Selection in the lab

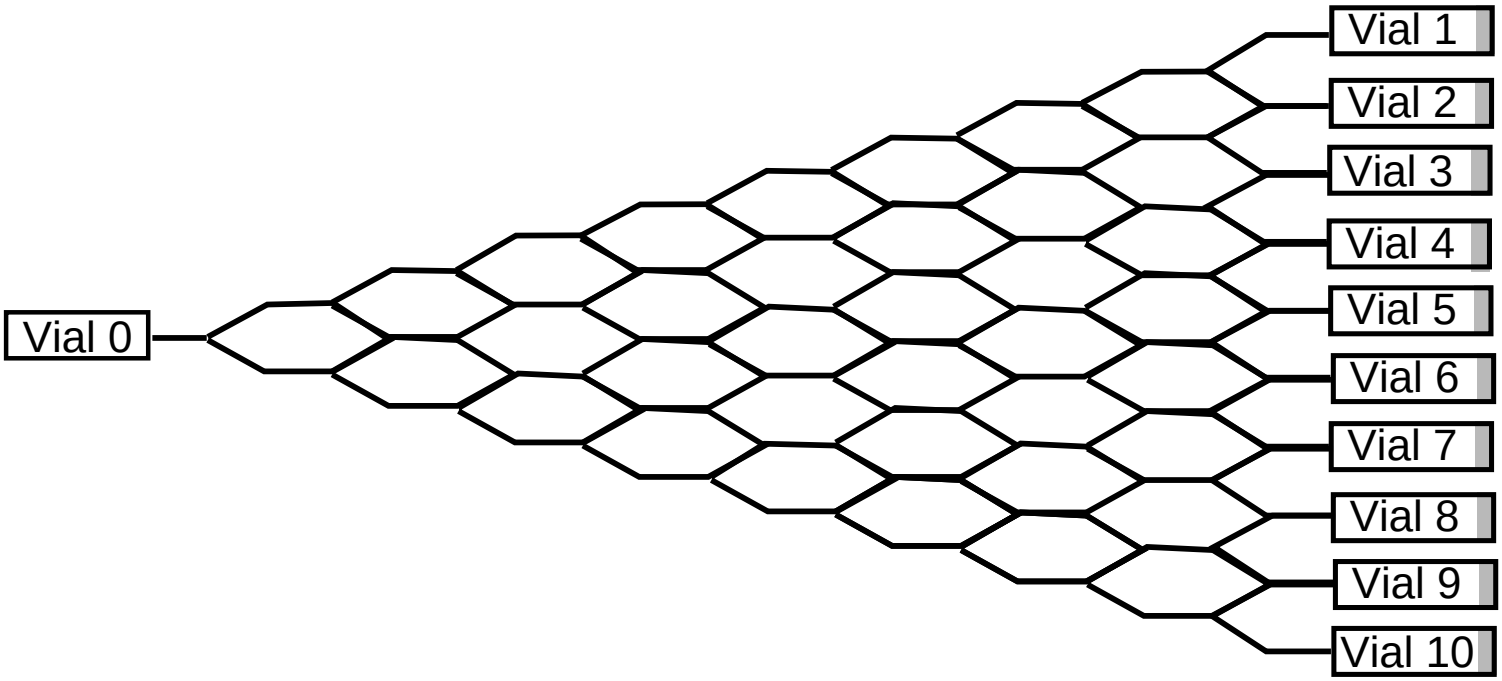
Drs. Denis Bourguet and Michael Whitlock are currently performing an evolution experiment in the fruitfly, *Drosophila melanogaster*, at UBC.

A large population of *D. melanogaster* has been maintained in the lab since 1970, after having been collected in Benin (Western Africa). This lab population has experienced very similar lab conditions for nearly thirty years.

As part of a larger experiment, Bourguet and Whitlock started ten lines of flies from this population in February 1997:

- Five control lines
- Five selected lines

Every generation, 500 male flies of a line were placed in the starting vial (lacking food) of a maze.



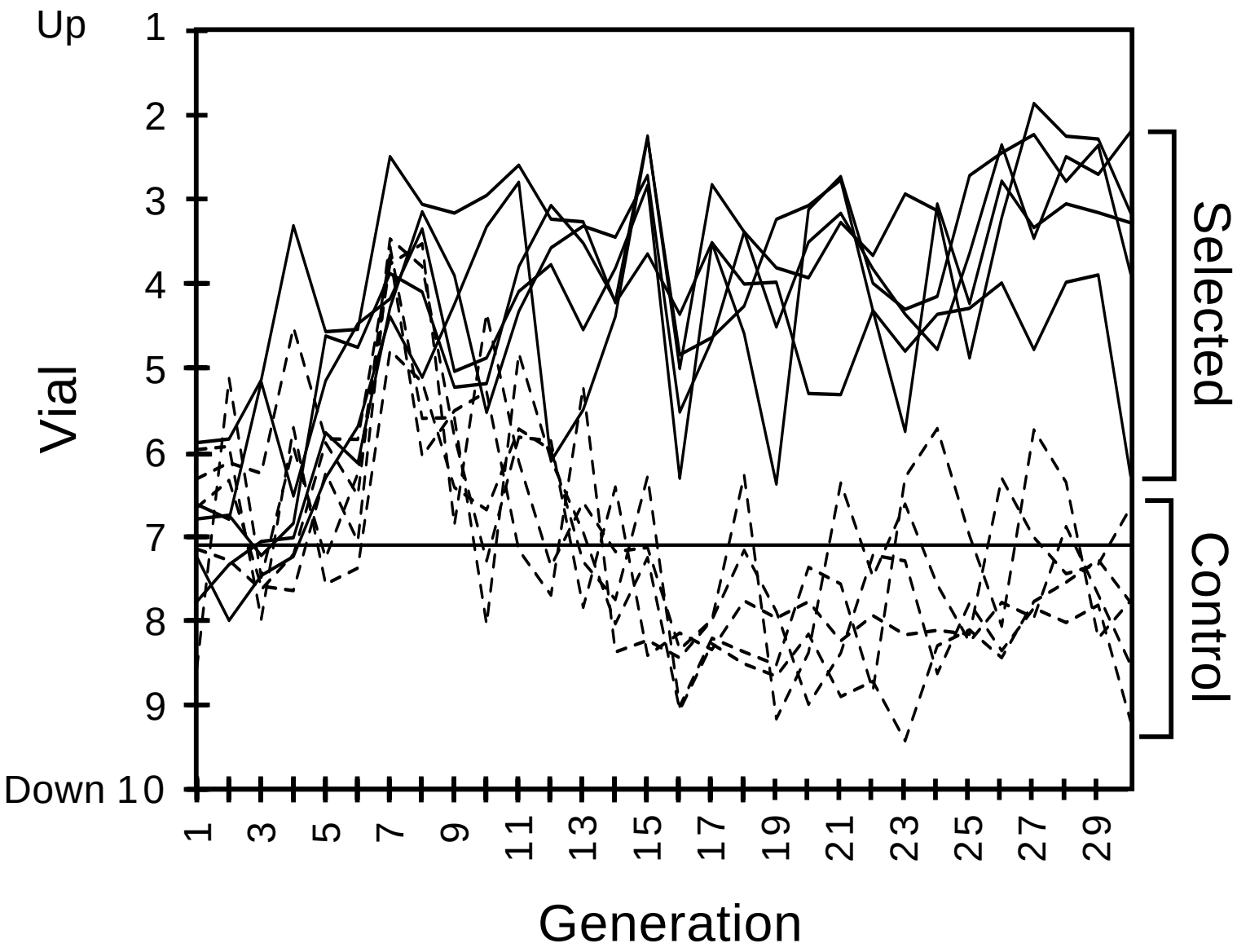
After 24 hours, flies were collected from the ten vials (with food) at the other end of the maze.

Control lines: Flies from the ten output vials were mixed and 50 males were randomly chosen to reproduce.

Selected lines: The fifty males in the highest vials were selected to reproduce (selection for "positive geotaxis").

In each case, the fifty chosen males were allowed to mate with 50 unselected females from the same line, and the process was repeated with their offspring.

What do you expect to have happened?

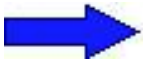


After only 30 generations, there was a response to selection such that flies from the selected lines pass through the maze to a vial that is, on average, 4.25 vials higher than flies from the control lines!

(For those who have had Bio 300, $p<0.001$ by a two-sampled t-test.)

What would have happened if:

- flies did not vary in the way they went through the maze?
- selected flies were chosen from the vials at random like the controls?
- the behavior in passing through the maze of the offspring was uncorrelated with the behavior of the parents?

 Evolution by natural selection occurs whenever

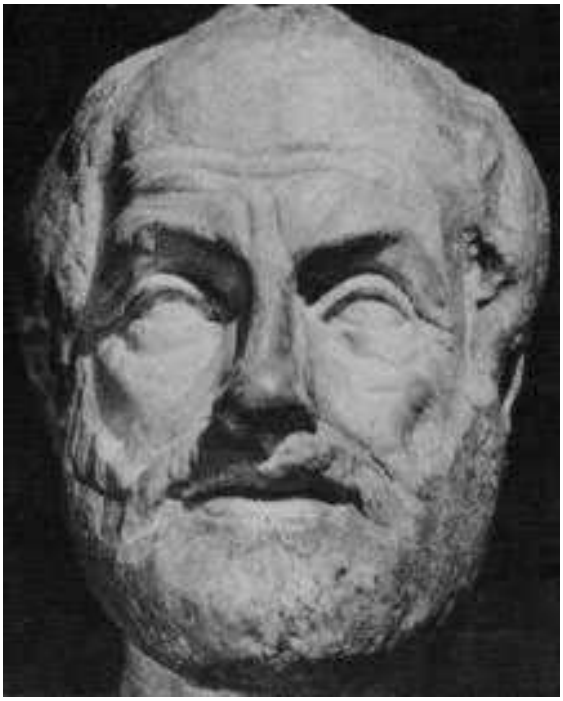
- individuals vary in some trait (VARIANCE)
- individuals with some trait values are more likely to live and/or reproduce than other individuals (SELECTION)
- parents have offspring with trait values that are similar to their own (HERITABILITY)

DEFINITIONS:

Evolution is the change in form and/or behavior of organisms between generations (= Descent with Modification in Darwin's terms).

Natural selection is the process whereby some individuals contribute more offspring to the next generation as a consequence of their carrying a trait or traits favorable to survival or reproduction.

History of Evolutionary Thought



The Chain of Being

Aristotle (384-322 BC) was a Greek philosopher, who examined the natural world for evidence of a divine order. Aristotle devised a hierarchical arrangement of natural forms, termed the "Scala Naturae" or Chain of Being.

Species were arranged in a linear fashion along a scale: God, man, mammals, egg-laying animals, insects, plants, and non-living matter.

Aristotle’s ideas formed the basis for the western belief in a fixity of species, each of which has a typical form.

Father of Modern Taxonomy

Carolus Linnaeus (1707-1778 AD) classified organisms following a binomial system of nomenclature, giving each species a specific and generic name [e.g. Homo sapiens (Genus species)].

Although his classification was largely based on morphology, Linnaeus recognized a fundamental difference between organisms that could interbreed (within a species) and those that could not (different species).

His classification system departed from the chain of being and reflected a nested series of relationships. The modernized Linnaean system is: Kingdom, Phylum, Class, Order, Family, Genus, Species.

Linnaeus believed in a balance of nature, within which each species had its place. Since this balance was thought to reflect a divine plan, Linnaeus originally believed that species would neither change nor go extinct.

Linnaeus later recognized that new species may occasionally arise, particularly through hybridization. One of Linnaeus' students first described a new species formed by hybridization in toadflax.

Degénération

Georges-Louis Leclerc, Comte de Buffon (1707-1788 AD) believed that the origin of life and species followed a material process, and looked to the physical and biological world for clues to this process.

In 1766, Buffon argued that the relationships among species in the Linnaean system of classification reflected common descent ("degénération"), with divergence over time.

Buffon placed great emphasis on the physical environment, which was thought to direct (somehow) the organic changes leading to a new species. The speciation process was thus caused by individual migration to new geographical locations, wherein the environment would cause changes to the organic particles.

Nevertheless, Buffon thought that the extent of divergence was limited to within a family. Each family had its own internal "mold" to which every species conformed, but species could change over time to some degree.

Organic life beneath the shoreless waves
Was born and nurs'd in ocean's pearly caves;
First forms minute, unseen by spheric glass,
Move on the mud, or pierce the watery mass;
These, as successive generations bloom,
New powers acquire and larger limbs assume;
Whence countless groups of vegetation spring,
And breathing realms of fin and feet and wing.

Erasmus Darwin
The Temple of Nature (1802)

Erasmus Darwin (1731-1802) was a leading philosopher, naturalist, and physician in 18th century Britain, who wrote one of the first treatises on evolution: *Zoonomia: Or The Laws of Organic Life* (1794-1796). He was also the grandfather of Charles Darwin.

Erasmus believed in the self-improvement of a species, through a constant effort to adapt to the environment (*transformism* or *transmutation*). He argued that life consisted of "one living filament", connecting all current forms to one common ancestor.

Inheritance of Acquired Characters

Jean-Baptiste Lamarck (1744-1829) was a French professional naturalist, who also developed a theory of transformism.

Lamarck believed in organic progression, whereby organisms would evolve through a hierarchy of more and more advanced forms. At the base of this hierarchy, "simple" organisms were constantly arising via spontaneous generation.

> "Nature, in producing in succession every species of animal, and beginning with the least perfect or simplest to end her work with the most perfect, has gradually complicated their structure."

The mechanism by which organisms advanced and adapted to changing circumstances was described in Lamarck's *Philosophie zoologique*(1809):

Lamarck's "First Law":

The use or disuse of a structure would lead to its development or diminishment.

Lamarck's "Second Law":

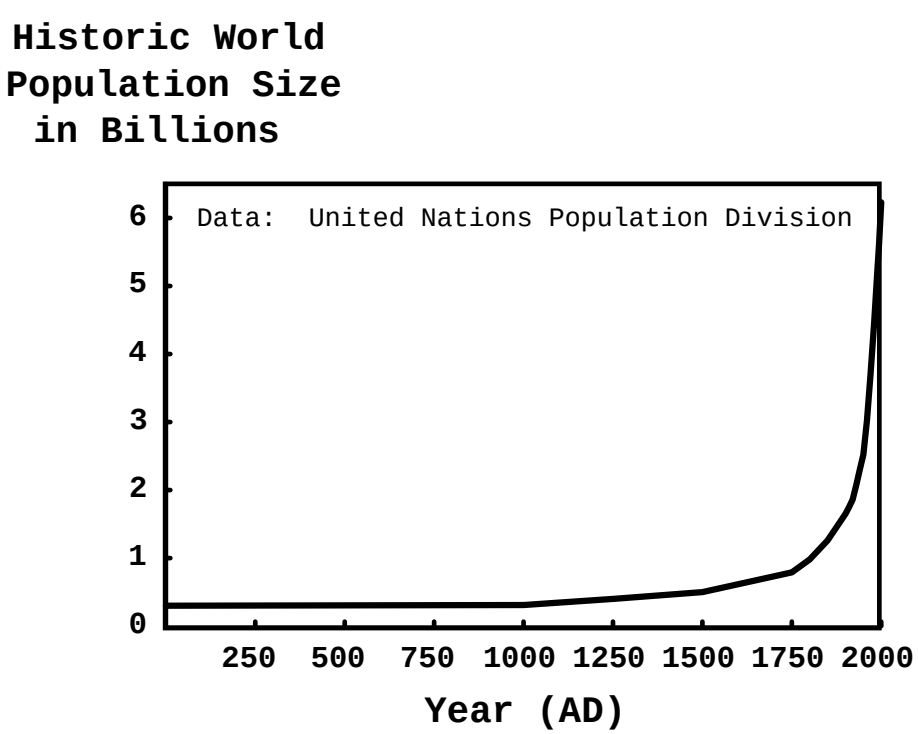
Such acquired characters could be passed to offspring (heritable).

Principle of Overproduction

Thomas Malthus (1766-1834) was an English clergyman, whose writings on population growth had a strong influence on the theory of evolution by natural selection developed by Charles Darwin and Alfred Russel Wallace.

In *An Essay on the Principle of Population* (1797), Malthus observed that most organisms produce far more offspring than can possibly survive.

Even when resources are plentiful, the size of a population tends to increase geometrically until the population outstrips its food supply. This led Malthus to believe that poverty, disease, and famine was a natural and inevitable phenomenon, leading to a "struggle for existence".



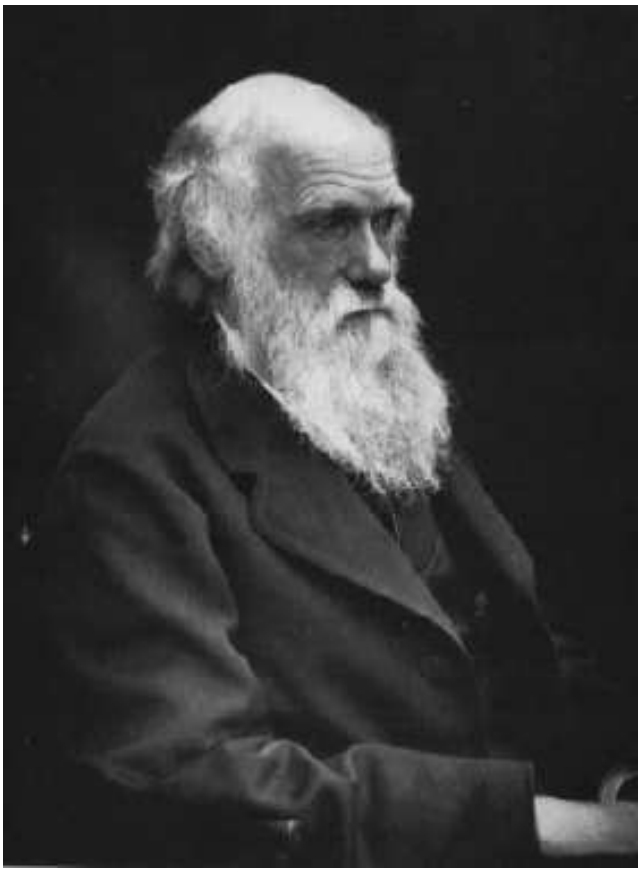
Uniformitarianism

Charles Lyell (1797-1875) was an English geologist, whose Principles of Geology influenced both Darwin and Wallace.

Lyell believed that the earth was constantly changing and that the processes that had molded the earth’s surface could be understood from current-day geological activities.

Lyell held a "uniformitarian" view of the world, meaning that the world was subject to gradual and continuous change. Yet, there was no progress or development over time in Lyell’s world-view. The earth simply remained at steady-state.

Lyell’s position suggested that the world had always been (roughly) similar to its current state. In particular, Lyell believed that the species composition of the world remained unchanged, with at least some members of all classes of organisms existing throughout the history of the earth.



Evolution by Natural Selection

Charles Darwin (1809-1882) was an English naturalist, a prolific writer, and a gentleman of private means that allowed him to focus on his life's work: the development of the theory of evolution by means of natural selection.

As a young man, Darwin was enlisted as a companion and naturalist aboard the H.M.S. Beagle (1831-1836). His voyage took him around the tip of South America to New Zealand and Tasmania.

Darwin read Lyell's Principles of Geology while on board the Beagle and came to accept Lyell's view that long-term geological processes were responsible for shaping the earth's surface in a gradual manner. Indeed, Darwin successfully applied uniformatarianism to explain the development of coral reefs.

Biogeography

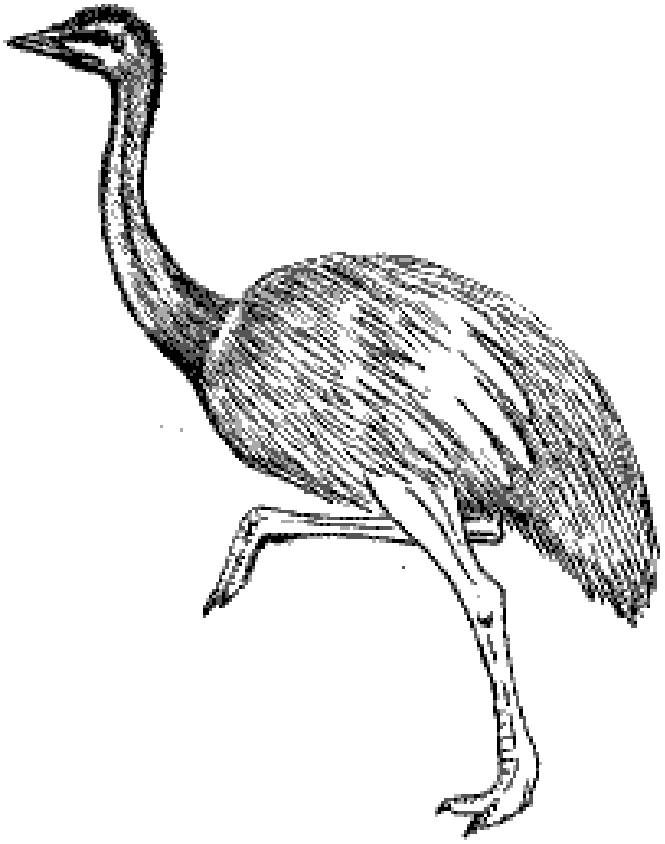
Perhaps the single-most important influence on Darwin's intellectual development was an appreciation, developed during the voyage, for biogeographical patterns.

Darwin noticed that two similar species (e.g. two species of rhea, a South American flightless bird) would often co-exist within a boundary zone. Within this zone, clearly neither species was superior and especially created to match the local circumstances. Instead, the species must compete with each other for survival within this territory.

Darwin also noticed that barriers, especially oceanic barriers, often led to distinctly different groups of species on different land masses. Why is the rhea of South America so different from the ostrich of Africa when the two have such similar lifestyles under such similar circumstances? Clearly, migration across oceans was limited, but why would a creator be limited by such boundaries?

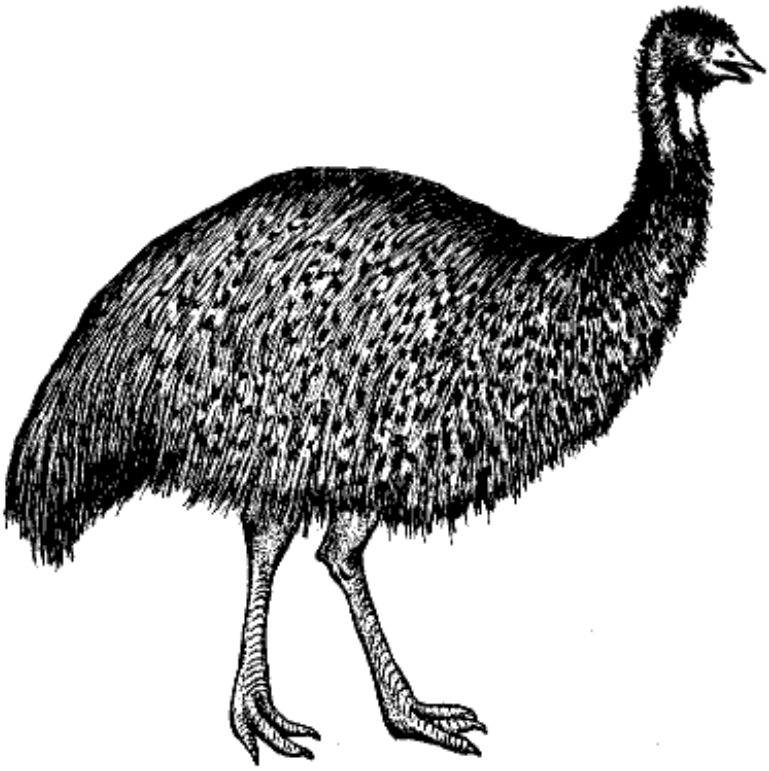
It was Darwin's visit to the Galapagos islands, hundreds of miles west of South America, that would most clearly lead to Darwin's views on the origin of species. Darwin puzzled over the fact that the giant tortoises from different islands were distinct from one another -- why would there be a separate creation of such similar forms on islands so close to one another?

Darwin was also impressed by the great diversity of finch species on the islands, although Darwin had difficulty classifying the finches and thought that his collection contained a wide assortment of birds including wrens, gross-beaks, and black-bird relatives. [It later turned out that all these specimens were closely related and represented a remarkable diversification of finch species].

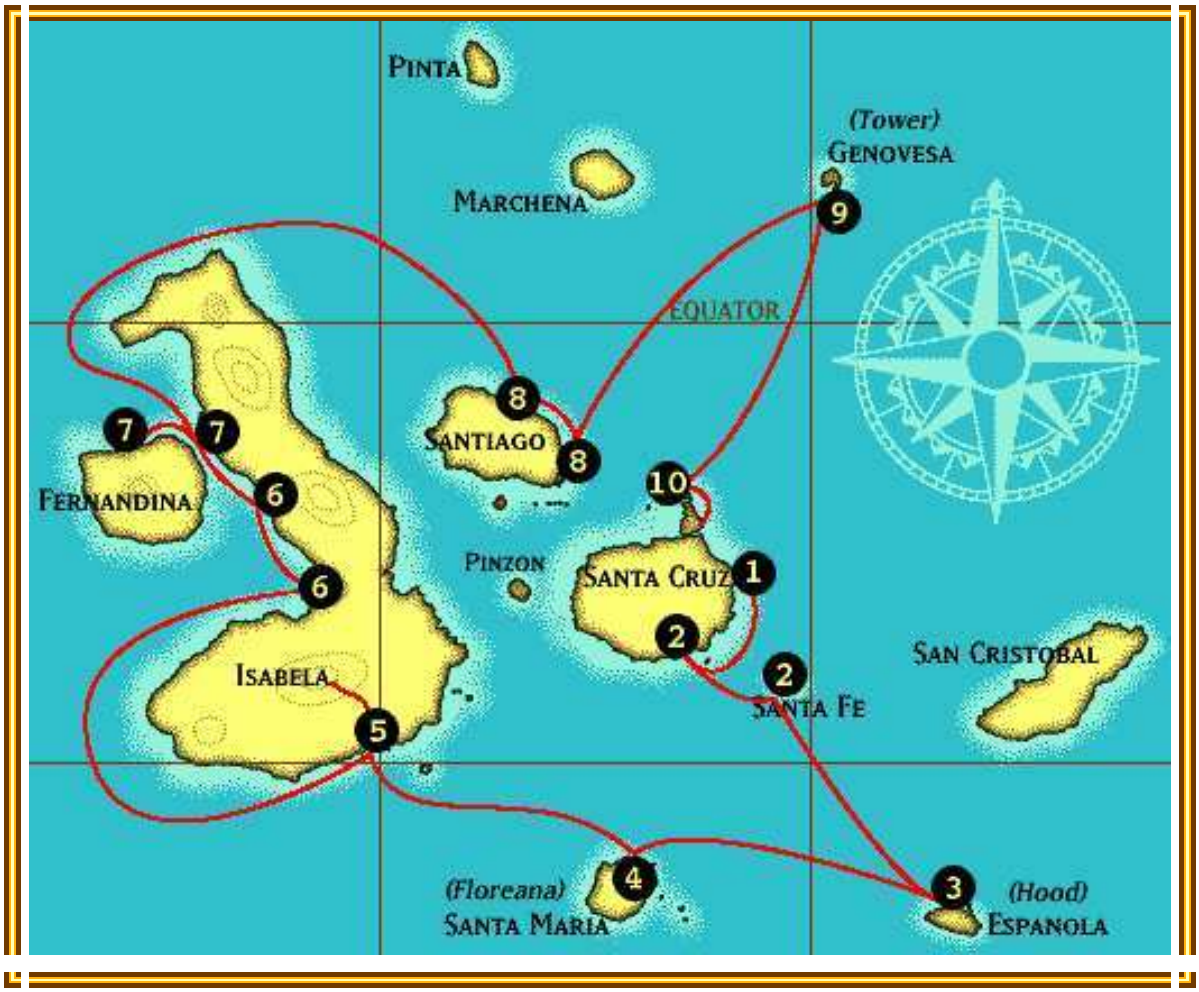


Rhea

Ostrich



Emu



Area:

Total area (sea): 45,000 square kilometers. *Total land area:* 7882 square kilometers

Size:

430 kilometers long, from Darwin Island in the northwest to Española in the southeast.

Islands:

13 major (above 14 square kilometers, 5 square miles); 8 smaller islands above .12 square kilometers. (.5 square miles); 40 named islets. Major islands include: **Española**, **Fernandina**, **Genovesa**, **Isabela** SouthWest/ SouthEast/ Central/ North, **Marchena**, **Pinta**, **Pinzon**, **San Cristóbal**, **Santa Cruz**, **Santa Fé**, **Santa Maria**, **Santiago**.

Darwin did recognize the differences between mockingbirds on different islands that he visited and wrote:

"When I see these Islands in sight of each other, and possessed of but a scanty stock on animals, tenanted by these birds, but slightly differing in structure and filling the same place in Nature, I must suspect they are only varieties....If there is the slightest foundation for these remarks the zoology of Archipelagoes - will be well worth examining; for such facts would undermine the stability of Species." Darwin's Ornithological Notes (1836), p. 262 (Barlow, 1963)

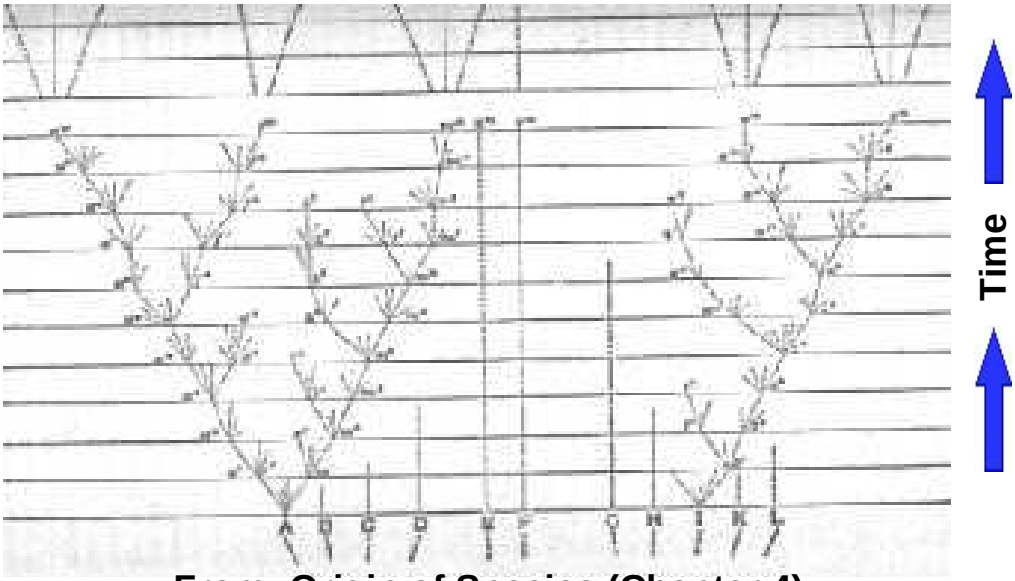
Returning to Britain, Darwin gradually developed his theory concerning natural selection. He recognized several critical facts:

- Variability exists within species
- Variant traits may be inherited [Darwin did not, however, know how.]
- From Malthus' principle of overproduction, many individuals must often die or fail to reproduce. In this "struggle for existence", variants that were slightly better suited to the environment would be more likely to survive.

It then follows logically that certain variants will be preserved over time over other variants and that populations will change over time in their composition. This is *evolution by natural selection*.

Thus, by 1838, Darwin came to believe in transmutation, which was neither directed by the will of the individual (as believed by Lamarck) nor by direct oversight of a creator.

As each population changes by natural selection, geographically isolated populations would become more and more different from one another ("divergent"). Darwin believed that this would initially lead to different varieties within a species. Eventually, with sufficient time and divergence, evolution by natural selection would also lead to new species and higher taxanomic divisions, in an ever branching process.



From: Origin of Species (Chapter 4)

Natural Selection Co-Discovered

In 1858, Alfred Russel Wallace (1823 - 1913) sent Darwin a manuscript from the Malay archipelago (Indonesia) describing Wallace's independent discovery of evolution by natural selection.

Wallace, unlike Darwin, was relatively poor and supported himself and his world-wide travels by capturing and selling specimens. Like Darwin, Wallace travelled to South America and throughout the Pacific.

Observing biodiversity and biogeography first hand was also critical to the development of Wallace's views. (QUOTE)

Charles Lyell and Joseph Hooker quickly arranged for both Darwin's and Wallace's views to be presented at the meetings of the Linnean Society in London in 1858.

The next year, Darwin published The Origin of Species by Means of Natural Selection. The depth and breadth of Darwin's book, developed over twenty years of thought and research, revolutionized science.

"At the time in question I was suffering from a sharp attack of intermittent fever... One day something brought to my recollection Malthus's "Principles of Population", which I had read about twelve years before. I thought of his clear exposition of "the positive checks to increase" - disease, accidents, war, and famine - which keep down the population of savage races to so much lower an average than that of more civilized peoples. It then occurred to me that... as animals usually breed much more rapidly than does mankind, the destruction every year from these causes must be enormous in order to keep down the numbers of each species... Vaguely thinking over the enormous and constant destruction which this implied, it occurred to me to ask the question, Why do some die and some live? And the answer was clearly, that on the whole the best fitted live. From the effects of disease the most healthy escaped; from enemies, the strongest, the swiftest, or the most cunning; from famine, the best hunters or those with the best digestion; and so on. Then it suddenly flashed upon me that this self-acting process would necessarily improve the race, because in every generation the inferior would inevitably be killed off and the superior would remain
- that is, the fittest would survive."

-- Wallace (1905; My Life) describing his discovery in 1858

Mendelian Genetics

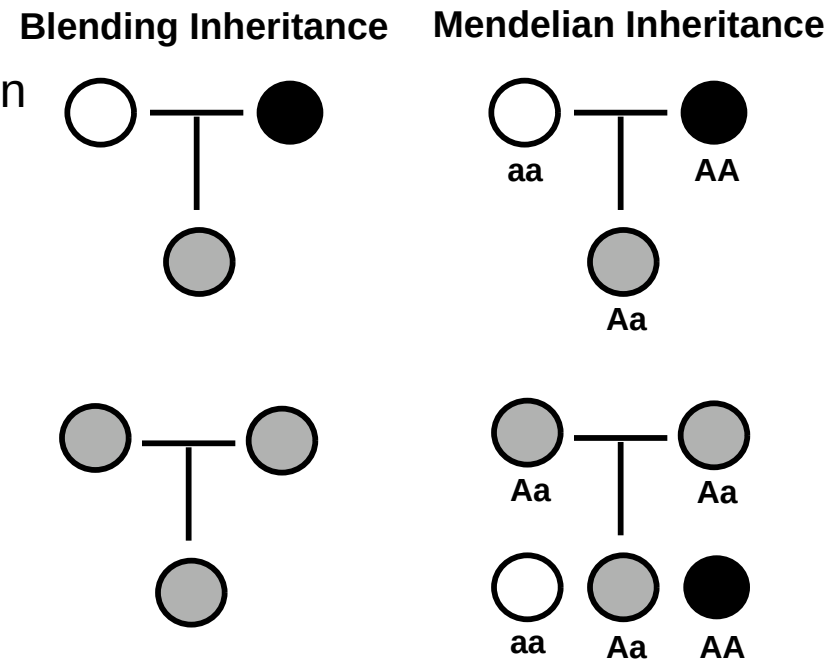
The greatest weakness in the theory of evolution by natural selection was the fact that Darwin and Wallace knew neither how variation among individuals was generated nor how it was inherited.

The rediscovery of Gregory Mendel's work by Carl Correns and Hugo deVries in 1900 clarified the laws of inheritance, at least for discrete characters (such as pea color).

Mendelian rules explain why offspring tend to resemble parents.

Mendelian genetics has one important characteristic: variation is not lost over time simply due to reproduction.

