Summary of Introduction to Evolutionary Biology

v0.7

Gleb Ebert

November 8, 2017

Preface

This document aims to summarize the lecture Introduction to Evolutionary Biology as it was taught in the autumn semester of 2017. The focus lies on concepts and examples from the lecture are only included when deemed necessary. Unfortunately I can't guarantee that it is complete and free of errors. You can contact me under **glebert@student.ethz.ch** if you have any suggestions for improvement. The newest version of this summary can always be found here: `https://n.ethz.ch/~glebert/`

1 Introduction

Definition: Evolution means biological change over time
Technical basis: Phenotypes of individuals that are encoded by heritable genotypes vary in a population and their frequencies change

1.1 History

Aristotle: Ladder of nature / perfection
Carl von Linné: Systematic classification of life
James Hutton & Charles Lyell: Gradual long-term processes shaped earth (Uniformitarianism)
Jean-Baptiste de Lamarck: Inheritance of acquired characteristics (Lamarckian evolution)
Charles Darwin: Evolution is descent with modification and results in survival of the fittest

1.2 Microevolution

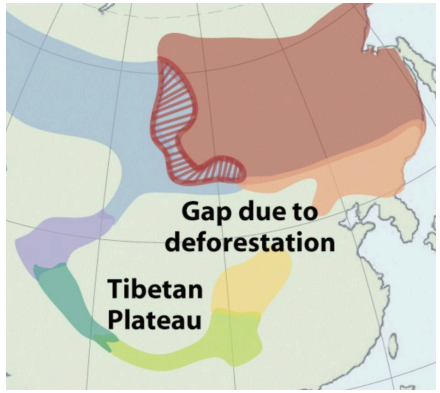
direct observation: small time-scales → short-term changes

Evidence of Microevolution:

1) Observation from natural populations
   - Bacterial adaptation to antibiotic stress
   - Soapberry bug adaptation to fruit
2) Observation from living anatomy
   Vestigal and rudimentary traits
   - Kiwi wings
   - Human coccyx (Steissbein)
   - Human arrector pili muscle (Haaraufrichter-Muskel)
   - (Appendix might be safe house for good gut bacteria)

1.3 Speciation

Speciation is the process that results in one species splitting into two or more. One example for this are ring species. They occur when one species spreads slowly around a geographical area to which they don't spread. By the time they meet up again the differences between the populations are too big to interbreed. One example for this is the Siberian Greenish Warbler (the hatched area is where no interbreeding occurs with two populations present).



1.4 Macroevolution

indirect observation: long time-scales → long-term changes

Evidence of Macroevolution:

1) Successions & Extinctions
   Law of succession: pattern of correspondence between fossile and recent forms from the same locale
   Comparative anatomy: Georges Cuvier argued that certain species are extinct. Recent macrofauna is only a fraction of all that ever existed
2) Transitional forms
   Darwinian evolution predicts intermediate forms between a species and its ancestor (e.g. *Microraptor gui* and *Archaeopteryx* between dinosaurs and modern birds
3) Homologies (Owen: "the same organ in different animals under every variety of form and function") can be found through comparative anatomy and comparative embryology. The similarity is due to inheritance from a common ancestor. They are phenotypically and genetically defined and enable the use of model organisms.
   Some molecular homologies are
   - the universal genetic code: bases and codons
   - the small-subunit (SSU) ribisomal RNA genes

2 Natural selection

Natural selection is the process underlying adaptive evolution

2.1 Darwins postulates of evolutionary change

Evolutionary change over time is a deductive implication of four postulates.

1) All populations contain variable individuals
2) Variation among individuals is, at least in part, heritable
3) Some individuals are more successful at surviving and reproducing than others
4) Survival and reproduction of individuals are not random; but individuals with the most favorable variation given the environment are those better at surviving.

These postulates can be tested in real world populations (e.g. Darwin's finches from the Galapagos islands). One has to be careful to not misinterpret biasing factors. For example heritability measures can be skewed by misidentified paternity, misidentified maternity, food quality or maternal effects such as egg quality.

2.2 Definition

Natural selection acts on individuals (more specifically, phenotypes), its consequences occur in populations as allele frequency changes.

2.3 Darwinian vs. Lamarckian evolution

Different processes proposed for the same pattern.

1.	No initial variation	Initial variation
2.	Individuals adapt during their lifetime	Selection acts on individuals
3.	Inheritance of aquired changes/characters (IAC)	Inheritance of surviving alleles if environment leads to adaptation
	⇒ individuals and populations evolve	⇒ populations evolve

Epigenetic modifications are attached to the genetic code and can be passed on to the offspring (up to two generations). Thus their evolutionary relevance is short-term.

2.4 Limits

- Natural selection acts on existing traits
- Natural adaptation does not lead to perfection
- Natural selection is non-random, but it is not progressive
- Natural selection is blind to the future, but tells us tales from the past

2.5 "Perfection" in nature

William Paley argued, that the vertebrate eye is too perfect and complicated to have resulted from natural processes. Thus it must be a creation of a conscious designer. By looking closely at eyes from various chordates one can see that there is a lot of variation in the complexity of the eye. Thus Paleys argument is wrong.

