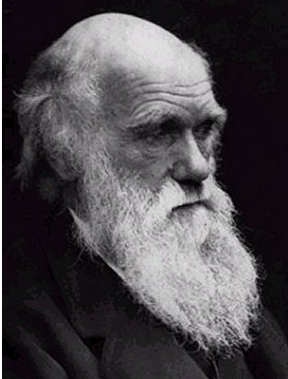
Evolution without selection

Joachim Hermisson

Mathematics and Biosciences Group

Mathematics & MFPL, University of Vienna

Evolutionary processes



Darwin

selection
theory →

1920 – 1930
„modern
synthesis“

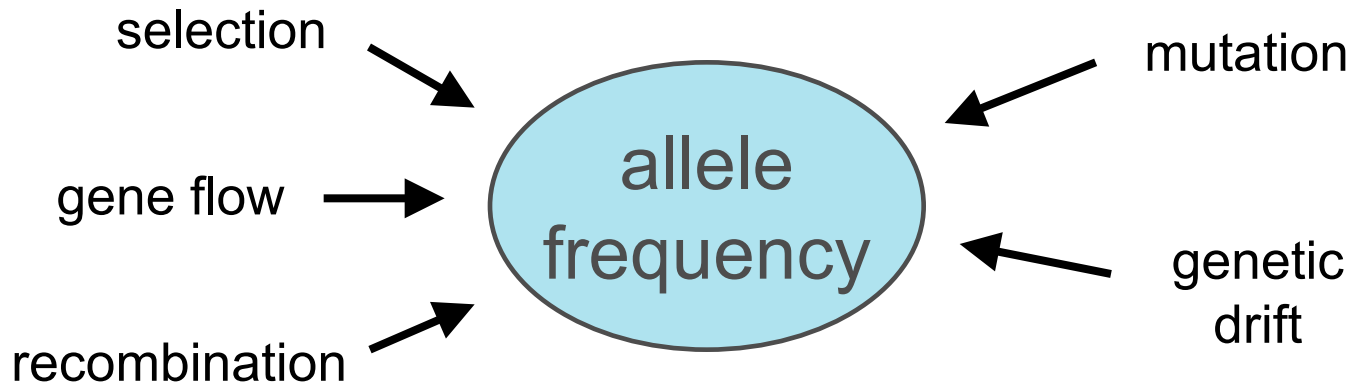
← laws of
inheritance



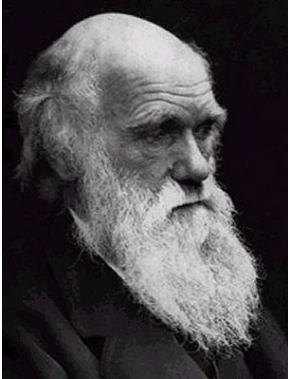
Mendel

Population Genetics

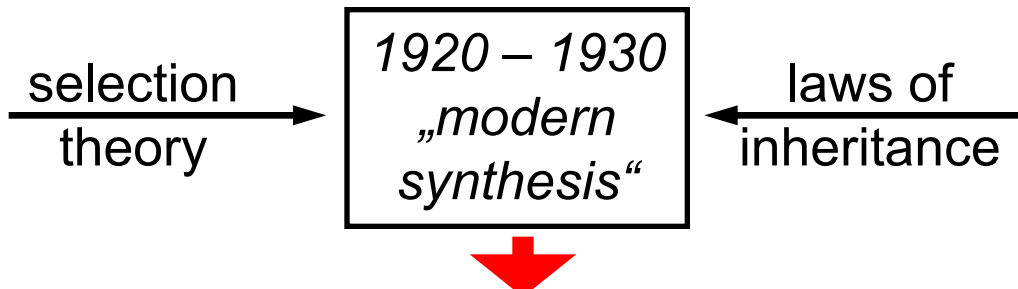
describes change of allele frequencies
under the elementary evolutionary processes:



