

Physalia ML

Day4a: introduction to natural selection

Matteo Fumagalli

Intended Learning Outcomes

In this session you will learn

- to describe genetic drift,
 - to interpret the change of allele frequencies over time,
 - to appreciate the effect of population size on drift,
 - to define fitness and selection coefficient.
-

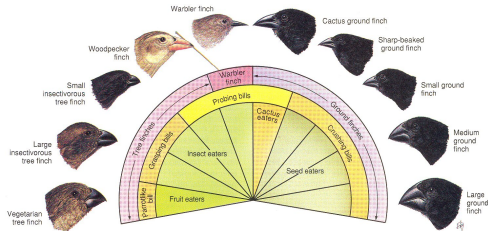
So far so good

- random mating
- genetic drift

What if different alleles affect survival?

Natural selection

- heritable traits that increase the **fitness** become more common in the population
- mutations evolve accordingly to their **effect on the fitness** of the carrier
- **functionality** is the prerequisite for selection to be effective



Fitness

We know that selection occurs because different individuals have different fitness, but what exactly do we mean by **fitness**?

Fitness

The word fitness in an evolutionary context can be defined as:

the expectation of the number of descendant genes at the same stage of the life cycle in the next generation.

In the next part, I will use *fitness* to indicate a property of genotypes, not of alleles or phenotypes.

Absolute and relative fitness

- **Absolute fitness:** measured as the change in abundance of a genotype from one generation to the next (assuming infinite population size, no mutation, and non-overlapping generations). In other words, the fittest genotypes will increase in abundance compared to less fit genotypes.

Absolute and relative fitness

- **Absolute fitness:** measured as the change in abundance of a genotype from one generation to the next (assuming infinite population size, no mutation, and non-overlapping generations). In other words, the fittest genotypes will increase in abundance compared to less fit genotypes.
- **Relative fitness:** calculated by dividing all fitness values by the largest value, meaning the fittest genotype always has a relative fitness of 1.

Fitness and selection

- Fitness is a property of a particular genotype.
- Selection is a process leading to different expectations of how gene copies are transmitted to the next generation.

If different individuals of a population have different fitness then we say that selection is operating.

If different individuals have the same fitness then we say that there is no selection, or equivalently, that the population is evolving **neutrally**.

Fitness and selection

We need to consider the **fitness** of each *genotype*.

For instance, for one locus with two alleles A and a in a diploid, each genotype has its fitness:

- ω_{AA} for genotype AA
- ω_{Aa} for genotype Aa
- ω_{aa} for genotype aa

Fitness and selection

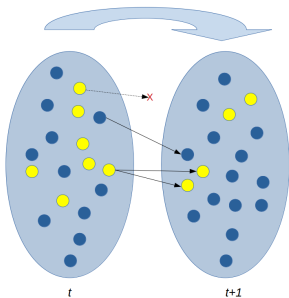
The strength of selection (or **selection coefficient**), often represented by the symbol s , is defined from the fitness values.

For example, if Aa is not the fittest genotype then the selection coefficient against heterozygotes can be interpreted as the deficit from a relative fitness of 1, so that

$$\omega_{Aa} = 1-s$$

What is the effect of selection on changes in allele frequency?

Selection and drift combined



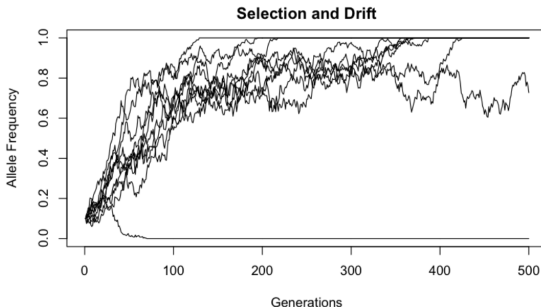
With drift only, all individuals had the same fitness.

The effect of high fitness is to make an individual more likely to be the parent of offspring in the next generation.