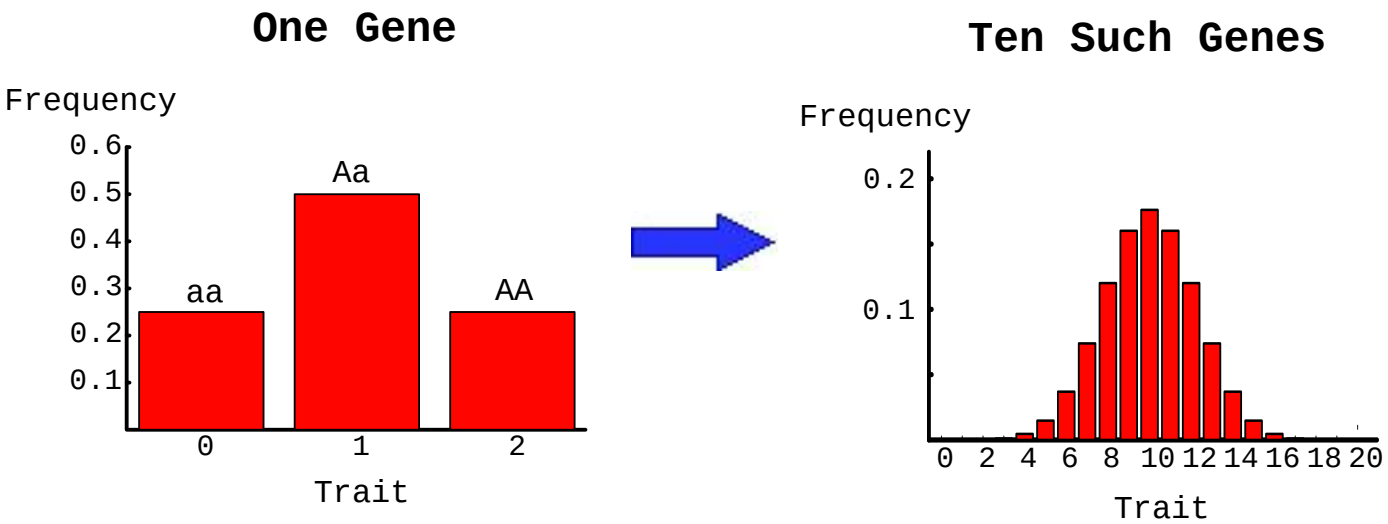
In contrast, with blending inheritance, offspring are the average of their parents and a diverse parental generation will be followed by a less diverse offspring generation.



Yet, it remained unclear whether Mendel's rules applied to continuously varying traits, such as height and weight.

Uniting Mendelian and Quantitative Genetics

In 1918, Ronald Aylmer Fisher (1890-1962) demonstrated that a large number of Mendelian characters (genes) influencing the same trait would lead to a nearly continuous distribution of trait values. The frequency distribution of traits would then look approximately normal (i.e. like a bell), as is the case for traits such as height and weight.



(We'll talk more about Fisher and other important 20th century evolutionary biologists throughout the term.)

Lines of Evidence for Evolution

Evolution

A cumulative change in the characteristics of organisms or populations from generation to generation.
(Academic Press Dictionary)

Natural selection

The process whereby some individuals contribute more offspring to the next generation as a consequence of their carrying a trait or traits favorable to survival or reproduction.

Artificial selection

The process of selection whereby traits in an organism are deemed favorable and are selectively bred by humans. (Humans specify who survives and reproduces.)

Homology

A similarity between species that is not functionally necessary but that results from inheritance from a common ancestor.

Vestige

A bodily part or organ that is degenerate or imperfectly developed in comparison to one more fully developed in closely related forms. (Webster's Dictionary)

Numerous lines of evidence exist for evolution. These may be categorized as follows (following Ridley):

1. Direct observation of change in natural populations
   - HIV
   - Blackcap migratory patterns
   - Seed morphology on islands
2. Direct observation of change under artificial selection
   - Habitat selection in Drosophila
   - Oil content in corn
   - Abdominal bristle number in Drosophila
   - Thorax length in Drosophila
   - Weight in mice
3. Homologous traits
   - Genetic code
   - Cell structure
   - Pentadactyl limbs
   - Stingers and ovipositors
   - Vestigial organs
4. Homologies may be hierarchically classified (nested)
   - Primates
   - Birds
5. Evidence for evolutionary change in fossil record
   - Trilobites
   - Foraminiferans

1. Evolution in Natural Populations.

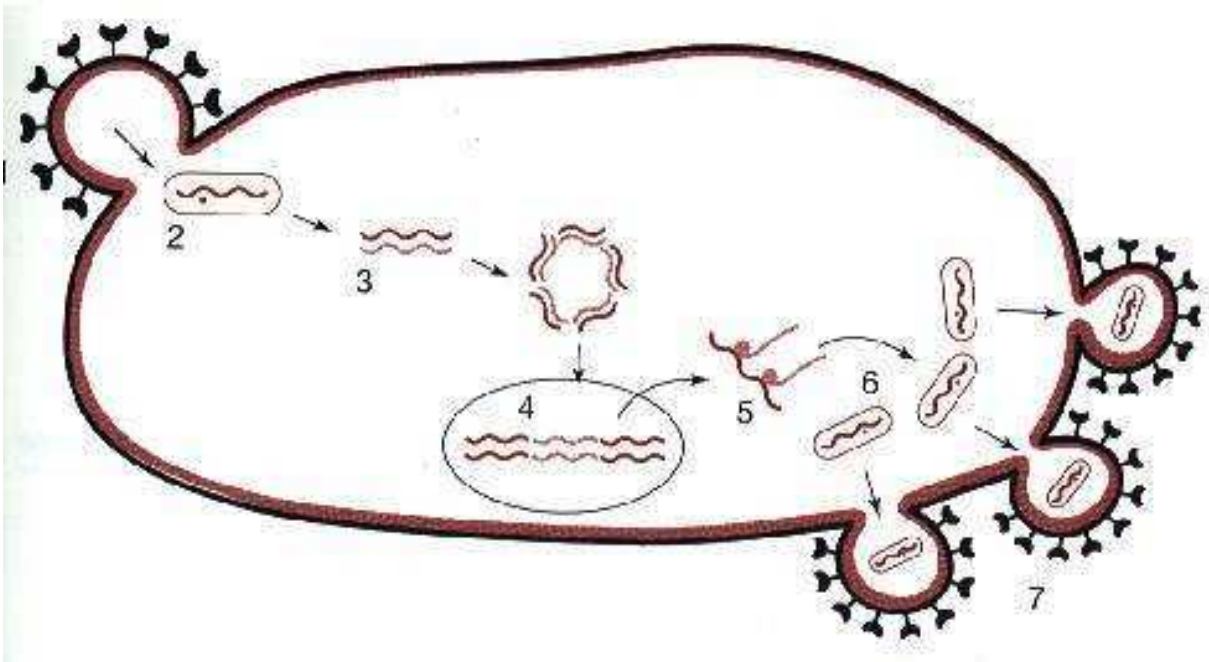
Human Immunodeficiency Virus (HIV)

HIV is the retrovirus that causes acquired immune deficiency syndrome (AIDS) in humans.

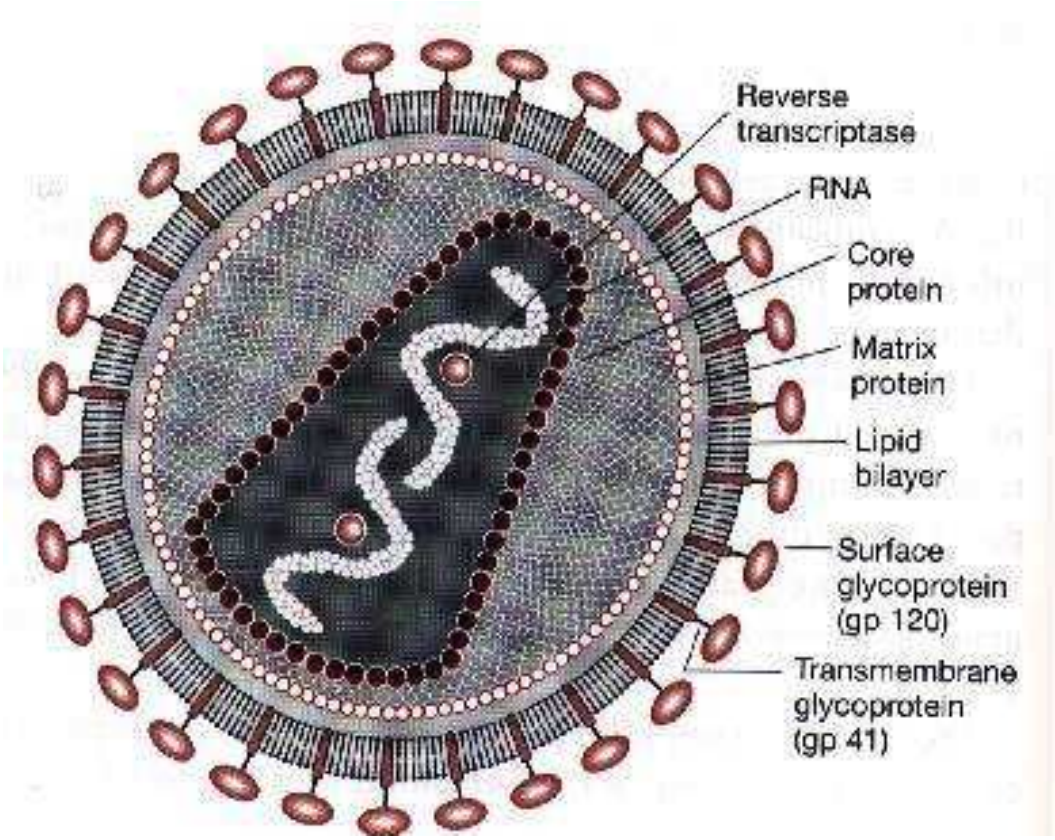
As a retrovirus, the virus particle (called a *virion*) contains RNA not DNA. When a retrovirus infects a host cell (1), the RNA becomes translated into DNA (2-3) by a protein called "reverse transcriptase" that is encoded in the viral RNA.

This DNA then enters the nucleus and integrates into the host cell's nuclear genome at a site that is actively transcribed (4). The DNA of the virus is subsequently transcribed and translated by the host cell (5), producing both the RNA and proteins needed to produce new virions (6).

Finally, new virions burst out of the host cell (7).



HIV contains only nine genes. We'll discuss the evolution of two of these genes: *env* (the gene that produces the outer surface glycoproteins of HIV) and *pol* (the gene that produces reverse transcriptase).

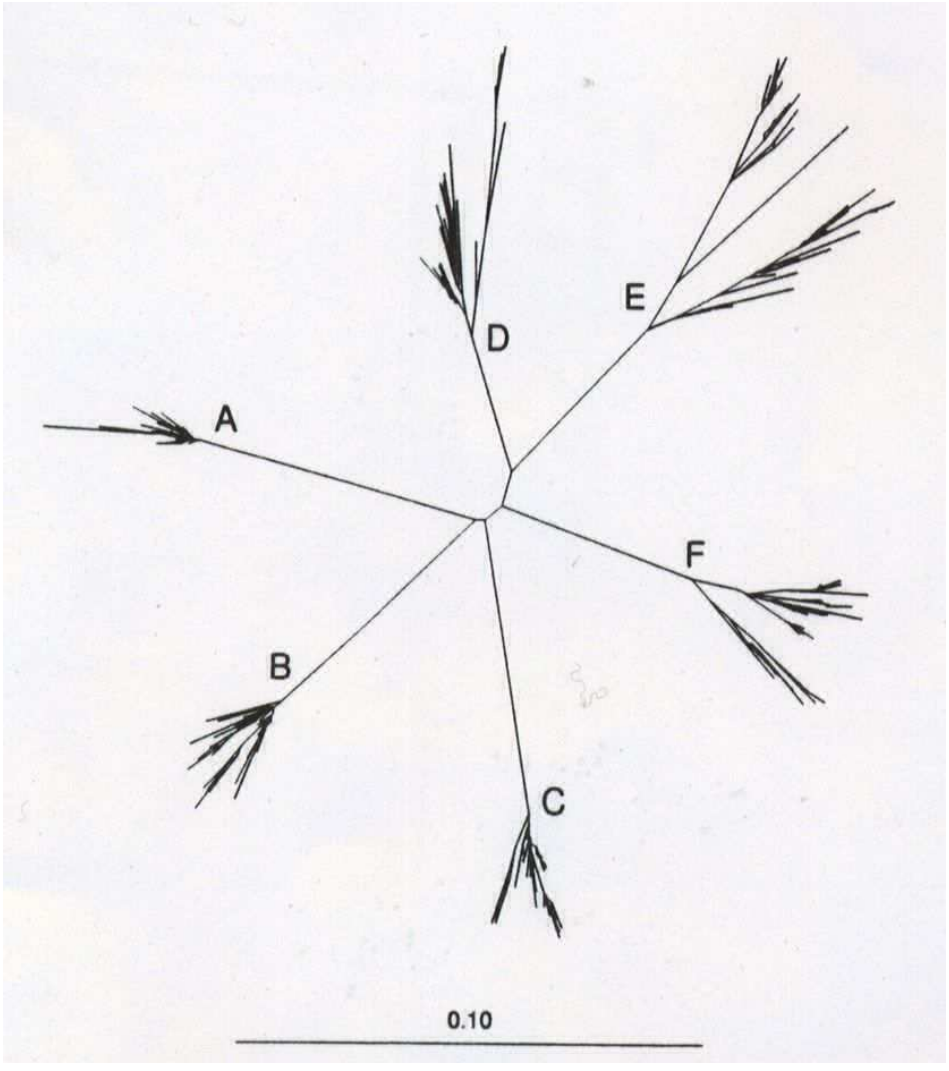


HIV has a high mutation rate and evolves very rapidly, thwarting the defenses of the immune system and the efficacy of drug therapy.

Ganeshan et al. (1997) sequenced regions of the *env* gene from six children who were infected with HIV during pregnancy. Sequences were determined at four different time points from ~12 viral clones each time.

Initially, all viruses within a child were similar. Over time, however, the viral sequences changed.

Ganeshan et al. illustrated these changes using a *phylogenetic tree*, which places sequences that are similar to one another close together. [Branch distances between two sequences on the tree is related to the number of sequence differences between them.]



Observations:

- Sequences from within all six children (letters A-F) clustered together,
- Sequences within a child changed over time rather than remaining constant, leading to the branching patterns.

This phylogeny demonstrates that each child was infected by a slightly different HIV virus and that the viral population then changed and diverged within each child (=evolution).

But why?

Interestingly, children D-F were slow to progress to disease, suggesting that they had mounted an adequate immune response to the virus.

In these children, the virus appears to have evolved at a higher rate (witness the longer branches).

Two possible explanations:

A higher mutation rate.
Selection.

How could we tease these apart?

Genetics review:

Mutations within a gene can be classified as:

- Synonymous (silent) = A mutation to a new codon that codes for the SAME amino acid
- Non-synonymous = A mutation to a new codon that codes for a DIFFERENT amino acid

If originally a DNA codon is CTC (coding for glutamic acid), a mutation in the third position to CTT would still code for glutamic acid (SYNONYMOUS), but a mutation to CTA would code for aspartic acid (NON-SYNONYMOUS).

If all mutations are neutral (not selected), then synonymous and non-synonymous mutations should occur at the rate expected given the genetic code.

If changing the amino acid of a protein is detrimental to the functioning of the protein, however, non-synonymous mutations would be eliminated by selection and would be less likely to remain within the population.

➡ Non-synonymous mutations less frequently observed.

If changing the amino acid of a protein is beneficial, however, non-synonymous mutations would be promoted by selection and would be more likely to remain within the population.

➡ Non-synonymous mutations more frequently observed.

Ganeshan et al. found that the rate of non-synonymous mutations in children D-F was significantly higher than the rate of synonymous mutations.

This suggests that, in these children, natural selection actively favored mutations in the *env* gene that changed the amino acid sequence of the viral surface proteins.

This is consistent with the hypothesis that children who are slow to progress to AIDS have an active immune system.

Active immune systems strongly select for HIV viruses that evolve different surface proteins, thus evading the immune system.

In the children who progressed to AIDS rapidly, Ganeshan et al. observed the opposite pattern. This suggests that the virus was already capable of evading the immune system and was not under selection to change.

Azidothymidine (AZT) is a drug designed to combat HIV by mimicking the nucleotide thymidine (T).

Reverse transcriptase uses AZT instead of T when translating RNA into DNA, which then blocks the growing DNA chain.

Initially, AZT was quite effective at halting the deterioration of the immune system in people with AIDS.

Within a few years, however, AZT stopped working in many of these patients.

Researchers found that the gene that encodes reverse transcriptase (*pol*) had changed over time (= evolved) within these patients.

The changes conferring resistance to AZT altered amino acids in the active site of the reverse transcriptase enzyme.

Often, the same changes occurred in different patients (= convergent evolution).

HIV replicates about 300 times per year, producing a large population of virus particles, many of which carry new mutations.

Viruses that contain mutations that decrease the affinity of reverse transcriptase for AZT are much more likely to replicate.

→ In the presence of AZT, natural selection favors those variants of HIV that are least likely to use AZT when translating RNA into DNA.

We can't stop HIV from evolving, but what might slow down the evolution of HIV?

SOURCES:

- Ganeshan et al. (1997) Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. J. of Virology 71:663-677.

- Freeman and Herron (1998) Evolutionary Analysis. Prentice Hall.

1. Evolution in Natural Populations

Migration Patterns in the Blackcap

The blackcap (*Sylvia atricapilla*) is a European songbird that, until the 1950's, migrated exclusively to the Mediterranean during the winter months.

Since then, about 10% of blackcaps from Germany and Austria have changed their migratory routes and have begun to overwinter in Britain.

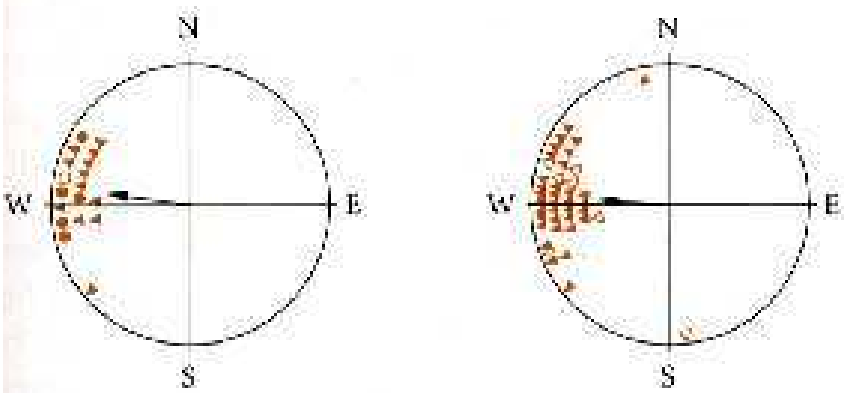
Now, several thousand blackcaps migrate each winter in a northwesterly direction from Continental Europe to the UK.

While this information alone is enough to know that *change* has occurred, it does not demonstrate the basis of this change nor the mechanisms causing the change.

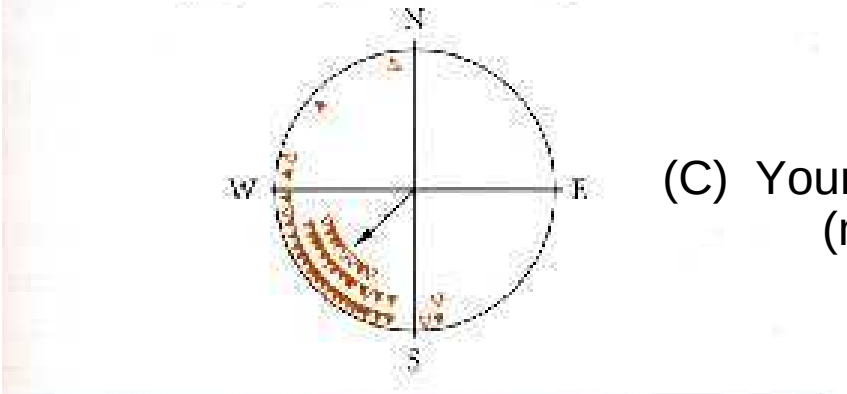
Berthold et al. (1992) took overwintering birds from the UK and brought them to a part of Germany where the local birds migrate to Spain. They mated the captive UK birds and hand-raised their offspring (= F1) as well as some local nestlings.

In the autumn, they placed the different birds in cages exposed to the night sky and marked where the birds attempted to fly.

(A) Adults from Britain
(mean 279°)



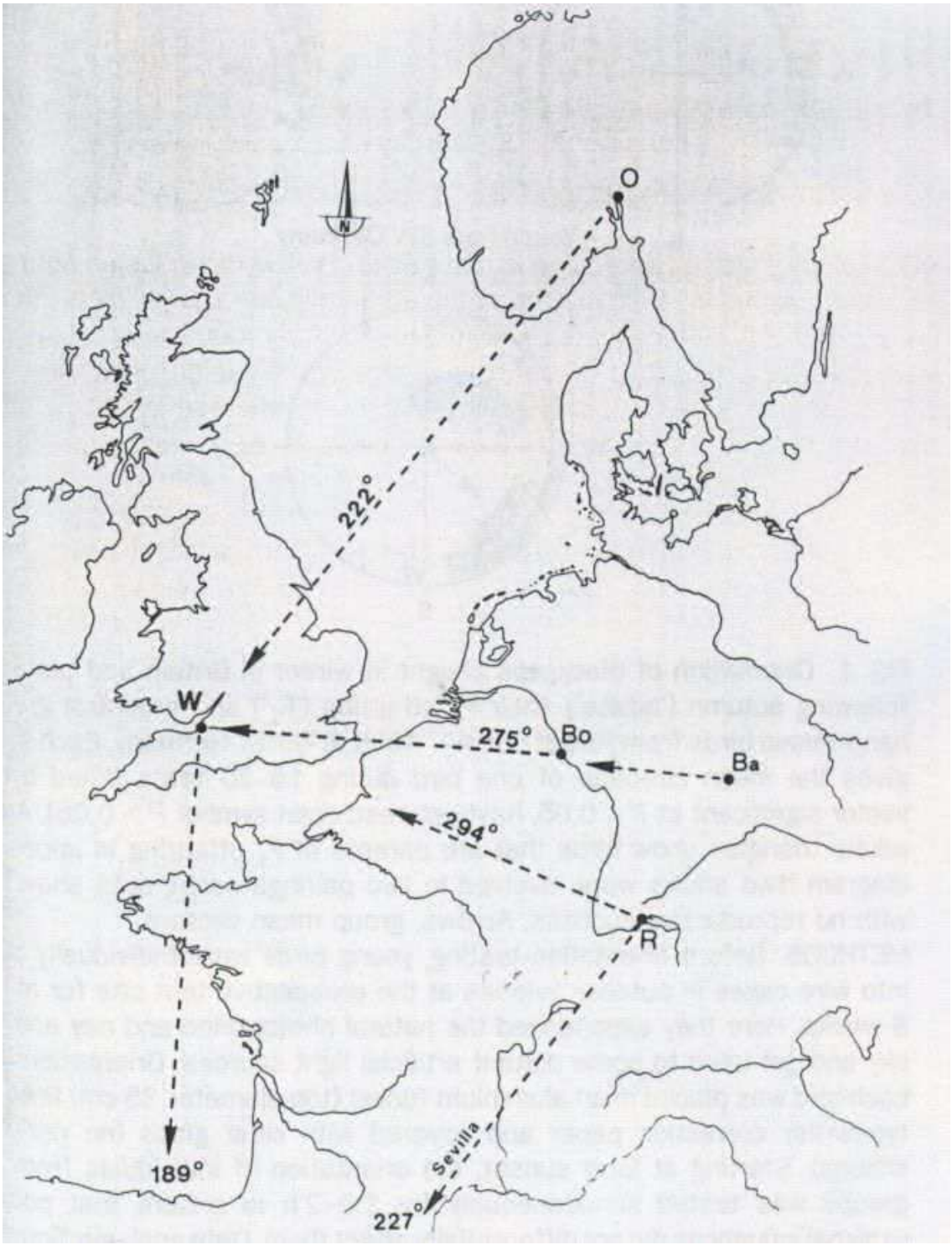
(B) F_1 offspring of British Adults
(mean 273°)



(C) Young from SW Germany
(mean 227°)

Both the British caught birds and their offspring raised in Germany moved to the WNW, whereas the German birds moved towards the SW.

These directions correspond to the new British migration route and the old Mediterranean route, respectively.



Furthermore, of the F1 birds, siblings were more likely to fly in a similar direction than non-sibling birds, providing additional evidence that the changes in migratory behavior are genetic.

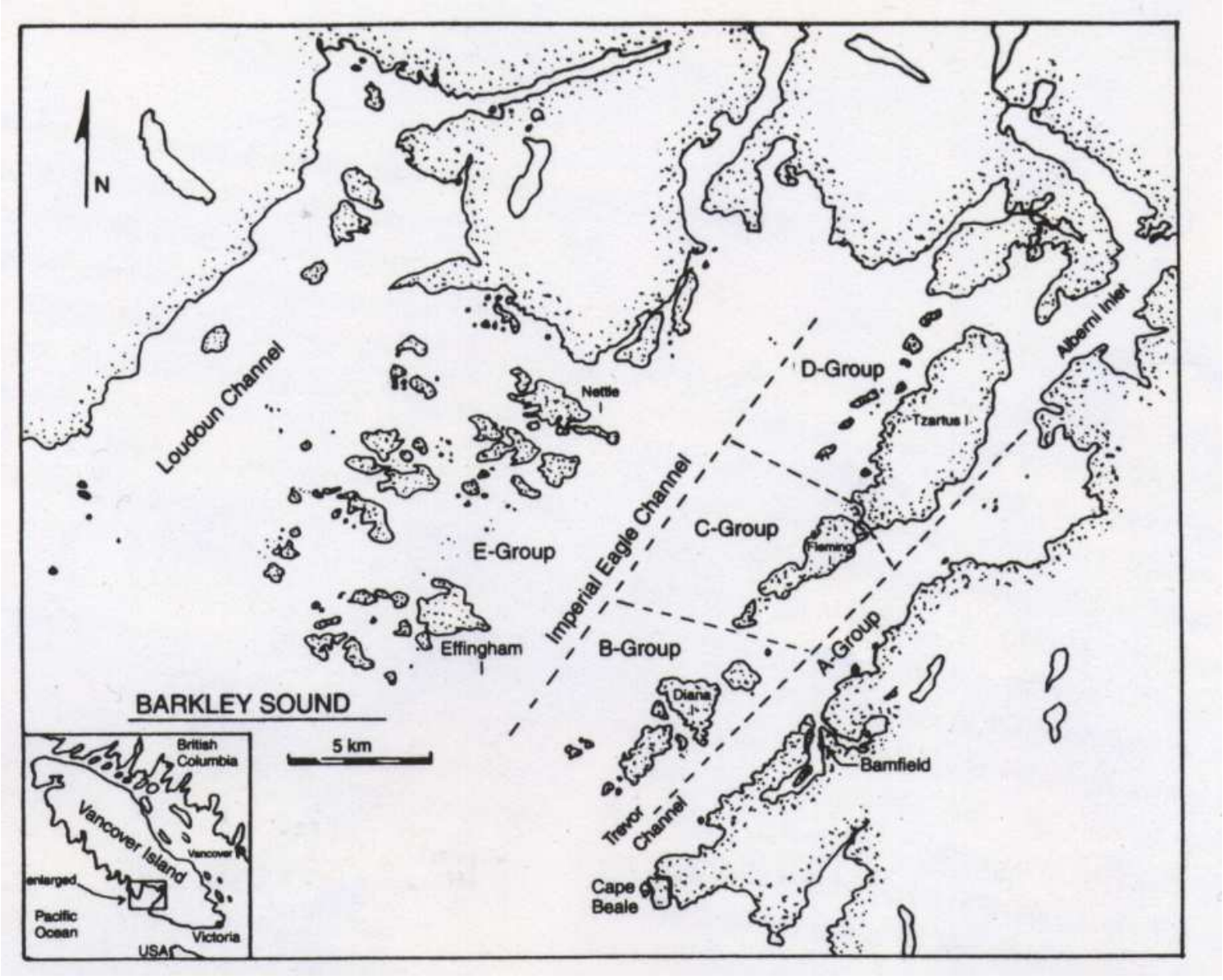
Although Berthold et al. do not know what forces have caused this evolutionary change, they suggest that natural selection has favored those blackcaps that have taken advantage of the warmer winters (climate change) and lowered competition in Britain.

SOURCES:

- Berthold et al. (1992) Rapid microevolution of migratory behaviour in a wild bird species. Nature 360: 668-670.
- Blackcap picture: http://www.univ-lehavre.fr/cybernat/pages/faunoi.htm

1. Evolution in Natural Populations
Dispersal in Island Plant Populations

Cody and Overton (1996) studied changes in seed size and shape that occurred in weedy, wind-dispersed plants located on the islands off the westcoast of Vancouver Island.



For eight summers between 1981 and 1991, Cody and Overton censused the plant populations of 200 islands and a region of the mainland.

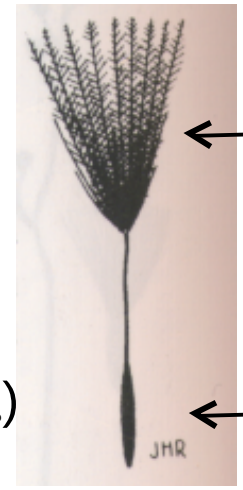
Extinction and recolonization events occurred frequently on the islands.

For a wind-dispersed plant, how might seeds that successfully colonize an island differ from the majority of seeds on the mainland?

After a plant has colonized an island, how might seed morphology evolve over time?

Cody and Overton studied two species that were common and whose seed dispersal was directly related to seed morphology: *Hypochaeris radicata* and *Lactuca muralis*·

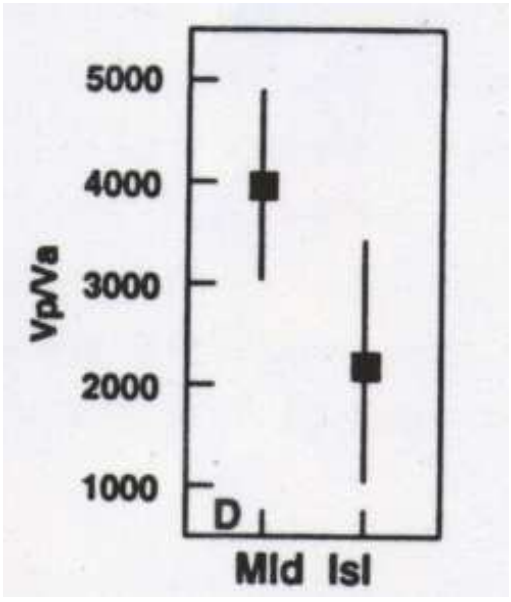
Both are asteraceae that produce achenes consisting of a feathery pappus and a seed.



Achene of Hypochaeris

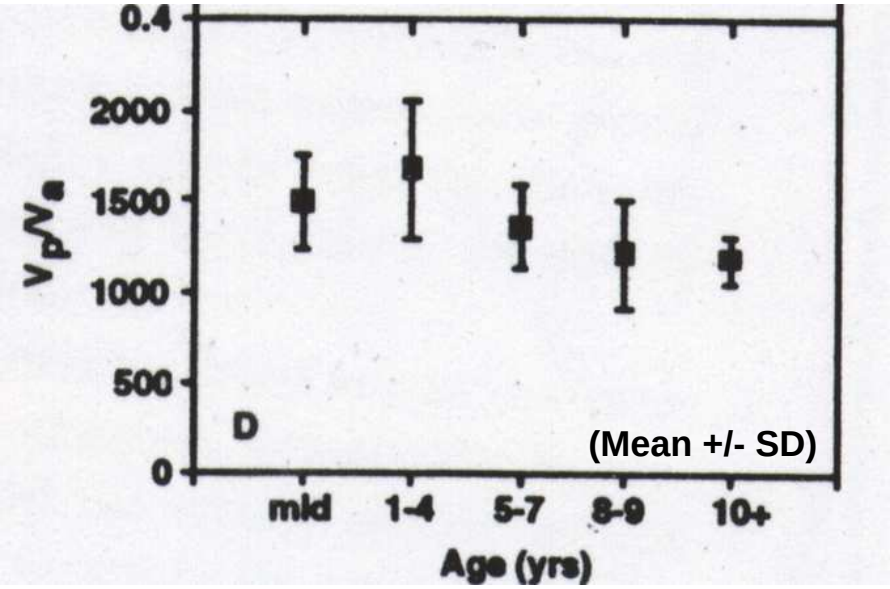
Dispersal ability of the achenes in these two species isrelated to the volume of the pappus (V_P) divided by thevolume of the achene (V_A).

Old island populations of *Hypochaeris radicata* have a significantly ($p<0.01$) lower dispersal ability than mainland populations.



(Mean +/- SD)

For *Lactuca muralis*, enough extinction and recolonization events have been observed to break down the changes in seed shape by the number of years that the plant has been on an island.



As predicted, plants that have recently arrived on an island produce seeds with greater dispersal ability than the mainland, but dispersal ability significantly decreases ($p<0.01$) as the age of the population increases.

The seed morphology of the island plant populations studied by Cody and Overton have evolved in a manner consistent with strong natural selection for dispersal ability among colonists and against dispersal ability among residents of these islands.

SOURCES:

- Cody and Overton (1996) Short-term evolution of reduced dispersal in island plant populations. J. of Ecology 84:53-61.
- Achene picture: Pojar and MacKinnon (1994) Plants of Coastal British Columbia. Lone Pine Publishing.

2. Evolution under Artificial Selection

Oil Content in Corn

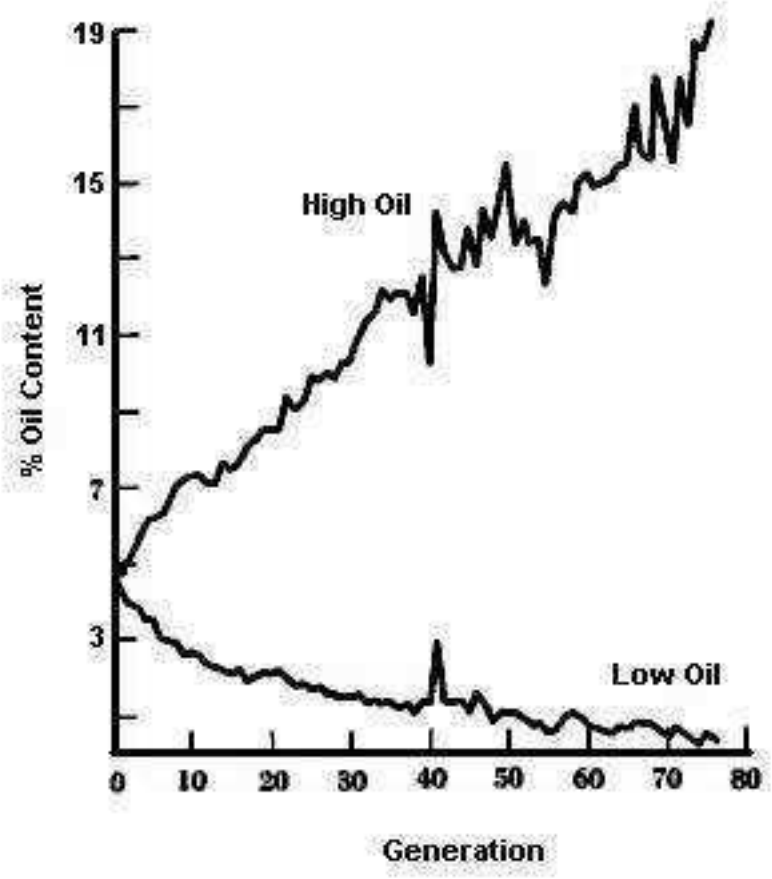
Artificial selection has been carried out on a variety of traits in a number of organisms.

Although some examples only a biologist could love, other examples have had an important impact on agriculture, including selection on

- birth weight, growth rate, and milk production in cows
- egg production in chickens
- back-fat in pigs
- grain yield in wheat

One of the longest running studies where evolutionary change has been documented began in corn in 1896 at the University of Illinois.