Evolutionary processes



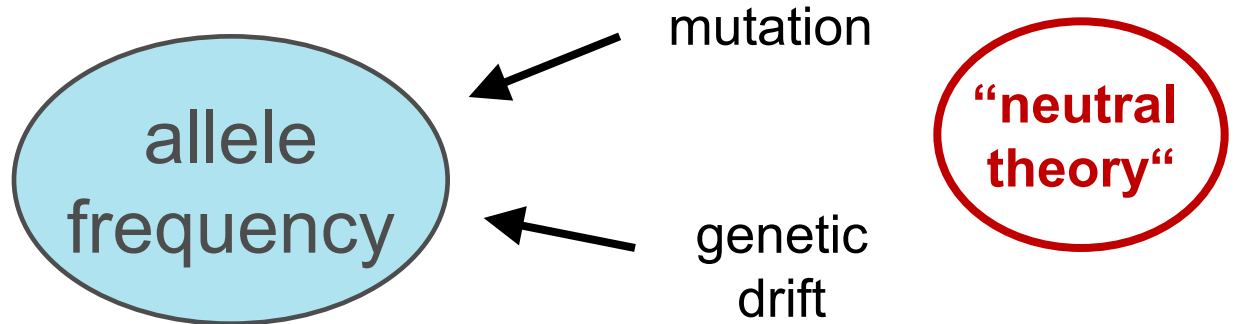
Darwin



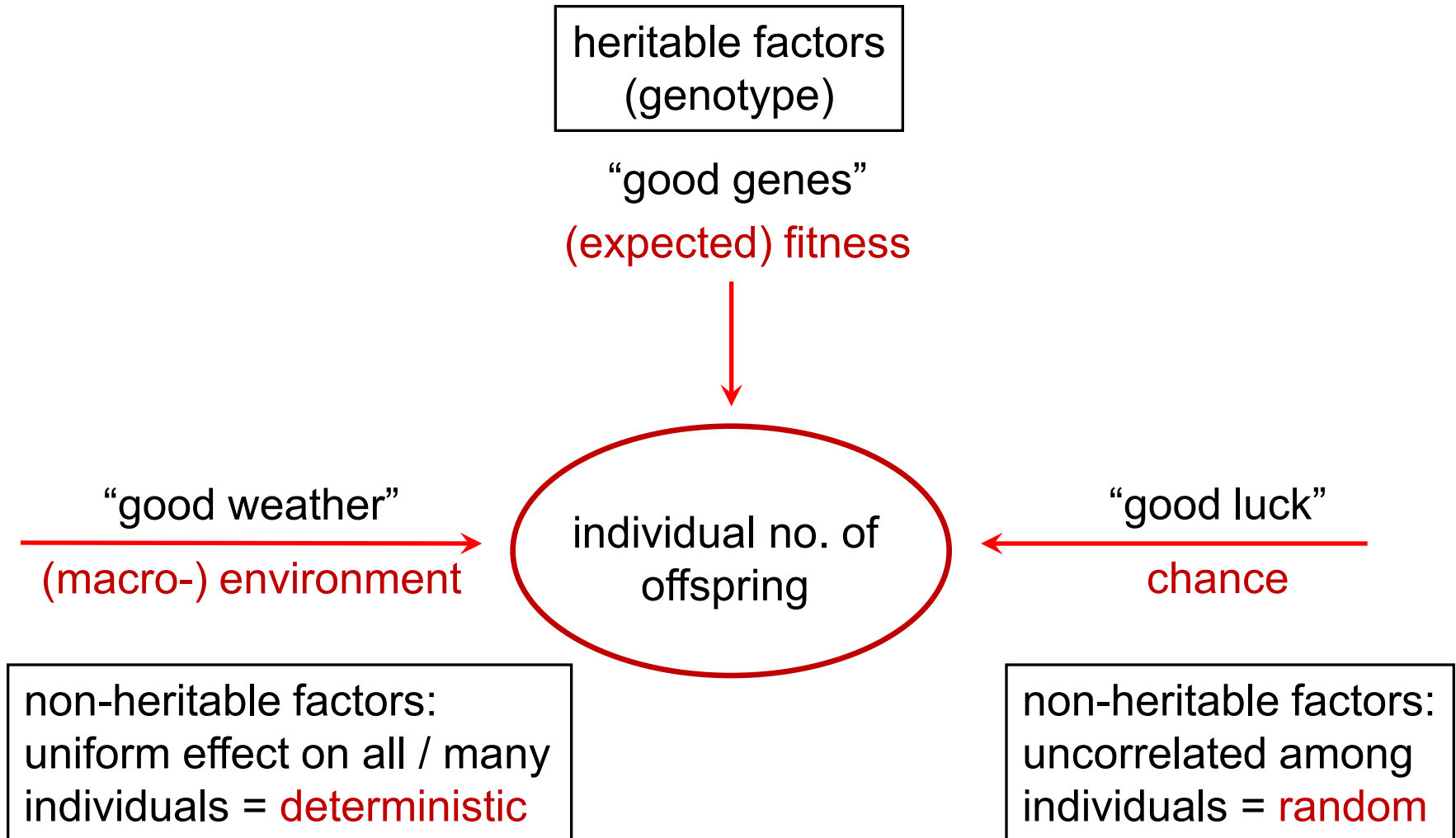
Mendel

Population Genetics

describes change of allele frequencies under the elementary evolutionary processes:

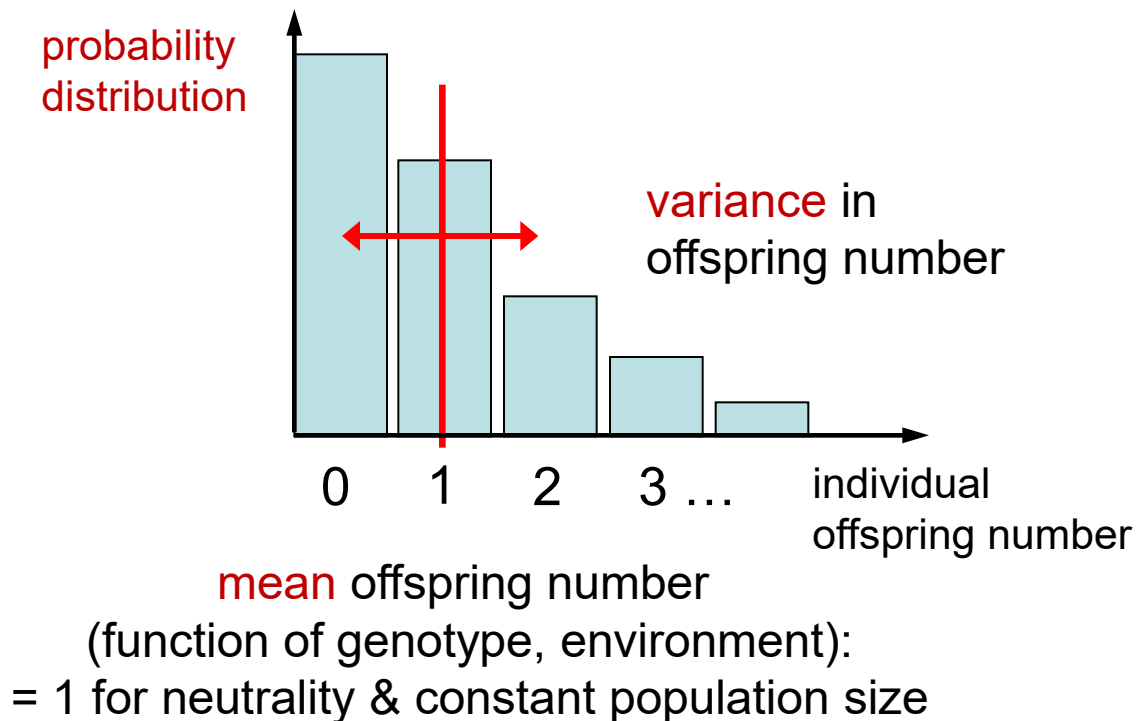


What is random genetic drift?



