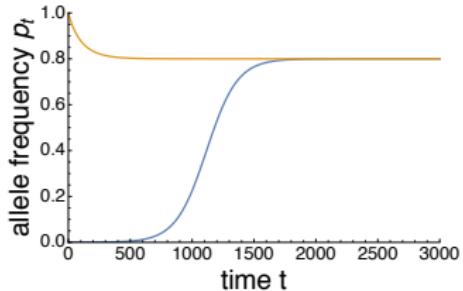
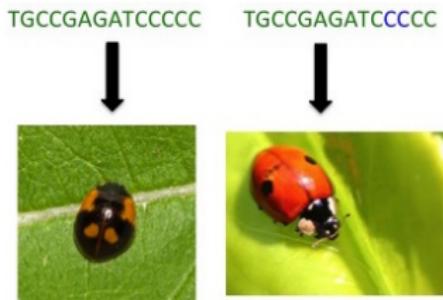


Random genetic drift

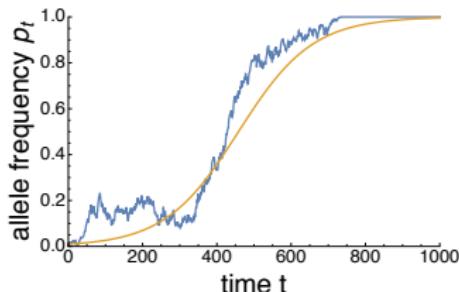
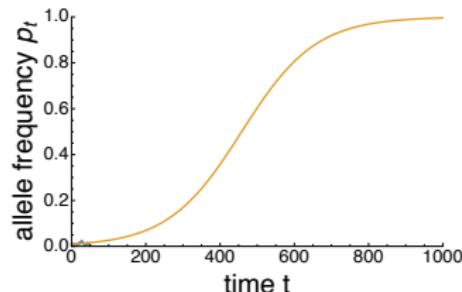
Selection and mutation as evolutionary forces



In the presence of selection and mutation, the evolutionary dynamics reaches an selection-mutation equilibrium in the long-term.

Deterministic and stochastic allele frequency change

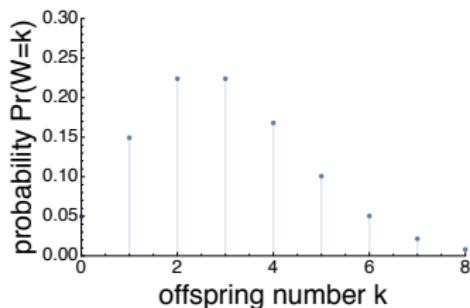
Evolutionary dynamics can be **deterministic** or **stochastic** (with random or chance components).



- There is a **variance** in allele frequency change under stochastic evolutionary dynamics
- This is so because allele frequency p is actually a **random variable**.

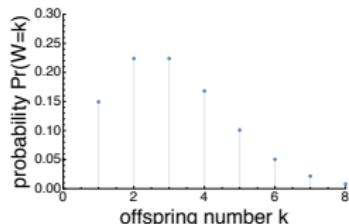
Fitness, fecundity, and survival are random variables

- The number of offspring $W \in \{0, 1, 2, \dots\}$ produced by an individual is in general a **random variable**.
- We have a probability $\Pr(W = k)$ that an individual produces k offspring.



Fitness, fecundity, and survival are random variables

$$\Pr(W = k) = \frac{e^{-w} w^k}{k!}$$



Poisson distribution with mean 3.

- $w = E_{\mathcal{P}}[W]$: mean (or expected) fitness of an individual given population state \mathcal{P} ¹.
- $\sigma^2 = \text{Var}_{\mathcal{P}}[W]$: variance in fitness of an individual.²

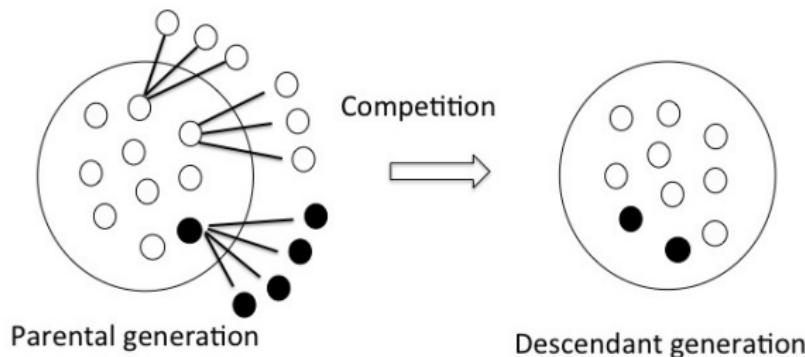
¹The expectation $E_{\mathcal{P}}[\cdot]$ has subscript \mathcal{P} , which means that we condition on the parental generation; that is, w is the average fitness of an individual given that we know the state of the parental generation (genotypes, phenotypes).