Two lines of corn were artificially selected. In one line, those plants with a high oil content were used as seed in the next generation. In the second line, plants with a low oil content were used as seed.



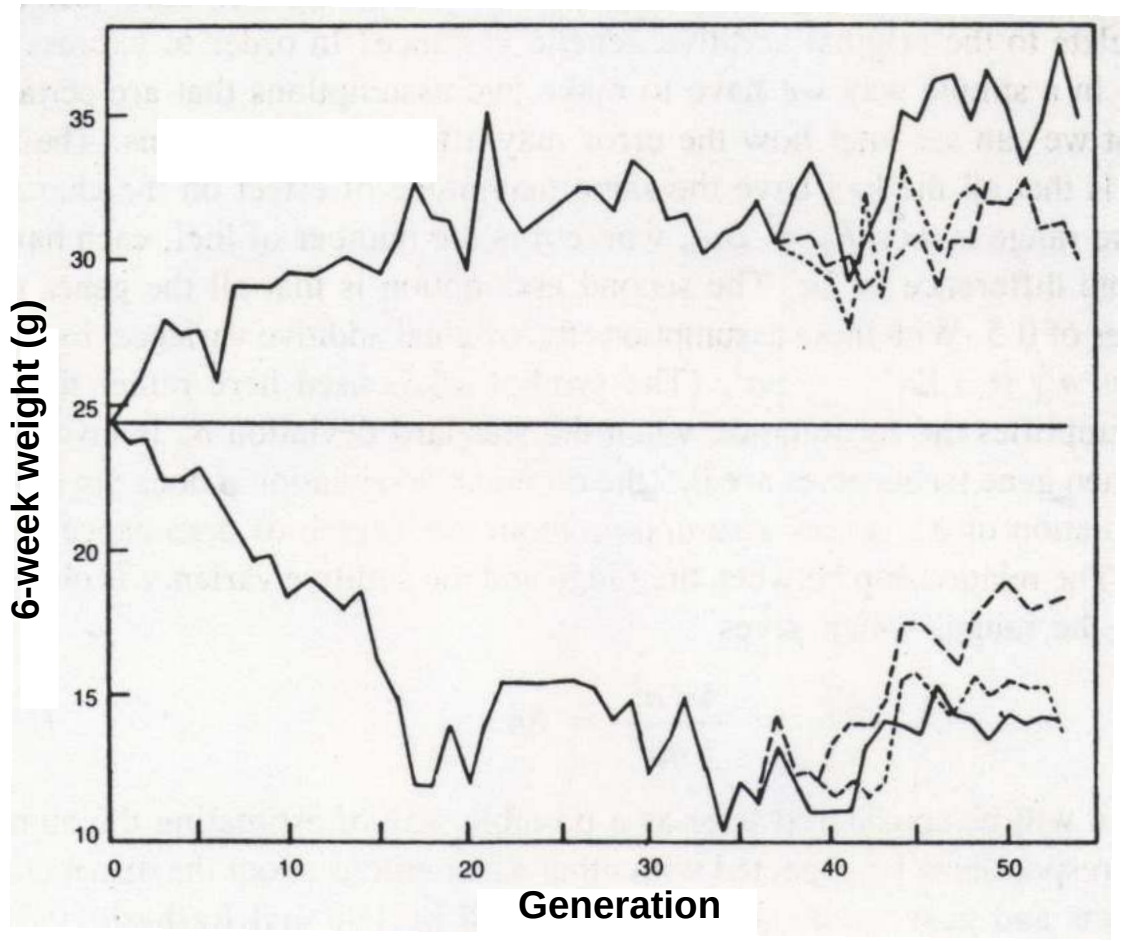
[Dudley and Lampert (1992)]

(Why has the lower line tapered off?)

Body Weight in Mice

More complicated traits also respond to artificial selection.

For example, the weight of mice at six weeks of age was selected, again in two separate lines (for heavier and for lighter mice).



[Roberts (1966)]

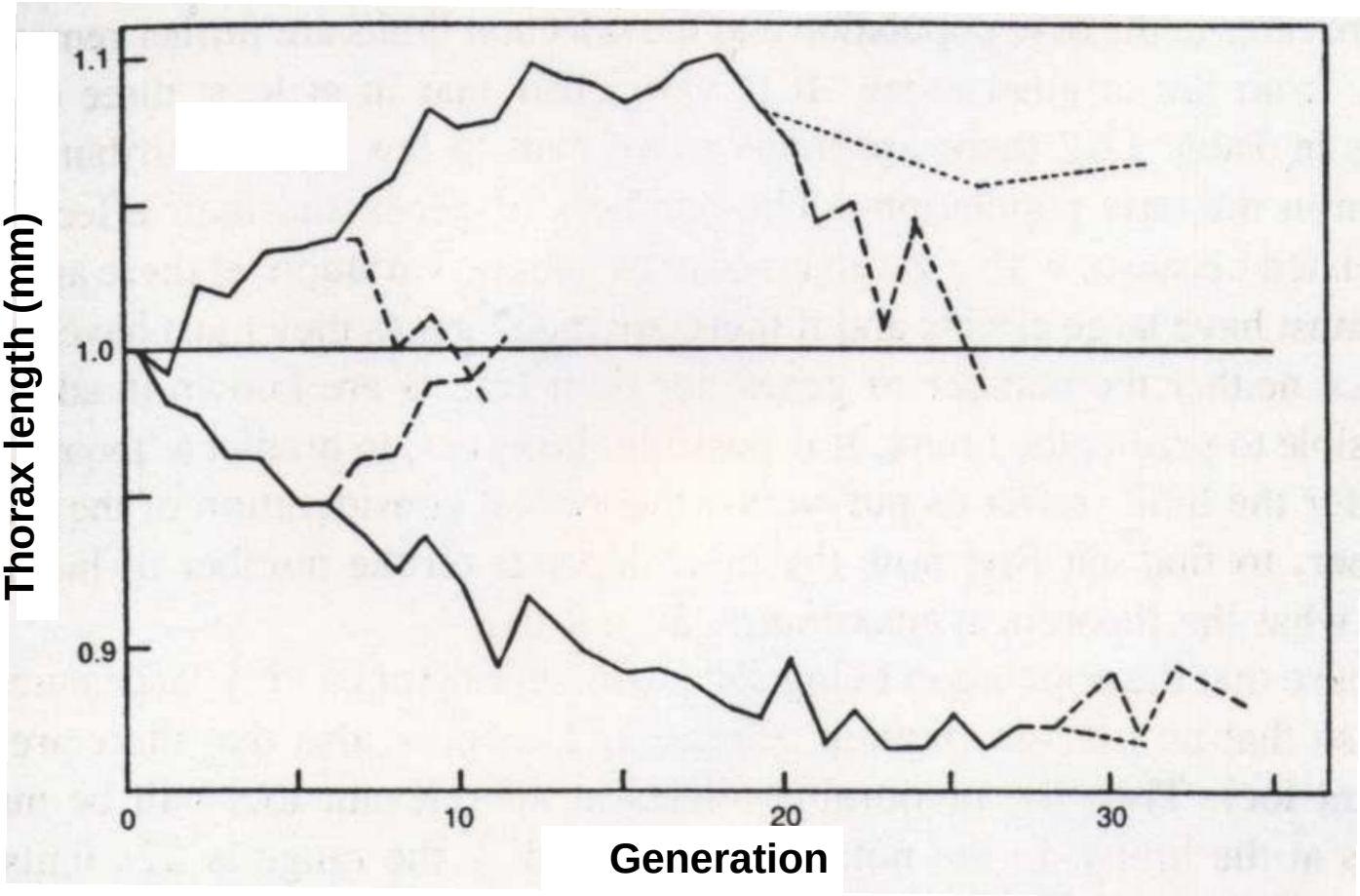
(The dashed lines refer to a subset of the two lines in which artificial selection was *reversed*.)

(The dotted lines refer to a subset of the two lines in which artificial selection was stopped.)

Thorax Length in Drosophila

Thorax length in adult Drosophila melanogaster shows a similar response to selection.

Here two separate lines were selected: one for longer flies and one for shorter flies.



[Robertson 1955]

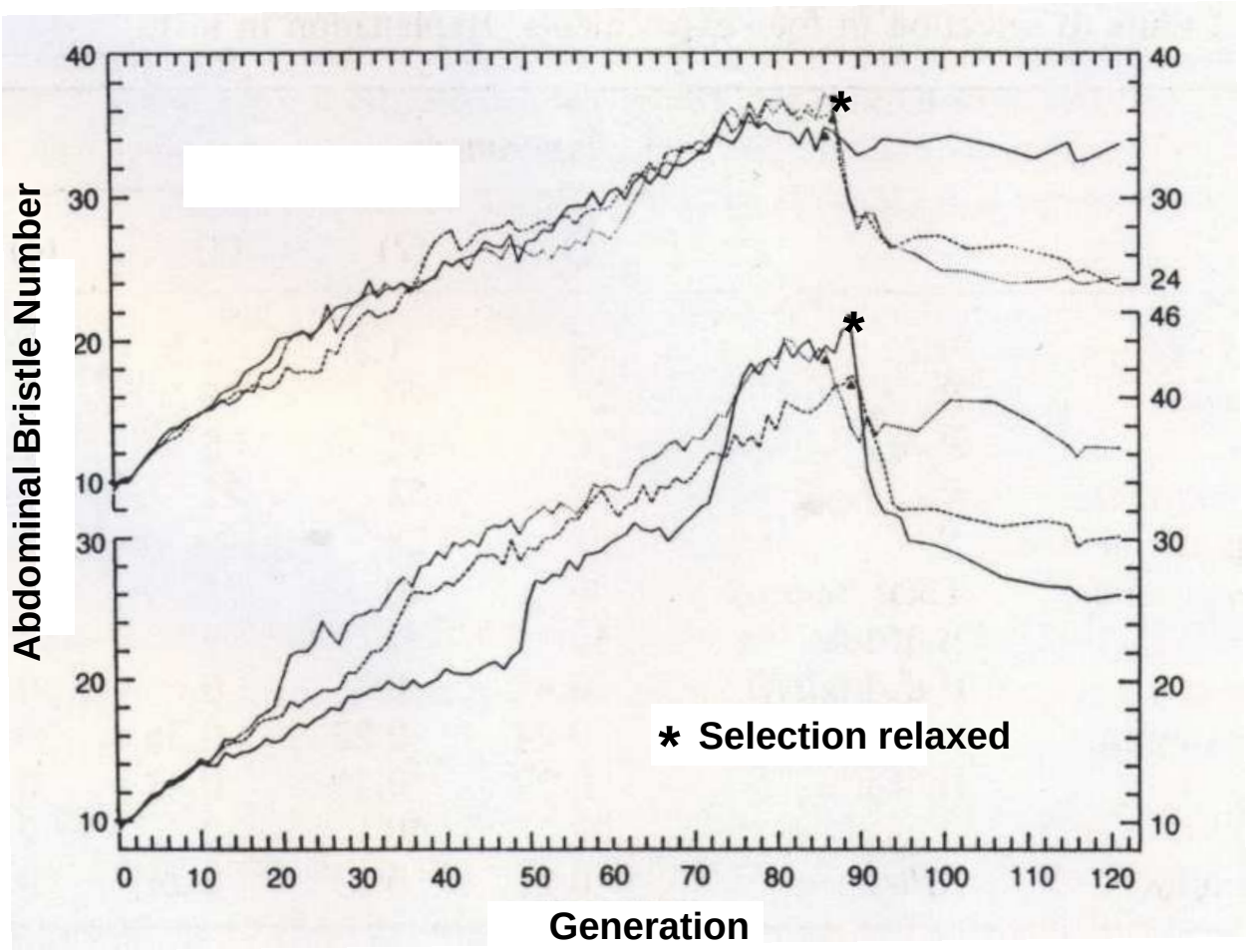
(The dashed lines refer to a subset of the two lines in which artificial selection was *reversed*

(The dotted lines refer to a subset of the two lines in which artificial selection was stopped.)

Abdominal Bristle Numbers

The number of bristles is a trait that is fairly easy to score in Drosophila melanogaster.

Yoo (1980) selected for an increased number of bristles in six replicate lines.



The number of bristles changed from around ten to around forty in 90 generations!

After about 90 generations, selection was then stopped (*).

(Why would the number of bristles decrease after artificial selection was stopped?)

➡ As long as the initial population is genetically variable, artificial selection is almost always successful and the trait under selection changes over time.

Even if you start with a genetically homogeneous population, artificial selection will still work, but it takes longer since selection can only act on the new mutations that occur.

For example, Mackay et al (1994) selected on abdominal bristle number in a highly inbred line of Drosophila (=extremely low in genetic variability).

Nevertheless, over 120 generations the high and low lines differed by 12 bristles on average!

OK - So a trait changes over time under selection, but could that ever lead to two different species?

Habitat Selection in Drosophila

Rice and Salt (1988, 1990) designed an experiment to test the hypothesis that selection could drive speciation among sympatric populations.

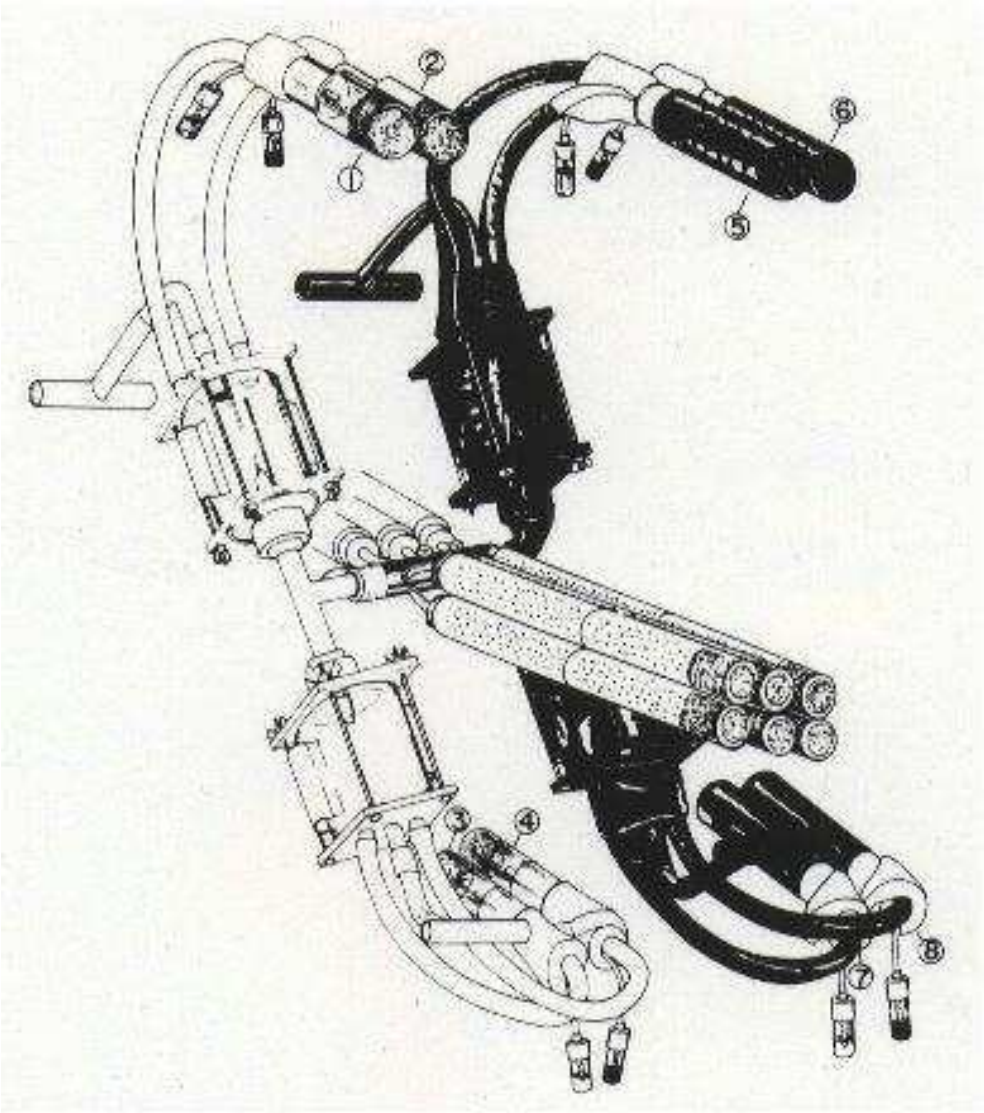
Definitions: from Bush and Howard (1986)

Species: "A group of populations whose evolutionary pathway is distinct and independent from that of other groups...achieved by the group's reproductive isolation from other groups."

Sympatry: "Two populations are sympatric if individuals of each are physically capable of encountering one another...with moderately high frequency."

Allopatry: "Allopatric populations are separated by uninhabitated space...across which migration (movement) occurs at a very low frequency."

Rice and Salt constructed an ingenious maze within which larval flies were placed.



Within the maze, flies could choose

- lightness or darkness by moving left or right [selection for phototaxis]
- up or down [selection for geotaxis]
- acetaldehyde (white vial) or ethanol (black vial) [selection for chemotaxis]

In addition, flies were collected from the eight "habitats" during three time periods: early (E), middle (M), and late (L) [selection for development time].

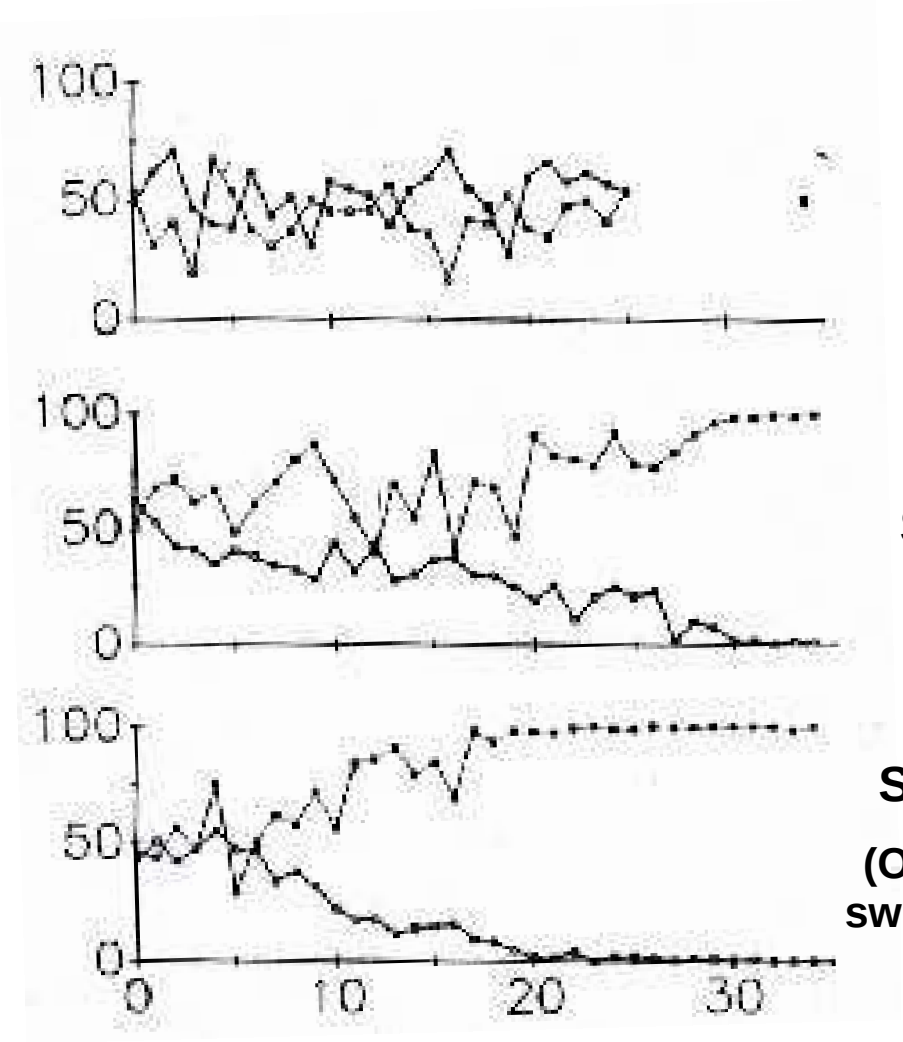
Flies were allowed to mate within the maze (females tend to mate only when they have located food).

For the control lines, all flies within the habitats were mixed and 120 females chosen.

For the selected lines, 60 females flies were drawn from habitat 5E (dark, up, acetaldehde) and 60 from habitat 4L (light, down, ethanol).

Larvae from the experimental females were mixed and placed together in the maze to start the next generation. Controls were run through the maze separately.

(Offspring of mothers collected from 5E and half of the controls were raised on a chemical that turned their eyes brown.)



Control

Selection

Selection +

(Offspring that switched habitats destroyed)

Over the 35 generations of the experiment, habitat specialization evolved in the selected lines:

- the offspring of females collected from habitat 5E (solid squares) were more likely to return to 5E
- the offspring of females collected from 4L (empty squares) were more likely to return to 4L.

No habitat specialization evolved in the control flies.

Since females tend to mate near the food vials, gene flow between 5E and 4L flies had virtually ceased by the end of the experiment.

The first step of speciation had occurred!

SOURCES:

- Rice and Salt (1988) Speciation via disruptive selection on habitat preference: Experimental evidence. American Naturalist 131: 911-917.
- Rice and Salt (1990) The evolution of reproductive isolation as a correlated character under sympatric conditions: Experimental evidence. Evolution 44: 1140-1152.

3. Evolution Makes Sense of Homologies

Richard Owen (1848) introduced the term *homology* to refer to structural similarities among organisms.

To Owen, these similarities indicated that organisms were created following a common plan or archetype.

That is, although each species is unique, the plans for each might share many features, just as the design plans for a Honda Civic and a Honda Prelude might be similar.

Nevertheless, if every organism were created independently, it is unclear why there would be so many homologies among certain organisms, while so few among others.

It is also hard to make sense of the fact that homologous structures can be inefficient or even useless.

Why would certain cave-dwelling fish have degenerate eyes that cannot see?

Darwin made sense of homologous structures by supplying an evolutionary explanation for them:

> *A structure is similar among related organisms because those organisms have all descended from a common ancestor that had an equivalent trait.*

Ridley uses a specific definition of homology:
"A similarity between species that is not functionally necessary."

I interpret this as: "A similarity between species that exists despite several plausible alternative traits that would function equally well."

3. Evolution Makes Sense of Homologies
The "Universal" Genetic Code

5’	T	C	A	G	3’
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)	**T**
	TTC Phe (F)	TCC Ser (S)	TAC Tyr (Y)	TGC Cys (C)	**C**
	TTA Leu (L)	TCA Ser (S)	**TAA Stop**	**TGA Stop** [*Trp (W)*]	**A**
	TTG Leu (L)	TCG Ser (S)	**TAG Stop**	TGG Trp (W)	**G**
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)	**T**
	CTC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)	**C**
	CTA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)	**A**
	CTG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)	**G**
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)	**T**
	ATC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	**C**
	ATA Ile (I) [*Met (M)*]	ACA Thr (T)	AAA Lys (K)	AGA Arg (R) [*Stop*]	**A**
	ATG Met (M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R) [*Stop*]	**G**
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)	**T**
	GTC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	**C**
	GTA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	**A**
	GTG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	**G**

[*Italics indicate the mammalian mitochondrial code.*]

The genetic code for protein-coding genes is nearly universal in eukaryotes and prokaryotes.

The exceptions include most mitochondrial genomes and some nuclear ones (e.g. Mycoplasma and Tetrahymena).

Even in these cases, the genetic code is quite similar.

Millions of alternative genetic codes exist, so why do all organisms have nearly the same one?

Since the anti-codon is at the opposite end from the amino acid binding site of a tRNA and does not interact with the binding site, there is no chemical necessity for a codon to be assigned to a particular amino acid.

The genetic code is homologous among living organisms: it is similar despite the fact that there exist many equally good genetic codes.

Under the hypothesis that evolution has occurred, however, the similarity among all genetic codes makes sense:

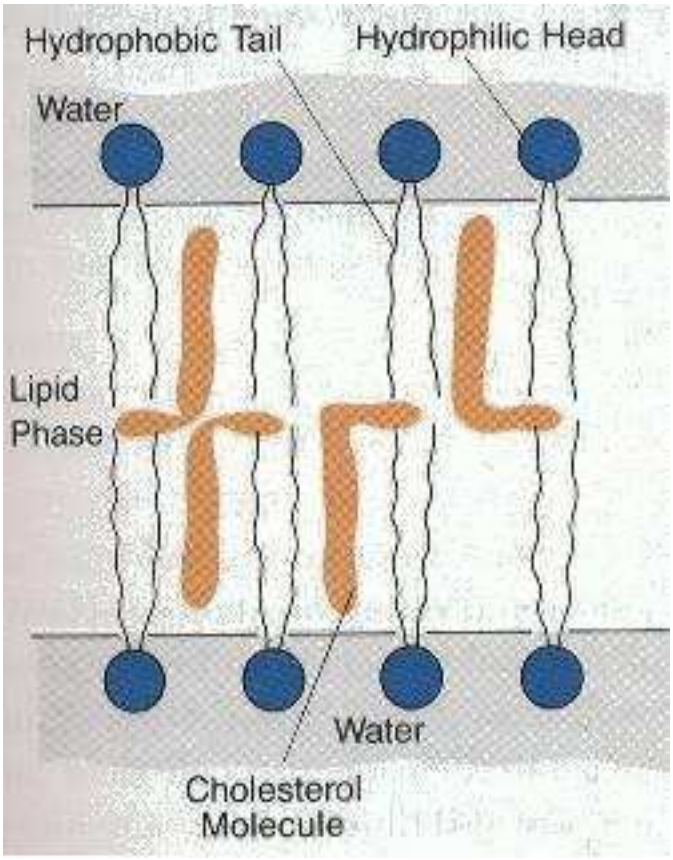
> The common ancestor to all known organisms had a genetic code similar to what we see today.
>
> Over the ages, the genetic code has passed unchanged (or nearly so) from parents to offspring, because mutations to the genetic code would have been disastrous (changing the amino acid sequence of all proteins produced).

(What would an evolutionist think if an organism were found today with an entirely different genetic code?)

3. Evolution Makes Sense of Homologies

Plasma Membrane

Similarly, the plasma membranes of all organisms, eukaryotic and prokaryotic, are structurally similar, consisting of a phospholipid bi-layer.



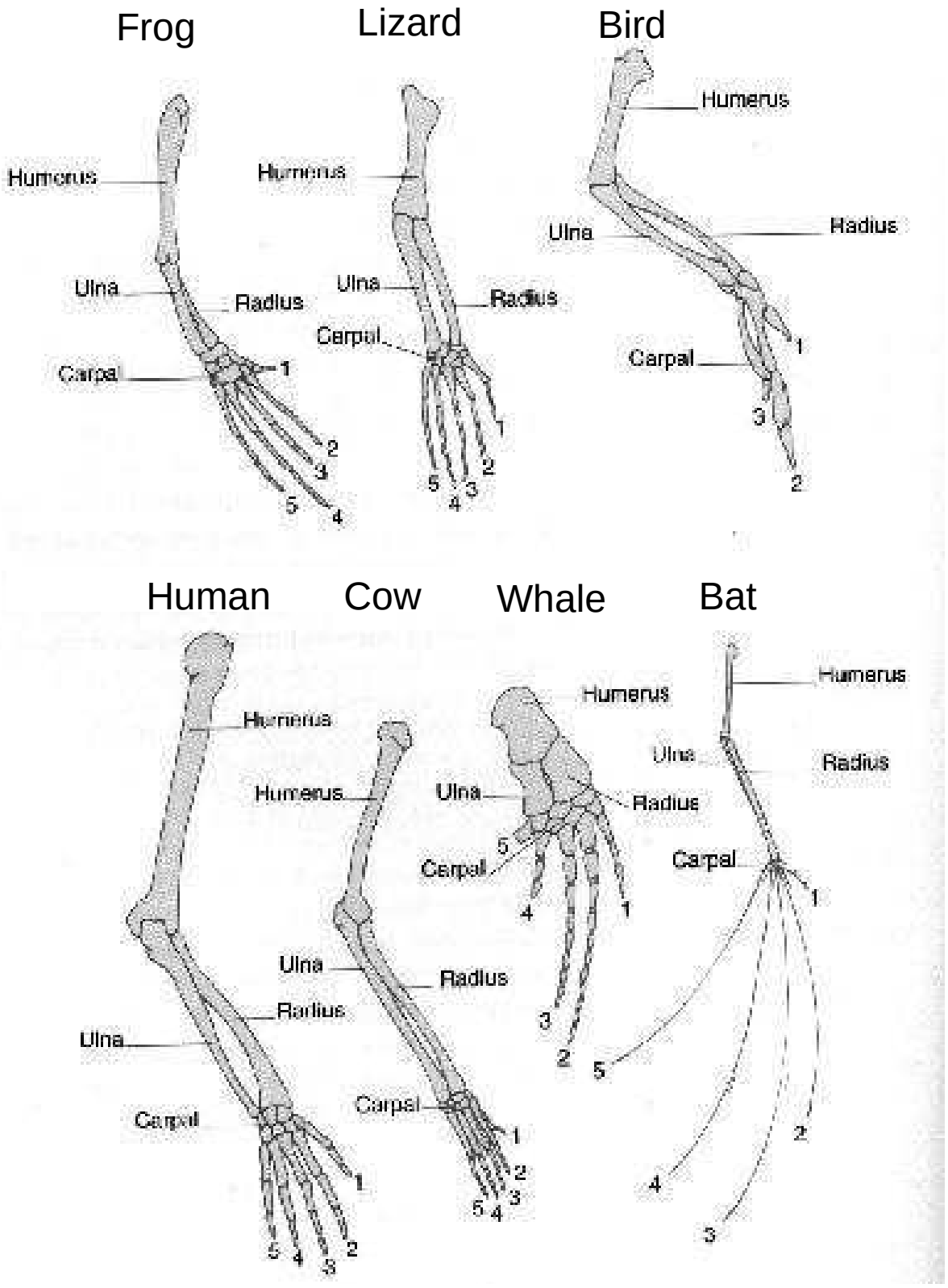
Many other possible membrane structures exist.

The hydrophobic fatty acid tails could be joined. There could be three hydrophobic fatty acid chains. Other hydrophilic groups could be involved besides glycerol phosphoric acid.

The similarity of the plasma membrane (as well as other cell structures) suggests that all living cells have descended from an ancestor with a similar membrane structure.

3. Evolution Makes Sense of Homologies
Pentadactyl Limb

One of the classic examples of a homologous structure is the pentadactyl (= five digit) limb.



All tetrapods (= four legged) have limbs with five digits, at least at some stage in development.

Certain tetrapods lose some of these digits during development, as in the bird wing shown here.

But if the bird wing does not need five digits, why do five initially develop in the growing embryo?

The most plausible explanation is that while the five digits are not functionally necessary, they represent a genetic artefact inherited from the ancestors of birds.

3. Evolution Makes Sense of Homologies

New Functions Evolve from Pre-existing Structures

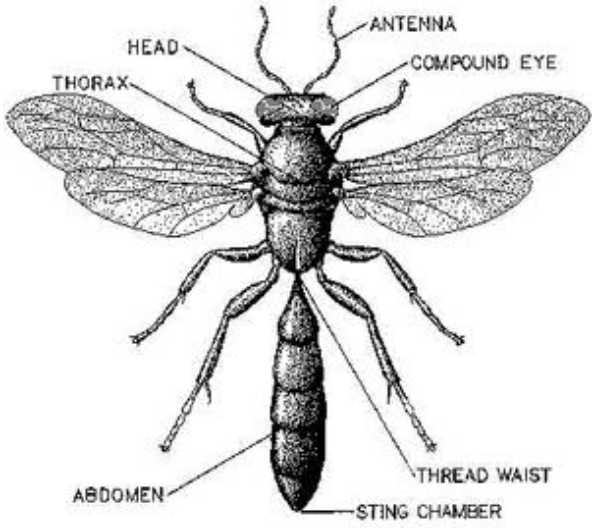
Homologous structures teach us an important lesson about evolution:

> Evolution works primarily by modifying pre-existing structures.

That is, even when two species function in completely different ways, they often use homologous structures to carry out those functions.

For example, birds and bats fly rather than run on all fours, yet their wings are modified fore-limbs rather than completely novel structures.

Similarly, the stinger of wasps and bees is a modified ovipositor, rather than an entirely new structure. (Explaining why only female bees sting!!)



The organ pip mud dauber, Typragilum politum.
Illustration by Bernice B. DeMarco.

3. Evolution Makes Sense of Homologies

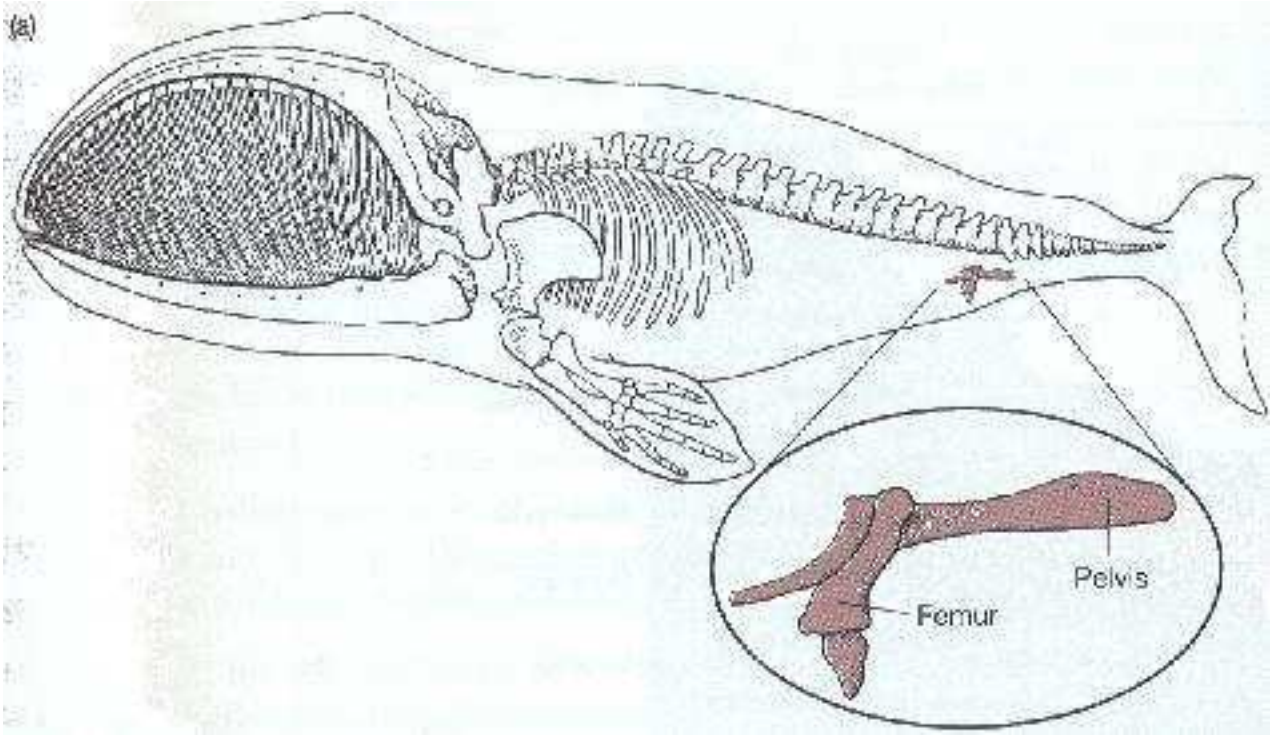
Vestiges

Structures that are functionless in a species but homologous to a functioning structure in other species are particularly difficult to explain except under the theory of common descent.

Such structures are known as vestigial structures.