Nevertheless, it is still possible that a fit individual will have no offspring.

Selection and drift combined

If allele A belongs to a genotype with high fitness, its frequency will still drift but with a "tendency" towards fixation.



Is it still possible that the A allele is lost? : jupyter-notebook: selection (3) time: 12:20 - 10mins QandA

Intended Learning Outcomes

In this session you have learnt

- to describe genetic drift,
- to interpret the change of allele frequencies over time,
- to appreciate the effect of population size on drift,
- to define fitness and selection coefficient.

Intended Learning Outcomes

In this session you will learn

- to describe all different types of natural selection,
 - to appreciate the effect of novel mutations on allele frequency,
 - to understand the concept of gene flow.
-

So far we have learnt how random mating and genetic drift change allele frequencies in time.

What if

- different alleles affect survival?
- new mutations arise?
- migrants move between populations?

In other words, what is the effect of

- **selection**
- mutation
- gene flow

on the change of allele frequency?

Types of selection

If an allele is **dominant** then the heterozygote has the same *phenotype* as the homozygote for the dominant allele.



Types of selection

If an allele is **dominant** then the heterozygote has the same *phenotype* as the homozygote for the dominant allele.



If an allele is **recessive** then the heterozygote has the same *phenotype* as the other homozygote.



Types of selection

If A is **dominant** then the heterozygote has the same *fitness* as the homozygote for that allele.



Types of selection

If A is **dominant** then the heterozygote has the same *fitness* as the homozygote for that allele.



If an allele is **recessive** then the heterozygote has the same *fitness* as the other homozygote.



The **selection coefficient** s is defined as $\frac{\omega_a}{\omega_A} = 1 - s$ for haploids with alleles A and a , with A being dominant.

The **selection coefficient** s is defined as $\frac{\omega_a}{\omega_A} = 1 - s$ for haploids with alleles A and a , with A being dominant.

$f_A(t)$ depends on the product between s (selection coefficient, or "strength") and t (time).

If s is small, then the value of t to generate a certain change in f_A is *inversely* proportional to s .

If $s > 0$ then A -bearing individuals have the advantage, the opposite is true for $s < 0$.

jupyter-notebook: selection (1)

Additive selection

What if a recessive allele is not totally masked by the dominant allele?



If A is dominant but the traits have **additive** fitness then the heterozygote has a fitness value that is intermediate between the two homozygotes*.

* You do not need to know any more than that. This enters an area called *quantitative genetics* and there are entire courses on that topic.

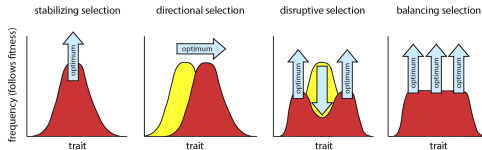
The types of selection just described can be defined as:

- additive selection: $\omega_{Aa} = 1 - s$; $\omega_{aa} = 1 - 2s$
- dominant advantageous allele: $\omega_{AA} = \omega_{Aa}$
- recessive advantageous allele: $\omega_{Aa} = \omega_{aa}$

with A being the advantageous allele.

If the selection coefficient does not change in time, we have several possible scenarios:

- **directional** selection (additive, dominant or recessive): one allele is favoured,
- **heterozygote advantage**: a special case of balancing selection where multiple alleles are maintained in the population,
- **heterozygote disadvantage**: a special case of disruptive selection where the extremes of a trait are favoured,
- stabilising selection: variation is reduced.



If the direction of selection changes over time, we have fluctuating selection.

Directional selection

A is the **advantageous allele** if $\omega_{aa} \leq \omega_{Aa} \leq \omega_{AA}$.

f_A will increase every generation and eventually reach 1 (i.e. **fixation** of A and **loss** of a).

The rate of change in allele frequency depends on s , selection coefficient.

Even small s can change allele frequency substantially over many generations.