²For the Poisson distribution, the variance equals the mean $\text{Var}_{\mathcal{P}}[W] = w$.

Expected (or mean) fitness and variance in fitness

Mean fitness depends on the survival probability and effective fecundity:

$$w = E_{\mathcal{P}}[W] = s + f_e$$



There is also a variance $\sigma^2 = \text{Var}_{\mathcal{P}}[W]$ in the fitness of an individual.

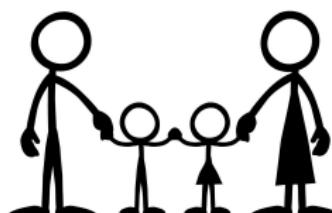
Two major sources of stochasticity effects (or random, or hazard, or chance effects)

- ① **Demographic stochasticity** refers to stochastic effects (chance effects) that affect each individual independently from each other, such as developmental noise or any source of variation specific to an individual.
- ② **Environmental stochasticity** refers to stochastic effects (chance effects) that affect the fitness of **several or all individuals** in the population simultaneously, such as fluctuations in resources, weather, diseases, or predation.

The consequence of both of these stochastic factors is that there will be a variance in fitness.

Demographic stochasticity

Family size is likely to be determined by demographic stochasticity.



- Local environmental effects affect individuals or families independently from effects on other families.
- Each individual is subjects to hazards that are independent of that experienced by others.
- The Poisson distribution is a good approximation to the family size distribution in humans.

Environmental stochasticity



- Fires, epidemics, environmental catastrophes affect many individuals simultaneously.
- This entails that there will be a covariance in the fitness of pairs, triplets, etc. of individuals.

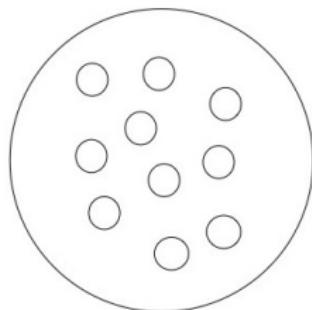
Consequence of demographic stochasticity

- ① Reproductive and survival variance (stochasticity) affects population dynamics.
- ② Reproductive and survival variance (stochasticity) affects evolutionary dynamics. It results in an additional fundamental evolutionary force, which is called **random genetic drift**.

We will investigate the effect of demographic stochasticity on allele frequency change, i.e., on the evolutionary dynamics.

Monomorphic population

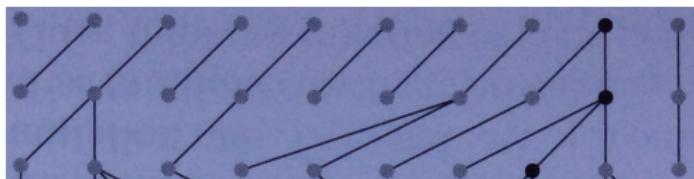
To understand genetic drift we start by assuming a monomorphic population and that the population size is constant (say it has reached a carrying capacity).



In the presence of stochasticity, a population is monomorphic if all individuals have the same fitness distribution. Then individuals are exchangeable.

The coalescence probability

- To understand genetic drift we start by introducing the **coalescence probability** Φ .
- This is the probability that two individuals randomly sampled in the offspring generation descend from the same individual in the parental generation.



In a population of constant size some individuals will reproduce more than others, just because of chance effects.

The coalescence probability

The coalescence probability is

$$\Phi = \frac{n\sigma^2}{n(n-1)} = \frac{\sigma^2}{n-1}$$

- $n\sigma^2$: number of ways of sampling two offspring from the same parent.
- $n(n-1)$: number of ways of sampling two offspring.