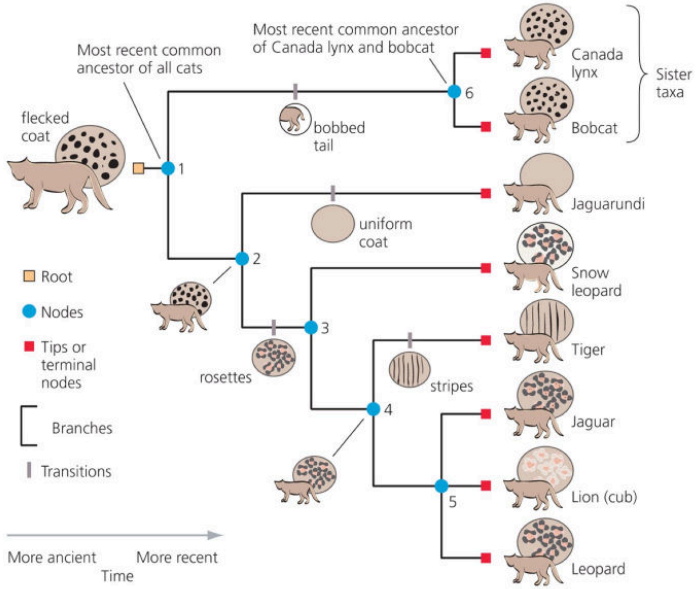
3 Phylogenetics

3.1 Evolutionary trees

Evolutionary trees aim to display the course of speciation over time and the relation between species. They start with one common ancestor (the root) and end with multiple species (tips or terminal nodes). Dichotomies (splits of branches into two) in between are called nodes. If phylogeny can't be resolved to dichotomies branches split into three or more subbranches. This phenomenon is called polytomy. Branches can have transitions where a specific characteristic evolved. The first group or species to split of is called basal.

Trees come in many forms depending on the author and on the kind of data they aim to display. They can be left-right, top-down or even circular.

One should not forget that evolutionary trees are merely hypotheses.



3.2 Phylogenetic inference

- Plesiomorphy: A characteristic that is shared between a species and its ancestor
- Apomorphy: A characteristic that is different from an ancestor
- Synapomorphy: Apomorphic and shared between multiple sister taxa. Also called homology
- Autapomorphy: Apomorphic and different from sister taxa
- Monophyletic group: All descendants of one ancestor (at least two taxa). Also called clade.
- Paraphyletic group: A subset of descendants of one ancestor

Classical taxonomic groups are not necessarily monophyletic (e.g. prokaryotes, dicots and fish).

3.3 Tree reconstruction

Pitfalls of trees inferred form phenotypes alone:

- Phenotypes are influenced by genotypes and environment
- Only genotypes are heritable
- Phenotypic similarity due to convergence (analogy / homoplasy; e.g. camera eye in mollusks and teleosts)

Combination of phenotypic and genotypic data (molecular markers) leads to better results. When using molecular markers each nucleotide is seen as an independent character.

If multiple possible trees exist, one can assume that the most parsimonious (least evolutionary steps) is the correct one. If multiple equally parsimonious trees turn up one estimates uncertainty. This is done by bootstrapping which is the generation of data sets made up from the original data set. Data points can be repeated multiple times or be absent altogether. When analyzing these data sets one gets a phylogenetic tree for each. The most likely tree will be the one that comes out the most after the analysis. The certainty of a certain clade is placed at its node and given in percent. This number is the percentage of the replicates in which that particular clade appeared and is also called the bootstrap support of the clade. High bootstrap support means that the clade is a winner across our artificial data set.

Reversals or "back-mutations" can remove synapomorphies.

3.4 Answering questions with phylogeny

The following examples illustrate some of the problems that phylogeny can help to solve.

- By looking at when body lice evolved from hair

lice one can estimate when humans started wearing clothes (around 107,000 years ago).
- Forensic scientists were able to conclude which patients got HIV from their dentist and which got it else where.
- In his E. coli long-term experimental evolution project (LTEE) Richard Lenski was able to show that the mutation rate in his bacteria evolved.

4 Science history

4.1 Ancient Greece

- Hippocrates: Importance of good notes and categorization
- Socrates: Dialectic inquiry (finding an answer by initiating a group discussion with questions)
- Plato: Ideal spiritual forms of real world things
- Aristotle: Emphasis on empirical things (Scala Natura)

4.2 Deduction and Induction

Induction infers generalizations base on individual instances. Reasoning in which the premises of an argument are considered to support its conclusion but do not guarantee its truth.
Deduction is the process of deriving consequences of what is assumed. Given the truth of the assumptions, a valid deduction guarantees the truth of the conclusion.