Heterozygote advantage

If $\omega_{Aa} > \omega_{AA}, \omega_{aa}$ we define

$$\begin{aligned}\frac{\omega_{aa}}{\omega_{Aa}} &= 1 - s_{aa} \\ \frac{\omega_{AA}}{\omega_{Aa}} &= 1 - s_{AA}\end{aligned}$$

f_A will tend to the same value regardless of its initial frequency.
In fact, selection won't eliminate either allele (it is a special case of **balancing selection**).

Heterozygote disadvantage

If $\omega_{Aa} < \omega_{AA}, \omega_{aa}$, it results in the fixation of one of the two alleles (which one?).

It is an example of **disruptive selection** that removes low-frequency alleles.

We have learnt how **selection** changes allele frequencies in time. You also know how **genetic drift** changes allele frequencies in time.

What is the combined effect of selection and drift?

jupyter-notebook: selection (3)

For a population of size N , the fixation probability u of a new mutation with selection coefficient s can be defined as:

- strongly deleterious, if $2Ns \ll -1$ then $u \approx 0$
- nearly neutral, if $-1 < 2Ns < 1$ then $u \approx 1/(2N)$
- strongly advantageous, if $2Ns \gg 1$ then $u \approx 2s$

Strongly advantageous mutations are not necessarily fixed. Slightly deleterious alleles have a small but non-zero chance of being fixed.

Whether an allele is strongly selected or nearly neutral depends on both the selection coefficient and the population size.

What is the effect of

- selection
- **mutation**
- gene flow

on the change of allele frequency?

Mutations

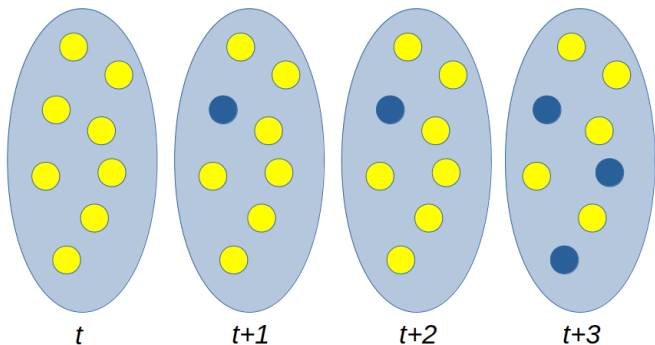
New mutations arise to produce new genetic variation that genetic drift can act on:

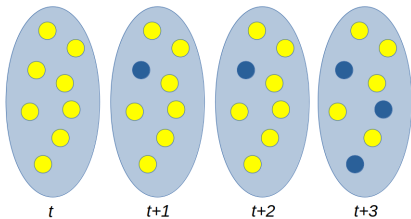
- deletions,
- insertions,
- translocations,
- point mutations.

Any of these mutations can be represented with a di-allelic model (e.g. presence/absence) if we assume that multiple mutations cannot occur in the same location*.

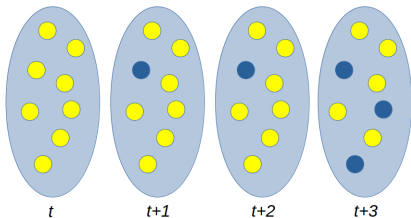
* This assumption is also called the *infinite site model*.

Assume that the yellow a allele in each individual randomly mutates to blue A with probability μ (called the **mutation rate**) in each generation.



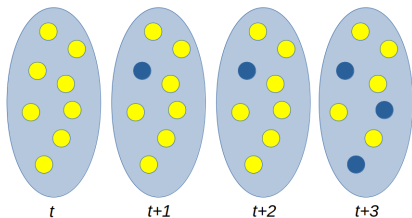


What is the expected allele frequency at the next generation given the current allele frequency and the mutation rate?