The coalescence probability

The simplest concrete example of a **coalescence probability** is

- A population of large and constant size n so that $w = 1$.³
- Offspring number follows a Poisson distribution so that $\sigma^2 \approx 1$ and $n - 1 \approx 1$.

Then

$$\Phi \approx \frac{1}{n}$$



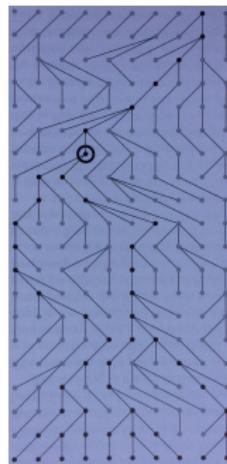
³This can be obtained if density-dependent competition is for space or we have a large carrying capacity.

Coalescence leads to identity-by-descent

Let us now trace back down the generations a line of genes (or lines of family names to take a non-genetic example).

Let q_t be the probability that two individuals randomly sampled in the population at time t have a common ancestor. This satisfies

$$q_t = \Phi + (1 - \Phi) q_{t-1}$$



In fact q_t is the probability that the two individuals belong to the same “dynasty” (or line of descent or lineage).

Coalescence leads to identity-by-descent

Let us assume that all individuals are initially different at time $t = 0$ so that they each carry a different allele (or a different family name), i.e., $q_0 = 0$. Then the probability that two individuals have a common ancestor has explicit solution

$$q_t = 1 - (1 - \Phi)^t$$

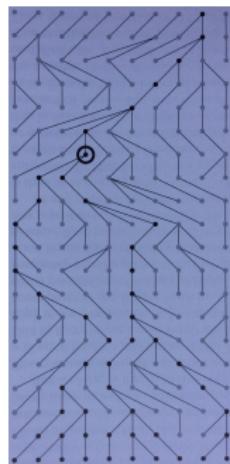
Since Φ is smaller than one, the only equilibrium $q^* = \lim_{t \rightarrow \infty} q_t$ satisfying the equation is

$$q^* = 1$$

Every individual has the same common ancestor.

Coalescence leads to identity-by-descent

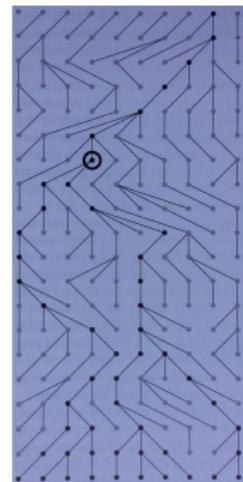
- The population becomes totally homogeneous!
- Hence if individuals have initially different alleles all diversity is lost!
- Hence if individuals have initially different family names all diversity is lost!



The process is entirely driven by chance effects.

Mitochondrial Eve: matrilineal **most recent common ancestor**

- We inherit **mitochondria** (organelle that supply chemical energy) only from Mom and the DNA does not recombine. Hence, we can think of mitochondrial DNA as a gene.
- Mitochondrial Eve: the most recent woman from whom all presently living humans inherit their mitochondrial DNA through Mom in an unbroken tree. This is **most recent common ancestor**.



Mitochondrial Eve lived between 200000 and 500000 years ago in Africa.⁴

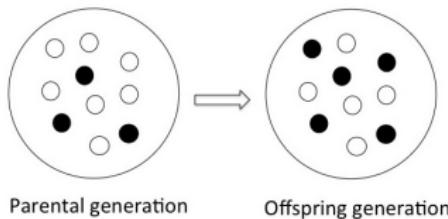
⁴Because genes tend to evolve independently, there is not a single individual that is the common ancestor to all individuals on this time scale.

We now look at the process from the perspective of allele frequency change

Our aim is to characterize

$$\Delta p = p' - p$$

the change in the frequency of allele A when the frequency p' in the offspring generation is considered as a random variable and can fluctuate owing to reproductive variance.



We focus on the change Δp given parental population state \mathcal{P} (thus knowing parental allele frequency p).

Stochastic allele frequency

The change in frequency of allele A can be written

$$\Delta p = \underbrace{\Delta p_{\text{deter}}}_{\text{"deterministic effect"}} + \underbrace{\Delta p_{\text{drift}}}_{\text{"stochastic effect"}}$$

① The term

$$\Delta p_{\text{deter}} = E_{\mathcal{P}}[p'] - p$$

is deterministic and is equivalent to effect of selection and mutation $\Delta p_{\text{deter}} = \Delta p_s + \Delta p_\mu$.