What is random genetic drift?

How do we define drift ?



What is random genetic drift?

How do we define drift ?

Change in allele counts due to variation in offspring number that is

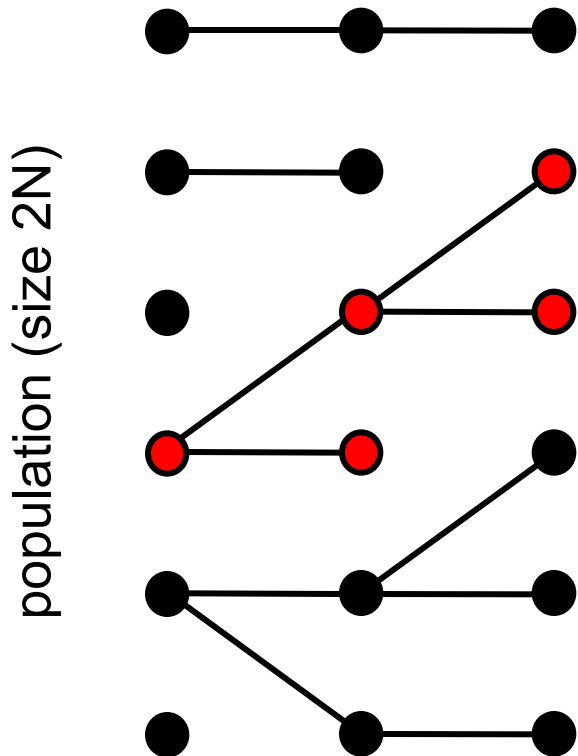
- independent of external factors
 - independent of genotypes (heritable factors)
- } *random*

Effects of drift **in a finite population**:

- allele **frequencies** in finite populations are not fixed, but undergo random changes from one generation to the next
- allele frequencies in sub-population diverge after population split
- alleles may get lost from a population or reach fixation

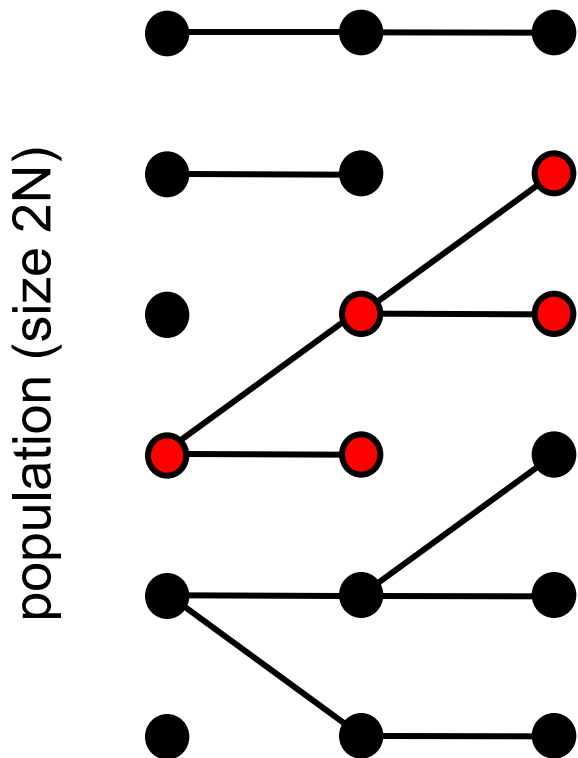
How to quantify this? – need a **model**

The Wright-Fisher model: Drift as binomial sampling



- Discrete time, non-overlapping generations
- Single haploid locus with two alleles
- Constant haploid population size $2N$
- Alleles in the next generation are sampled from an *infinite gamete pool* of the current generation
 - “Offspring randomly choose a parent” and inherit his/her genotype
 - Sampling with replacement \Rightarrow ?

The Wright-Fisher model: Drift as binomial sampling



Binomial distribution

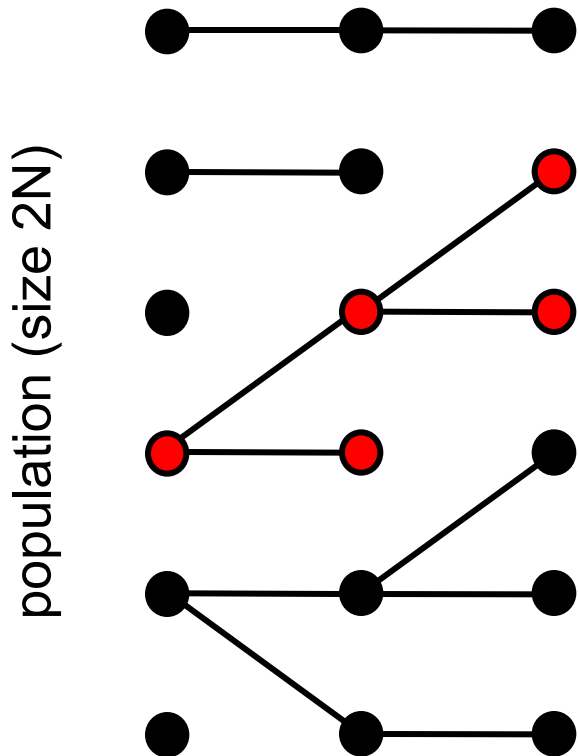
single parent: probability for k offspring
(# trials $n = 2N$; success prob. $p = 1/2N$)