4.3 Aristotle

- pupil of Plato
- big influence on future course of western science through founding of fields of logic, biology and psychology
- generated one of the most impressive systems of thought
- heavy emphasis on rules of deductive logic
- lack of rigorous experimental methods
- four causes
  1) Material cause
  2) Formal cause
  3) Efficient cause: like modern understanding of cause-effect relation
  4) Final cause: the purpose for which a thing exists of an action is done. The future end something is supposed to serve

5 Co-option

Co-option (or exaptation) is the process by which a structure or system with an original function adds or changes to a new function.

5.1 Myxococcus xanthus

Myxococcus xanthus is a social soil bacteria. It can secrete antibiotics and lytic enzymes to externally digest other microbes. These bacteria show social behavior in the form of "wolf pack hunting" and sporulating through fruiting bodies (only a minority of the colony becomes a spore).
Myxococcus shows two types of motility: S-motility drives swarming on soft agar surfaces, while A-motility contributes to swarming on relatively firm agar surfaces. S-motility requires pili and fibrils. If one knocks out pilin production the bacteria can regain the ability to swarm on firm agar. This is called evolved cooperative motility (ECM). Through various test it was shown, that instead of redeveloping pili, Myxococcus adapted A-motility to regain the ability to make fibrils.

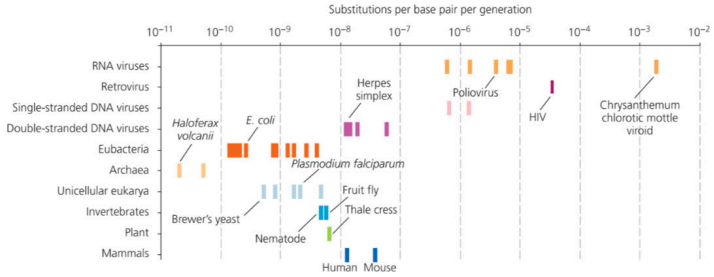
6 Sources of phenotypic variation

6.1 Genetic mutation

genetic differences between individuals cause phenotypic variation

6.1.1 Mutation rates

Base-pair matching during synthesis leaves only one unrepaired mismatch every 100 million times. This results in around 38 new mutations per human gamete or 76 per baby. Also a replacement generation of the human population would average 83 mutations per bp site.
The base substitution rate in various organisms varies by several orders of magnitude.



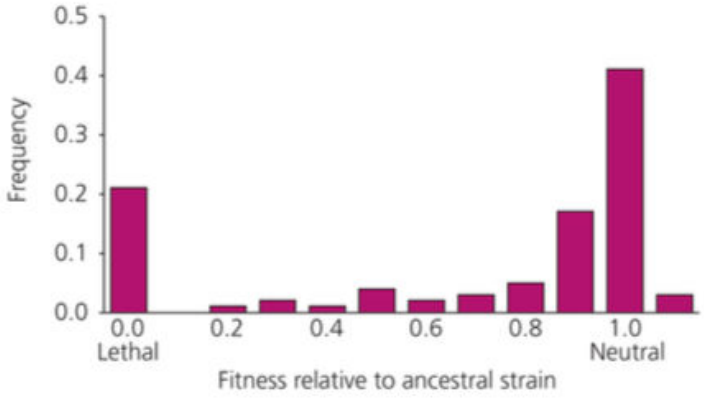
Mutation rates can be calculated from Mutation Accumulation Lines:

1) maintain 25 replicate lineages of one common ancestor
2) grow for 500 generations while choosing randomly which individual is transferred (no selection); in control lineages transfer many individuals per generation (selection) ⇒ mutations accumulate and reduce fitness
3) sequence one chromosome from ancestor and from individual taken from each replicate lineage
4) average mutation rate per bp per generation

$$= \frac{\text{total \# of mutations}}{\text{\# of bps} * \text{\# of generations} * \text{\# of lineages}}$$

6.1.2 Distribution of mutation fitness effects

Most mutations are neutral or negative in their effect on fitness. A frequency-fitness diagram might look like this:



6.2 Environmental variation

environmental variation in space/time cause an individual organism's phenotype to vary