② The term

$$\Delta p_{\text{drift}} = p' - E_{\mathcal{P}}[p']$$

is stochastic—the **genetic drift** term—resulting in deviations around the mean in the offspring generation.

Stochastic allele frequency

The change in frequency of allele A is

$$\Delta p = \Delta p_{\text{deter}} + \Delta p_{\text{drift}}$$

On average the change due to drift is nil

$$E_{\mathcal{P}}[\Delta p_{\text{drift}}] = 0$$

and so on average the expected change in frequency is due to deterministic forces, mutation and selection

$$E_{\mathcal{P}}[\Delta p] = \Delta p_{\text{deter}} = \underbrace{\Delta p_s}_{\text{selection}} + \underbrace{\Delta p_\mu}_{\text{mutation}}$$

Stochastic allele frequency

$$\Delta p = \Delta p_{\text{deter}} + \Delta p_{\text{drift}}$$

But there is also a variance in the fluctuations

$$\text{Var}_{\mathcal{P}}[\Delta p] = \underbrace{\text{Var}_{\mathcal{P}}[\Delta p_{\text{deter}}]}_0 + \underbrace{\text{Var}_{\mathcal{P}}[\Delta p_{\text{drift}}]}_{\text{"stochastic effect"}}$$

which is the variance in allele frequency in the offspring generation given the parental population (since the variance of a constant, like an average, is zero):

$$\text{Var}_{\mathcal{P}}[\Delta p_{\text{drift}}] = \text{Var}_{\mathcal{P}}[p']$$

The square root $\sqrt{\text{Var}_{\mathcal{P}}[p']}$ quantifies the typical displacement around the mean due to stochastic effects.

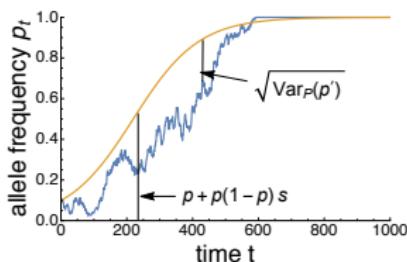
Stochastic allele frequency

Allele frequency change is thus typically of the form

$$\Delta p = \Delta p_{\text{deter}} \pm \sqrt{\text{Var}_P[p']}$$

"on average" "typical deviation"

As an example, here is the frequency p_t of A at time t under selection and drift is

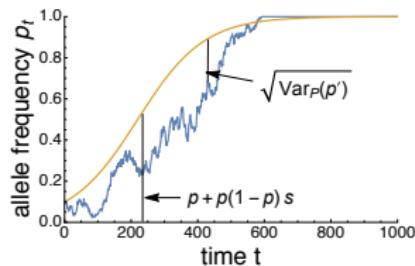


For the deterministic process (in yellow) we have $w_A = 1 + s$ and $w_B = 1$. Parameters $n = 10000$ and $s = 0.1$.

The coefficient of variation

From the expectation and variance we can compute the **coefficient of variation** in allele frequency

$$c_v = \frac{\sqrt{\text{Var}_{\mathcal{P}}[p']}}{E_{\mathcal{P}}[p']}$$



- The coefficient of variation measures fluctuations relative to the mean.⁵
- If the coefficient of variation is close to zero, then the variance does not affect much allele frequency dynamics that is close to that of the deterministic process.

⁵In the figure for the deterministic process (in yellow) we have $w_A = 1 + s$ and $w_B = 1$. Parameters $n = 10000$ and $s = 0.1$.

Genetic drift alone is an evolutionary force

In the absence of fitness difference (and mutation) the expected allele frequency in the offspring generation is that in the parental generation (no systematic force is affecting allele frequency)

$$\Delta p_{\text{deter}} = 0$$

On average there is no change in allele frequency

$$E_{\mathcal{P}}[\Delta p] = \underbrace{E_{\mathcal{P}}[\Delta p_{\text{deter}}]}_0 + \underbrace{E_{\mathcal{P}}[\Delta p_{\text{drift}}]}_0 = 0$$

This is called **neutral evolution**.