$$\Pr[k] = \binom{2N}{k} \left(\frac{1}{2N}\right)^k \left(1 - \frac{1}{2N}\right)^{2N-k}$$

mean: $E[k] = n \cdot p = 1$

variance: $\sigma_k^2 = n \cdot p(1 - p)$
 $= 1 - (1/2N) \approx 1$

The Wright-Fisher model: Drift as binomial sampling



Binomial distribution

population: allele frequency $p_t = \frac{i}{2N}$

probability for $p_{t+1} = \frac{j}{2N}$ in next generation:

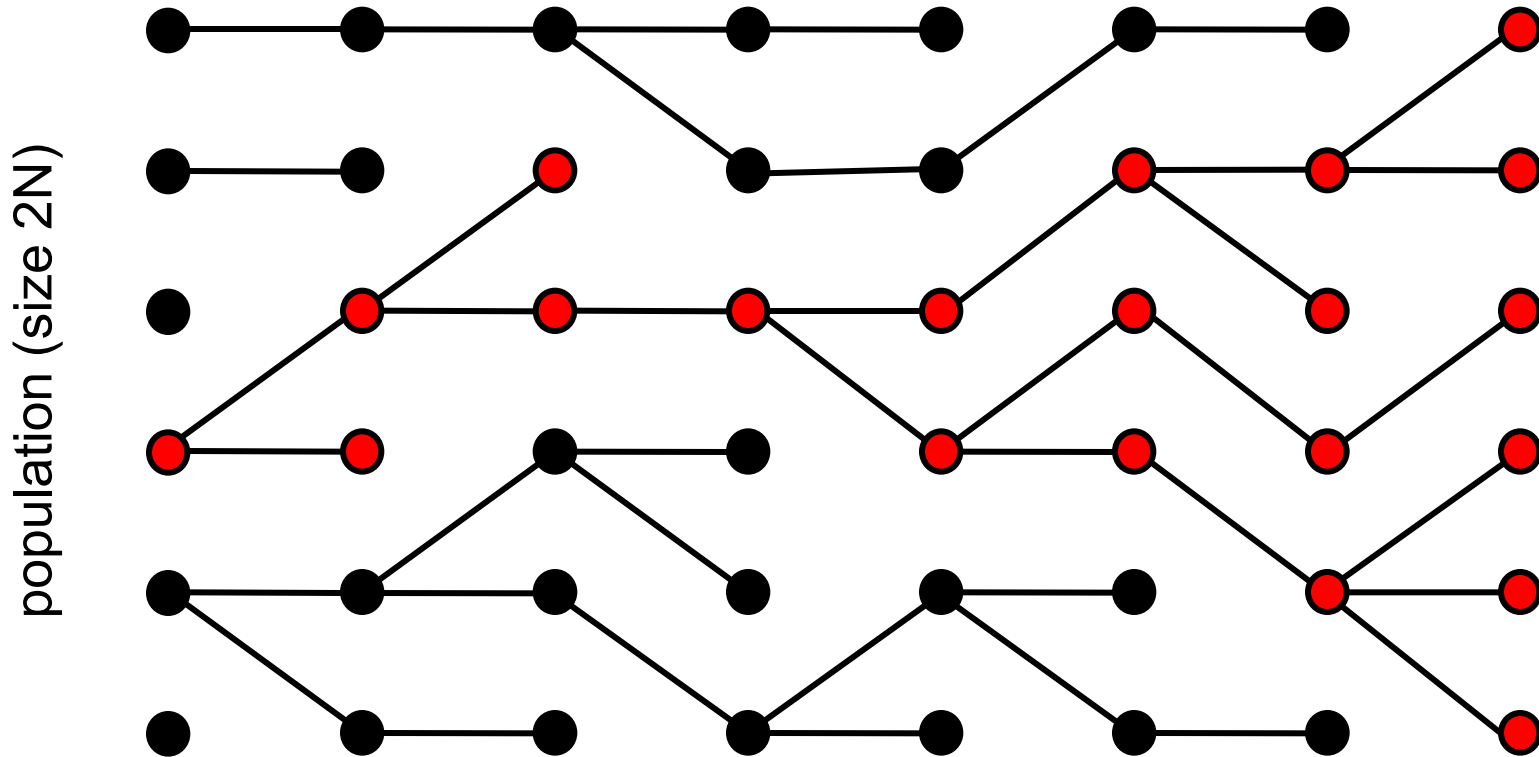
$$P_{ij} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}$$