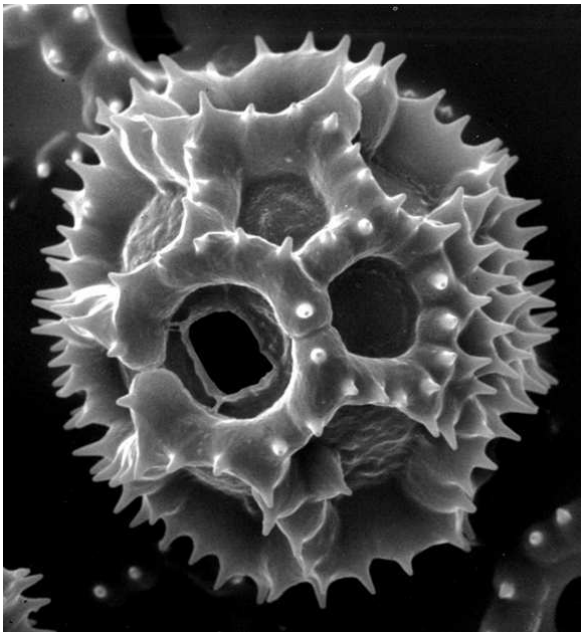
For example:

The pelvic girdle of a whale



The eye bulbs of blind, cave-dwelling creatures, such as the grotto salamander (Typhlotriton spelaeus):

The anthers and pollen of asexual dandelions:



Dandelion (Taraxicum officinale)
http://usda-apmru.tamu.edu/pollen/graphics/Taraxicum_officinale.htm

Vestigial structures are extremely illogical if each creature were independently created, but make sense if organisms inherit traits from their ancestors with gradual modification over time.

4. Homologies are Nested

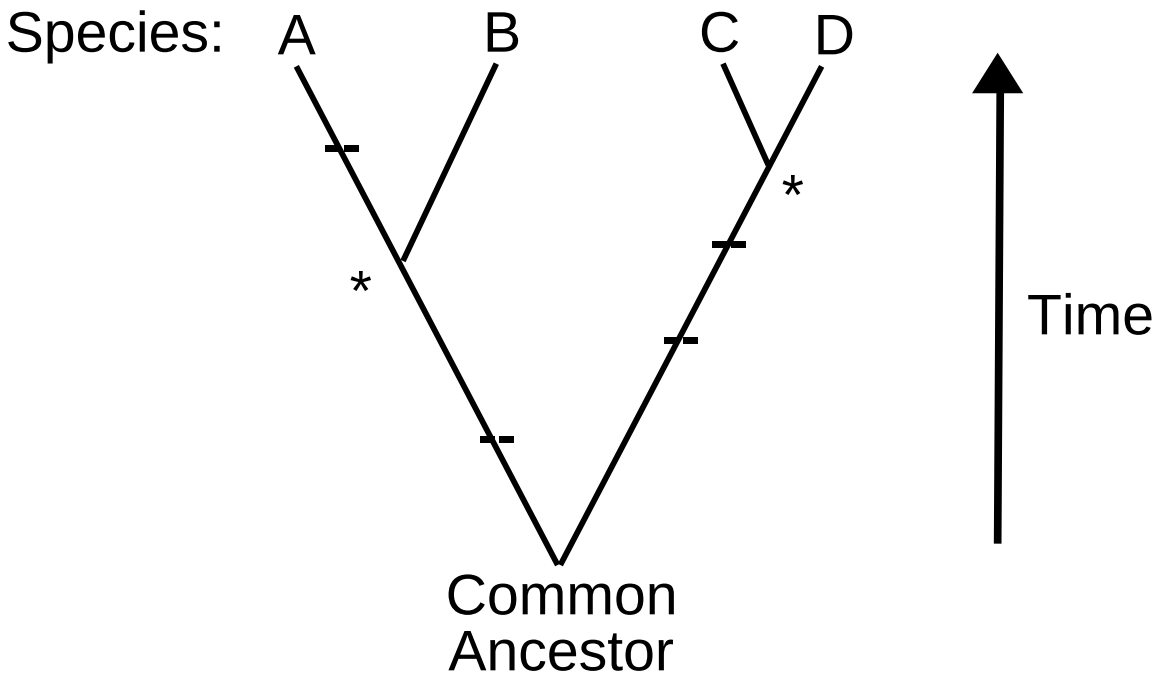
One of the striking features about similar structures are that they cluster.

Certain species share many similarities at every level of organization (e.g. humans and chimpanzees), whereas other species only share certain nearly universal homologies (e.g. humans and Escherichia coli).

This observation makes little sense for created objects, since a creator could mix and match features observed in any organism.

By contrast, under the theory of evolution, we would expect the number of shared homologies to be high for closely related species and to decrease over time as species diverge from each other.

Consider a sequence of DNA in four species:



(* = Speciation event. -- = Mutation)

Here, species C and D would have identical sequences, species A and B would only differ by one mutation, whereas A-C and A-D would differ by four mutations and B-C and B-D by three mutations.

In fact, "relatedness" among organisms is determined by how many (and which) features they share.

4. Homologies are Nested

Primate Phylogeny

Although different traits may reflect the evolutionary history of a group of species slightly differently, there should be rough correspondence.

For example, morphologically, humans and chimps are much more similar than are humans and gibbons.

If this is because humans and chimps are more closely related (=have had a common ancestor more recently), then we would expect the DNA sequences of humans and chimps to be more similar as well.

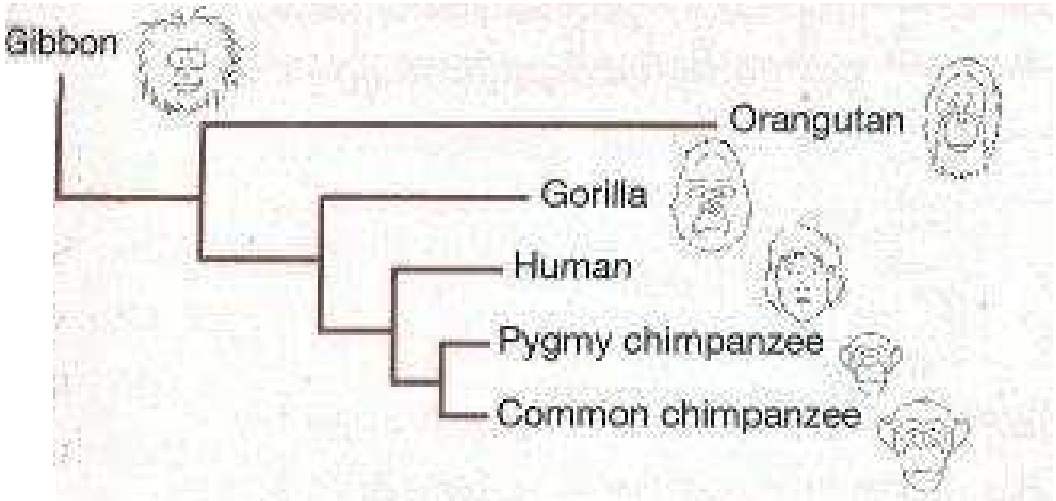
Three phylogenetic trees were reconstructed based on the DNA sequences of:

(a) 4700 bp of mitochondrial DNA,

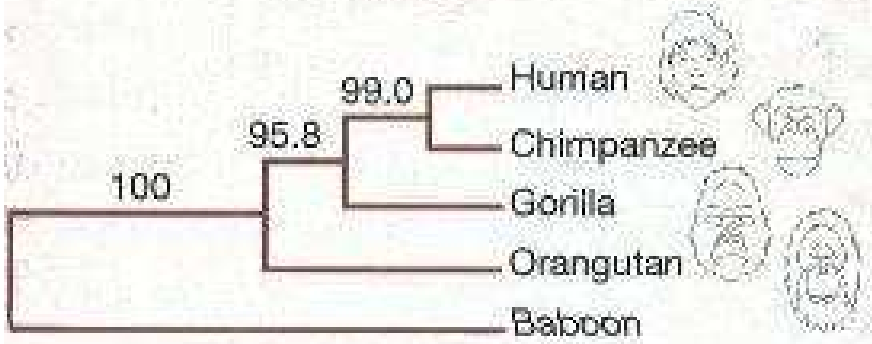
(b) the testis specific protein on the Y chromosome,

(c) noncoding regions of the β-globin gene.

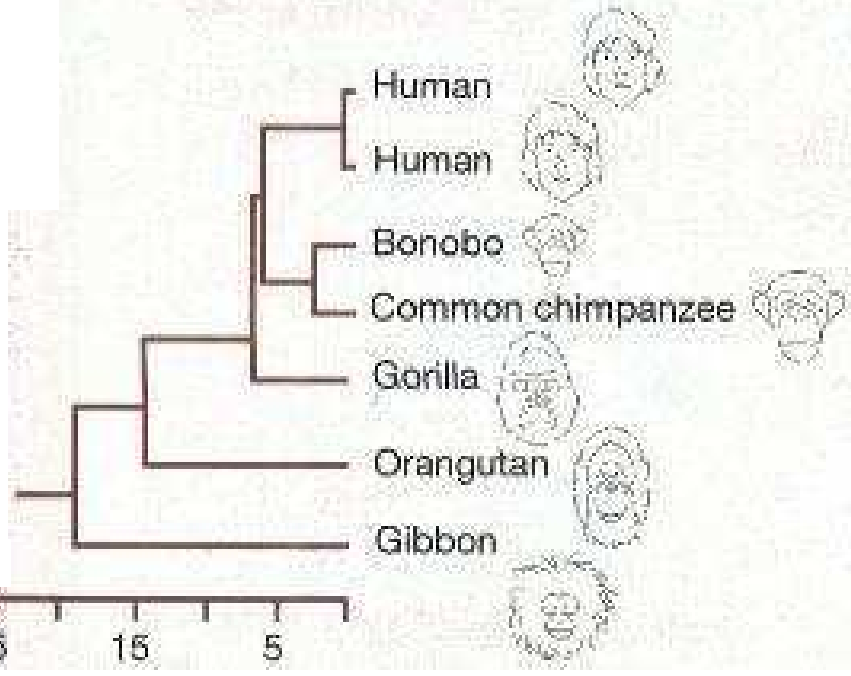
(a)



(b)



(c)



MYA: 25 15 5

All three trees show that human and chimp DNA sequences are more similar on average than are human and gibbon or human and orangutan sequences.

4. Homologies are Nested

Bird Phylogeny

Homologous characters can also be identified in fossil organisms, where they also tend to exhibit a nested series rather than non-overlapping categories.

Among dinosaur fossils, some taxa share few traits in common with birds while others share several.

In particular, maniraptors share more traits in common with modern birds than do any other type of dinosaurs.

Similarly, coelurosauria besides maniraptors share some, but not as many features of birds.

This nested series is thought to represent the evolutionary trajectory of dinosaurs along the lineage that led to birds.

Here are some of the traits that categorize the various groupings of archosaurs (including crocodiles, dinosaurs, and birds):

Archosauria

- hard-shelled eggs
- single openings in each side of the skull in front of the eyes (antorbital fenestrae)
- openings in the lower jaw (the mandibular fenestra)
- a high narrow skull with a pointed snout
- teeth set in sockets

Dinosauria

- reduced fourth and fifth digits on the hand
- foot reduced to three main toes
- three or more vertebrae composing the sacrum
- an open hip socket

Saurischia

- a grasping hand
- asymmetrical fingers
- long, mobile neck
- pelvis with a pubis that points downward and forward

Therapoda

- hollow, thin-walled bones

Coelurosauria

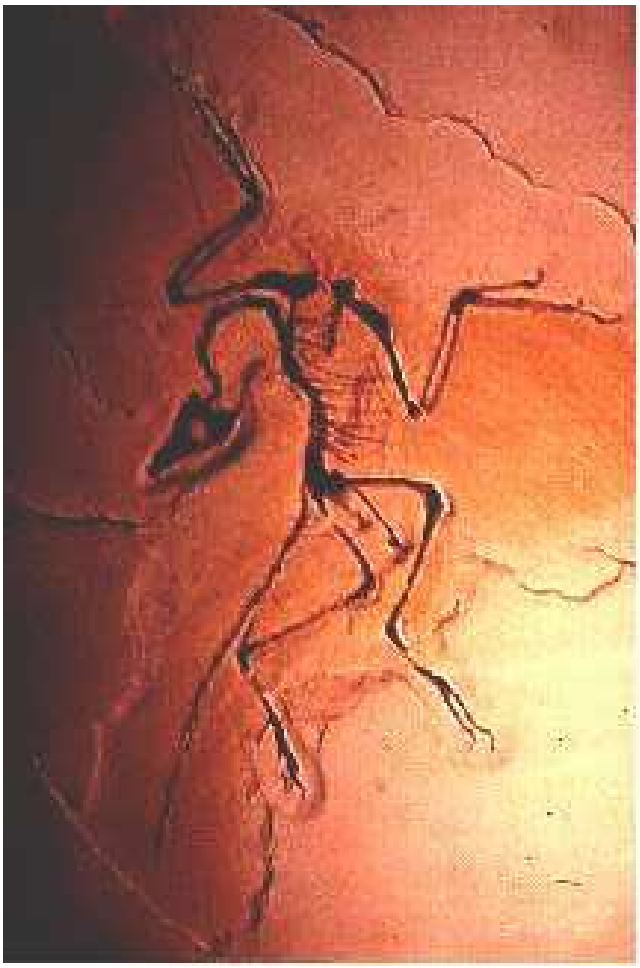
- elongated arms
- well-developed hinge-like ankles
- (some coelurosaurs have open nests like birds in which eggs were brooded)

Maniraptora

- the semilunate carpal present in wrist
- modified forelimb (elevated forelimb capable of folding)
- a fused collar bone and sternum
- pelvis with a pubis that points downwards

Aves

- feathers
- wings
- wishbone
- opposable big toe



The earliest known fossil bird (Archaeopteryx) shares these features, but also has many features of therapods including:

- teeth
- extensive tail vertebrae
- claws
- lack of a keeled breastbone

Archaeopteryx probably did not fly. It lacked the prominent keel of modern birds upon which flight muscles attach. (It may have been a glider.)

Crocodilians

Birds

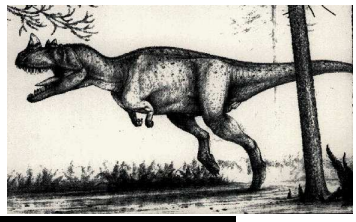
Archaeopteryx

Deinonychus

Aves

Compsognathus

Maniraptora



Ceratosaurus

Coelurosauria

Diplodocus

Therapoda

Triceratops

Saurischia

Dinosauria

Archosauria

5. Evidence for Evolution from the Fossil Record

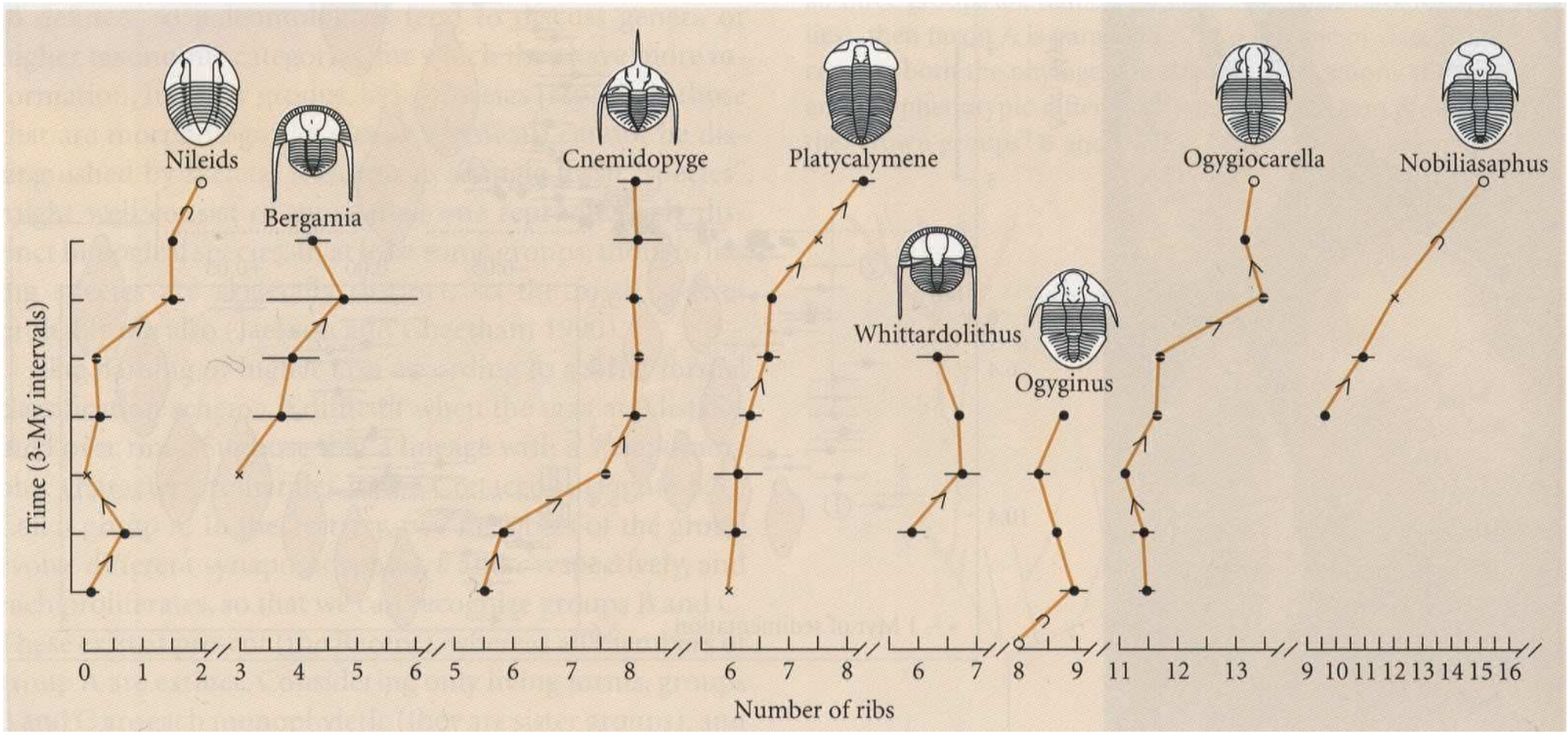
Trilobite Evolution

Although the fossil record is often poor and incomplete, there are certain deposits where sedimentary layers remain in a nearly continuous series.

Fossils from these series provide direct evidence of evolutionary change.

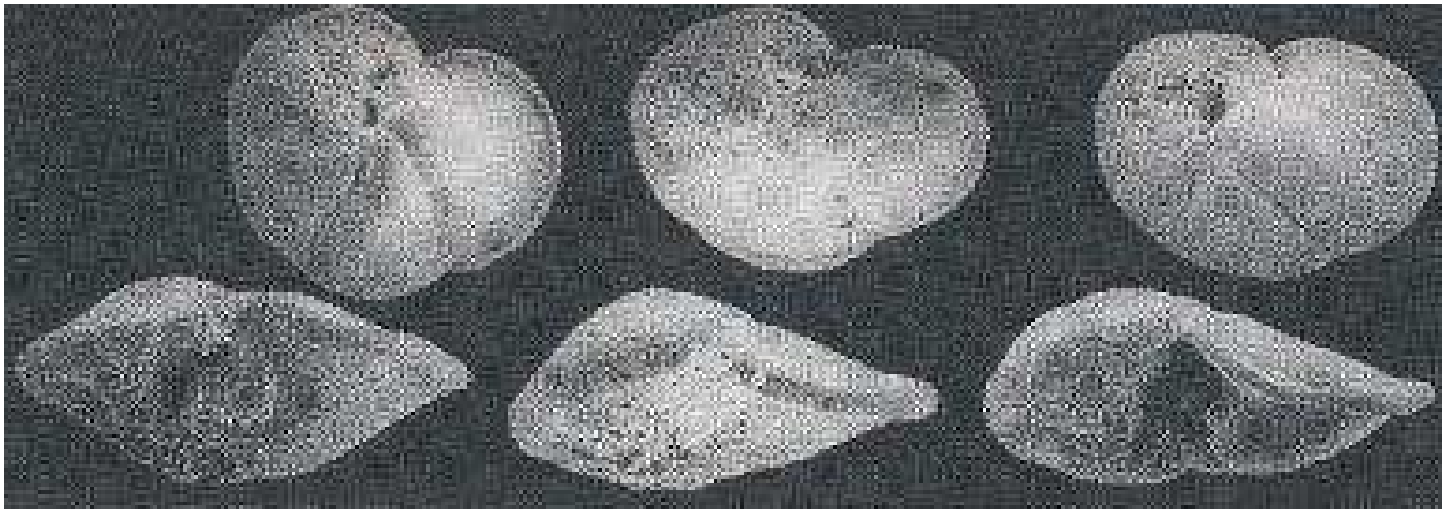
Sheldon (1987) examined a series of sedimentary layers from the Ordovician period (500 MYA) containing trilobite fossils (extinct marine arthropods).

Some of these changes over time were so large that the animals at the end of the series are assigned to a new genus!


Nileids
Bergamia
Cnemidopyge
Platycalymene
Whittardolithus
Ogyginus
Ogygiocarella
Nobiliasaphus
Time (3-My intervals)
0 1 2 3 4 5 6 5 6 7 8 6 7 8 6 7 8 9 11 12 13 9 10 11 12 13 14 15 16
Number of ribs

5. Evidence for Evolution from the Fossil Record
Foraminiferan Evolution

An even finer scale analysis was performed by Malmgren et al. (1983) on a species of foraminiferan (shell-bearing protozoans) from 10MYA to recent times.

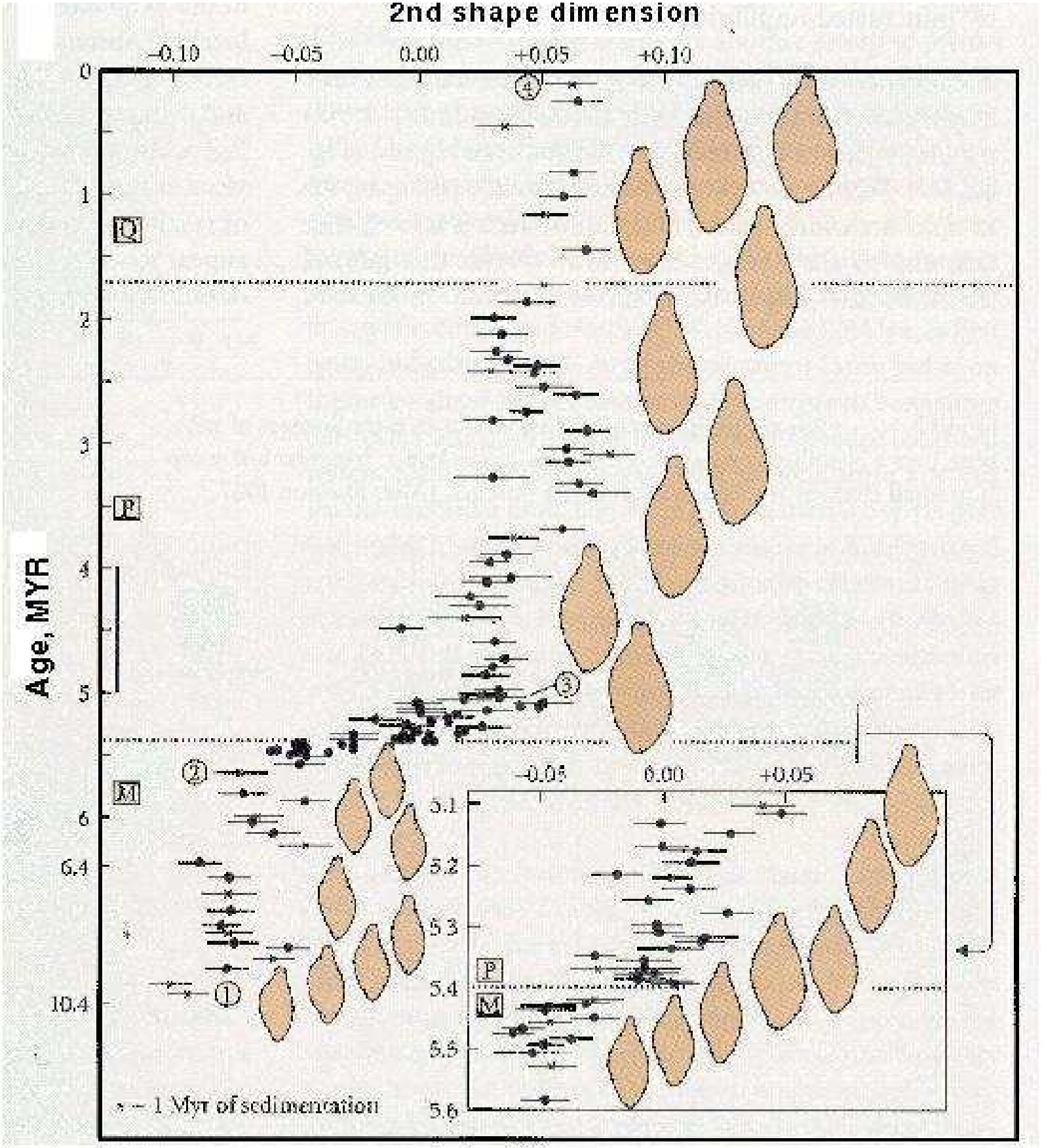


[Three epochs are represented:

Miocene (M),	Pliocene (P),	Pleistocene (Q)
23.8-5.2 MYA;	5.2-1.8 MYA;	1.8 MYA - 10,000 YA].

Over this period, the fossil shells evolved a larger, thicker shell, with a more pronounced ridge.

Although the fossil record demonstrates that change occurred in a continuous manner (=without breaks or jumps), the rate of change was not always the same: shape changed most around the Miocene/Pliocene boundary.


2nd shape dimension
−0.10
−0.05
0.00
+0.05
+0.10
Age, MYR
0
1
2
3
4
5
6
6.4
10.4
Q
P
M
1
2
3
4
−0.05
0.00
+0.05
5.1
5.2
5.3
5.4
5.5
5.6
P
M
× – 1 Myr of sedimentation

These changes were large enough that the lineage is assigned to the species *Globorotalia plesiotumida* in the Miocene, but to the species *Globorotalia tumida* afterwards.

The fossil record demonstrates evolutionary changes do occur.

The disadvantage of the fossil record is that it is generally difficult to determine the selective forces that may have contributed to these changes.

The advantage of the fossil record over present-day observations of evolution is that higher order evolutionary changes may be tracked (e.g. the origin of new species, new genera, etc).

SOURCES:

- Pentadactyl limbs: Ridley (1997) Evolution.
- Whale, salamander, primate trees: Freeman and Herron (1998) Evolutionary Analysis.
- Membrane photo: Wessells and Hopson (1988) Biology.
- Dinosaur information: UC Museum of Paleontology
- Trilobite and foraminiferan fossil record: Futuyma (1998) Evolutionary Biology.

Mechanisms of Evolutionary Change

Biologists have gained an understanding of evolutionary change from primarily two sources:

- Observing evolutionary change
- Deducing how evolutionary change must happen given the biological processes at work

Making these Deductions

We will look at a number of models that have been constructed to investigate how populations should change over time under particular forces:

- selection
- mutation
- random genetic drift (= change due to sampling error)

Results from these models provide us with a clearer understanding of how evolutionary change is accomplished and with specific, quantitative predictions that can be more easily tested.

Forms of Selection

To determine how evolution will occur under natural selection requires specifying how selection acts on a trait.

There are three general classes of selection:

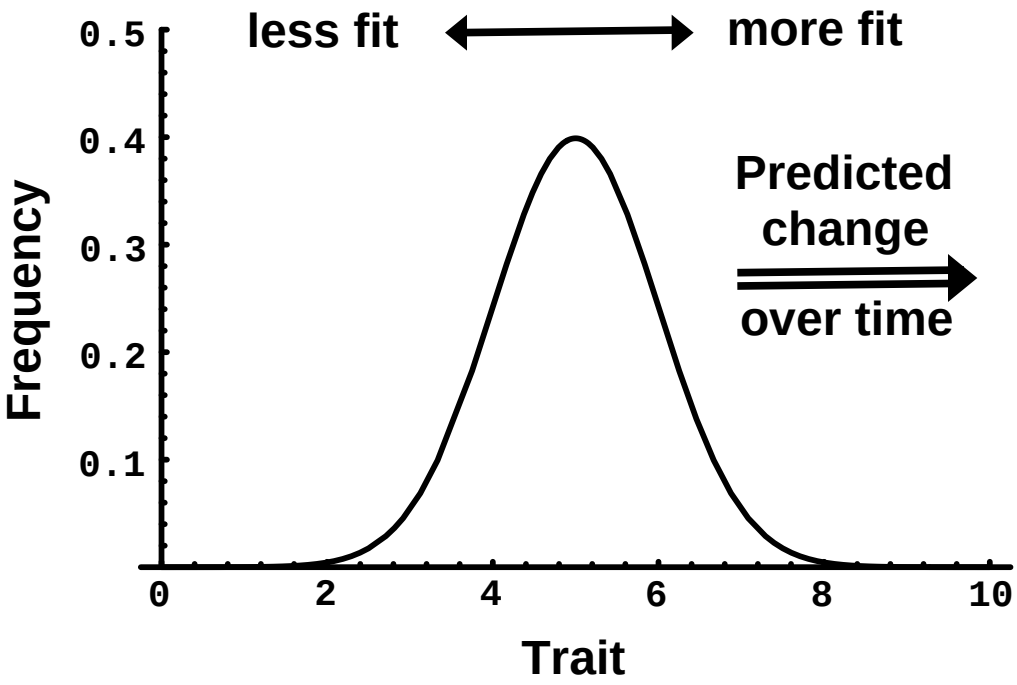
Directional selection = Selection for a higher or lower value of a particular trait

- **UP =** Individuals with more of a certain trait are more likely to survive and reproduce
- **DOWN =** Individuals with less of a certain trait are more likely to survive and reproduce

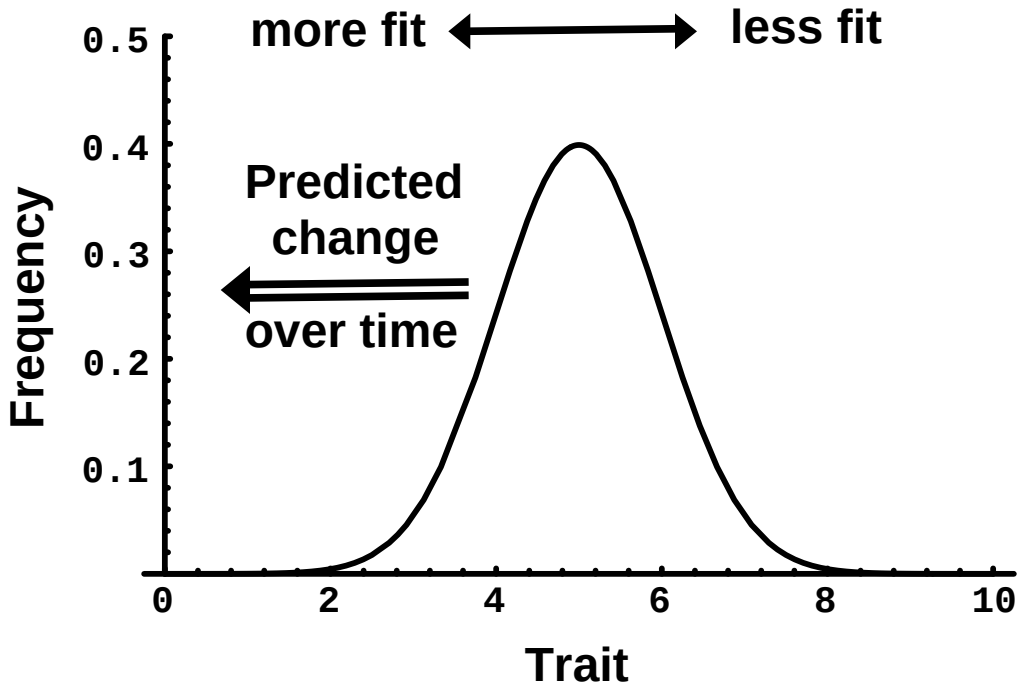
Stabilizing selection = Selection for intermediate values of a trait over extreme values

Disruptive selection = Selection for extreme values of a trait over intermediate values

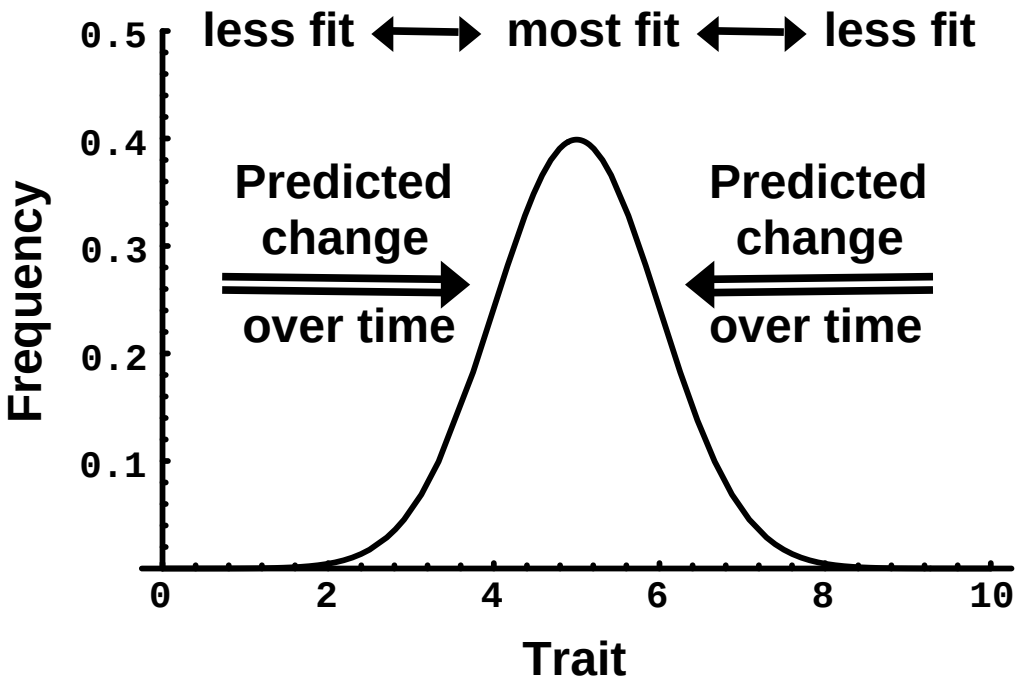
Directional Selection: UP



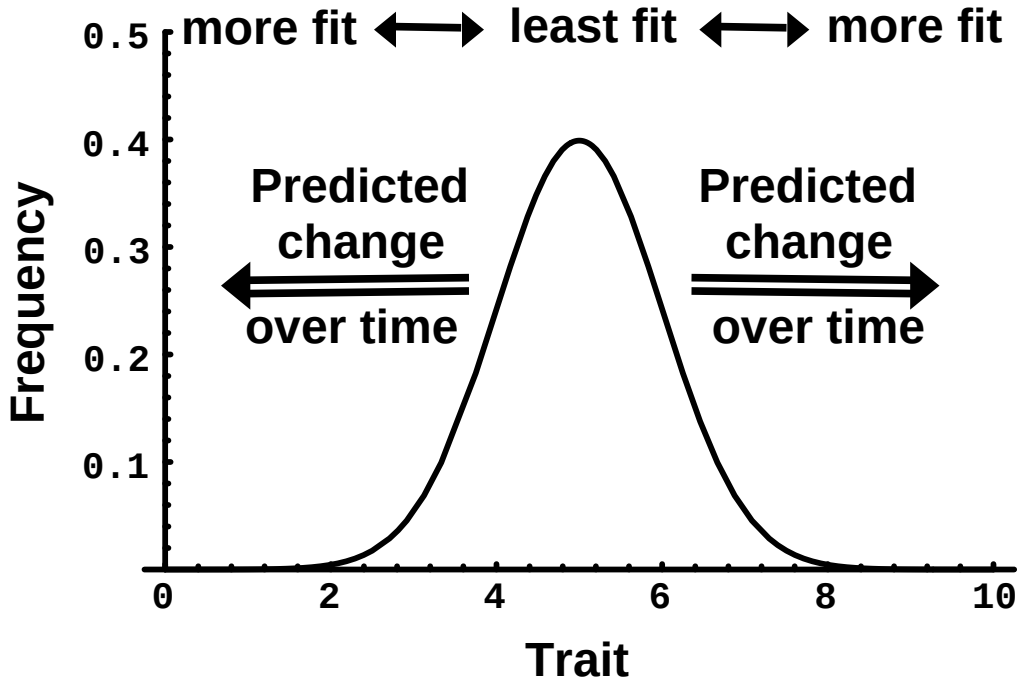
Directional Selection: DOWN



Stabilizing Selection



Disruptive Selection



Models are particularly important in cases like disruptive selection:

The population can be pulled up or pulled down depending upon the relative strengths of selection. Under certain special circumstances, the population might even split apart leading to both an up and a down line.