Genetic drift alone is an evolutionary force

In the absence of natural selection we have

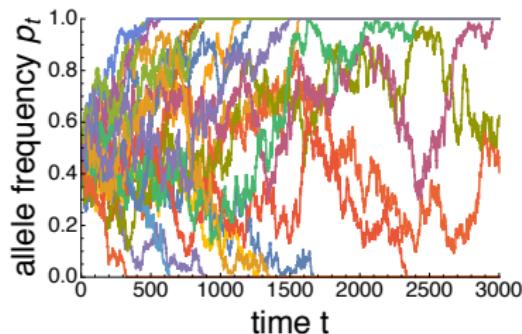
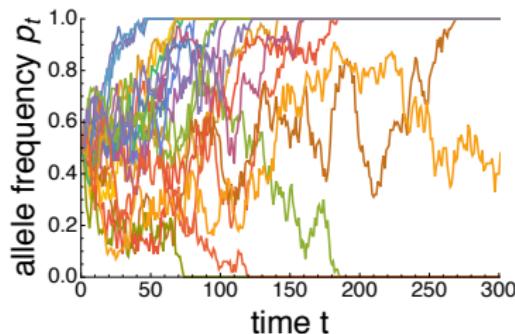
$$\Delta p \pm \underbrace{\Phi p(1-p)}_{\text{"typical deviation"}}$$

This is the variance $p(1-p)$ in A frequency among individuals in the parental generation times the coalescence probability Φ .

- Only if alleles descend from the same ancestor can there be fluctuations in allele frequency around the mean and thus $\text{Var}_{\mathcal{P}}[p'] > 0$.

Initial frequency of A given by $p_0 = 0.5$

Allele frequency p_t of A as a function of time t with a coalescence probability of $1/n$.

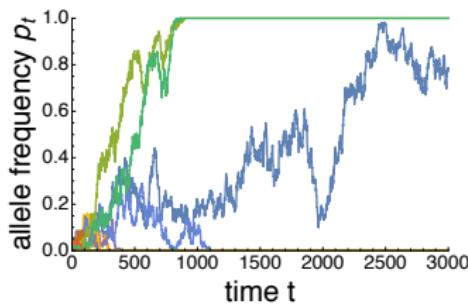
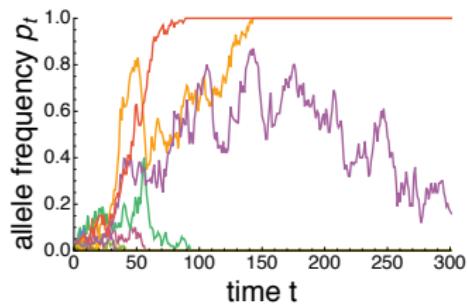


20 realization of the evolutionary process with $p_0 = 0.5$. Parameters: $n = 100$ (right panel) and $n = 1000$ (left panel) with reproduction following a Poisson distribution and keeping population size constant.

Approximately half the realizations result in the fixation of A. This is the **survival of the luckiest**.

Initial frequency of A given by $p_0 = 1/n$ (one A individual)

Allele frequency p_t of A as a function of time t (in the absence of selection and drift) with a coalescence probability of $1/n$.



Parameters: $n = 100$, $p_0 = 0.01$, and 100 realizations (left panel) and $n = 1000$, $p_0 = 0.001$, and 1000 realizations (right panel). Reproduction follows a Poisson distribution and keeping population size constant.

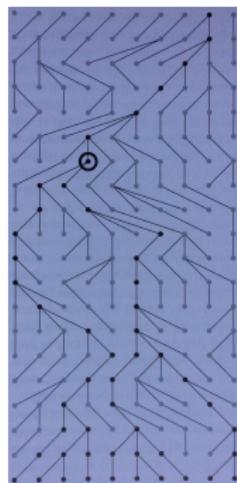
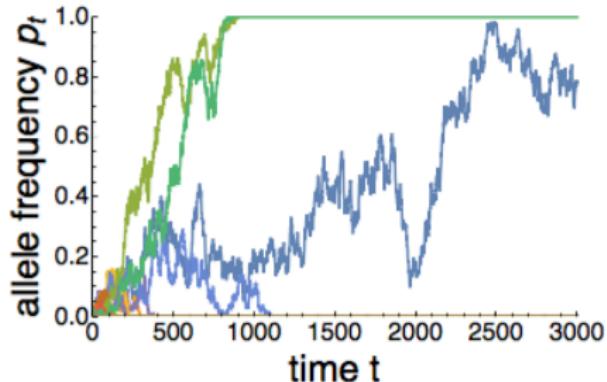
Few realizations results in the fixation of allele A, most go extinct.⁶

⁶Probability that A fixes is $1/n$ and the average time to fixation is about $2n$.