6.3 Genotype x environment variation

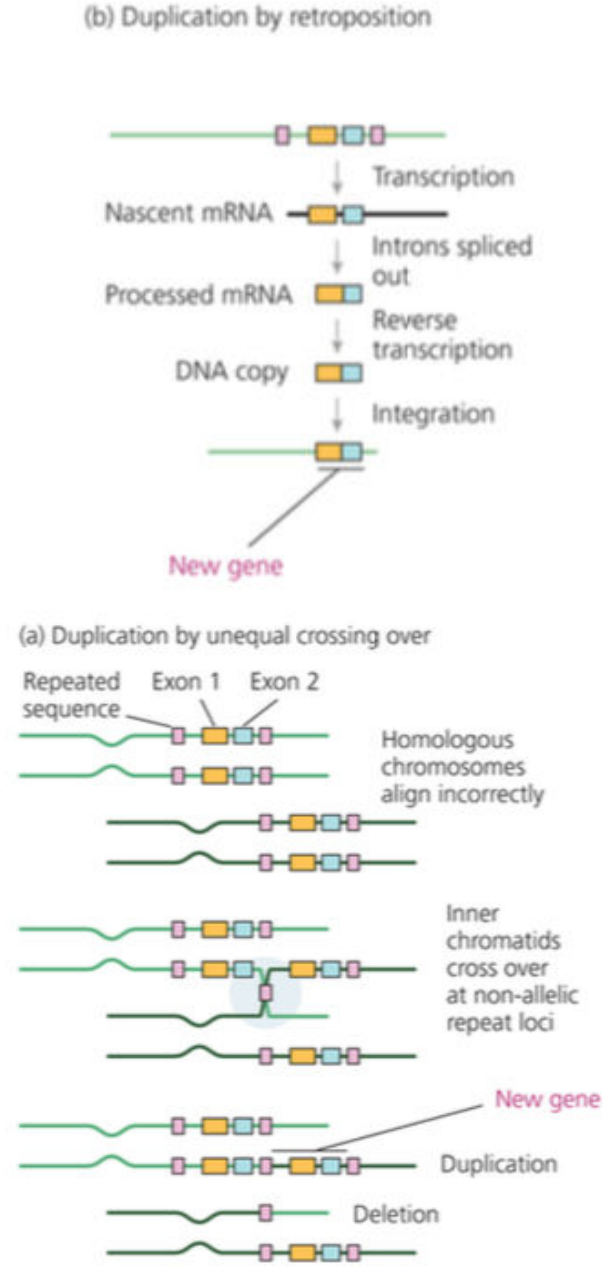
genetic differences cause distinct individuals to respond to environmental variation differently

6.4 vocabulary to know

complementary base pairs	neutral alleles
point mutation	gene duplication
transition	unequal crossing over
transversion	pseudogenes
replacement (non-synonymous) substitutions	paralogs, paralogous
silent site (synonymous) substitutions	orthologs, orthologous
loss-of-function mutations	linkage
indels	polymorphism
selection coefficient	frequency
polyploidy	chromosome inversion
genome duplication	

6.5 New genes

New gene copies can arise either by retro-transposition, which results in genes without introns or regulatory regions in eukaryotes, or by un-equal crossing over during meiosis, which results in new copies with introns and regulatory regions.



By contrasting the number of duplicate genes vs. age of several species Michael Lynch arrived at an estimated duplication rate of 0.01 per gene per million years. Considering that most organisms have a lot of genes a big amount of new genes arise every 1 million years. These new gene copies can have different fates:

- lost
- deactivated and regained as pseudogene
- diverge to serve function distinct from parent gene

6.6 Inversions

An inversion is an event when a piece of a chromosome is flipped around and refuse the other way around. They can enforce allele linkage due to lack of crossing over. Thus inversions that lock in advantageous allele combinations can be maintained by selection.

6.7 Whole genome duplication

Whole genome duplications are most commonly found in self-fertilizers. They result in a massive new input of genetic material that can evolve.

6.8 More impactful mutations

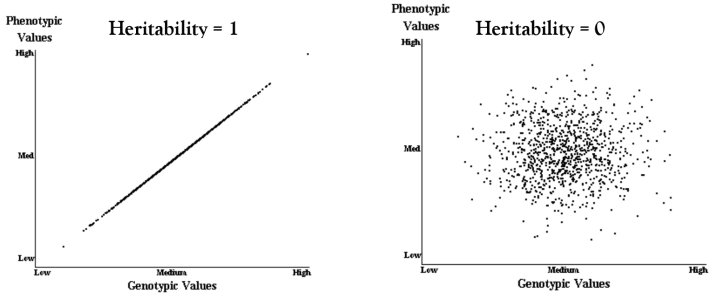
Name	Description	Mechanism	Significance
Point mutation	Base pair substitutions in DNA sequences	Chance errors during DNA synthesis or during repair of damaged DNA	Creates new alleles
Chromosome inversion	Flipping of a chromosome segment, so order of genes along the chromosome changes	Breaks in DNA caused by radiation or other insults	Alleles inside the inversion are likely to be transmitted together, as a unit
Gene duplication	Duplication of a short stretch of DNA, creating an extra copy of the sequence	Unequal crossing-over during meiosis or retrotransposition	Redundant new genes may acquire new functions, by mutation
Genome duplication	Addition of a complete set of chromosomes	Errors in meiosis or (in plants) mitosis	May create new species; massive gene duplication

7 Population Genetics

Population genetics integrate Darwin's theory of evolution by natural selection and the Mendelian laws of inheritance. The field seeks to document and explain changes in allele, genotype and phenotype frequencies. It offers a quantifiable definition of evolution: It happens by allele frequency changes in the population.