The distributions drawn above illustrate *quantitative* traits, such as height or weight, that have a *normal (bell-shaped) distribution*.

Such traits are generally influenced by a large number of genes.

Before discussing models of many genes, we will start with simpler models of one genetic locus.

Dynamics without Selection

Before we learn how selection affects a population, we need to understand what happens in the *neutral* case, when all individuals are equally fit, i.e. equally capable of survival and reproduction.

Consider a diploid population with two alleles (*A* and *a*) at one locus.

Let

- P[t] = frequency of *AA* individuals
- Q[t] = frequency of *Aa* individuals
- R[t] = frequency of *aa* individuals

I use [t] to denote frequencies in the t^{th} generation.

NOTE: P[t]+Q[t]+R[t] = 1 since the frequencies of all non-overlapping categories must always sum to one.

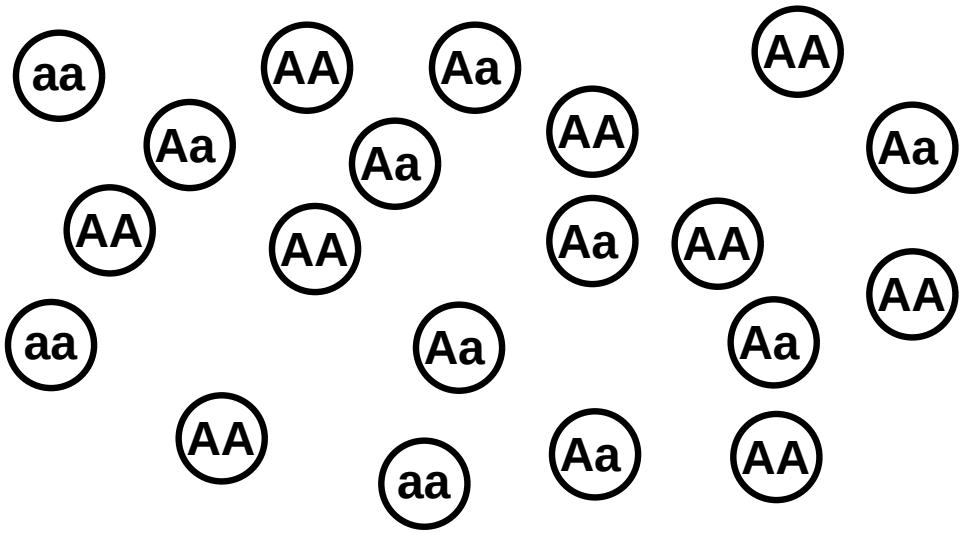
The frequency of allele *A*, p[t], in this population is given by the frequency of *AA* plus half of the frequency of *Aa* (since only half of a heterozygote's genes are *A*):

$$p[t] = P[t] + 1/2\ Q[t]$$

Similarly, the frequency of allele *a*, q[t], in this population is given by half of the frequency of *Aa* plus the frequency of *aa*:

$$q[t] = 1/2\ Q[t] + R[t]$$

NOTE: $p[t]+q[t] = P[t]+Q[t]+R[t] = 1$.



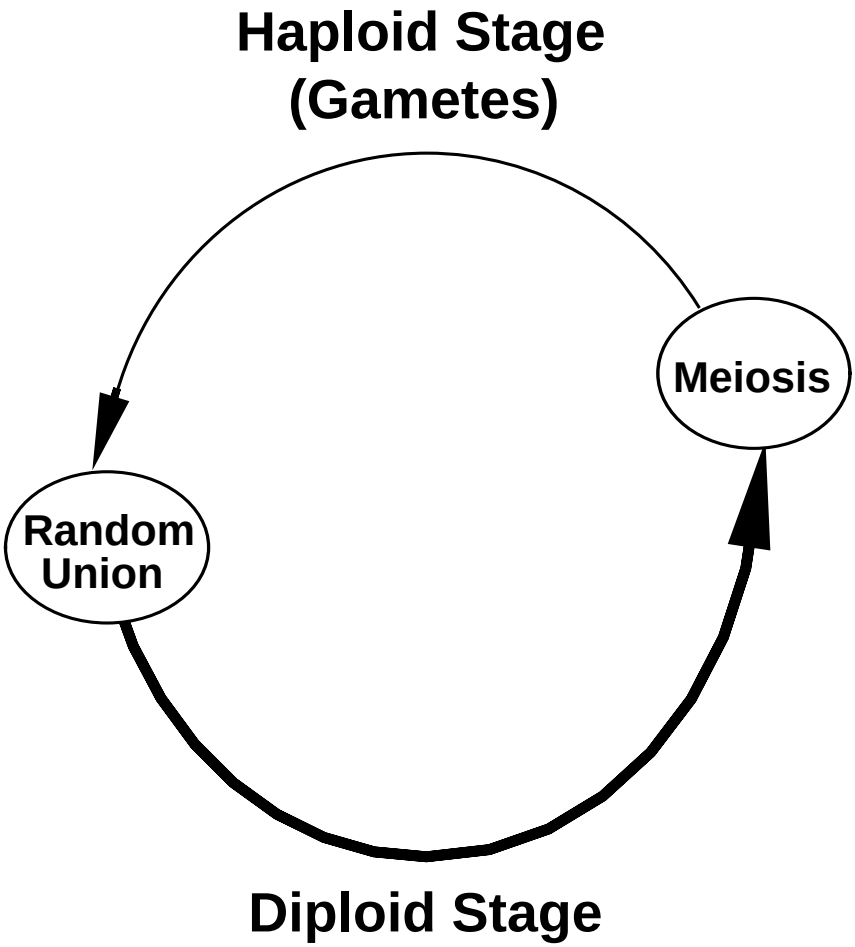
Number AA = 9	**Frequency AA = 9/20 = 0.45 = P[t]**
Number Aa = 8	**Frequency Aa = 8/20 = 0.40 = Q[t]**
Number aa = 3	**Frequency aa = 3/20 = 0.15 = R[t]**
Total number = 20	**Total frequency = 1**

Number A = 26	**Frequency A = 26/40 = 0.65 = p[t]**
Number a = 14	**Frequency a = 14/40 = 0.35 = q[t]**
Total number = 40	**Total frequency = 1**

Let us model what happens to these genotype frequencies over one generation.

We will assume:

- No selection
- No mutation
- Large population size (= ignore sampling error)
- Random mating of gametes



When meiosis occurs and the adult population produces gametes, *AA* individuals produce 100% *A* gametes, *Aa* individuals produce 50% *A* gametes, and *aa* individuals produce 0% *A* gametes.

Therefore, the frequency of *A* among the gametes is:

$$100\%\ P[t] + 50\%\ Q[t] + 0\%\ R[t] = P[t] + 1/2\ Q[t]$$

But we already have seen P[t] + 1/2 Q[t]: it was the allele frequency, p[t], in the previous generation.

Similarly, the frequency of *a* among the gametes will be q[t].

➡ When all adults are equally fertile, the allele frequencies among the gametes equals the allele frequencies among the adults that produced them.

Under the assumption that gametes randomly unite, there are four possibile combinations:

- First gamete is *A* and second gamete is *A* (probability $p[t]^2$)
- First gamete is *A* and second gamete is *a* (probability $p[t]\ q[t]$)
- First gamete is *a* and second gamete is *A* (probability $q[t]\ p[t]$)
- First gamete is *a* and second gamete is *a* (probability $q[t]^2$).

Hence, among the offspring of the next generation, the frequency of the three genotypes equals:

- $P[t+1] = p[t]^2$
- $Q[t+1] = 2\ p[t]\ q[t]$
- $R[t+1] = q[t]^2$

Point 1: After one generation of random mating, the genotypes within the population are at the frequencies we would expect them to be if the alleles within an individual are uncorrelated.

These frequencies, expected in the absence of selection and with random mating, are known as the *Hardy-Weinberg frequencies*

Among these diploid individuals, notice that the frequency of *A* is the same as it was in the previous generation:

$$p[t+1] = P[t+1] + 1/2\ Q[t+1]$$

$$= p[t]^2 + p[t]\ q[t]$$

$$= p[t]\ (p[t] + q[t]) = p[t]$$

(Show that q[t+1] = q[t].)

Point 2: In the absence of selection and mutation, allele frequencies remain constant. Meiosis and random mating do not change allele frequencies.

(In the cartoon population illustrated above, the population was not at Hardy-Weinberg. Were there too many or too few heterozygotes?)

Models of Selection

Selection can act at many different stages in an organism's life cycle.

- survival of juveniles
- survival of adults
- mating success
- fertility
- selection among gametes

The sum total effect of selection within a generation is measured by *fitness*:

Absolute Fitness = The average number of offspring of a given type per parent of the given type.

One can speak of the fitness of an individual or a genotype or an allele.

Fitness can also be measured on a relative scale:

Relative Fitness = The average contribution to the offspring generation relative to the contribution of another type.

How does selection affect allele frequencies over time?

How fast do allele frequencies change?

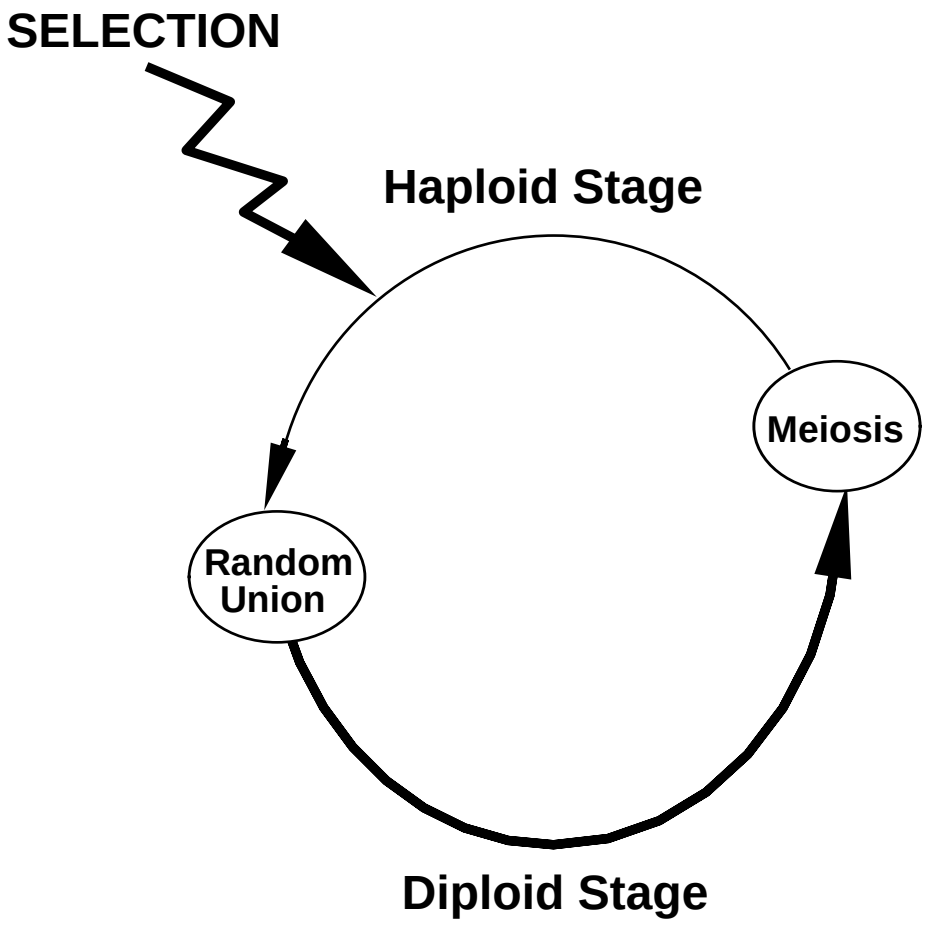
What is the end-point of selection?

Does fitness always increase over time?

One-Locus Haploid Model of Selection

We will derive the simplest model where selection acts only in the haploid stage.

This would be appropriate for haploid organisms (bacteria, certain protozoans and algae) and also for diploid organisms (like humans) for genes that experience selection only in the haploid (i.e. gamete) phase.



At the beginning of the haploid stage in generation t, the frequency of allele *A* is p[t] and of *a* is q[t].

$$(p[t] + q[t] = 1)$$

Selection acts during the haploid stage and is measured by:

- W_A = relative fitness of *A*
- W_a = relative fitness of *a*

Mathematically, the relative fitnesses are used to *weight* the frequency of each allele; alleles with higher fitness have more representation in the next generation.

After selection, the frequency of *A* alleles becomes

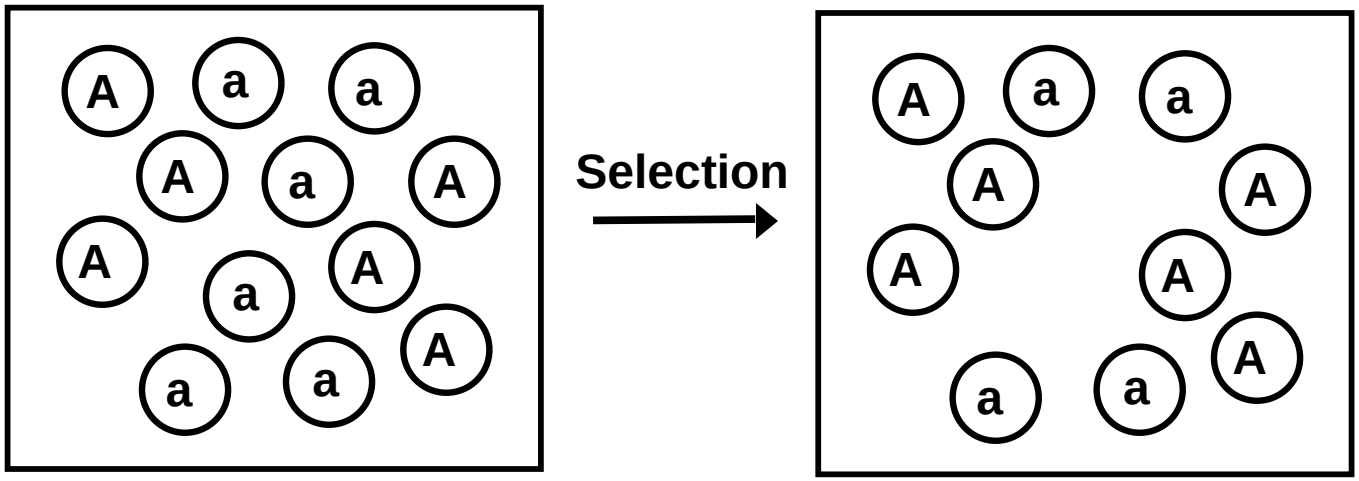
$$\frac{p[t]\, W_A}{p[t]\, W_A + q[t]\, W_a}$$

and the frequency of *a* alleles becomes

$$\frac{q[t]\, W_a}{p[t]\, W_A + q[t]\, W_a}$$

(The denominator in these equations is the mean RELATIVE fitness in the population, $\overline{W}[t]$.)

Survival of A = 1 Survival of a = 2/3



Before selection

6 A : 6 a

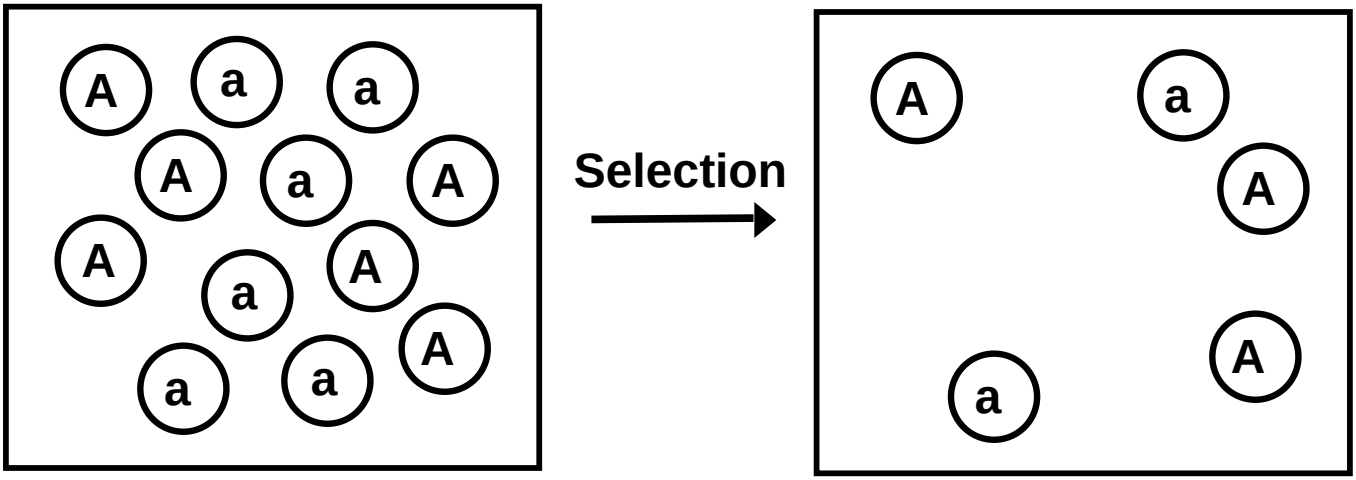
p = ?

After selection

6 A : 4 a

p = ?

Survival of A = 1/2 Survival of a = 1/3



Before selection

6 A : 6 a

p = ?

After selection

3 A : 2 a

p = ?

The survival probabilities of *A* and *a* are very different in the top and bottom examples, but the relative fitnesses are the same (fitness of *a* is 2/3 the fitness of *A*) as are the frequencies after selection.

➡ Relative fitnesses, not their absolute values, determine changes in allele frequencies.

(Show that multiplying the relative fitnesses by a constant C does not change the frequency of *A* after selection.)

We next assume that the surviving haploid adults mate at random to produce diploid offspring that undergo meiosis to generate the next generation of haploids.

But we already know that random mating and meiosis do not change allele frequencies.

Therefore the allele frequency among the haploids in the next generation is the same as at the end of selection in the previous generation:

$$p[t+1] = \frac{p[t]\, W_A}{p[t]\, W_A + q[t]\, W_a}$$

The change in allele frequency over one generation with selection in the haploid stage is then:

$$\Delta p[t] = p[t+1] - p[t] = \frac{(W_A - W_a)\, p[t]\, q[t]}{\bar{W}[t]}$$

(Challenge: Show that this is true.)

This tells us a few important things about selection in the haploid model:

1. **Directional Selection (UP):** If *A* is more fit than *a,* p[t] increases over time unless p[t] or q[t] equals zero.

2. **Directional Selection (DOWN):** If *A* is less fit than *a,* p[t] decreases over time unless p[t] or q[t] equals zero.

3. The amount of change from one generation to the next is proportional to p[t] q[t], which is a measure of the amount of *genetic variance* at the locus and is highest when p[t] = q[t] = 1/2.

These equations only tell us where the population will be after one generation, but we would like to know what happens over longer periods of time.

In this model, there is a simple trick that simplifies matters. If we divide p[t+1] by q[t+1], we get:

$$\frac{p[t+1]}{q[t+1]} = \frac{p[t]\, W_A}{q[t]\, W_a}$$

➡ The ratio of p[t] to q[t] changes by W_A/W_a each generation.

In particular, after t generations,

$$\frac{p[t]}{q[t]} = \frac{p[0]}{q[0]} \left(\frac{W_A}{W_a} \right)^t$$

Since q[t] = 1-p[t], we can solve this formula for p[t]:

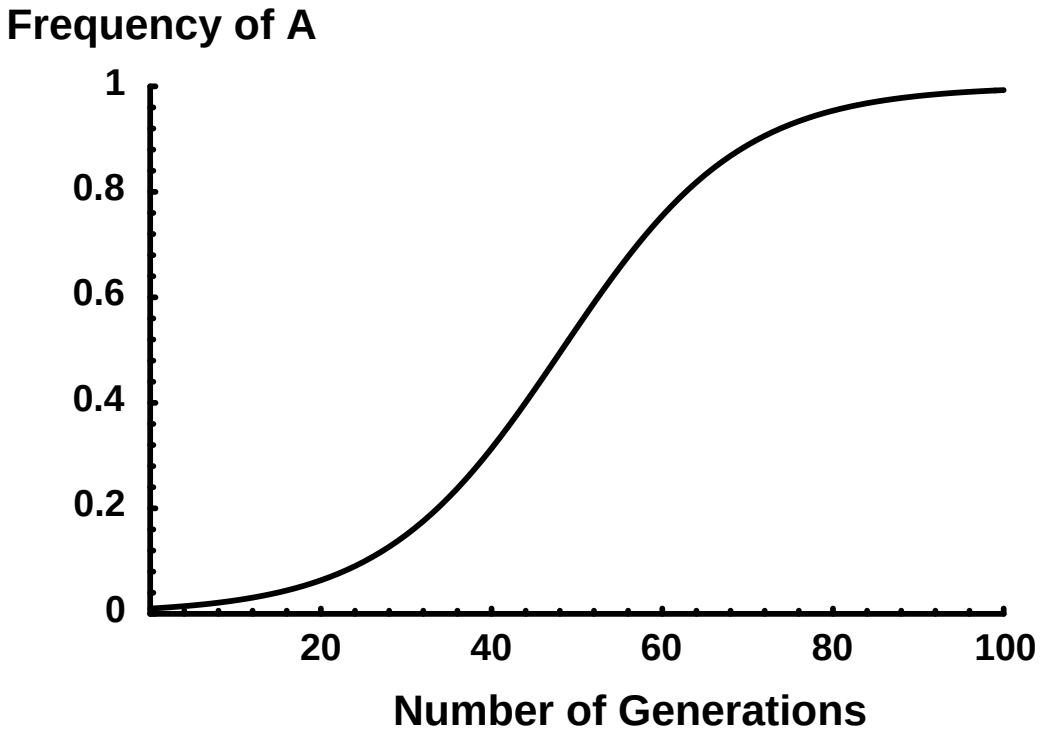
$$p[t] = \frac{p[0]\, W_A^{\,t}}{p[0]\, W_A^{\,t} + q[0]\, W_a^{\,t}}$$

This allows us to predict the allele frequency in any future generation!

Using these equations, we can answer several types of questions.

What would the frequency of *A* be after 100 generations of selection if *A* is 10% more fit than *a* and if one in every hundred alleles is initially *A*?

How does the allele frequency change over time?



If *A* changes in frequency from 0.001 to 0.01 in 10 generations, how much must it be favored?

If *A* changes in frequency from 0.01 to 0.001 in 10 generations, how much must it be disfavored?

The above equations depend only on the relative fitnesses of the alleles. Therefore, we can measure fitness relative to the fitness of one of the alleles (*a*).

$$W_a = 1$$

$$W_A = 1 + s$$

s is known as a *selection coefficient*, which measures the amount by which fitness of a type (here *A*) differs from the fitness of a reference type (*a*).

How long would it take for 95% of the alleles to be *A* if *A* is initially present in 5% of the population and if the selection coefficient favoring allele *A* is s = 0.001? 0.01? 0.1?

Some general principles:

- If $W_A > W_a$, allele *A* will rise in frequency to fixation (p = 1).
- If $W_A < W_a$, allele *A* will decrease in frequency until lost (p = 0).
- If $W_A = W_a$, allele *A* will stay constant in frequency (= neutral case).
- The spread of a beneficial allele always follows an S-shaped curve, increasing slowly when the allele is rare and also when it is common. (Why is this?)
- Weakly favored alleles take longer to spread than strongly favored alleles.
- VERY ROUGHLY, for *A* to go from low frequency to high frequency, it takes on the order of:
  - 100 generations if s is around 0.1.
  - 1000 generations if s is around 0.01.
  - 10,000 generations if s is around 0.001.

(As a rule of thumb, if s is 10 times smaller, it takes 10 times longer to observe the same amount of change in the frequency of *A*.)

Fitness in the One-Locus Haploid Selection Model

Natural selection invokes an image of a population becoming ever more fit. Does the mean fitness of a population always increase over time?

Definitions:

The *mean fitness* (= average fitness) of a population is calculated as the frequency of each type in the population times its fitness.

$$\bar{W}[t] = p[t]\, W_A + q[t]\, W_a$$

The *variance in fitness* of a population is calculated as the frequency of each type in the population times the square of its fitness minus the mean fitness:

$$\mathrm{Var}(W[t]) = p[t]\left(W_A - \bar{W}[t]\right)^2 + q[t]\left(W_a - \bar{W}[t]\right)^2$$

$$= p[t]\, q[t]\, (W_A - W_a)^2$$

(Note: A bar over a quantity is used to denote an average.)

Is *W[t+1]* always greater than or equal to *W[t]*? First, note that:

$$\bar{W}[t+1] = p[t+1]\,W_A + q[t+1]\,W_a = \frac{p[t]\,W_A^2 + q[t]\,W_a^2}{p[t]\,W_A + q[t]\,W_a}$$

Therefore, the change in mean fitness from one generation to the next is:

$$\bar{W}[t+1] - \bar{W}[t] = \frac{p[t]\,W_A^2 + q[t]\,W_a^2}{p[t]\,W_A + q[t]\,W_a} - \left(p[t]\,W_A + q[t]\,W_a\right)$$

$$= \frac{p[t]\,W_A^2 + q[t]\,W_a^2 - \left(p[t]\,W_A + q[t]\,W_a\right)^2}{p[t]\,W_A + q[t]\,W_a}$$

$$= \frac{p[t]\,q[t]\,(W_A - W_a)^2}{p[t]\,W_A + q[t]\,W_a}$$

$$= \frac{Var(W)}{\bar{W}[t]}$$

This quantity is never negative, demonstrating that the mean fitness of a population will only increase or remain the same over time.

Fitness in the One-Locus Haploid Selection Model

➡ Fundamental Theorem of Natural Selection.

"The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time."

-- R. A. Fisher (1930)
The Genetical Theory of Natural Selection

Remark 1: Remember the assumptions!! One genetic locus, discrete generations, constant fitnesses, no mutation, no migration, no meiotic drive. Fisher's fundamental theorem does NOT always hold in more complicated models.

Remark 2: Variance in fitness will be highest (and mean fitness will increase most rapidly) when gene frequencies are near 1/2 and when fitnesses are very different.

Problem: While mean RELATIVE fitness may increase over time, this does not mean that the mean ABSOLUTE fitness of the population is increasing.

In fact, as Fisher (1930) noted, the mean absolute fitness of a population cannot increase indefinitely since the world would soon become overrun with a species that was growing exponentially.

He argued, that in the long-term, the mean absolute fitness of a population must hover around one (one offspring per parent). This was achieved, in his view, by a balance between the "progress" of natural selection and a "deterioration" of the environment:

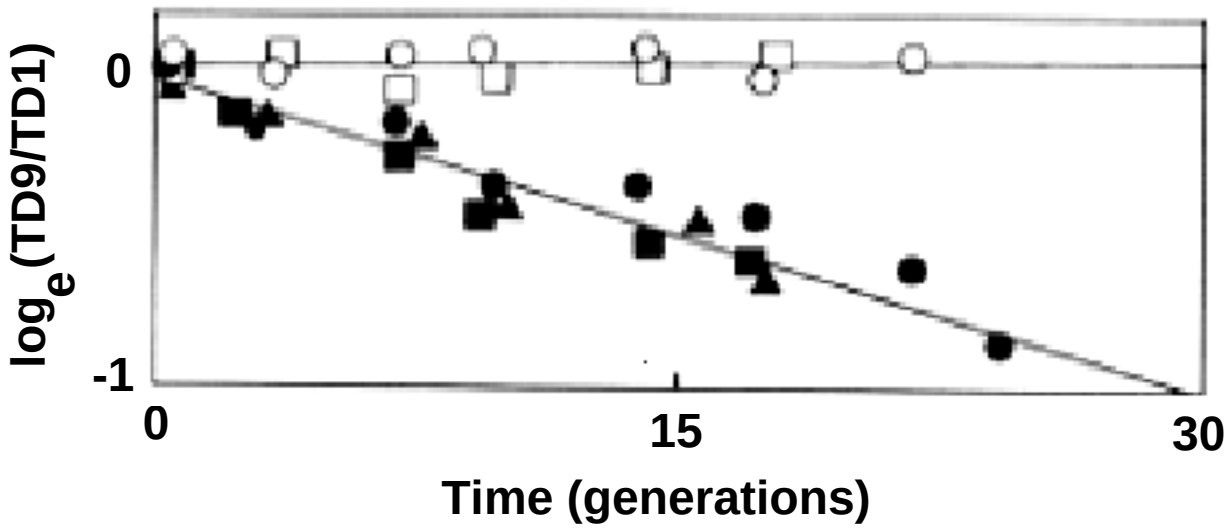
> "Against the rate of progress in fitness must be set off, if the organism is, properly speaking, highly adapted to its place in nature, deterioration due to undirected changes either in the organism [mutations], or in its environment [geological, climatological, or organic]."
>
> -- R. A. Fisher (1930)
> The Genetical Theory of Natural Selection

EXAMPLE: One-Locus Haploid Selection

Dykhuizen and Dean (1990) report the results from an experiment with the haploid bacteria, *Escherichia coli* .

They took two strains (TD9 and TD1) that had a genetic difference in the lactose pathway.

They competed the strains in a chemostat with glucose as the limiting nutrient (OPEN SYMBOLS; two replicates) and with lactose as the limiting nutrient (FILLED SYMBOLS; three replicates).



Using the haploid model for selection, show why it makes sense that the log of the ratio of the frequencies of TD9 to TD1 falls along a straight line over time.

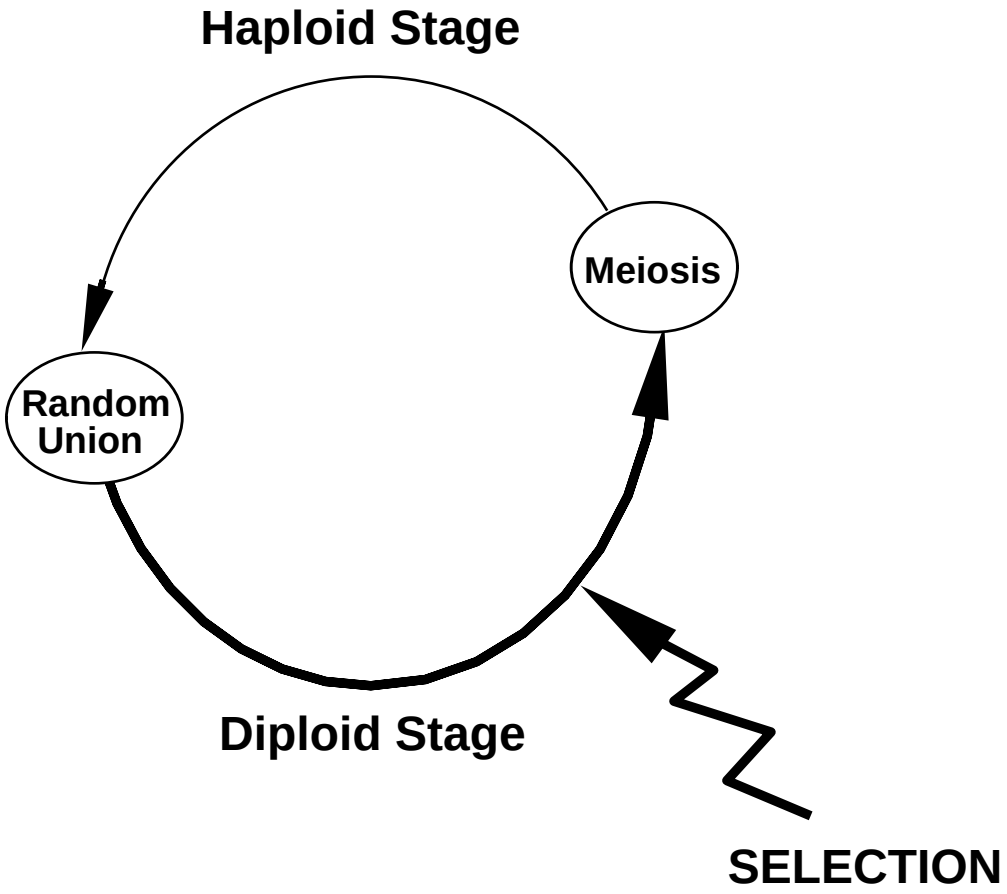
Dykhuizen and Dean conclude that there are no significant fitness differences between the strains when glucose is supplied, but that selection acts against the TD9 strain with s=-0.033 when lactose is supplied. From the graph, confirm this estimate for s.

Using highly controlled and simplified chemostat experiments, it is difficult to accurately measure selection coefficients lower than 0.005 in magnitude. What factors limit the accuracy of fitness measures?

One-Locus Diploid Selection Model

We turn now to a model describing selection among diploid organisms.

We assume that selection acts in a large population only in the diploid phase (not in the haploid phase), with random mating, no migration, and no mutation.



Let p[t] equal the frequency of the *A* allele among adult diploids.

Since fair meiosis does not alter allele frequencies and since there is no selection in the haploid phase, p[t] will also equal the frequency of the *A* allele among the haploid gametes of these adults.

With either random union of these gametes in a gamete pool or with random mating of individuals, the diploid offspring will be in the Hardy-Weinberg proportions:

$$AA : Aa : aa$$

$$p^2[t] : 2\, p[t]\, q[t] : q^2[t]$$

We wish to determine the effects of selection on diploid individuals with the relative fitnesses:

$$W_{AA} : W_{Aa} : W_{aa}$$

Selection weights the frequency of each genotype by its relative fitness:

$$p^2[t]\, W_{AA} \;:\; 2\, p[t]\, q[t]\, W_{Aa} \;:\; q^2[t]\, W_{aa}$$

Dividing each of these by their sum gives the frequency of the three genotypes after selection.

Since the frequency of *A* equals the frequency of *AA* individuals plus half the frequency of *Aa* individuals, the frequency of the *A* allele among adult diploids in the next generation (t+1) equals:

$$\mathbf{p[t+1] = \frac{p[t]^2 W_{AA} + p[t]\, q[t]\, W_{Aa}}{p[t]^2 W_{AA} + 2\, p[t]\, q[t]\, W_{Aa} + q[t]^2 W_{aa}}}$$

(The denominator is the mean relative fitness, $\overline{W}[t]$, with diploid selection.)

This equation describes the change in allele frequency among diploid adults from one generation to the next.

We will not fully analyse this equation, but we will show how selection in the diploid phase works graphically and discuss some important results.

Forms of Selection in the Diploid Model

In the diploid model, all three general forms of selection discussed earlier (directional, stabilizing, and disruptive) are represented.

- **Directional Selection:**
  - UP: $W_{AA} > W_{Aa} > W_{aa}$
  - DOWN: $W_{AA} < W_{Aa} < W_{aa}$
- **Stabilizing Selection:**
  - $W_{AA} < W_{Aa} > W_{aa}$

("Heterozygote advantage", "Overdominance")

- **Disruptive Selection:**
  - $W_{AA} > W_{Aa} < W_{aa}$

("Heterozygote disadvantage", "Underdominance")

(Which forms of selection were present in the haploid selection model?)

The effects of selection on a diploid population depend strongly on which form of selection is operating.

Directional Selection

These plots show directional selection favoring the *A* allele for different strengths of selection.