mean: $E[p_{t+1}] = p_t$

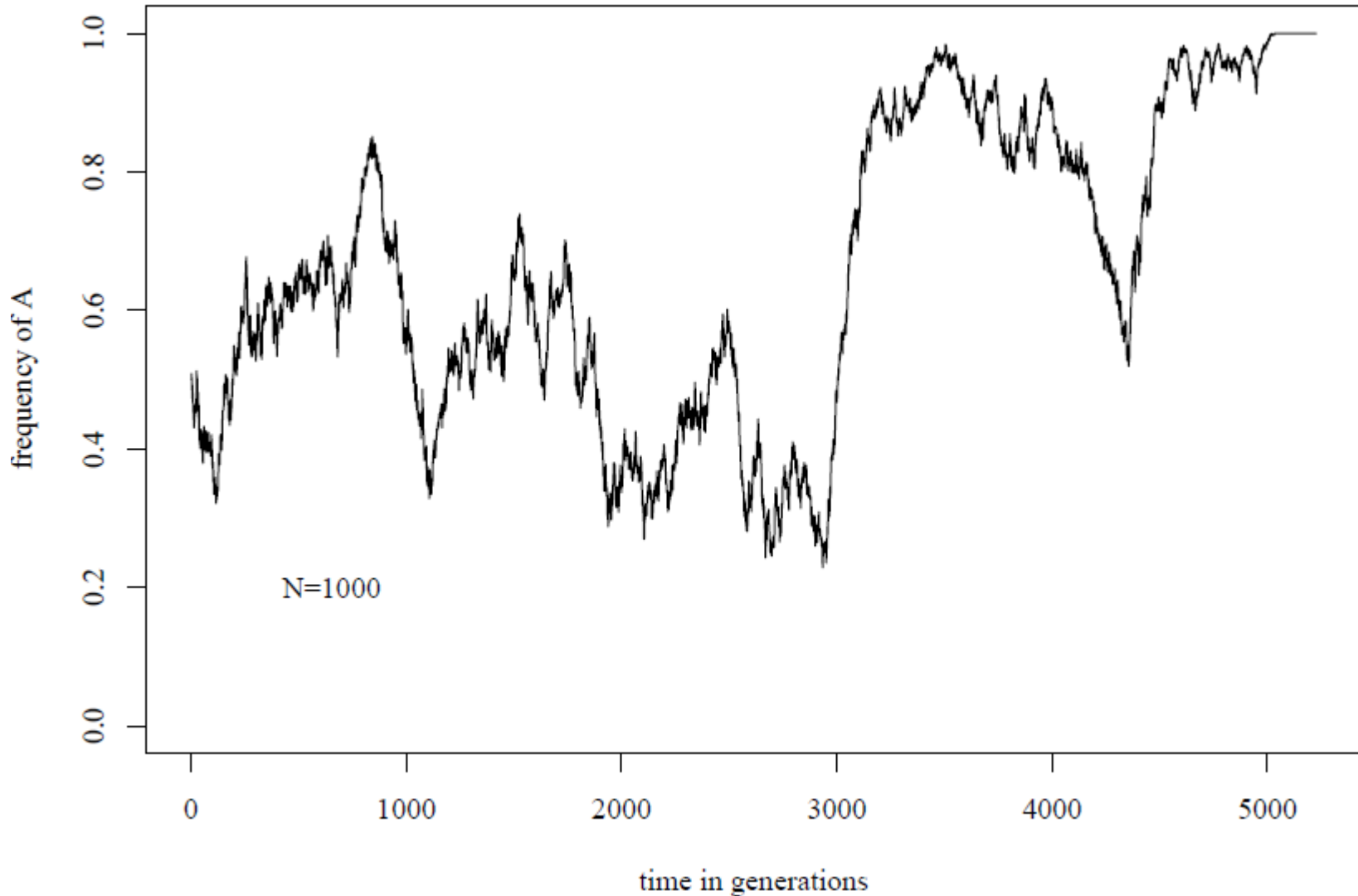
variance: $\text{Var}[p_{t+1}] = \frac{p_t(1 - p_t)}{2N}$

➤ drift stronger in small populations

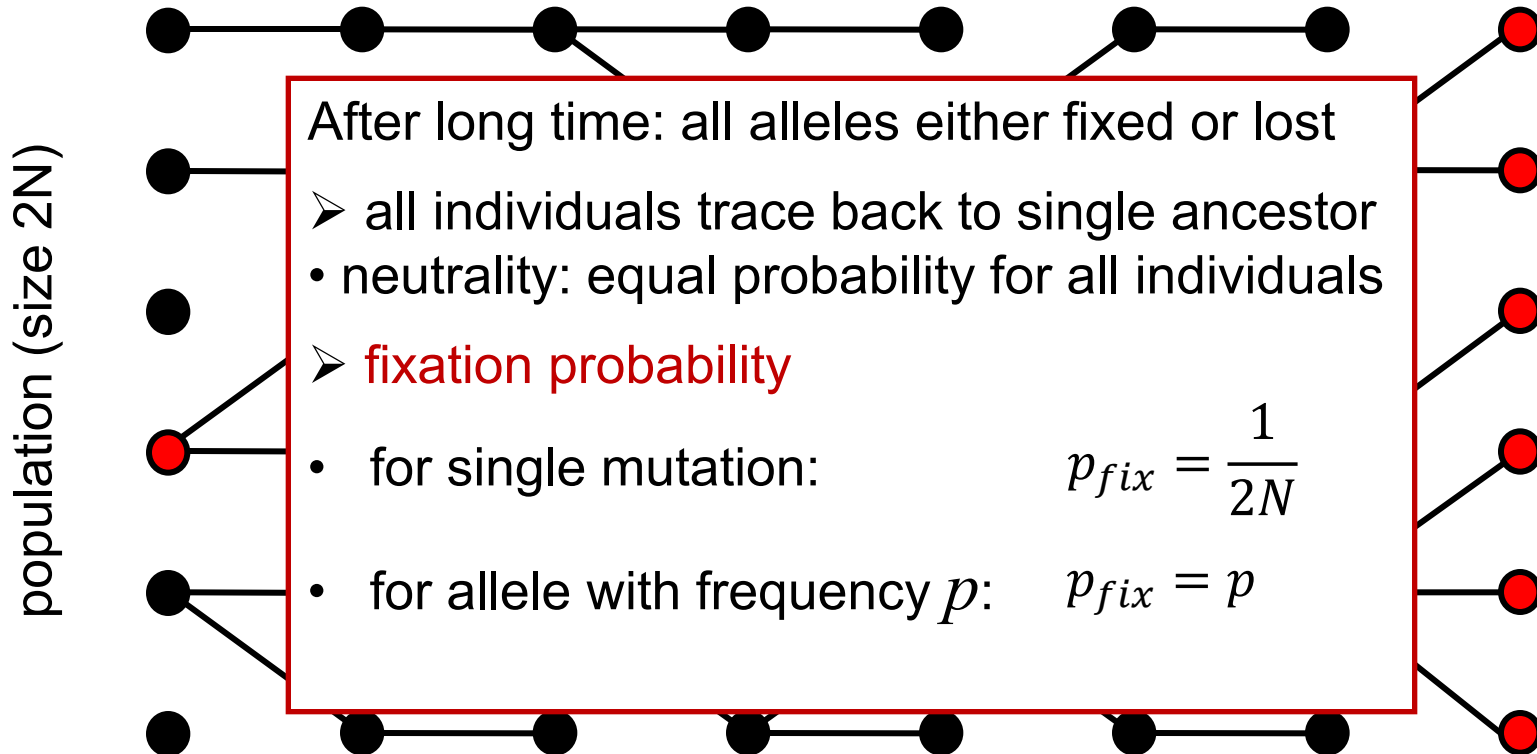
The Wright-Fisher model: Fixation of neutral alleles