7.1 Heritability

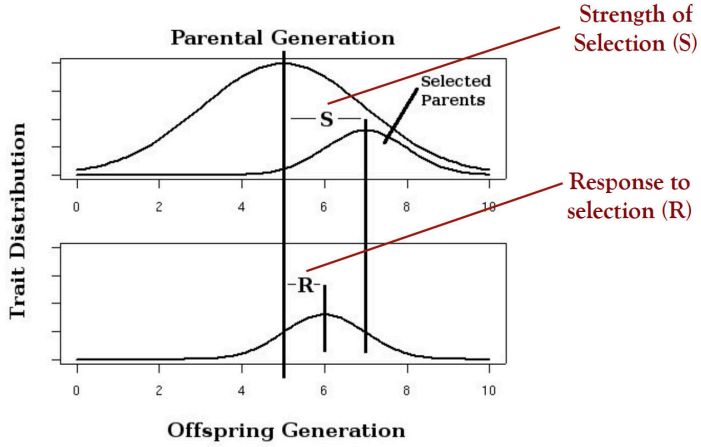
$$\text{Heritability} = \frac{\text{Variance (Genotype)}}{\text{Variance (Phenotype)}}; \quad h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$



Influence of Environment → 0

Influence of Environment → 1

7.1.1 Selection Experiments



H = R/S

7.2 Calculating Genotype and Allele Frequencies

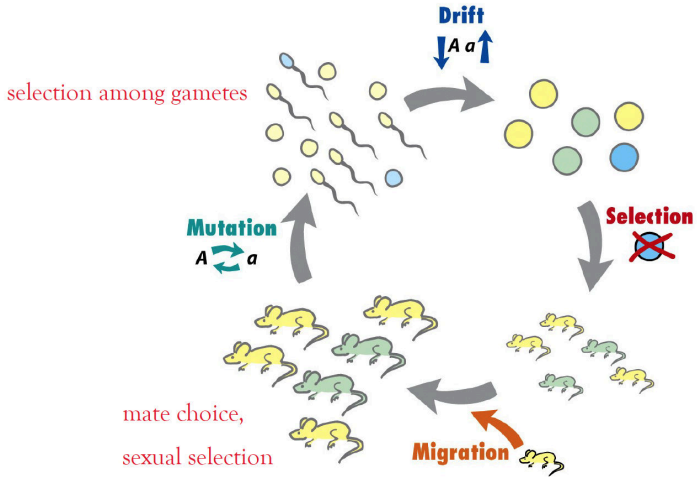
	AA	Aa	aa
f(Genotyp)	$\frac{36}{100}$	$\frac{48}{100}$	$\frac{16}{100}$

$$f(A) = \frac{36+36+48}{200} = 0.6; \quad f(a) = \frac{48+16+16}{200} = 0.4$$

$$f(A) + f(a) = 1$$

7.3 Evolution of Populations

The forces causing evolution in real populations:

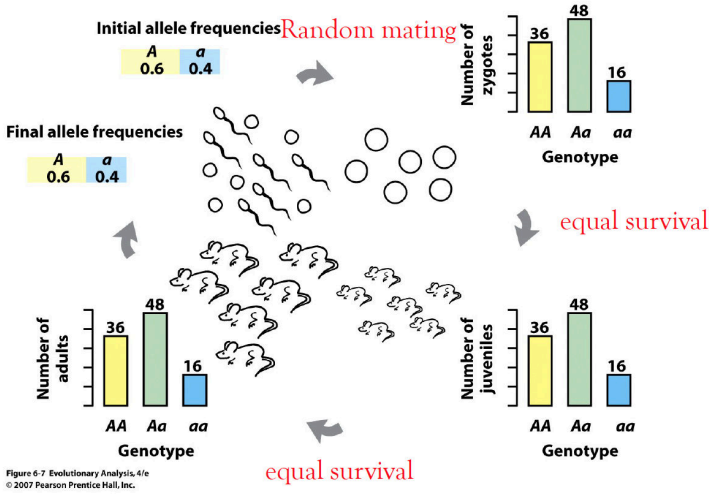


7.3.1 Hardy-Weinberg Population and Principle

The ideal population:

1) infinite population size
2) random mating among individuals
3) no evolution:
   1) no selection
   2) no mutation
   3) no migration
   4) no chance of evolution (genetic drift)

→ Does our population evolve?
If not: the population is in Hardy-Weinberg Equilibrium.
(The HWP is a null model against which we test if a population is evolving or not.
Simples case with two alleles (A and a): $p^2+2pq+q^2 = 1$
Binomial equation: $(p+q)^2 = 1$
More general: $(SUM(p_n))^2 = 1$