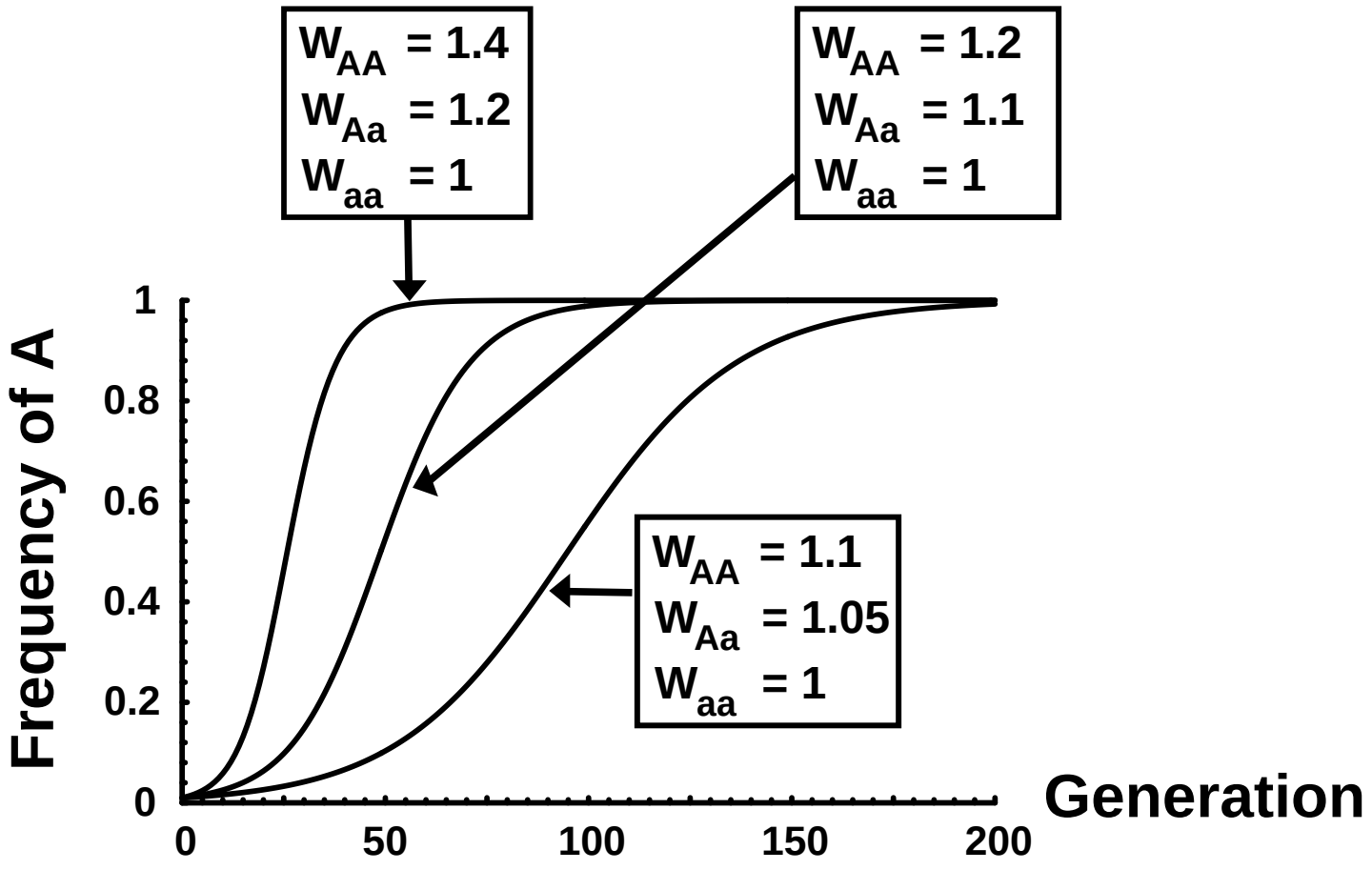
Notice that as the fitnesses become more similar, it takes longer for *A* to rise from low frequency to high frequency.

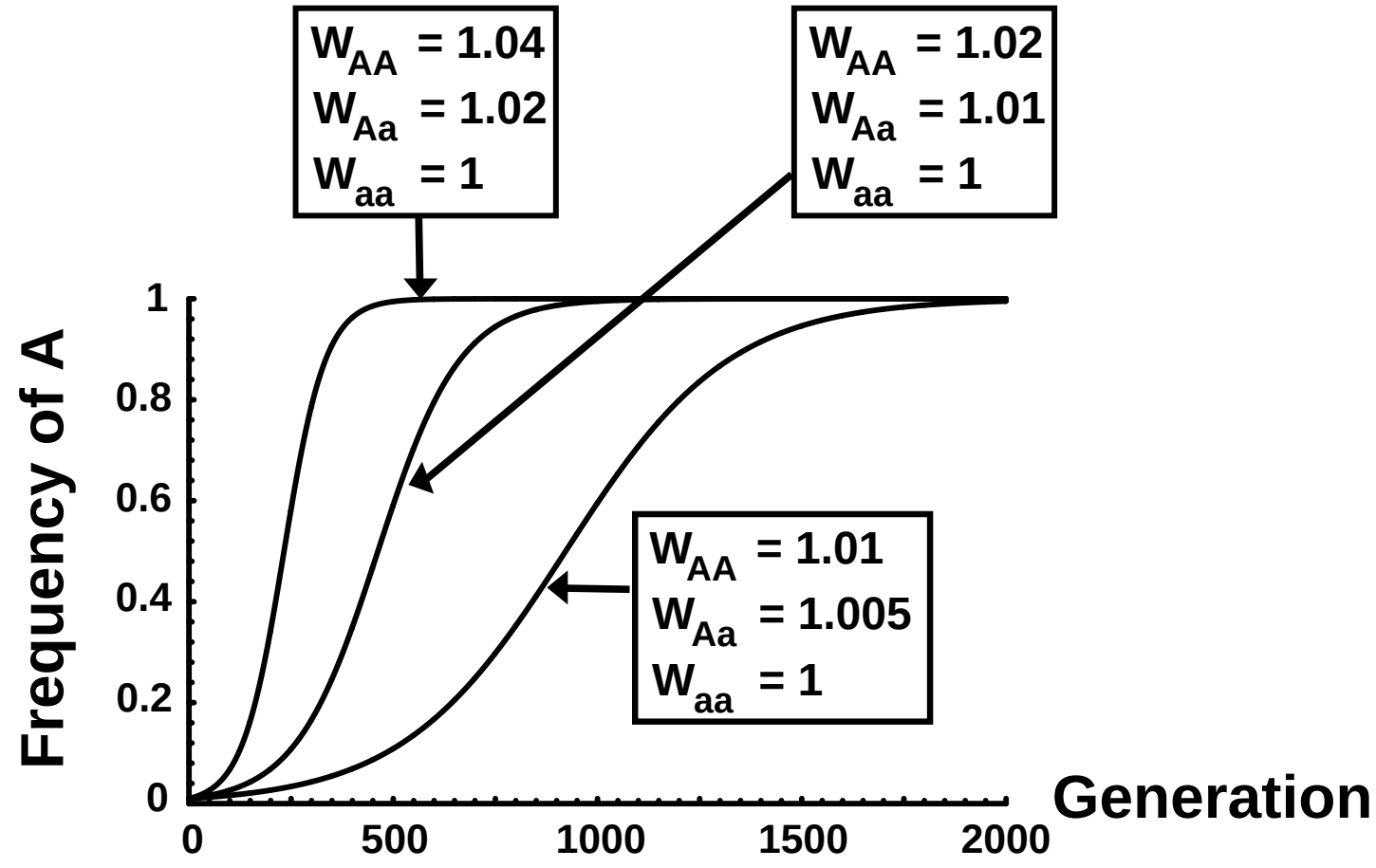
ROUGHLY, as the fitness differences decrease by a factor of x, the time it takes for *A* to spread increases by a factor of x.

Fitnesses in the diploid model with directional selection are often defined relative to one homozygote as:

- $W_{AA} = 1 + s$
- $W_{Aa} = 1 + h\,s$
- $W_{aa} = 1$

s is again called the *selection coefficient*, while h is called the *dominance coefficient*.


W_{AA} = 1.4
W_{Aa} = 1.2
W_{aa} = 1
W_{AA} = 1.2
W_{Aa} = 1.1
W_{aa} = 1
W_{AA} = 1.1
W_{Aa} = 1.05
W_{aa} = 1
Frequency of A
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
Generation

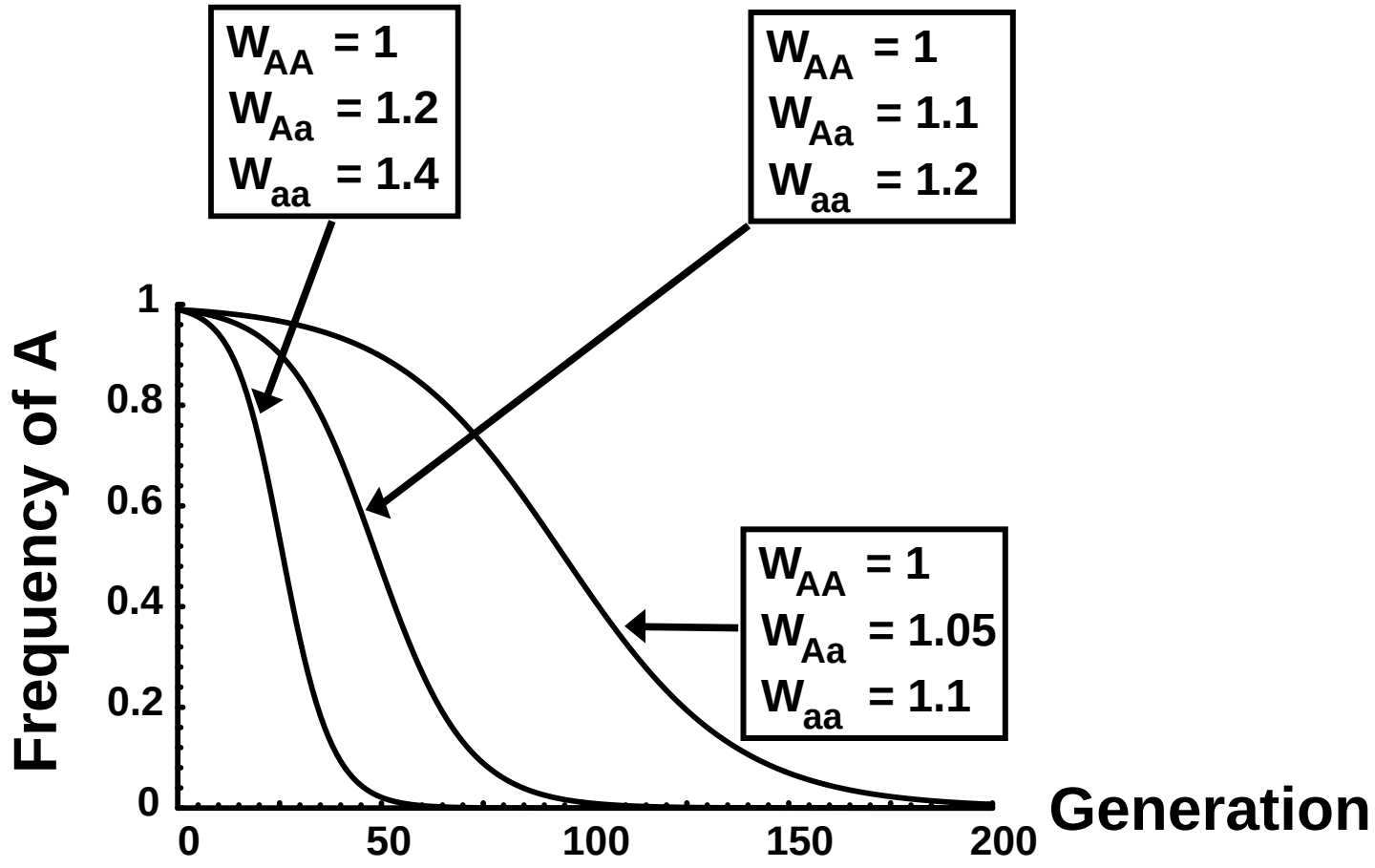

W_{AA} = 1.04
W_{Aa} = 1.02
W_{aa} = 1
W_{AA} = 1.02
W_{Aa} = 1.01
W_{aa} = 1
W_{AA} = 1.01
W_{Aa} = 1.005
W_{aa} = 1
Frequency of A
0
0.2
0.4
0.6
0.8
1
0
500
1000
1500
2000
Generation

In this case, h describes the dominance of the *A* allele:

- *A* is *dominant* (h = 1)
- *A* is *partially dominant* (1/2 < h < 1)
- *A* is *additive* (h = 1/2)
- *A* is *partially recessive* (0 < h < 1/2)
- *A* is *recessive* (h = 0)

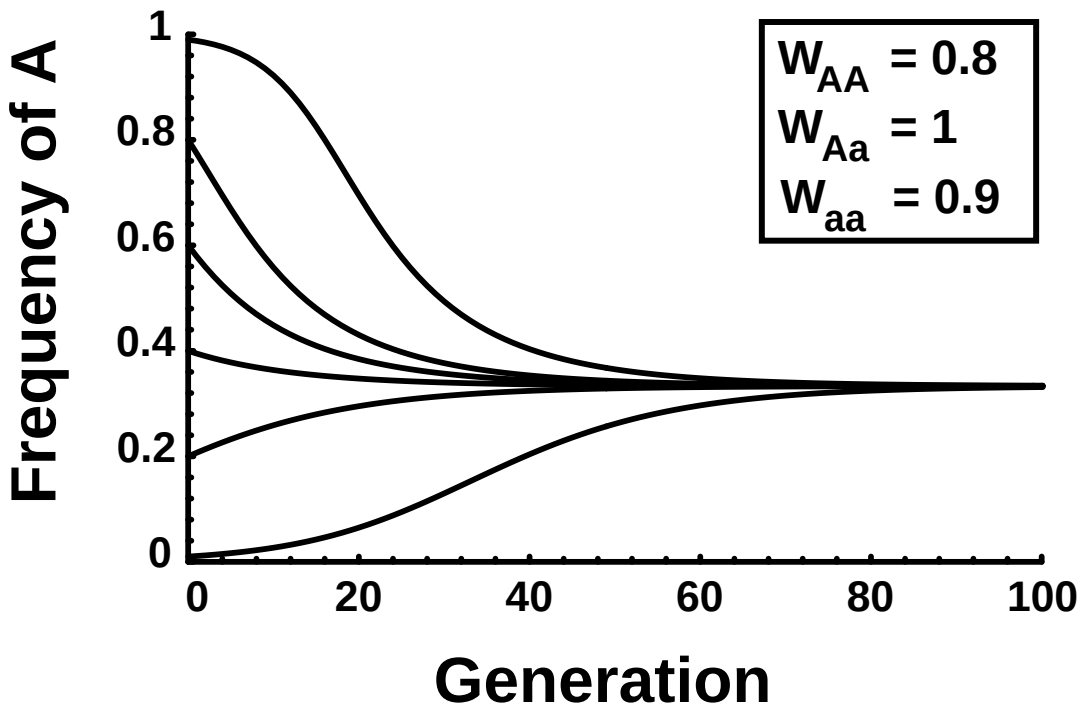
In the above graphs, I've used the same value of h in every case (h=1/2). For your homework, you will explore how the dominant and recessive cases behave.

If $W_{AA} < W_{Aa} < W_{aa}$, directional selection favors the *a* allele:



Heterozygote Advantage

The frequency of *A* behaves quite differently when the heterozygote has the highest fitness:



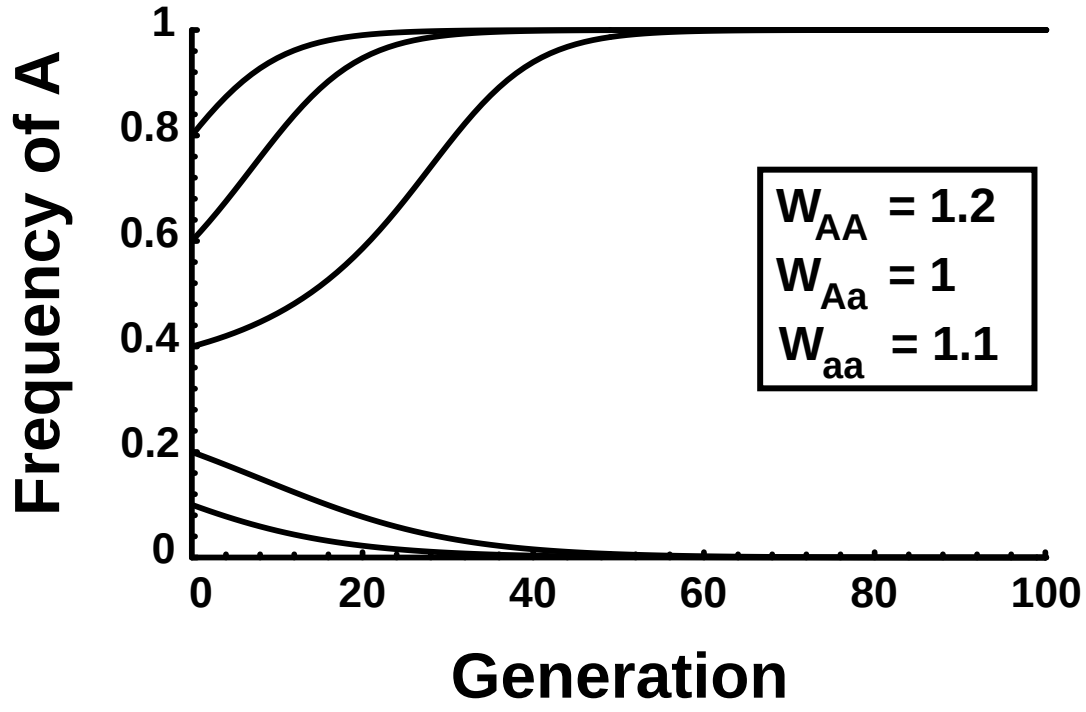
Regardless of the starting frequency of *A*, the system approaches an intermediate frequency of *A*, where it remains forevermore.

This value is called a *polymorphic equilibrium*.

The exact value of the polymorphic equilibrium depends on the fitnesses.

Heterozygote Disadvantage

When the heterozygote has the lowest fitness, the *A* allele rises in frequency if it is common or decreases in frequency if it is rare:



The cut-off between where *A* rises or drops is called an *unstable polymorphic equilibrium*.

The exact value of this cut-off depends on the fitnesses.

[NOTE: The population never splits in two.]

Equilibria

An equilibrium is a value of the allele frequency where the allele frequency remains constant over time, i.e., where p[t+1]=p[t].

The equilibria for the diploid model can be found by solving:

$$p[t+1] = \frac{p[t]^2 W_{AA} + p[t]\,q[t]\,W_{Aa}}{p[t]^2 W_{AA} + 2\,p[t]\,q[t]\,W_{Aa} + q[t]^2 W_{aa}} = p[t]$$

Doing the algebra, p remains constant over time only if p=0, p=1, or:

$$p = \frac{W_{Aa} - W_{aa}}{(2W_{Aa} - W_{AA} - W_{aa})}$$

This last equilibrium only falls between 0 and 1 when

$$W_{AA} < W_{Aa} > W_{aa}$$

OR

$$W_{AA} > W_{Aa} < W_{aa}$$

With directional selection, there is no polymorphic equilibrium (no value where p[t+1]=p[t] between 0 and 1).

The frequency of *A* will approach the polymorphic equilibrium in the first case (heterozygote advantage), but will be repelled from it in the second case (heterozygote disadvantage).

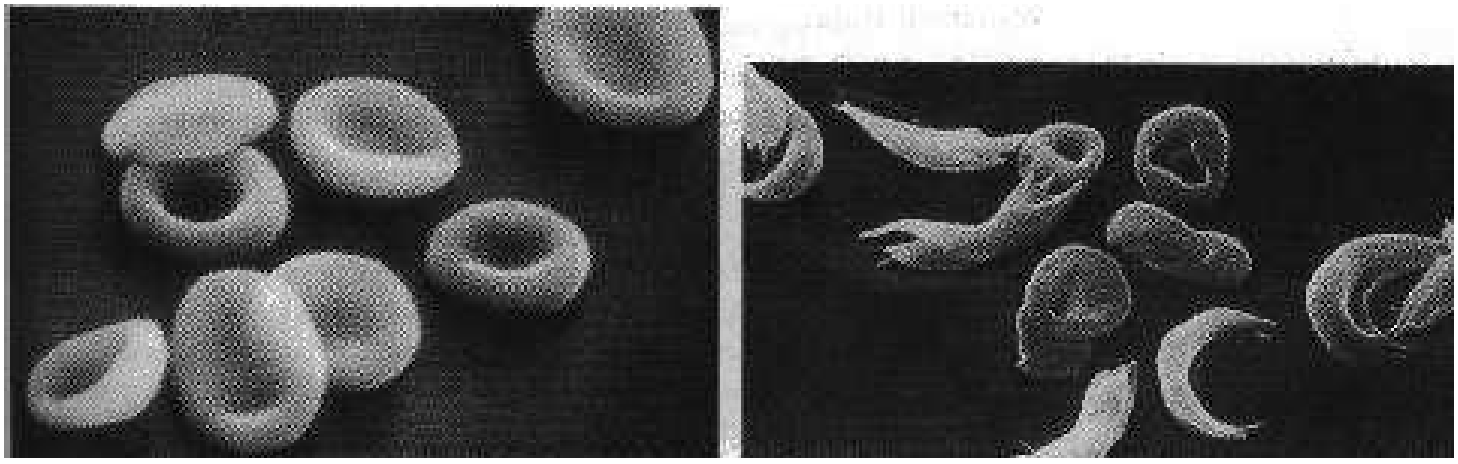
What is the value of the polymorphic equilibrium in the graphs shown above?

Example: Sickle Cell Anemia

Sickle-cell anemia is a human disease affecting the shape and flexibility of red blood cells.

It is caused by a single mutation in the sixth amino acid of the β chain of hemoglobin.

The mutant form of hemoglobin (S) tends to crystalize and form chains, causing distortions in red blood cells:



Homozygous carriers of *S* experience the greatest degree of sickling and tend to suffer severe anemic attacks.

Heterozygous carriers of *S* also suffer from sickling of red blood cells, but to a lesser degree.

However, heterozygous carriers are less likely to die from malaria, since red blood cells infected with the parasites causing malaria tend to sickle and be destroyed.

The fitness of each genotype can be estimated from from the extent to which genotypic frequencies among adults depart from Hardy-Weinberg.

For the Nigerian population studied:

- $W_{AA} = 0.88$
- $W_{AS} = 1$
- $W_{SS} = 0.14$

What is the expected equilibrium frequency of the non-mutant allele (*A*)?

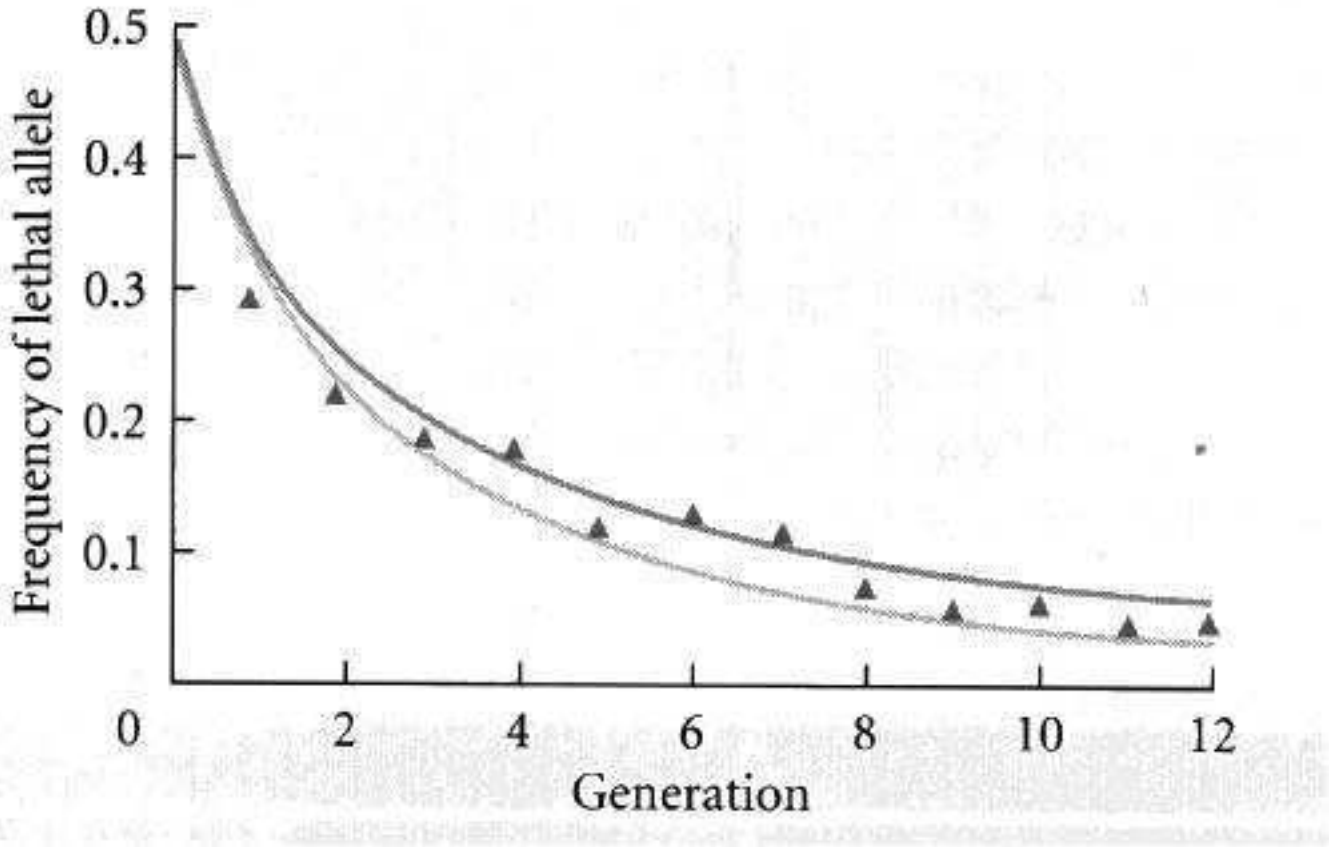
The frequency of the non-mutant allele (*A*) in this population was 0.877, which matches extremely well!

Example: Decline of a recessive lethal allele

Dawson (1970) studied the decline of a recessive lethal mutation (W_{aa}= 0) in a population of flour beetles.

He started a laboratory population of beetles with an initial frequency of the mutation of 1/2.

The frequency of *a* was measured and observed to decrease over time:



The triangles are the observations, the top line shows the expected decline in *a* assuming that W_{Aa}= 1, and the bottom line shows the expected decline in *a* assuming that W_{Aa}= 0.9.

The data match the expected decline for a recessive (or nearly recessive) lethal allele.

Summary

Haploid Model

- If $W_A > W_a$, population converges to fixation on *A*
- If $W_A < W_a$, population converges to fixation on *a*
- Mean relative fitness always increases or stays the same

Diploid Model

- If $W_{AA} > W_{Aa} > W_{aa}$, population converges to fixation on *A*
- If $W_{AA} < W_{Aa} < W_{aa}$, population converges to fixation on *a*
- If $W_{AA} < W_{Aa} > W_{aa}$, population converges to polymorphic equilibrium
- If $W_{AA} > W_{Aa} < W_{aa}$, population converges to fixation on *A* or on *a*
- Mean relative fitness always increases or stays the same **(Challenge!)**

SOURCES:

- Sickle Cell example: Freeman and Herron (1998) Evolutionary Analysis.
- Tribolium example: Futuyma (1998) Evolutionary Biology.

Two-Locus Models

The human genome is thought to have about 100,000 genes, Drosophila about 10,000 genes, and even bacteria contain thousands of genes per cell!

Clearly, one-locus models of selection are highly simplified depictions of how populations evolve over time.

Exact analyses with multiple loci are, however, extremely difficult if not impossible to obtain.

Even with only two loci, the dynamics are complicated and not completely understood. Results are limited in scope, focusing on particular fitness schemes.

Nevertheless, results from two-locus models are very important in determining what properties of the one-locus model are unique and might not apply to the real-world situation in which a number of loci collectively interact to guide the formation of the individual.

Consider two loci, **A** and **B**, with two alleles each:

$\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{B_1}$, $\mathbf{B_2}$, at frequencies $\mathbf{p_{A_1}}$, $\mathbf{p_{A_2}}$ and $\mathbf{p_{B_1}}$, $\mathbf{p_{B_2}}$.

There are four possible combinations of these alleles on a chromosome:

Chromosome type:	$\mathbf{A_1\ B_1}$	$\mathbf{A_1\ B_2}$	$\mathbf{A_2\ B_1}$	$\mathbf{A_2\ B_2}$
Frequency:	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$

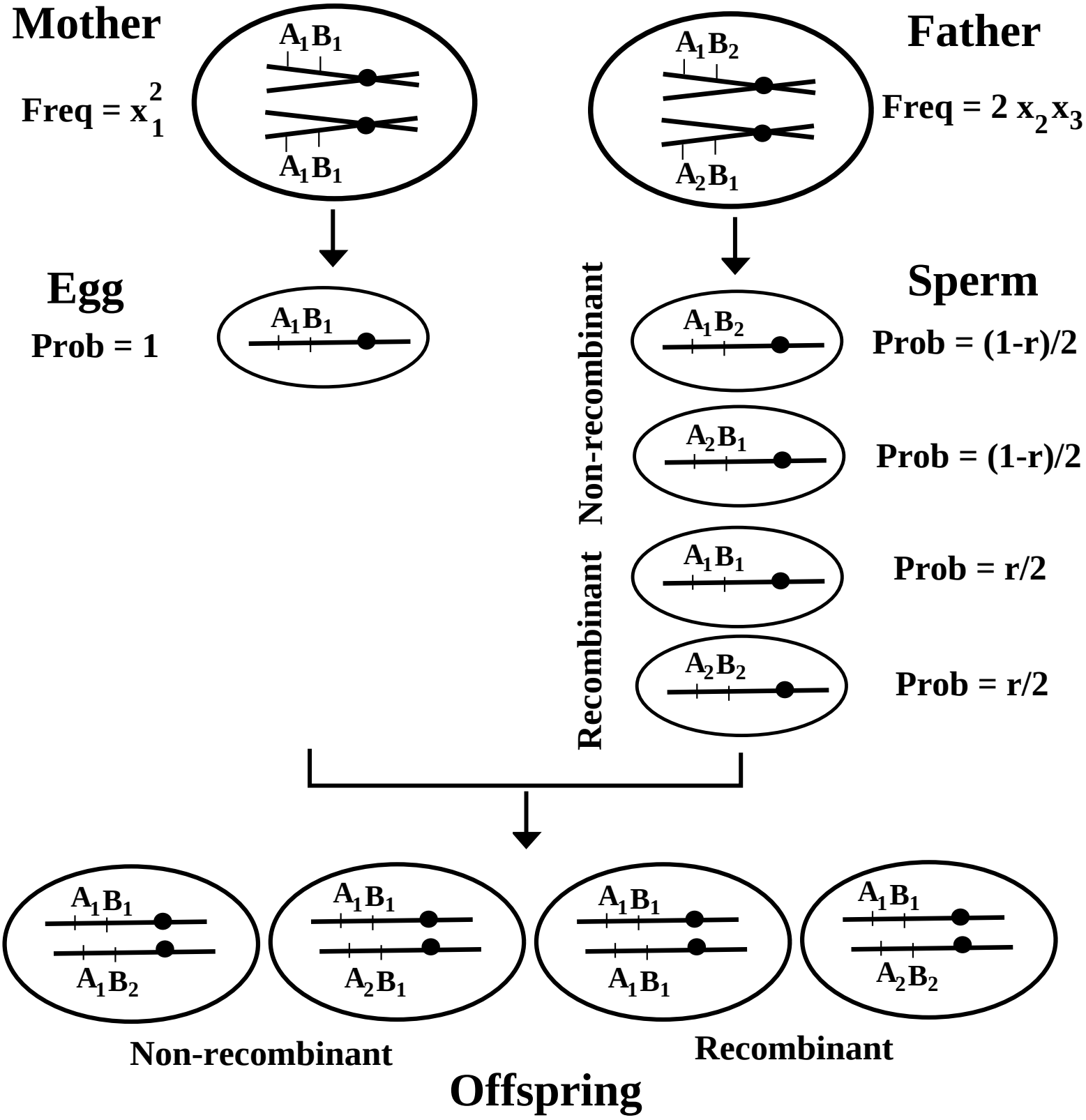
[Note: $\mathbf{x_1} + \mathbf{x_2} + \mathbf{x_3} + \mathbf{x_4} = 1$]

There are two important new concepts in the two-locus model: *Recombination* and *linkage disequilibrium*.

Recombination

As you know from genetics, recombination occurs during meiosis in sexual organisms to generate gametes carrying new combinations of alleles.

We specify the rate of recombination between two loci by r.



[Note: Recombination may occur in any individual but it only changes the type of offspring produced if the parent was a *double heterozygote*.]

Linkage Disequilibrium

Linkage disequilibrium, on the other hand, measures whether an allele at one locus is associated (or correlated) with an allele at a second locus.

Linkage disequilibrium, D, is measured by $\mathbf{x_1\ x_4 - x_2\ x_3}$.

Positive D implies that the chromosomes $\mathbf{A_1\ B_1}$ (x_1) and $\mathbf{A_2\ B_2}$ (x_4) are more common than expected.

Negative D implies that the chromosomes $\mathbf{A_1\ B_2}$ (x_2) and $\mathbf{A_2\ B_1}$ (x_3) are more common than expected.

In a randomly mating population, linkage disequilibrium measures the difference between observed and expected chromosome frequencies:

$$|D| = |OBS - EXP|$$

$$D = x_1 - p_{A_1}\ p_{B_1}$$

$$-D = x_2 - p_{A_1}\ p_{B_2}$$

$$-D = x_3 - p_{A_2}\ p_{B_1}$$

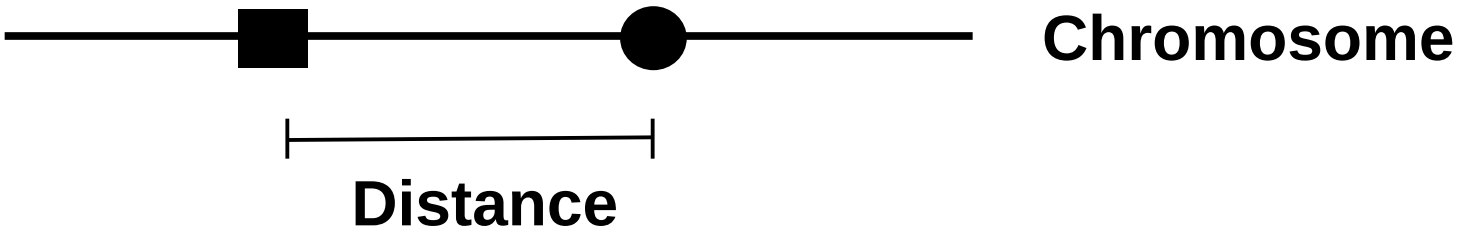
$$D = x_4 - p_{A_2}\ p_{B_2}$$

Comparing r and D

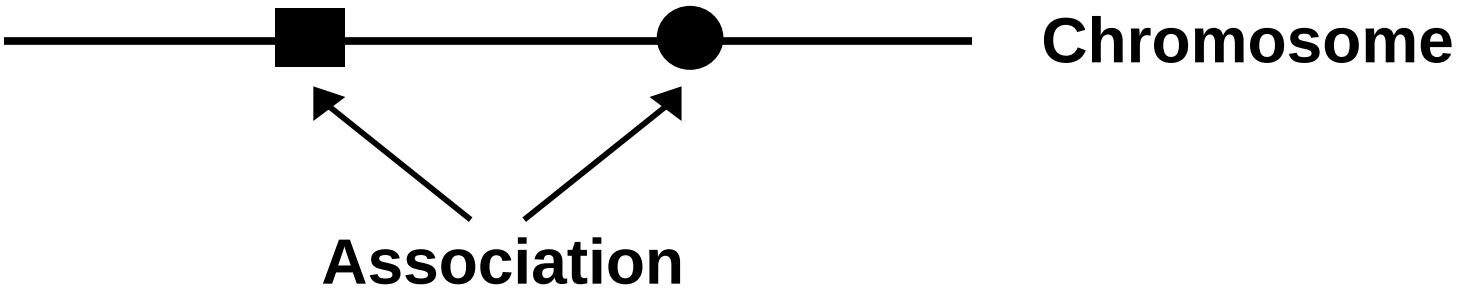
Recombination rate (r) is a measure of the distance between two loci and equals the probability that a gamete contains a chromosomal combination not found in the parents.