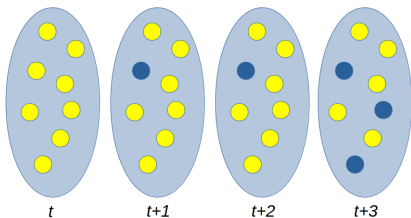


What is the expected allele frequency at the next generation given the current allele frequency and the mutation rate?

In the absence of other forces (e.g. genetic drift and selection), an equilibrium will be reached at:
mutation rate to blue / (mutation rate to blue + mutation rate to yellow)



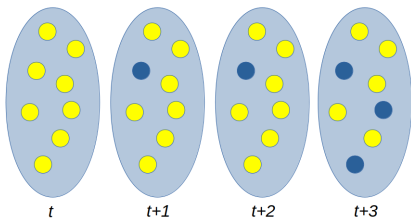
What is the expected allele frequency at the next generation $E[f_A(t+1)]$ given the current allele frequency $f_A(t)$ and the mutation rate μ ?



What is the expected allele frequency at the next generation $E[f_A(t+1)]$ given the current allele frequency $f_A(t)$ and the mutation rate μ ?

$$E[f_A(t+1)] = f_A(t) + \mu f_a(t) \quad (1)$$

Why?

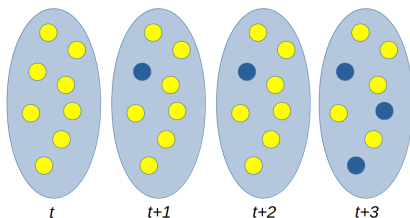


What is the expected allele frequency at the next generation $E[f_A(t+1)]$ given the current allele frequency $f_A(t)$ and the mutation rate μ ?

$$E[f_A(t+1)] = f_A(t) + \mu f_a(t) \quad (1)$$

Why?

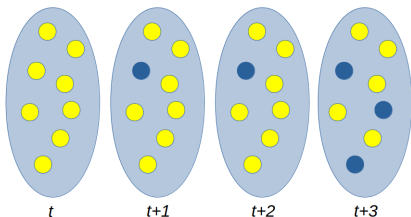
f_A will increase by the number of alleles a that mutate to A . This number is given by the rate μ times the frequency of a alleles to be mutated to A .



If mutations occur in both directions, e.g. mutations occur at rate $\mu_{a \rightarrow A}$ from a to A and $\mu_{A \rightarrow a}$ from A to a , then

$$E[f_A(t+1)] = (1 - \mu_{A \rightarrow a})f_A(t) + \mu_{a \rightarrow A}f_a(t) \quad (2)$$

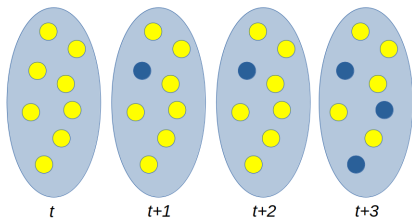
Why?



If mutations occur in both directions, e.g. mutations occur at rate $\mu_{a \rightarrow A}$ from a to A and $\mu_{A \rightarrow a}$ from A to a , then

$$E[f_A(t+1)] = (1 - \mu_{A \rightarrow a})f_A(t) + \mu_{a \rightarrow A}f_a(t) \quad (2)$$

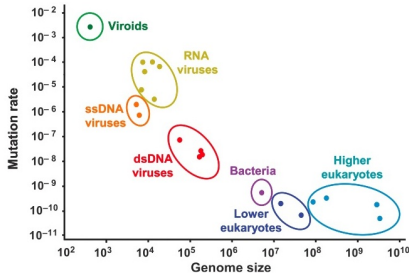
Why? f_A will increase by the number of alleles a that mutate to A but decrease by the number of alleles A that mutate to a .