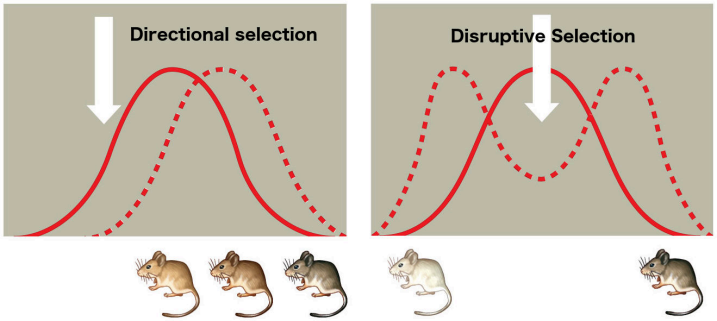
7.3.2 Violations of the HWE

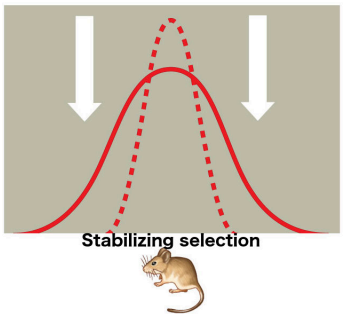
1) Allele frequencies change over time
2) Genotype frequencies do not meet expected values

7.3.3 Testing of the HWE

1) Calculate the observed allele frequencies
2) Calculate the expected number of each genotype under HWE
3) Compare expected and observed numbers
4) Optional but more exact: use a statistical chi-square test for significance

7.4 Selection





In frequency-dependent selection, the fitness of a phenotype depends on its frequency in the population:

- positive FDS: phenotype increases if it becomes most common in the population
- negative FDS: phenotype declines if it becomes most common in the population

7.4.1 Selection coefficient s

The larger the selection coefficient s is the faster allele frequencies change. Fitness w (relative reproductive success) is inversely correlated to s:

$$w = 1 - s$$

7.5 Calculating changes in genotype and allele frequencies due to selection

frequency of first allele: $p = f(A) = 0.5$
frequency of second allele: $q = f(a) = 0.5$
fitness of genotypes:
$w(AA) = 1.0;\ w(Aa) = 0.7;\ w(aa) = 0.6$

1) Calculate average fitness w for the whole population before selection:

$$w = p^2 * w(AA) + 2pq * w(Aa) + q^2 * w(AA)$$

2) Calculate genotype frequencies after selection:

$$f'(AA) = \frac{p^2 * w(AA)}{w}; \quad f'(Aa) = \frac{2pq * w(Aa)}{w}$$

$$f'(aa) = \frac{q^2 * w(aa)}{w}$$

3) Calculate allele frequencies among gametes after selection and random mating

$$f'(A) = p' = f'(AA) + 0.5 * f'(Aa)$$

$$f'(a) = q' = f'(aa) + 0.5 * f'(Aa)$$

$$\Delta p = p' - p; \quad \Delta q = q' - q$$

7.6 Selection can act on different genotypes

If a flour beetle population has two alleles (+ being dominant and viable while - is recessive and lethal). The genotype fitness will play out as follows:

$w(+/+)$	$w(+/-)$	$w(-/-)$
1	1	0

One can predict how the frequencies of the alleles will change:

Generation	+/+	+/-	-/-	$p = f(+)$	$q = f(-)$	$\overline{w}_{pop}$
0	0	1000	0	0.5	0.5	
1 (HWE)	250	500	250	0.5	0.5	1
1 (Selection)	250	500	0	0.6	0.33	0.75
2 (Selection)	436	436	0	0.75	0.25	0.89

7.6.1 Selection against recessive alleles

The section against a recessive lethal allele is less and less efficient the rarer it becomes.

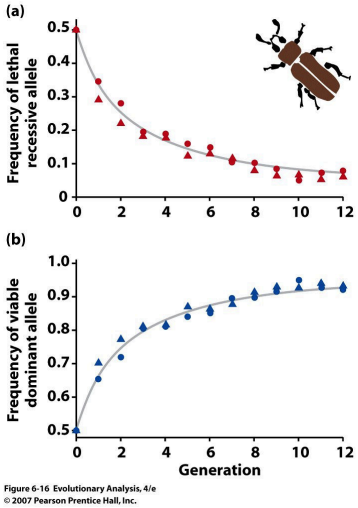


Figure 6-16 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

7.6.2 Selection against dominant alleles

A inferior dominant allele will be eliminated from the gene pool relatively quickly.

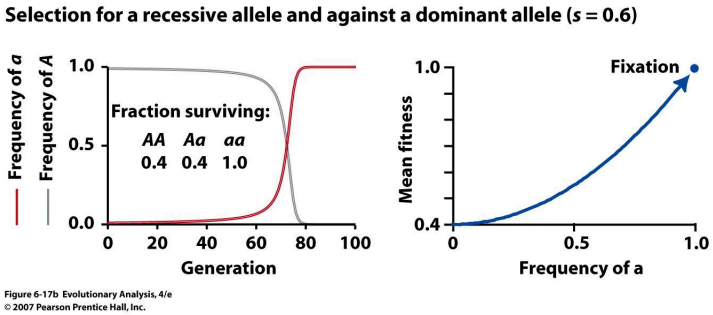


Figure 6-17b Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

7.6.3 Selection favoring heterozygotes

Heterozygote superiority (overdominance) leads to a stable equilibrium.