Linkage disequilibrium (D) is a measure of whether an allele at one locus tends to be found more often with an allele at another locus.

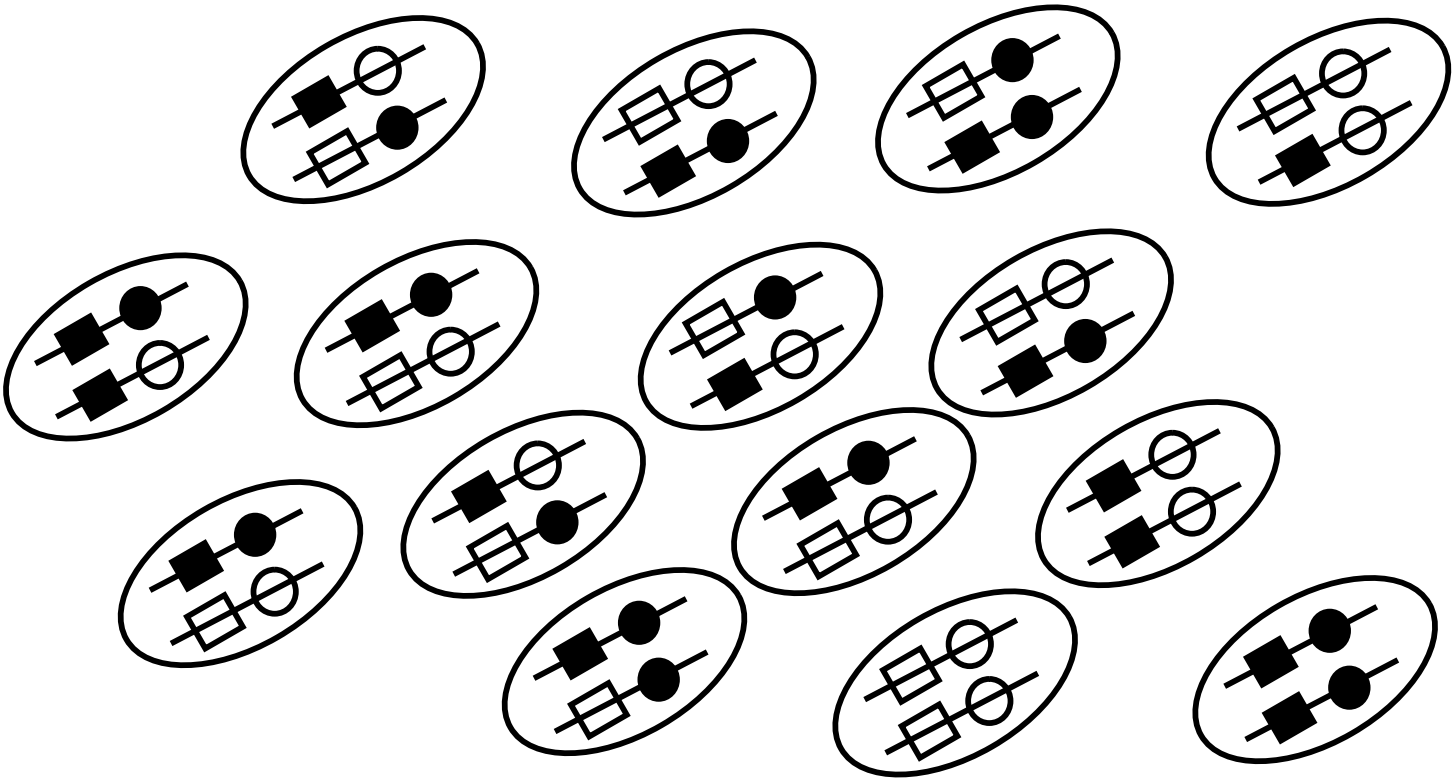
RECOMBINATION RATE:



LINKAGE DISEQUILIBRIUM:



Example



Among 15 diploid individuals (30 chromosomes):

- 10 chromosomes are solid square - solid circle
- 7 chromosomes are solid square - hollow circle
- 5 chromosomes are hollow square - solid circle
- 8 chromosomes are hollow square - hollow circle

- $\mathbf{x_1}$ **=**
- $\mathbf{x_2}$ **=**
- $\mathbf{x_3}$ **=**
- $\mathbf{x_4}$ **=**

- **D =**

Since D is positive, solid squares and solid circles are more likely to be found together on a chromosome than expected:

Frequency of solid squares = 17/30

Frequency of solid circles = 15/30

Expected frequency of solid square - solid circle chromosome = 17/30 * 15/30 = 0.28 (less than the observed 10/30).

(Similarly, hollow square - hollow circle chromosomes are more common than expected.)

Two-Locus Model without Selection

If all individuals are equally fit, chromosome frequencies change from one generation to the next according to:

$$x_1[t+1] = x_1[t] - r\,D[t]$$
$$x_2[t+1] = x_2[t] + r\,D[t]$$
$$x_3[t+1] = x_3[t] + r\,D[t]$$
$$x_4[t+1] = x_4[t] - r\,D[t]$$

In the absence of selection, $p_{A_1}[t+1]$ can be shown to equal $p_{A_1}[t]$. (CHALLENGE: Try to show this.)

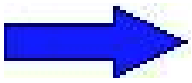
Point 1: In the absence of selection, allele frequencies remain constant.

Also notice that recombination only matters when there is linkage disequilibrium in the population. (Why?)

If D is positive, for example, recombination reduces the frequency of the two chromosomes that are more common than expected: $\mathbf{A_1 B_1}$ and $\mathbf{A_2 B_2}$.

In the absence of selection, D[t+1] = (1-r) D[t]. (CHALLENGE: Try to show this.)

Point 2: Linkage disequilibrium decays at a rate r every generation.

 After an amount of time t, the expected amount of disequilibrium is $D[t]=(1-r)^t D[0]$.

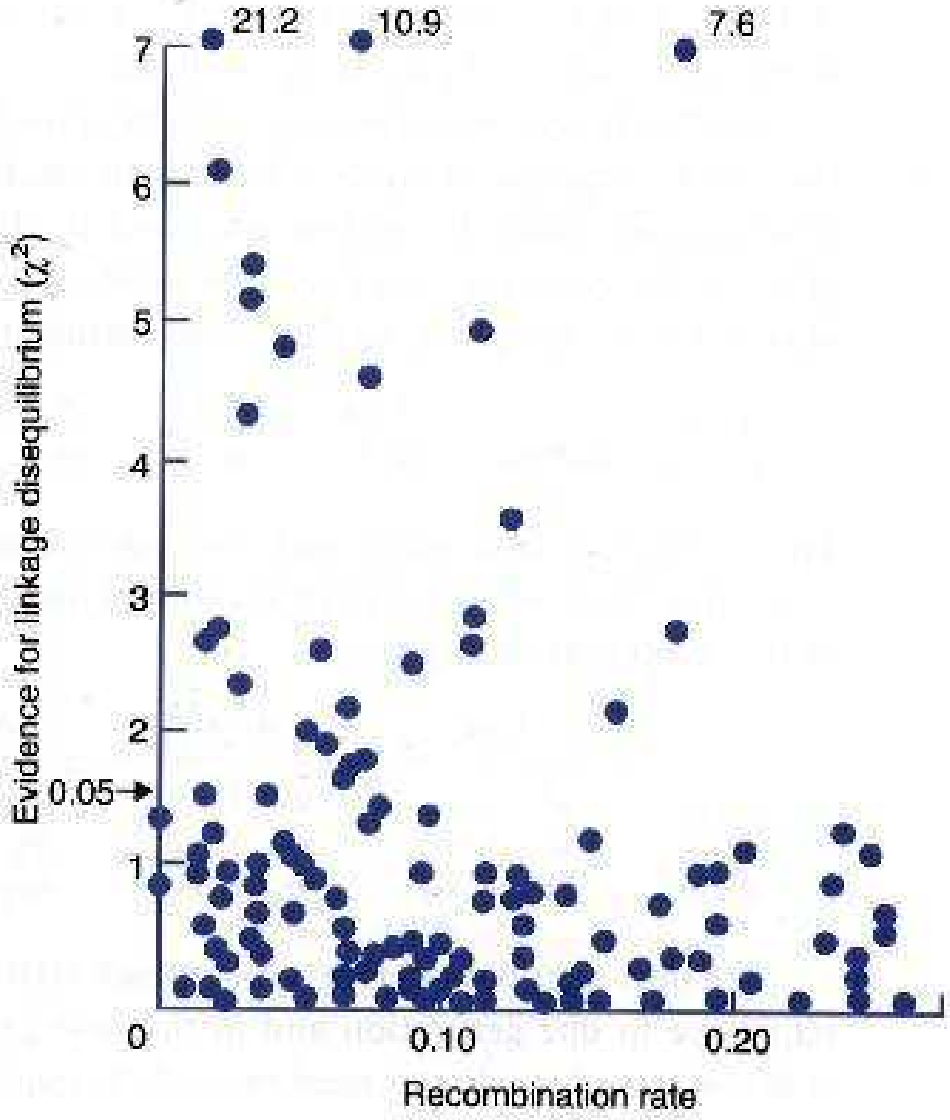
Therefore, after enough time has passed, we expect to see little linkage disequilibrium between two neutral (= not selected) loci unless they are very tightly linked.

[NOTE: Unlike Hardy-Weinberg, however, *linkage equilibrium* (D=0) is not attained in one generation.]

Example

Linkage disequilibrium was measured between several pairs of loci in *Drosophila melanogaster*.

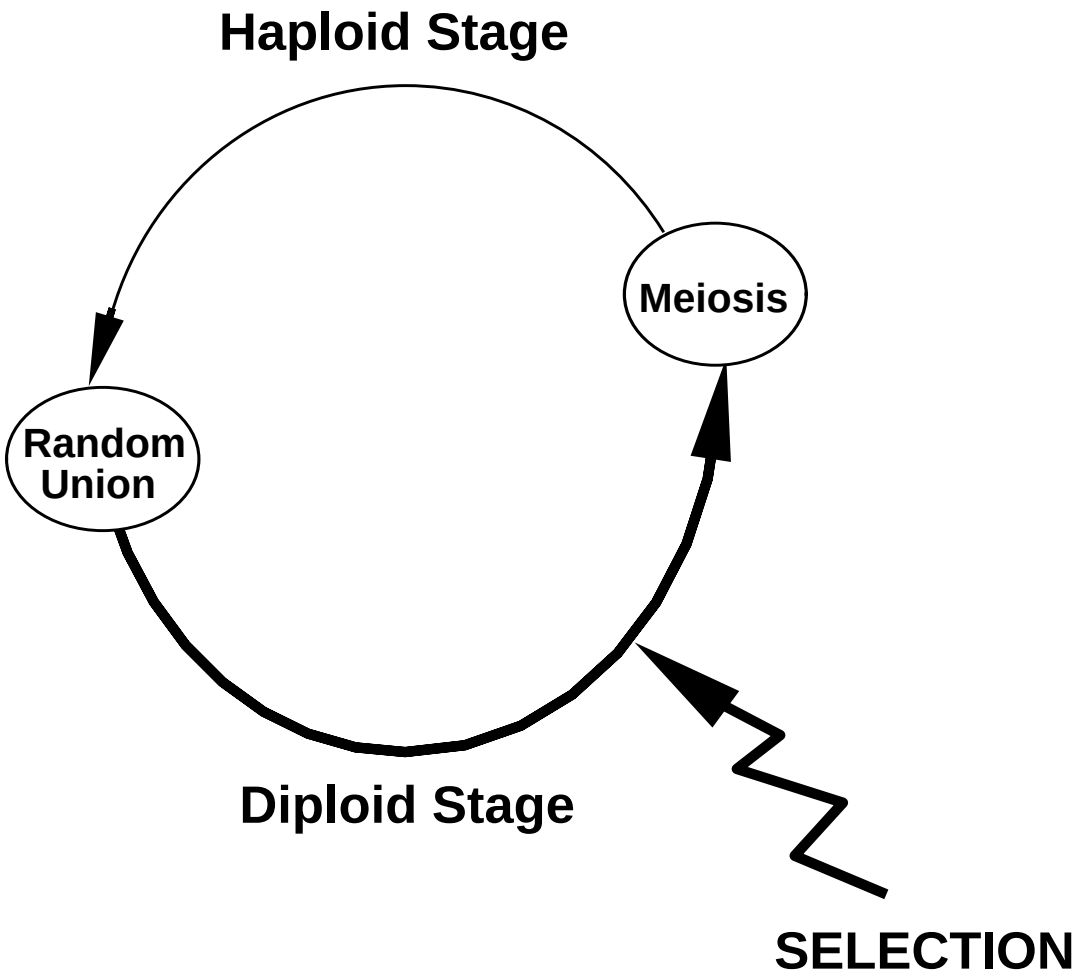
The statistical evidence for a non-zero D value is here plotted as a function of distance between the pair of loci:



Alleles at most pairs of loci are not significantly associated.

Two-Locus Selection Model

Next we'll consider selection acting upon two loci in the diploid phase of a life cycle:



where the survival of a diploid individual depends on its two-locus genotype:

		Chromosome from father			
	Fitness	A_1B_1	A_1B_2	A_2B_1	A_2B_2
Chromosome from mother	A_1B_1	w_{11}	w_{12}	w_{13}	w_{14}
	A_1B_2	w_{21}	w_{22}	w_{23}	w_{24}
	A_2B_1	w_{31}	w_{32}	w_{33}	w_{34}
	A_2B_2	w_{41}	w_{42}	w_{43}	w_{44}

Assuming that fitness depends only on the alleles carried (regardless of whether they're on maternal or paternal chromosomes), the equations describing evolution in the two-locus diploid model are:

$$x_1[t+1] = x_1[t] \sum_{j=1}^{4} x_j[t] \frac{w_{1j}}{\bar{w}} - \frac{w_{14}}{\bar{w}} r D[t]$$

$$x_2[t+1] = x_2[t] \sum_{j=1}^{4} x_j[t] \frac{w_{2j}}{\bar{w}} + \frac{w_{14}}{\bar{w}} r D[t]$$

$$x_3[t+1] = x_3[t] \sum_{j=1}^{4} x_j[t] \frac{w_{3j}}{\bar{w}} + \frac{w_{14}}{\bar{w}} r D[t]$$

$$x_4[t+1] = x_4[t] \sum_{j=1}^{4} x_j[t] \frac{w_{4j}}{\bar{w}} - \frac{w_{14}}{\bar{w}} r D[t]$$

(DO NOT MEMORIZE)

where

$$\bar{w} = \sum_{i=1}^{4} \sum_{j=1}^{4} x_i \, x_j \, w_{ij}$$

is the *mean fitness* of all members of the current population.

Important Properties of Two-Locus Model

Surprisingly, even with only two loci, we do not understand the general behavior of the chromosome frequencies over time under selection.

Several important results have been determined, which we will highlight.

Two Neutral Loci

- Disequilibrium decays at a rate r every generation.
- Allele frequencies do not change over time.
- Chromosome frequencies do change as the disequilibrium decays.

One Neutral Locus; One Selected Locus

- Selected locus evolves as in the one-locus selection model.

- Allele frequencies at neutral locus do not stay constant if there is linkage disequilibrium!

- If A_1 is favorable and D is positive, B_1 increases in frequency.

- If A_1 is favorable and D is negative, B_1 decreases in frequency.

➡ *GENETIC HITCHHIKING* = Allele frequencies change at the neutral locus because of an association (D) with the selected locus.

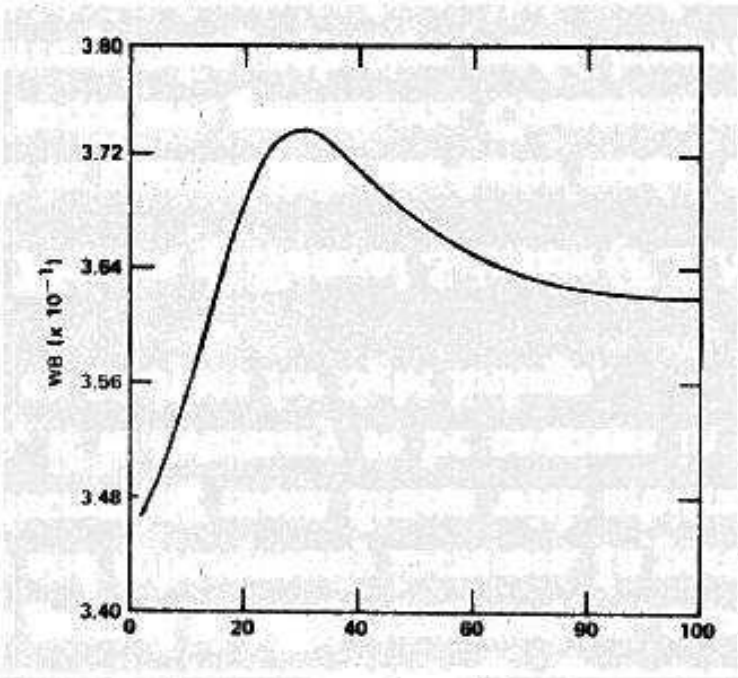
- The extent of hitchhiking will be large only when D is large and r is small (D decays slowly).

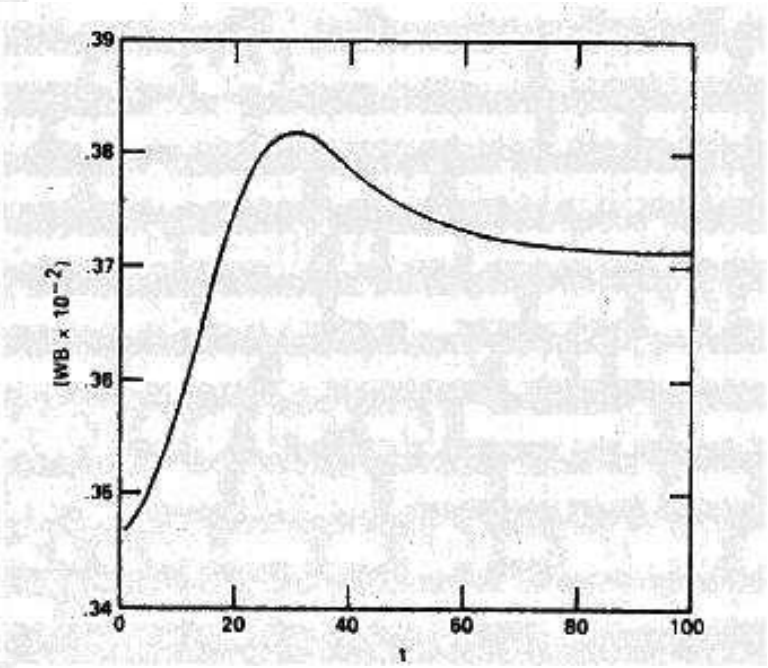
Two Selected Loci

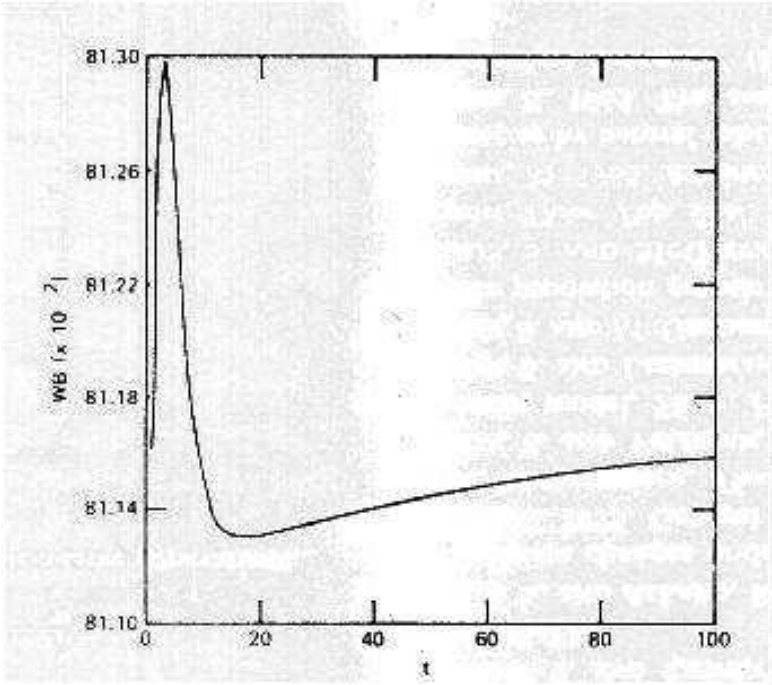
- Not fully analysed.
- It is thought that with purely directional selection the favorable genotype rises to fixation (NOT PROVEN).
- With heterozygote advantage at both loci,
  - If fitnesses at the two loci ADD together to determine fitness (additive model), the one-locus equilibrium predicts the two-locus equilibrium.
  - If fitnesses at the two loci MULTIPLY together to determine fitness (multiplicative model), the one-locus equilibrium predicts the two-locus equilibrium ONLY when r is large. When r is small, the population approaches a polymorphic equilibrium where linkage disequilibrium is maintained indefinitely.
- Mean fitness can decline over time!!
- Cycling can occur under some fitness schemes.

Mean fitness behavior in simulations of Karlin and Carmelli (1975):

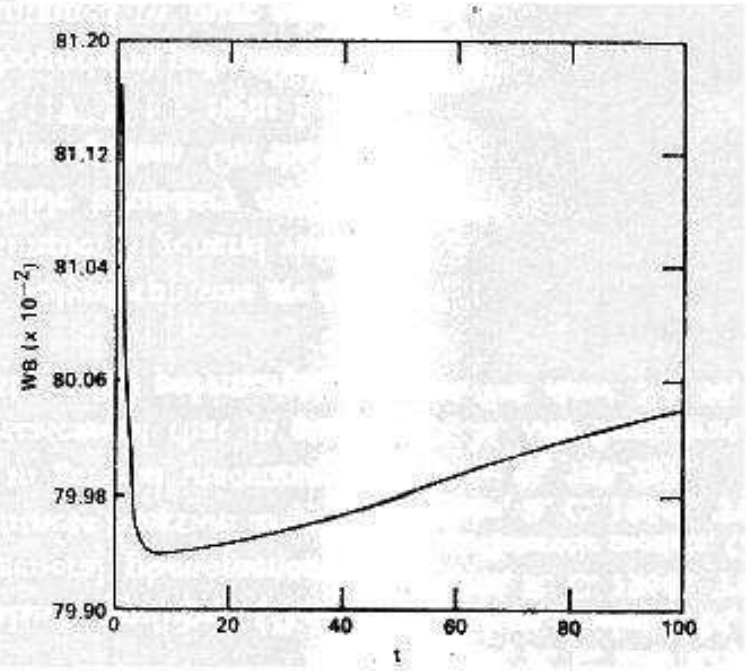
Mean Fitness







Time



Time

IMPORTANT CONCEPTS TO REMEMBER:

- Linkage disequilibrium measures associations among alleles at different loci.
- Linkage disequilibrium decays over time in the absence of selection.
- With selection, linkage disequilibrium may be generated and maintained by selection *even at equilibrium*.
- Linkage disequilibrium between a selected and a neutral locus can cause alleles at the neutral locus to change in frequency (hitchhiking).
- Mean fitness need not increase.

SOURCES:

- Drosophila disequilibrium figure: Ridley (1996) p. 207, originally from Langley (1977).
- Two-locus results: Karlin (1975) TPB, 7:364-398.
- Mean fitness decreases: Karlin and Carmelli (1975) TPB, 7:399-421.

Quantitative Genetics

The genetic basis of many traits is only poorly known.

We lack specific information about:

- The number of genes
- The position of genes within the genome
- The fitness effects of particular alleles
- The interactions among loci

Even if we had more information about the genetic basis of a trait, explicit models with multiple loci are astonishingly complex.

(Remember that even the two-locus model is not completely understood!)

➡ Evolutionary models fall into two main camps:

(a) Models that explicitly follow allele frequency changes at specific loci

- Behavior of these models is well understood only for one or two loci.

(b) Models that ignore genetic details (e.g. recombination rates, linkage disequilibria) and focus on the average effect of many genes

- Such simplified multi-locus models are easier to analyse. However, it remains unclear to what extent the genetic details matter.

The latter class of models are called:

Quantitative Genetic Models

Building a Quantitative Genetic Model

Phenotype: The visible [or measurable] properties of an organism that are produced by the interaction of the genotype and the environment.

- Webster's

In quantitative genetics, the phenotypic value (= **P**) of an individual (e.g. height) is attributed to the genotype of the individual and to its environment:

$$\mathbf{P = G + E}$$

The genotypic value (= **G**) is a measure of the influence of every gene carried by the individual on the phenotypic value.

The environmental deviation (= **E**) is a measure of the influence of the environment of an individual, scaled such that the average value of **E** is zero within a population.

An additional component is possible when the environment and genotype interact to influence the phenotype of an individual. This is called *genotype-by-environment interaction (**G**x**E**)*.
(We will ignore this complication.)

Example: (From Intermediate Genetics Course, by P. McClean at North Dakota State Univ).

Genetically uniform strains of wheat were grown in Casselton, North Dakota by Dr. Jim Anderson.

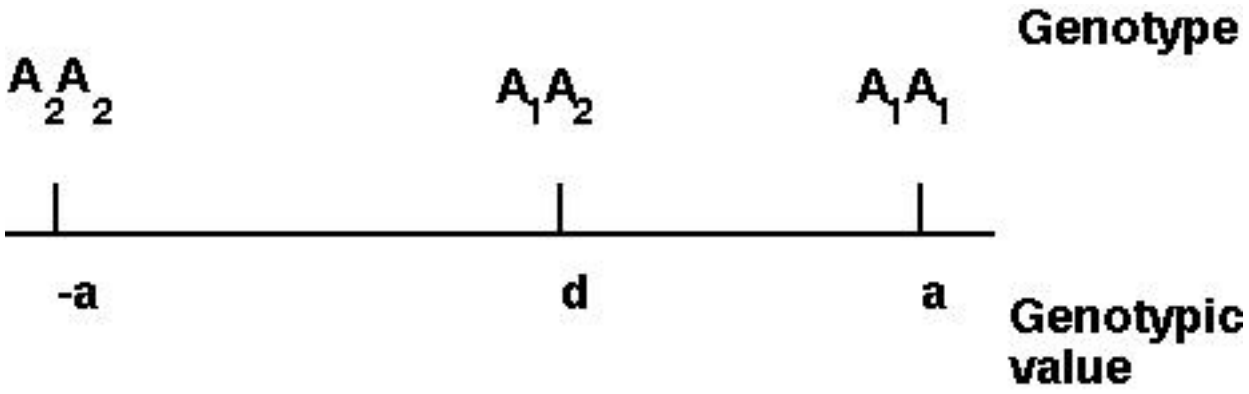
	Yield (bushels/acre)		
	Genotype		
Year	**Roughrider**	**Seward**	**Agassiz**
1986	47.9	55.9	47.5
1987	63.8	72.5	59.5
1988	23.1	25.7	28.4
1989	61.6	66.5	60.5
1990	0.0	0.0	0.0
1991	60.3	71.0	55.4
1992	46.6	49.0	41.5
1993	58.2	62.9	48.8
1994	41.7	53.2	39.8
1995	53.1	65.1	53.5

For the roughrider strain of wheat, the average yield (= the phenotype) over the ten years was 45.6 bushels/acre. This would be an estimate of the genetic (**G**) component of roughrider's phenotype.

In 1986, the environmental deviation was +2.3 for roughrider, such that **P = G + E** = 45.6 + 2.3 = 47.9.

From One-Locus to Many

Consider one locus with two alleles (A_1 and A_2):



Here we have arbitrarily assigned a scale to the genotypic values such that A_1A_1 individuals are given the value +a and A_2A_2 individuals are given the value -a.

(This notation follows that standardly used in the quantitative genetics literature.)

The mean phenotypic value at any point in time will depend on the allele frequencies:

$$\text{mean}(\mathbf{P}) = \text{mean}(\mathbf{G}) + \text{mean}(\mathbf{E}) = \text{mean}(\mathbf{G})$$

$$= p^2(a) + 2pq(d) + q^2(-a)$$

[Remember that **E** is measured such that mean(**E**)=0.]

If a parent reproduced clonally (like Dolly), then its offspring would inherit the complete genome of the parent. In this case, the parent and offspring would have the exact same value of **G**.

If a parent reproduces sexually, however, only one allele at each locus will be passed on to the offspring.

The next step is to measure the *average effect* on the phenotype of this one allele:

Average effect of an allele: The mean phenotype of individuals which received that allele from one parent when the other allele is chosen at random from the population, measured as a deviation from the population mean.

Since a randomly chosen allele will be A_1 with probability p and will be A_2 with probability q the average effect of allele A_1 would be:

$$p\,a + q\,d - \text{mean}(\mathbf{P})$$

IN OTHER WORDS

If a parent is homozygous at a locus, it cannot transmit this status to its children.

Only one allele is passed to the offspring, so whether the offspring will be homozygous or not depends on the allele frequency within the rest of the population.

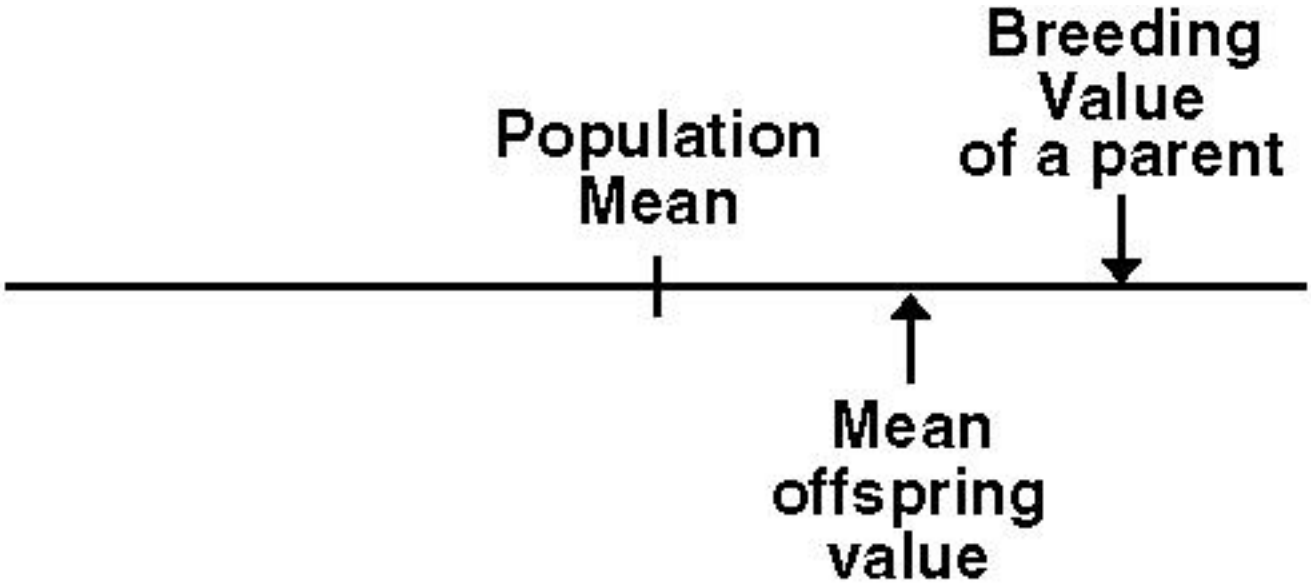
The average effect measures how offspring that inherit a specific allele differ from the population as a whole.

The advantage of defining the average effect of an allele transmitted from parent to offspring is that one can sum these effects up across many loci to estimate the average influence that a parent will have on their offspring through genetic inheritance!

(WARNING: This ignores interactions between loci.)

Breeding value of an individual (A): The sum of the average effects of all alleles carried by an individual.

Since only half of these alleles are transmitted from a parent to its offspring, the phenotype of offspring will, on average, be half the breeding value of the parent.



This provides a second definition of the breeding value: as twice the difference between the mean of the offspring of an individual and the mean of the population.

➡ MEASURABLE

If you take the genotypic value of an individual (**G**) and subtract off its breeding value (**A**), you are left with terms that measure interactions among the alleles that the individual happens to carry.

These interactions are of two sorts:

- Dominance interactions among alleles at a locus (**D**)
- Epistatic interactions among alleles at different loci (**I**)

All these terms are important in determining what an individual will look like:

$$P = G + E$$

$$= A + D + I + E,$$

but it is primarily the breeding value (**A**) that influences the phenotype of the offspring of this individual.

Example: To increase milk yield, dairy farmers estimate the breeding value of bulls from the average dairy production of the daughters of each bull. Those bulls with a higher breeding value are then used to produce the next generation of cows.

Say that the daughters of a particular bull (mated to several cows) produce 100 liters of milk per day in a herd with an average production of 75 liters. The breeding value of the bull will then be estimated at 125.

Say that a particular cow produces 100 liters of milk per day compared to an average of 75. Her daughters (when mated with different bulls) produce 80 liters of milk per day. The mother cow is estimated to have a breeding value of 85.

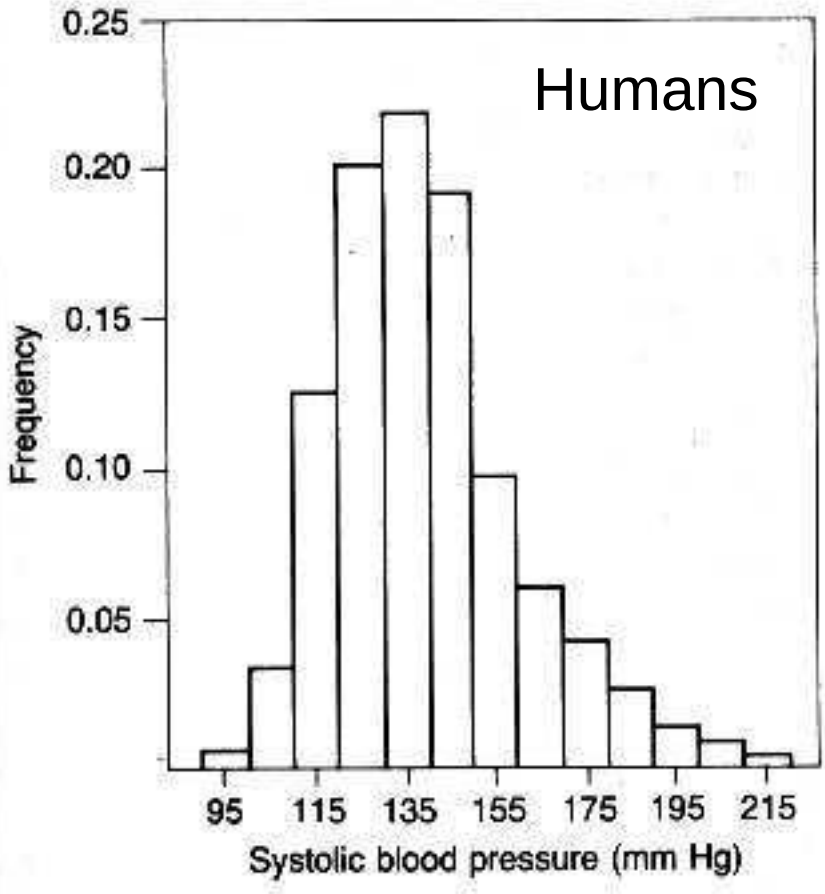
The difference between her actual yield (100) and her breeding value (85) will be due to rearing environment and/or interactions among alleles (at the same or different loci).