In the absence of other forces (e.g. genetic drift and selection), we can show that an equilibrium will eventually be established at:

$$f_A = \frac{\mu_{a \rightarrow A}}{\mu_{a \rightarrow A} + \mu_{A \rightarrow a}} \quad (3)$$

Mutation rate



It is important to note that:

- mutation is a weak force in higher organisms,
- with no genetic drift (and selection), it takes a long time for the allele frequency to reach equilibrium,
- we can often ignore recurrent mutations.

What is the effect of

- selection
- mutation
- **gene flow**

on the change of allele frequency?

Population subdivision

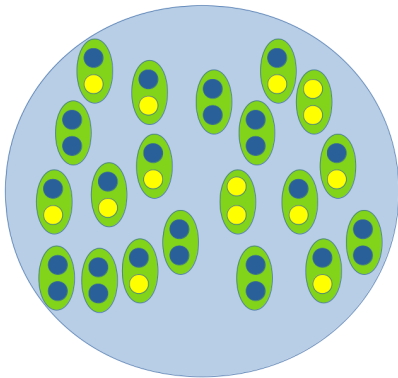
There is population subdivision, or **structure**, when the population is not randomly mating because of geographic or social structure.

Population subdivision is important to

- understand the effects of drift and natural selection,
- plan conservation strategies for rare or endangered species.

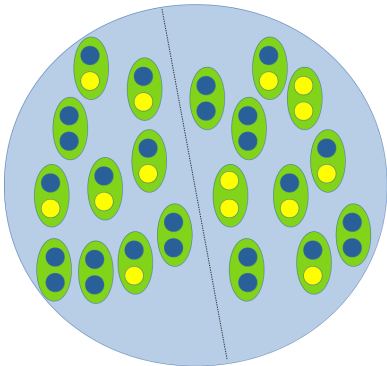
Population subdivision

Let's assume we have a population comprising of a certain number of individuals (i.e. diploid genotypes)



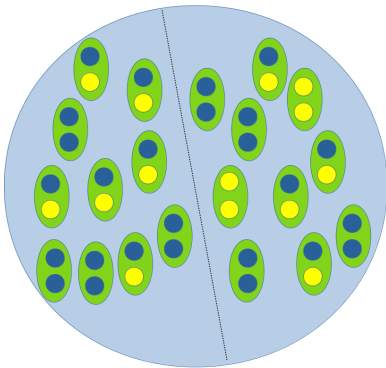
Population subdivision

At some point in time, a geographical/social barrier (dashed line in the figure) may prevent mating between individuals belonging to the two different groups/regions (left and right of the dashed line).



Population subdivision

The two populations will experience separate genetic drifts and their allele frequency will change accordingly.



What if some individuals can move from one population to another? What will the effect on the allele frequency be?

The model of genetic drift can be extended to include the effect of migration (and therefore of gene flow) and the allele frequency will change accordingly.

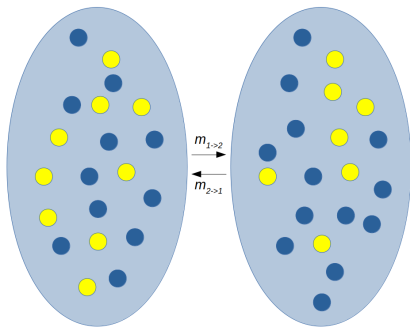
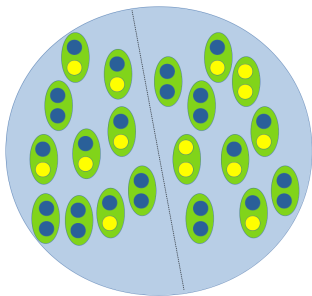


Figure 1: An individual from one population is replaced with an individual from the other with probability m (migration rate).

The model of genetic drift can include gene flow.



jupyter-notebook: subdivision

Intended Learning Outcomes

In this session you have learnt

- to describe all different types of natural selection,
 - to appreciate the effect of novel mutations on allele frequency,
 - to understand the concept of gene flow.
-