Genetic drift and common ancestry

It is genetic drift that drives identity-by-descent.



Each individual has a low probability of $1/n$ of becoming the common ancestor of the whole population but for sure one of them will become the common ancestor .

Very small fluctuations in large populations

The variance in allele frequency is

$$\text{Var}_{\mathcal{P}}[p'] \approx \frac{1}{n}$$

As population size becomes large, fluctuations vanish

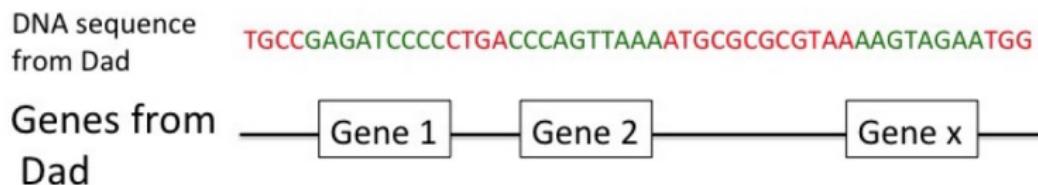
$$n \rightarrow \infty \qquad \frac{1}{n} \rightarrow 0 \qquad \text{Var}_{\mathcal{P}}[p'] \rightarrow 0$$

This is an example of the **law of large numbers**, which says that if the same (stochastic) experiment is repeated a large number of times (here the experiment is the reproduction of an individual), then the result should become close to the expected value for the experiment (here the expected reproduction).

Since every population is finite, genetic-drift will affect any population.

Neutral evolution

- Non-coding regions of DNA (non-functional) evolve neutrally.
- Because of redundancy of the genetic code, many nucleotides in the DNA have no phenotypic effect and evolve neutrally.
- Human ancestry (like Mitochondrial Eve) and that of other species is reconstructed by assuming neutral genetic markers (using the **molecular clock**).
- Some memes (cultural information) may evolve neutrally (e.g., the shape of pottery or its decoration).



Natural selection and genetic drift

In the presence of natural selection we have

$$\Delta p = \frac{p(1-p)s}{\bar{w}} \pm \underbrace{\sqrt{\text{Var}_{\mathcal{P}}[p']}}_{\text{"typical deviation"}}$$

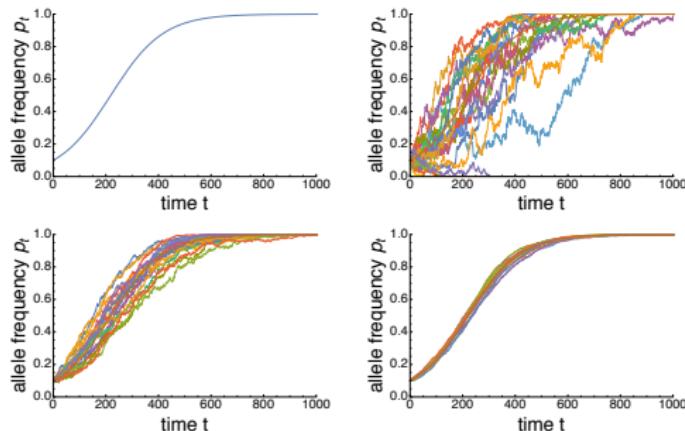
where

$$\text{Var}_{\mathcal{P}}[p'] \approx \Phi p(1-p)$$

which depends on the coalescence probability $\Phi \approx 1/n$.