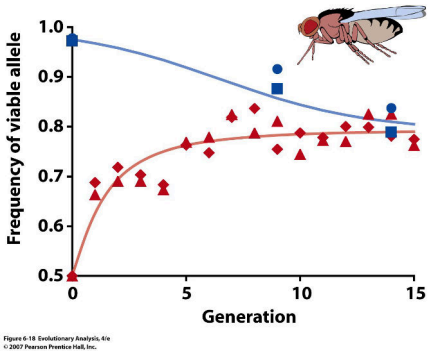
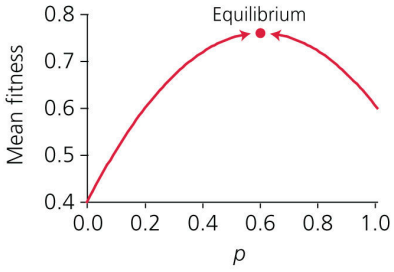


Figure 6-18 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

(b) Mean fitness as a function of p for overdominance



7.6.4 Selection against heterozygotes

Heterozygote inferiority (underdominance) leads to a unstable equilibrium.

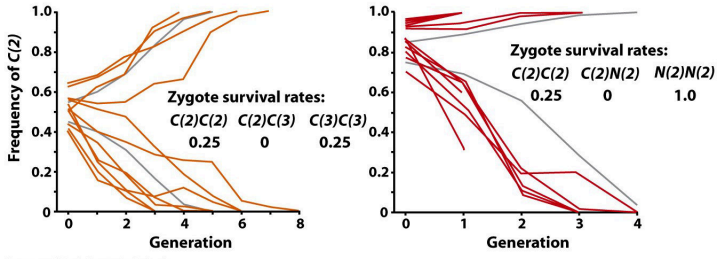
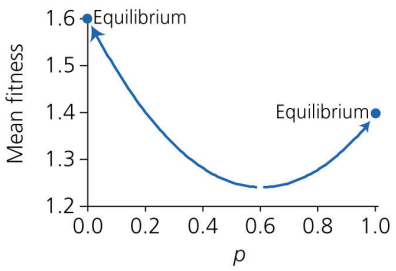


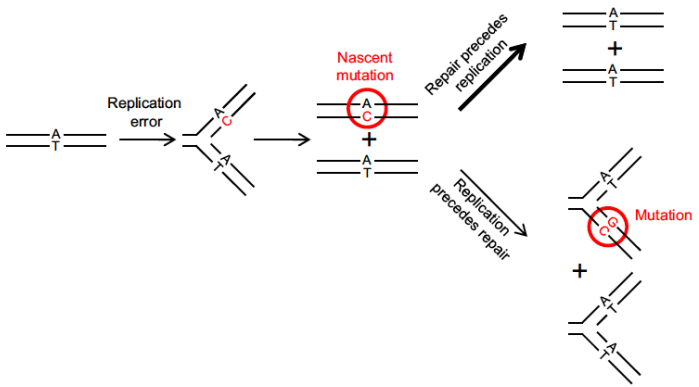
Figure 6-19f Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

(c) Mean fitness as a function of p for underdominance



7.7 Mutation

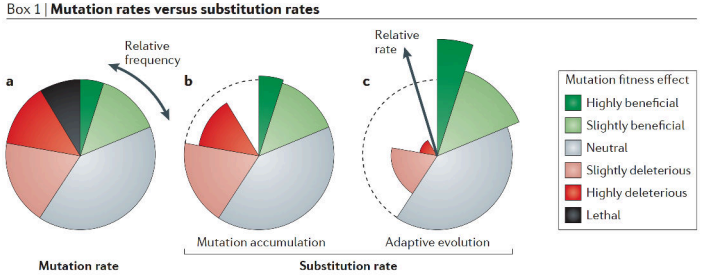
New mutations can arise by erroneous DNA replication



7.7.1 Mutation fitness effects

- Positive: rare, selected fro increase
- Neutral: more common, fate determined by drift of hitchhiking
- Negative: very common, selected against, level in population determined by mutation-selection balance

After selection the frequency of different fitness levels between mutation is different from the initial mutations.



7.7.2 Induced mutagenesis

Induced mutagenesis is a stress-induced reaction to interrupted DNA replication due to UV-irradiation damage, polymerase-inhibition by antibiotics, etc. It leads to a transient increase of mutability through error-prone polymerases, increased DNA uptake / recombination or activation of "sleeping" prophages in genome

7.7.3 Mutation rate estimation

$$\mu = \frac{\text{Mutations}}{\text{Time} * \text{Genetic target}}$$

Mutation is the ultimate source of genetic variation. Over time more and more alleles are added to a gene pool. In the short run the effect may be weak, but it can slowly cause substantial change in the long-term.