The Wright-Fisher model: Fixation of neutral alleles

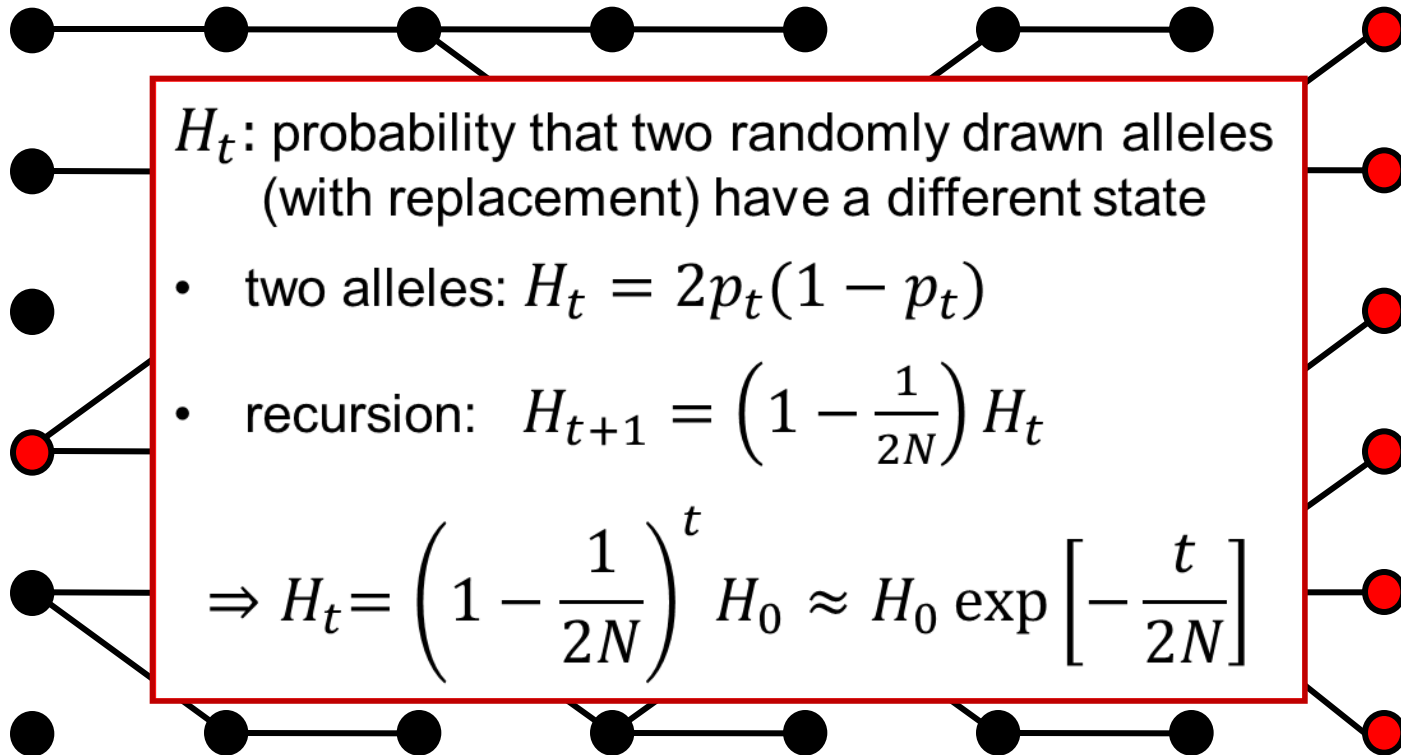


The Wright-Fisher model: Fixation of neutral alleles



The Wright-Fisher model: Heterozygosity

population (size $2N$)



The neutral theory

Drift and Mutation

Can drift (plus mutation) explain observed patterns of diversity ?