Selection and genetic drift (high initial A frequency)

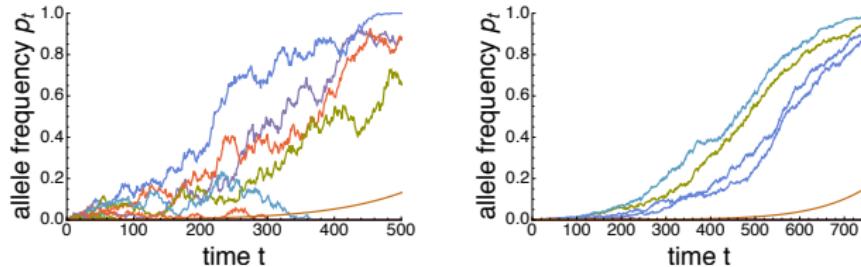
The larger population size the better the match to the deterministic process:



Parameters: $s = 0.01$, $p_0 = 0.1$ and deterministic process (north-left)
 $n = 1000$ (north-right), $n = 10000$ (south-left) and $n = 100000$ (south-right). Fitness is $w_A = 1 + s$ and $w_B = 1$; in the stochastic model reproduction follows a Poisson distribution with fixed population size.

Selection and genetic drift (one initial allele A)

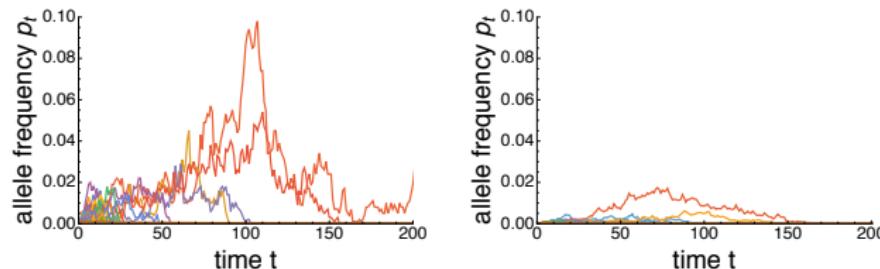
The spread of **beneficial alleles** ($s > 0$) that fix is accelerated but most of the time the beneficial allele actually gets extinct.



Parameters: $s = 0.01$, $p_0 = 1/n$, $n = 1000$ and 200 hundred runs (left panel), and $n = 10000$ and 750 runs (right panel). Fitness is $w_A = 1 + s$ and $w_B = 1$; in the stochastic model reproduction follows a Poisson distribution and keeping population size constant.

Selection and genetic drift (one initial allele A)

Deleterious alleles ($s < 0$) tend to increase in frequency but most of the time go extinct. In some cases though they can fixate.



Parameters: $s = -0.01$, $p_0 = 1/n$, $n = 1000$ and 200 hundred runs (left panel), and $n = 10000$ and 750 runs (right panel). Fitness is $w_A = 1 + s$ and $w_B = 1$; in the stochastic model reproduction follows a Poisson distribution.

Small isolated groups tend to acquire deleterious alleles by drift.⁷

⁷e.g., a certain number of disease-associated alleles are found predominantly Ashkenazi Jews, a Jewish diaspora population.

Selection and genetic drift: take home message 1

Suppose allele A is rare (say $p = 1/n$ so that there is only one initial copy arriving by mutation): then natural selection is a weak force. An allele that is favored by selection can easily go extinct owing to genetic drift.

- This is an example where the law of large numbers does not apply because the same experiment (here the reproduction of an individual carrying the A allele) is not much repeated, since only very few individuals carry allele A.
- But if there are constant mutations, eventually a beneficial mutation will appear that will spread (i.e., the experiment is repeated).

Hence, over the long-run mutations increasing survival and reproduction will enter the population and there has been plenty of time for evolution.

Selection and genetic drift: take home message 2

Suppose allele A has an appreciable frequency (say $p > 0.1$): then natural selection is a strong force and will generally dominate genetic drift in large populations. As a rule of thumb, the outcome of evolution will be mainly determined by natural selection if

$$s > 1/n$$

That is, if the selection coefficients exceed the coalescence probability.

- Now the law of large numbers does apply because the same experiment (here the reproduction of an individual carrying the A allele) is not repeated many times and dampens stochasticity.
- There is thus an interaction between natural selection and drift only for alleles in law number.

Summary so far: three fundamental evolutionary forces

$$\Delta p = \underbrace{\Delta p_s}_{\substack{\text{selection} \\ \text{deterministic}}} + \underbrace{\Delta p_\mu}_{\text{mutation}} + \underbrace{\Delta p_{\text{drift}}}_{\text{stochastic}}$$

- ① Natural selection.
- ② Mutation.
- ③ Random genetic drift.

At the ultimate level, every heritable part of phenotype at all level of biological organization in all species has been shaped by these three evolutionary forces.