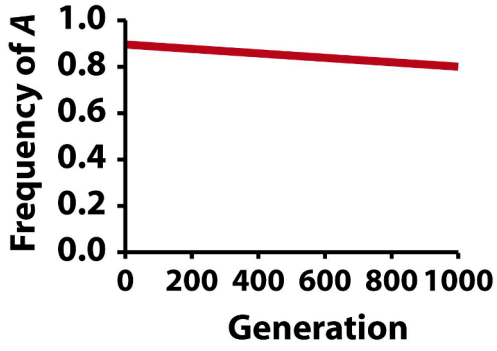


Figure 6-24 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

7.7.4 Mutation-Selection Balance

After a population reaches a fitness peak a balance will be apparent between

- the rate at which new deleterious mutations occur and
- the rate at which these deleterious mutations are selected against (negative selection)

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

$\hat{q}$ = equilibrium frequency of mutant allele
μ = rate of mutation to mutant allele
s = coefficient of selection against mutant allele

Examples of autosomal recessive alleles: SMA (spinal muscular atrophy) causes loss of muscle control through progressive atrophy (lower extremities first). Loss-of-function mutations lead to gradual death of motor neuron cells in parts of spinal chord. A high mutation rate to the loss-of-function allele leads to the maintenance of SMA at mutation-selection balance.
Cystic fibrosis can cause various complications with the most frequent cause of death being lung problems (at 80%). One in 25 people carries one copy of the recessive disease allele in CFTR. The CFTR protein is use by *Salmonella typhi* to invade gut cells which causes typhoid fever. The invasion can only happen in the wild type CFTR but not with the cystic fibrosis causing mutation. Meaning that there is a trade-off in CFTR between protection against causal agent of typhoid fever

and causing cystic fibrosis in homozygous carriers.

For dominant lethal alleles the mutation - selection balance looks as follows

$$\hat{q} = \mu$$

7.8 Genetic Drift

- random survival of alleles (sampling errors)
- most powerful in small populations
- strongest when natural selection is weakest
- cannot produce adaptations

Small population/sample sizes magnify the effects of change hence replication in experimental science.

7.8.1 Heterozygosity is reduced over time

The probability that any given allele will drift to fixation is equal to its frequency in the population:

$$P(fix) = \frac{x}{2N}$$

where x = total number of allele copies and
N = # of diploid individuals in population

7.8.2 Effective population size (N_E)

Effective population sizes are especially sensitive to unequal sex ratios in populations. The higher the inequality the lower the N_E and the higher the effect of genetic drift.

$$N_E = \frac{4 * N_{males} * N_{females}}{N_{males} + N_{females}}$$

7.8.3 Chance, Determinism, and Evolution

- Mutational input is largely random
- Evolution by selection is non-random
- Evolution by genetic drift is random

→ The power of drift is inversely proportional to the power of selection.

7.9 Migration

Migration homogenizes allele frequencies across populations if not opposed by other forces of evolution such as selection. F_{ST} values measure the degree to which separate populations are genetically distinct due to absence of gene flow.

7.10 Evolutionary change

The forces that cause evolutionary change:
Forces that create variation in evolving populations:

- mutation
- recombination

Forces that determine the fate of variation:

- selection
- genetic drift
- migration
- (indirectly: non-random mating, NRM)

7.11 Selfing & Inbreeding

Selfing (self-fertilization) reduces heterozygote frequency. Inbreeding can depress average fitness.

7.12 Evolution at multiple loci: linkage disequilibrium

Extension of Hardy-Weinberg analysis to two loci through tracking of not only allele but also chromosome frequencies.
There is genetic linkage between two loci if they are on the same non-recombining stretch of a chromosome

a in nuclear chromosomes of sexual diploids loci remain together after meiotic crossing
b in non-recombining organisms / organelles loci are on the same single chromosome

If two loci are linked, selection on one locus can affect the evolutionary fate of the other (a hitchhiking-effect)

7.12.1 Linkage equilibrium

The frequency of any haplotype can be calculated by multiplying the frequencies of the constituent alleles.

$$g_{AB} = f(AB) = f(A) * f(B); \quad g_{Ab} = f(Ab) = f(A) * f(b)$$

$$g_{aB} = f(aB) = f(a) * f(B); \quad g_{ab} = f(ab) = f(a) * f(b)$$

With g being the genotype frequency, one can calculate the coefficient of linkage disequilibrium D:

$$D = g_{AB}g_{ab} - g_{Ab}g_{aB}$$

Linkage happens due to relative physical location on chromosomes, but linkage equilibrium and disequilibrium are characteristics of populations:

- In a population in linkage equilibrium the frequencies of B and b alleles are the same on A- and a-bearing chromosomes (and vice versa).
- In a population in linkage disequilibrium the frequencies of B and b alleles are different on A- vs. a-bearing chromosomes (and vice versa).

Deviations from HW expectations suggests that one of three mechanisms causing linkage disequilibrium is at work:

- selection on multi-locus genotypes
- genetic drift
- population admixture (i.e. migration)