Motoo Kimura 1960's: Neutral theory of molecular evolution

The observed natural diversity at the molecular level is largely the result of neutral evolution: patterns of genetic variation and diversity are explained by drift and mutation



What neutral theory does **not** say:

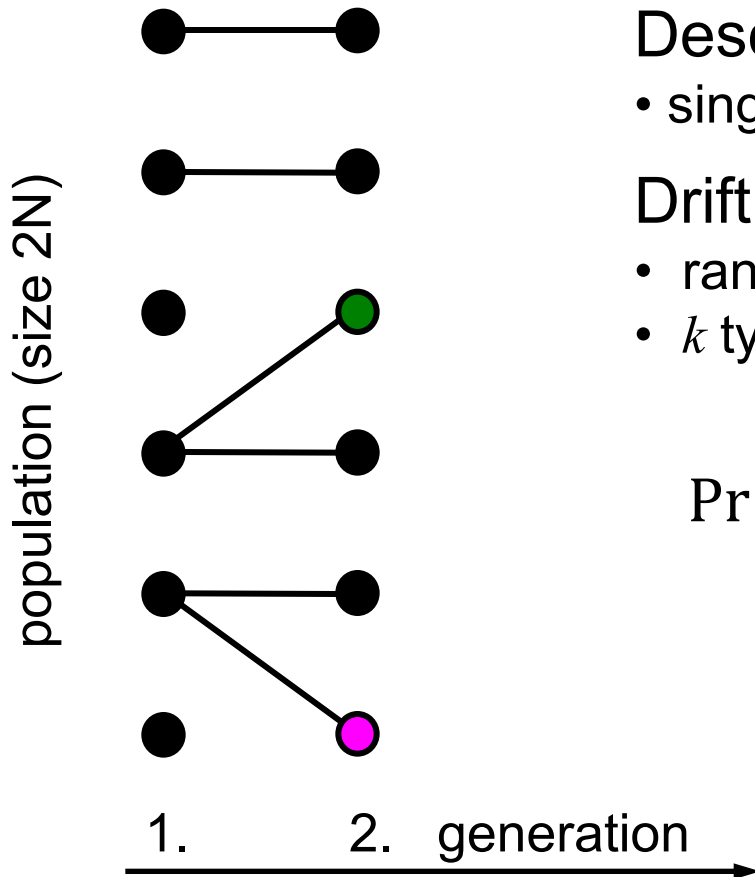
“Selection is not important for evolution”

- Purifying selection responsible for conserved DNA (eg in genes)
- Adaptive evolution due to positive selection
- **But:** adaptive diversity tiny relative to neutral diversity

Which patterns do we expect? – need a **model**

The neutral theory

Drift and Mutation



Description of neutral genetic variation

- single locus, multiple alleles

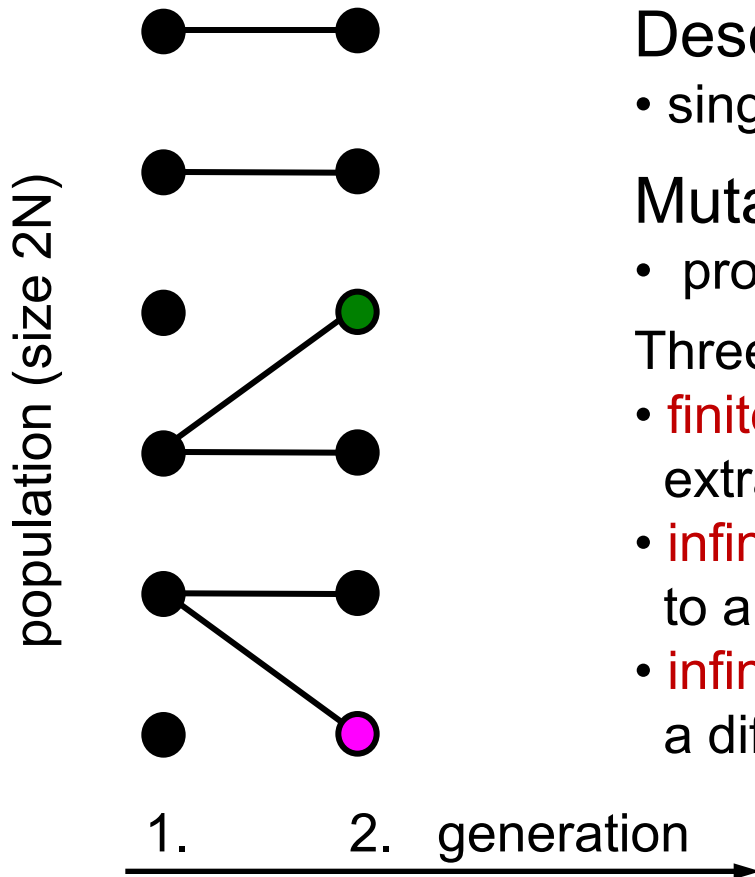
Drift:

- random sampling of parents
- k types: **multinomial** offspring distribution

$$\Pr[p'_1, \dots, p'_k] = \frac{N!}{\prod_i (p'_i N)!} \prod_i p_i^{p'_i N}$$

The neutral theory

Drift and Mutation



Description of neutral genetic variation

- single locus, multiple alleles

Mutation:

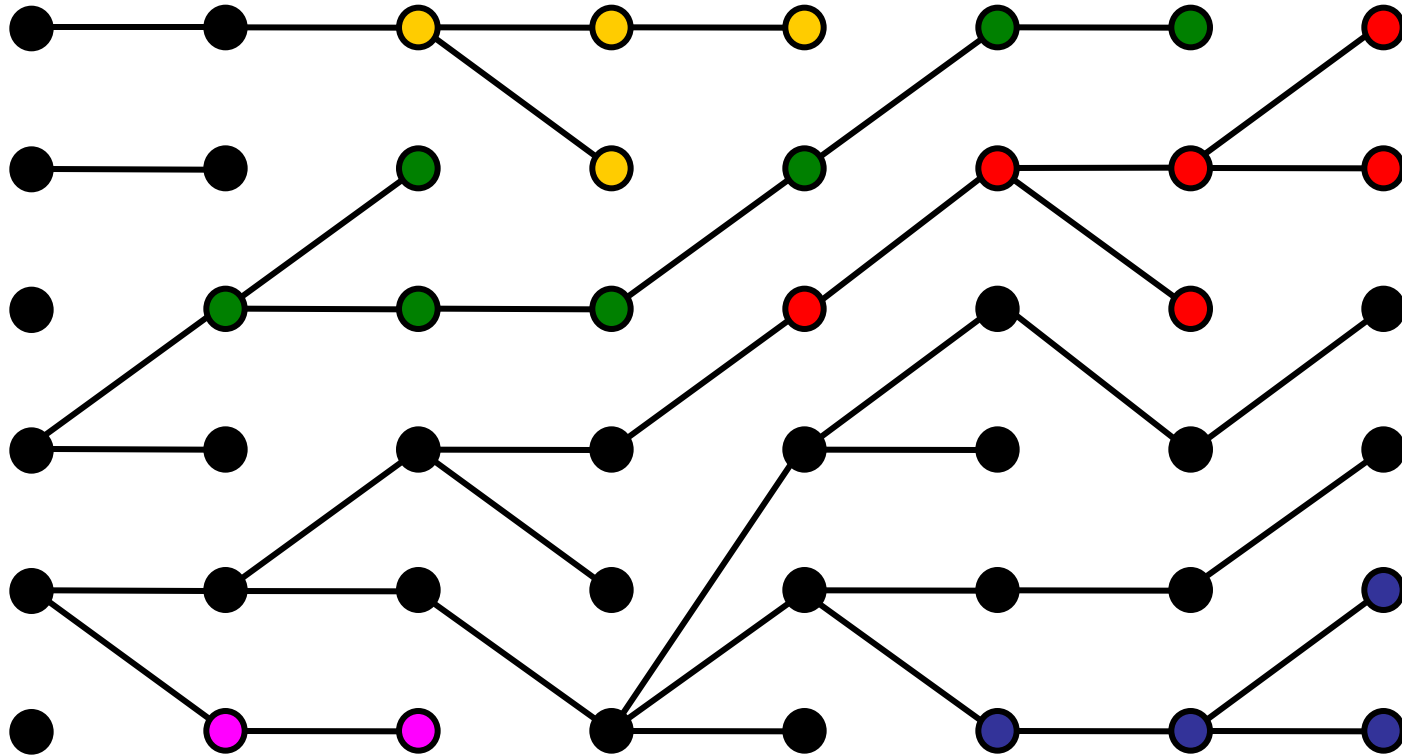
- probability u for each offspring

Three different mutation schemes:

- **finite alleles model**: like deterministic, include extra step in *infinite gamete pool*
- **infinite alleles model**: every mutation leads to a new allele (“new color”)
- **infinite sites model**: every mutation occurs at a different site and thus *remains visible*

The neutral theory

Drift and Mutation



The neutral theory

Drift and Mutation

1. Neutral variation: expected **heterozygosity** H / **nucleotide diversity** $E[\pi]$

➤ What is the probability that two randomly sampled alleles from a Wright-Fisher population have a different type?

- change in heterozygosity from generation $t \rightarrow t+1$

$$H_{t+1} = 2u + (1 - 2u) \left(1 - \frac{1}{2N}\right) H_t$$

- in equilibrium ($H = H_{t+1} = H_t$): $H \approx \frac{\theta}{\theta + 1}$, $\theta = 4Nu$

- nucleotide level ($\theta \ll 1$): $E[\pi] = H_{\text{nucleotide}} = \frac{\theta}{\theta + 1} \approx \theta$

The neutral theory

Drift and Mutation

1. Neutral variation: expected **heterozygosity** H / **nucleotide diversity** $E[\pi]$

$H \approx \theta = 4Nu$ should increase with mutation rate u and with pop. size N

➤ Although an increase is observed, there are strong deviations from the prediction of neutral theory:

E Coli	$H \approx 0.16$	$u \approx 10^{-10}$	$\rightarrow N \approx 10^8 ?$	$> 10^{10}$ in each human !
Drosophila	$H \approx 0.01$	$u \approx 3 \cdot 10^{-9}$	$\rightarrow N \approx 10^6 ?$	$> 10^{15}$ (?)
Homo	$H \approx 0.001$	$u \approx 3 \cdot 10^{-8}$	$\rightarrow N \approx 10^4 ?$	$\sim 10^9$

Reasons?

- population bottlenecks
- selection

The neutral theory

Drift and Mutation

1. Neutral divergence: **substitution rates**

➤ At which rate are neutral alleles substituted in a population?

• new mutational input per generation: $2Nu$

• fixation probability for each new mutant: $p_{fix} = \frac{1}{2N}$

➤ neutral substitution rate: $2Nu \cdot \frac{1}{2N} = u$

- **independent of population size !**
- basis for “molecular clock” estimates

The neutral theory

Drift and Mutation

Can drift (plus mutation) explain observed patterns of diversity ?

- Sparked the fierce adaptationist / neutralist debate
 - **Today**: selection seems to be very important even at the molecular level:
 - New mutations: many non-coding parts of the genome under selection (regulatory elements, etc)
 - Substitutions: large fractions seem to be driven by positive selection (> 50% in *Drosophila*)
 - **But**: neutral theory generally accepted null model of molecular evolution
- Foundation of **statistical genetics** as research field