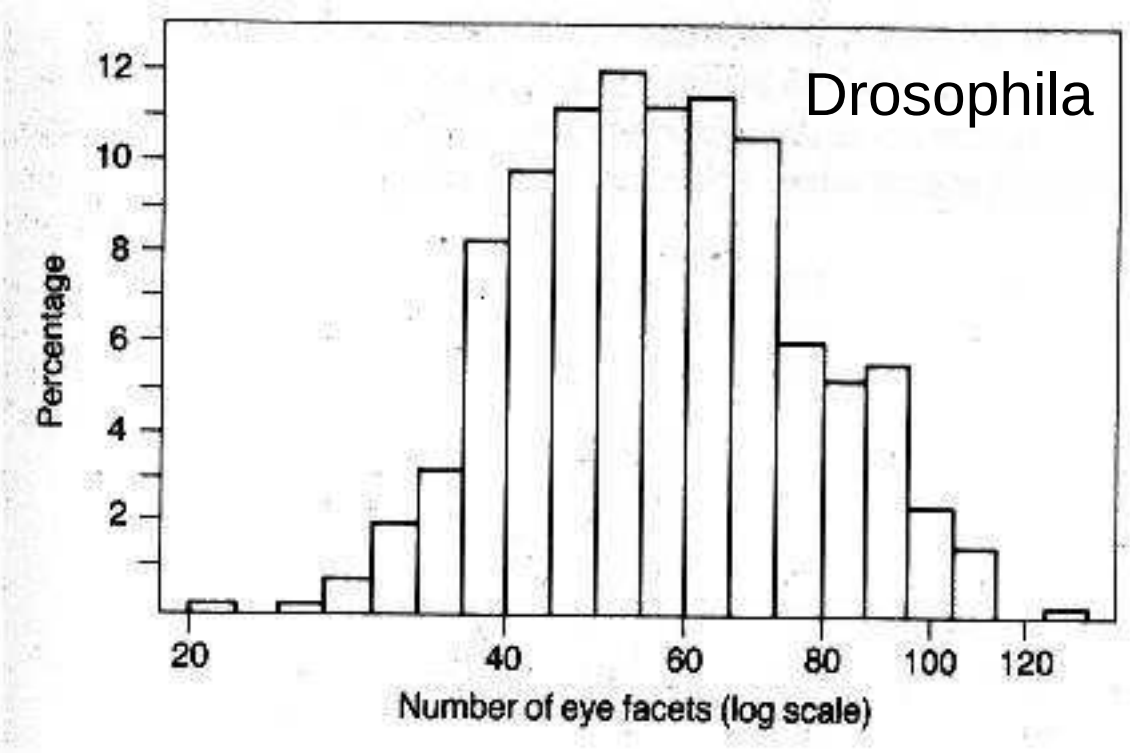
Describing a Population

Quantitative genetics is particularly concerned with describing the variation within a population and with estimating the genetic component of this variation.

The variance of phenotypic values (V_P) is an important measure in quantitative genetics.

For many traits, phenotypic values vary greatly within a population:





Just as the phenotype of an individual may be due to genetic and/or environmental influences, the variability within a population may be due to genetic or environmental differences among individuals.

The phenotypic variance is sub-divided into genotypic variance (V_G) and environmental variance (V_E):

$$V_P = V_G + V_E$$

(We are ignoring any correlation or interaction between genes and environment.)

The genotypic variance describes how much phenotypic variance is due to genetic differences within a population.

Let’s now ask how similar, on average, are parents to their offspring?

But just as the genotypic value of an individual is not passed in its entirety to offspring, only a component of the genotypic variance explains the resemblance of relatives within a population.

The genotypic variance is further sub-divided into:

- Additive genetic variance (V_A)
- Dominance variance (V_D)
- Epistatic variance due to interactions among loci (V_I)

$$V_G = V_A + V_D + V_I$$

The additive genetic variance (V_A) equals the variance in breeding values of individuals within a population.

Example: The dairy yields of 10 cows are:

75, 88, 52, 83, 82, 43, 100, 48, 79, 100.

The phenotypic mean is 75.

The phenotypic variance is estimated as 425.6, calculated from:

$$V_x = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

The breeding values of these same cows are:

74, 88, 56, 83, 82, 57, 93, 54, 84, 81.

The mean breeding value is 75.2.

The additive genetic variance is estimated as 205.5 (also calculated from the above formula).

Inheritance

These quantities can be used to determine how much relatives will resemble one another.

They also determine how much evolutionary change will be accomplished when certain parents reproduce while others do not.

Broad Sense Heritability (H^2): The extent to which variation in phenotype is caused by variation in genotype ($=V_G/V_P$).

Broad sense heritability will be 1 if all of the phenotypic variation within a population is due to genotypic differences among individuals.

Broad sense heritability will be 0 if all of the phenotypic variation is caused by environmental differences.

Example: The roughrider strain of wheat is genetically uniform. Therefore, all variation in yield among plants is due to environmental differences among plants. $H^2 = 0$.

➡ Heritability is measured in a particular population at a particular place and time.

Observing $H^2 = 0$ in roughrider does not mean that wheat yield would have no genetic component in other populations of wheat.

Narrow Sense Heritability (h^2): The extent to which variation in phenotype is caused by genes transmitted from parents (= V_A/V_P).

Narrow sense heritability will be 1 only if there is no variation due to genetic interactions (dominance or epistasis) or to environmental differences. When $h^2 = 1$, **P** = **G** = **A**.

Narrow sense heritability can be zero even if broad sense heritability is not (BUT NOT VICE VERSA) because all the genetic variation within a population may be due to dominance or epistasis.

Example: For the ten cows, V_P = 425.6 and V_A = 205.5. The narrow sense heritability is 205.5/425.6 = 0.48.

Example: Variation in human birth weight (from Penrose, 1954; Robson, 1955)

Cause of variation	% of total
Genetic	18
Additive	15
Non-Additive	1
Sex	2
Environmental	82
Maternal genotype	20
Maternal environment	24
Age of mother	1
Parity	7
Intangible	30

A very large portion of the phenotypic variability is environmental in origin.

The broad sense heritability of birth weight is only 0.18 and the narrow sense heritability is 0.15.

Looking at Relatives

(We'll assume that mating is random and that epistasis and gene-by-environment interaction can be ignored.)

A parent with genotypic value **G** will have an offspring whose genotypic value will, on average, be **A**/2 (half of the parent's breeding value).

It can be shown (see proof if interested) that the covariance between the phenotype of a parent and its offspring will therefore equal $1/2\ V_A$.

More importantly, the correlation between the phenotypic values of parents and offspring will equal:

$$\text{Corr[parent, offspring]} = 1/2\ V_A/V_P = 1/2\ h^2$$

One can therefore use the correlation between offspring and parents to estimate the narrow sense heritability:

$$h^2 = 2\ \text{Corr[parent, offspring]}$$

If both parents are measured, the correlation between the mid-parent values and the offspring value is:

$$\text{Corr[mid-parent, offspring]} = V_A/V_P = h^2$$

The correlation between two full-siblings is:

$$\text{Corr[full siblings]} = (1/2\ V_A + 1/4\ V_D)/V_P$$

The expected correlation between many other types of relatives has also been calculated.

Sample heritabilities (from Falconer, 1989):

		h^2
Man		
	Stature	0.65
	Serum immunoglobulin level	0.45
Cattle		
	Adult body weight	0.65
	% butterfat	0.40
	Milk-yield	0.35
Pigs		
	Back-fat thickness	0.70
	Weight gain per day	0.40
	Litter size	0.05

The Importance of Heritability

Narrow-sense heritability is particularly important in animal and plant breeding, because it predicts what the will happen in the next generation when particular parents are chosen to breed.

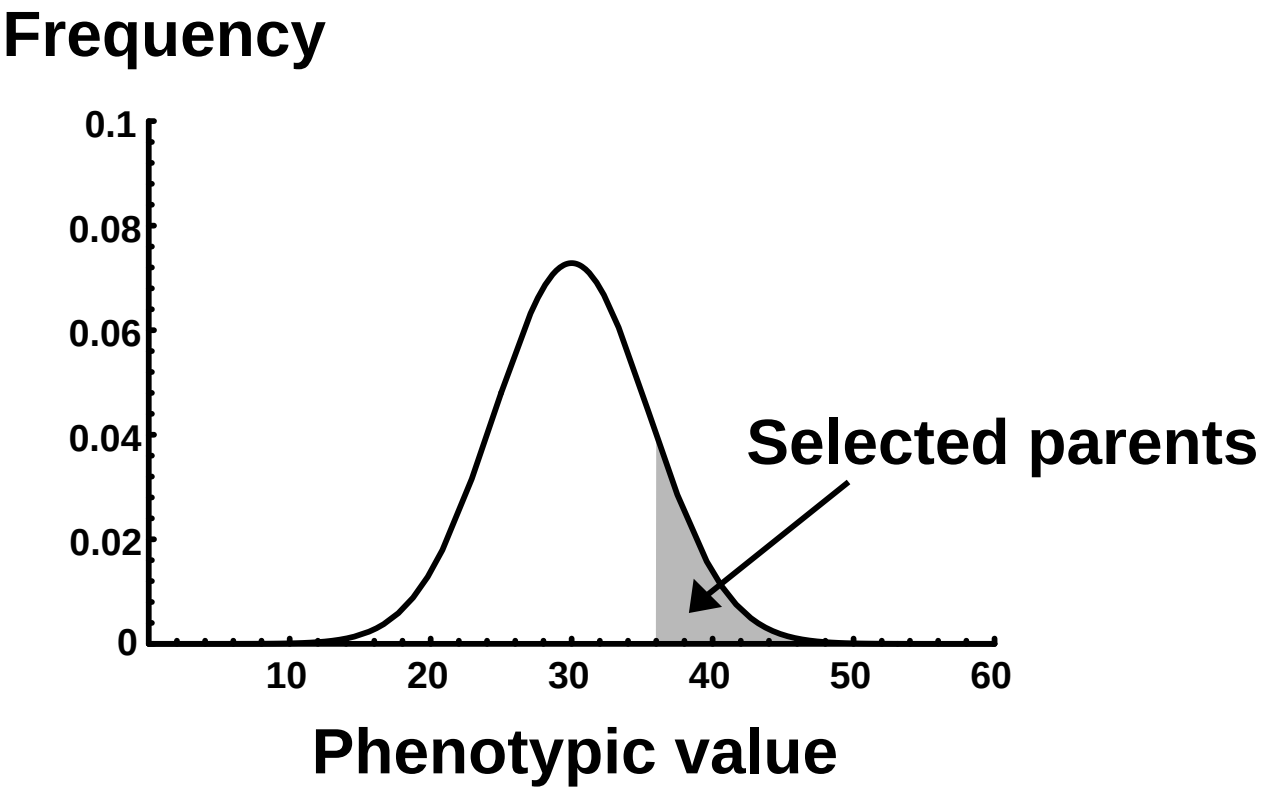
Selection differential (S): The phenotypic mean of parents chosen to breed minus the population mean.

Response to selection (R): The phenotypic mean of offspring of these parents minus the population mean.

The most important formula in quantitative genetics states that:

$$\mathbf{R} = h^2 \mathbf{S}$$

Characters with a high heritability (e.g. back-fat thickness in pigs) will respond rapidly to selection, whereas characters with low heritability (e.g. litter size in pigs) will respond slowly.



Population mean = 30

Mean of selected individuals = 38.8

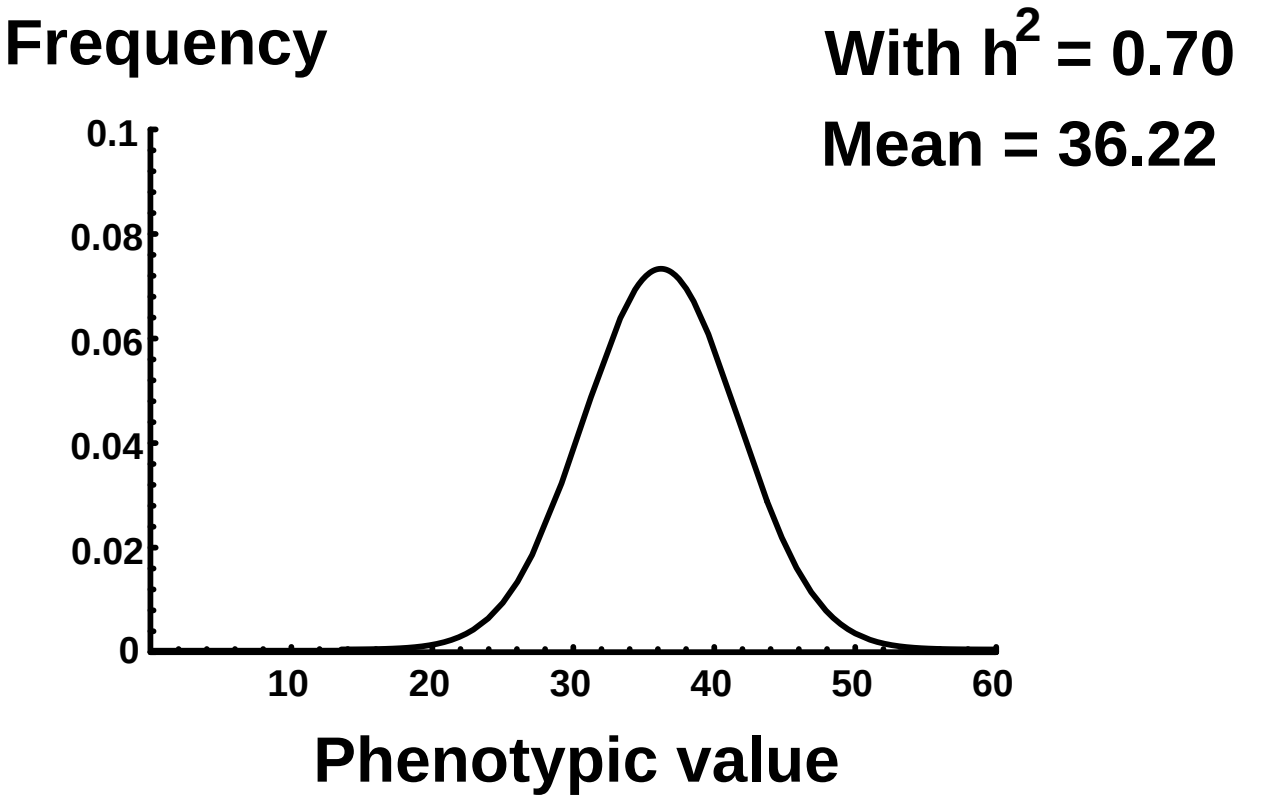
Selection differential = 38.8 - 30 = 8.88

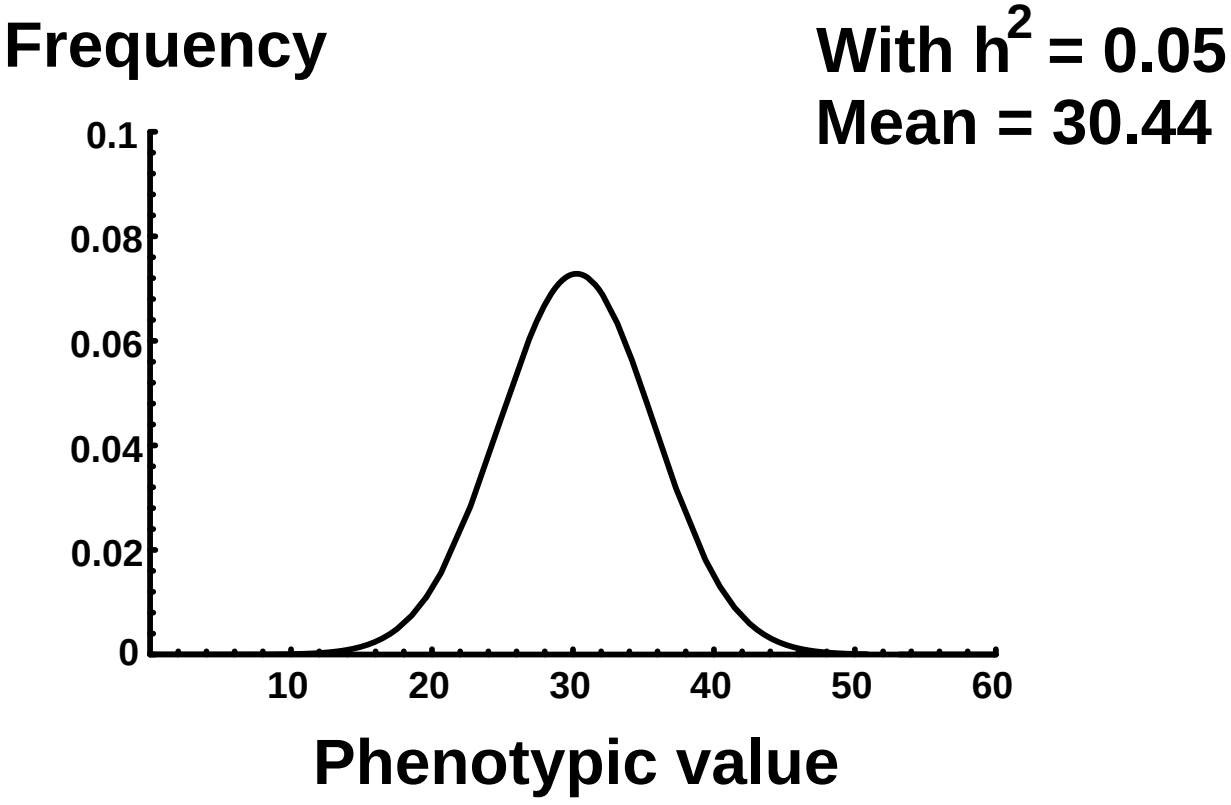
Response to selection:

with $h^2 = 0.70$ R =

with $h^2 = 0.05$ R =

EXPECTED OFFSPRING DISTRIBUTIONS





EXAMPLE: Clayton, Morris, and Robertson (1957) selected on abdominal bristle number in Drosophila melanogaster.

A previous estimate for heritability of bristle number within this population was 0.52.

The parental mean was 35.3 bristles.

Among those allowed to breed (those selected), the mean number of bristles was 40.6.

The selection coefficient was S = 40.6-35.3 = 5.3.

The response to selection was expected to be
R(exp) = 0.52*5.3 = 2.8.

The offspring generation actually had 37.9 bristles on average, for an *observed* response of
R(obs) = 37.9 - 35.3 = 2.6.

(Reasonably close!)

Long-Term Response to Selection

Heritability is not a constant attribute of a population.

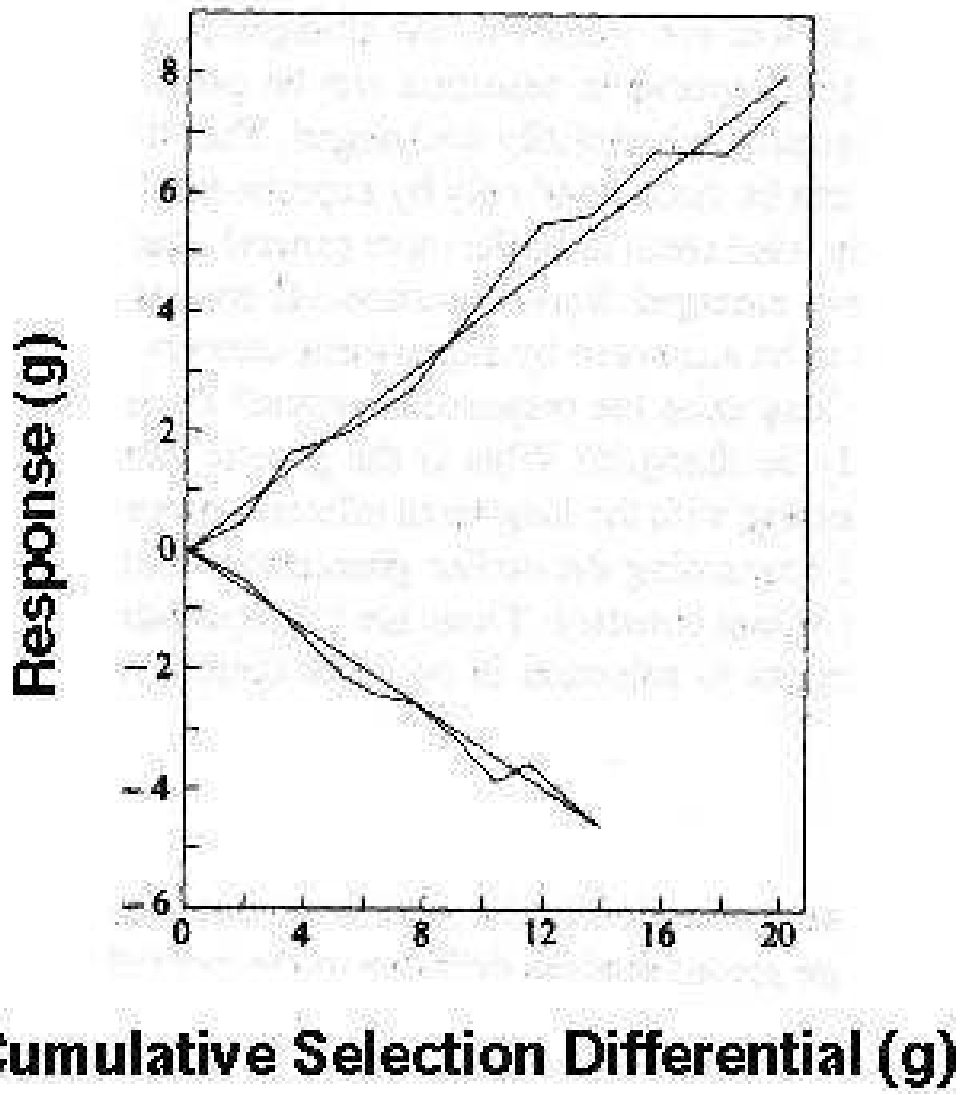
As the population evolves, the heritability of a trait will change as:

- allele frequencies change
- disequilibria change
- variance is reduced

Often, however, heritability remains roughly the same over a number of generations.

This can be seen if we plot response against the cumulative selection differential; if heritability remains constant, this gives a straight line.

EXAMPLE: Six-week weight in mice selected over ten generations (Falconer, 1973)

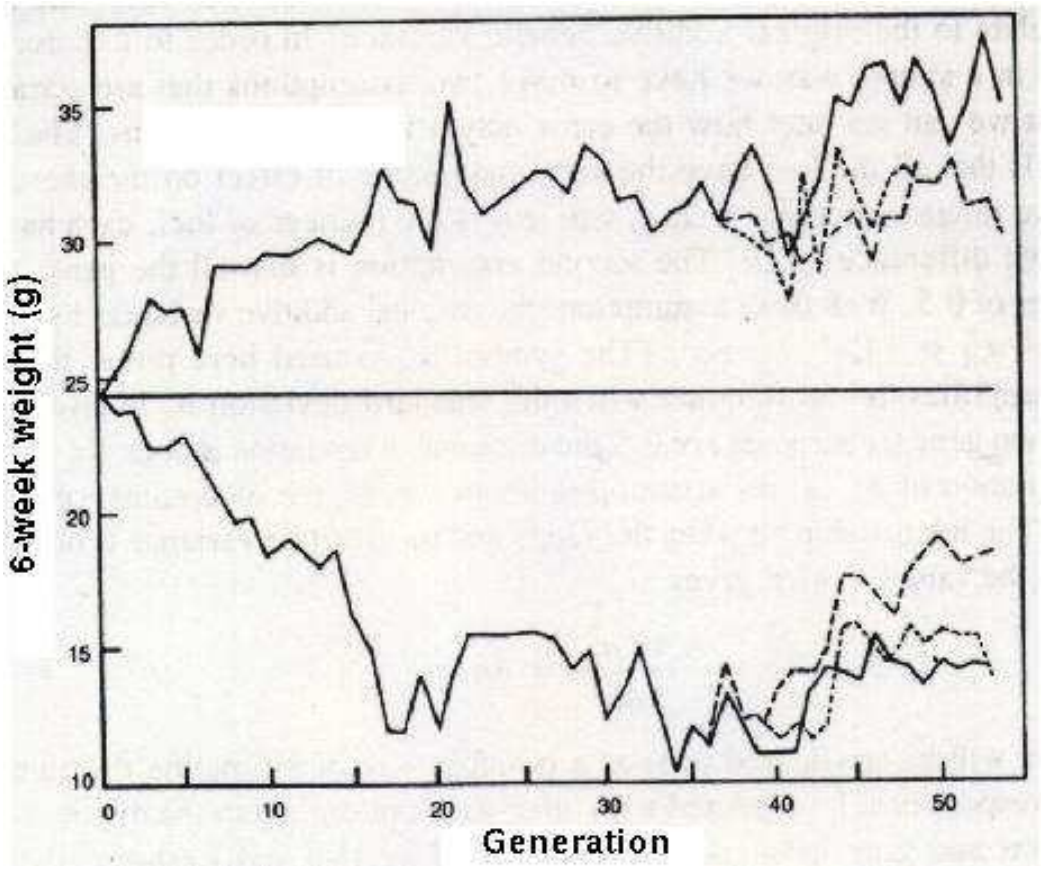


Here h^2 is estimated to be 0.398 in the case of upwards selection and 0.328 in the case of downwards selection.

Limits to Selection

After longer periods of time, the response to selection on a quantitative trait often reaches a plateau.

Mice weight in grams



The total response before a plateau is reached depends on many factors.

(1) The total response will be less when few individuals are chosen to breed, since less genetic variation is preserved among these individuals.

(2) The total response will be less when selection occurs rapidly because of genetic hitchhiking (some alleles that act in the opposite direction may get dragged along and fix, especially when S is high).

(3) The total response will be less if few loci contribute to the trait, since those few loci will go to fixation and since the array of possible combinations of alleles is much more limited.

In practise, animal and plant breeders have to juggle a number of factors (financial, logistic, biological) in an attempt to optimize the response of a bred population to selection.

SOURCES:

- Falconer, D. S. (1989) Introduction to Quantitative Genetics. 3rd Ed. Longman, London.
- Phillip E. McClean has an excellent set of web-pages teaching basic quantitative genetics.

Generating Novel Variation

Additive genetic variance in fitness (i.e. differences in the average fitness of alleles at a locus) is the basic fuel upon which evolution acts in a sexual population.

Yet, in most of the models we've discussed, selection exhausts this variation. Eventually:

One-Locus Model with Haploid Selection:

- *A* fixed (no genetic variation)
- *a* fixed (no genetic variation)

One-Locus Model with Diploid Selection:

- *A* fixed (no genetic variation)
- *a* fixed (no genetic variation)
- Polymorphic equilibrium (no difference in average fitness of *A* and *a*)

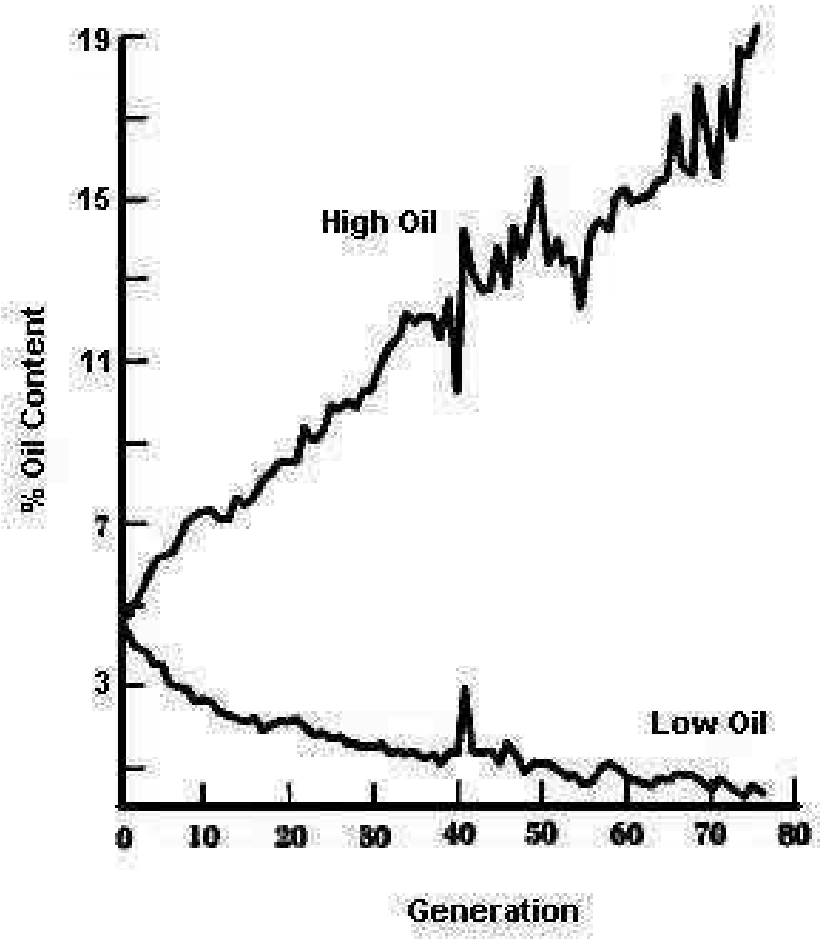
Two-Locus Model:

- Fixation at *A* and *B* locus (no genetic variation)
- Fixation at one-locus and polymorphic equilibrium at second (no difference in average fitness of alleles)
- Polymorphic equilibrium at both loci (additive genetic variance often absent)
- Cycling (additive genetic variance maintained)

Quantitative Genetic Model:

- Additive genetic variance is exhausted by selection

EXAMPLE: Oil content in corn.



Generation	Heritability (high line)	Heritability (low line)
1-9	0.32	0.50
10-25	0.34	0.23
26-52	0.11	0.10
53-76	0.12	0.15

(From Dudley, 1977; See Ridley p. 239).

What prevents evolution from grinding to a halt?

Mutation

Mutation is the ultimate source of novel genetic variation.

Mutation: Spontaneous change from one allele to another allele due to

- basepair substitutions
- deletions
- insertions
- inversions
- translocations

The error rate in DNA replication per generation varies from organism to organism:

Species	Genome size (basepairs)	Mutation rate per basepair	Mutation rate per genome
Bacteriophage lambda	$4.7\ 10^{4}$	$2.4\ 10^{-8}$	0.001
Escherichia coli	$3.8\ 10^{6}$	$4\ 10^{-10}$	0.002
Neurospora crassa	$4.5\ 10^{7}$	$5.8\ 10^{-11}$	0.003
Drosophila melanogaster	$4.0\ 10^{8}$	$2.3\ 10^{-9}$	0.93

(Estimates per generation.
From Drake 1974; See also Futuyma, 1998)

➡ The genome of an organism is not faithfully replicated from generation to generation.

EXAMPLE: Humans have a diploid genome size of 6.8 10^{9} basepairs. Assuming a generation time of 20 years and an est. mutation rate per year of 3.5 10^{-9} (Li, 1997), this suggests that **300-500** new mutations appear somewhere within the human genome each generation.

The spontaneous mutation rate also varies from gene to gene:

Species and locus	Mutations per 100,000 gametes (or cells)
Escherichia coli	
Streptomycin resistance	0.00004
Resistance to T1 phage	0.003
Arginine independence	0.0004
Drosophila melanogaster	
Yellow body	12
Brown eyes	3
Eyeless	6
Homo sapiens	
Retinoblastoma	1.2-2.3
Achondroplasia	4.2-14.3
Huntington's chorea	0.5

(From Dobzhansky, 1970; See also Futuyma, 1998)

The mutation rate also depends on the alleles involved:

e.g. Coat color mutations in mice (Russell 1963; Schlager and Dickie 1971)

- 11.2 10^{-6} per gene per generation (wildtype to mutant)
- 2.5 10^{-6} per gene per generation (mutant to wildtype)

Mutations disturbing wildtype function (*forward mutations*) often occur at higher rates than mutations restoring wildtype function (*back mutations*).

Roughly, in multicellular organisms, mutations occur at an approximate rate of 10^{-9} - 10^{-8} per basepair per year or 10^{-6} - 10^{-4} per gene per generation.

How does mutation affect the maintenance of variation?

One-Locus Mutation Model (No Selection)

In a neutral model without selection, mutations can maintain a large amount of genetic variation.

Let:

- μ is the mutation rate from *A* to *a*
- ν is the mutation rate from *a* to *A*

If p[t] is the frequency of *A* in the gamete pool after meoisis, then in the next generation in the absence of selection:

$$\mathbf{p[t+1]} = (1-\mu)\,\mathbf{p[t]} + \nu\ \mathbf{q[t]}$$

$$\mathbf{q[t+1]} = (1-\nu)\,\mathbf{q[t]} + \mu\ \mathbf{p[t]}$$

The only equilibrium of this equation is when p[t+1]=p[t]=$\hat{p}$:

$$\hat{\mathbf{p}} = \frac{\nu}{\mu+\nu}$$

$\hat{p}$=0 and $\hat{p}$=1 are NOT equilibria. Fixation is not stable when mutations recur.

However, the population approaches equilibrium at a rate of only $\mu + \nu$.

➡ **VERY SLOW!**

Example:

With $\mu = \nu = 10^{-6}$, the population will eventually reach the equilibrium of $\hat{p}=1/2$.

With a starting frequency of p[0]=0

- After 10,000 generations, p=0.0099
- After 100,000 generations, p=0.0906
- After 1,000,000 generations, p=0.4323

This occurs over such a long time frame that other forces such as selection (even very weak selection) or sampling error in finite populations are likely to overwhelm evolution of the system to $\hat{p}$.

One-Locus Diploid Model with Mutation and Selection

Mutations often cause changes in fitness.

In the vast majority of cases, these changes are deleterious (= reduce fitness), e.g. the many mutations causing severe human genetic diseases.

Only rarely will a mutation produce a more fit individual.

To what extent will selection be effective at eliminating deleterious mutations from a population?

NOTE: I will use "mutation" to refer to a new alteration in the DNA. I will use mutant to refer to the allele produced by mutation. A mutant allele may remain in a population long after it originated by mutation.

TABLE 3.9

X-Linked Traits: Estimates of Varying Validity of Birth Frequencies and Mutation Rates

No. in catalogue of X-linked genes (McKusick, 1968)	*Trait*	*Approx. relative reproductive fitness of trait bearers (f)*	*No. of families reported in literature*	*Orders of size of*	
				Birth frequency ($\times 10^{-6}$) (x)	*Mutation rate* ($\times 10^{-6}$) $\mu = \frac{1}{3}(1-f)x$
6	Duchenne type muscular dystrophy	0.0	Very many	200–220	70–90
7	Becker type muscular dystrophy	0.5	Many	9–12	1–2
8	Factor VIII Deficiency (Haemophilia A)[a]	0.6–0.9	Very many	100–120	20–40
9	Factor IX Deficiency (Christmas Disease—Haemophilia B)[a]	0.6–0.9	Very many	20–30	5–10
10	Hypogammaglobulinaemia	0.0	Many	10–15	3–5
11	Hurler's syndrome	0.0	Few	< 1.0	< 0.1
12	Late spondylo-epiphyseal dysplasia	High	~ 5	< 1.0	< 0.1
13	Wiskott-Aldrich syndrome (thrombocytopenia)	0.6–0.8	~ 10	< 1.0	< 0.1
14	Hypophosphataemia and Vitamin D-resistant rickets	0.3–0.5	Many	5–15	1–4
15	Hypoparathyroidism (Neonatal)	0.2–0.4	Few	< 1.0	< 0.1
16	Diabetes insipidus (Nephrogenic)	0.3–0.5	Few	< 1.0	< 0.1
17	Diabetes insipidus (Neuro-hypophyseal)	0.8–0.9	Few	1–5	~ 1.0
18	Oculo-cerebro-renal syndrome of Lowe	0	~ 10	< 1.0	< 0.1
19	Hypochromic anaemia (Pyridoxine responsive type)	High	Few	< 1.0	< 0.1
20	Fabry's syndrome (angiokeratosis corporis diffusum)	0.7–0.9	Few	2–5	< 1.0
22	Macular bullous dystrophy	Low	1	< 0.1	< 0.1
23	Keratosis follicularis spinulosa decalvans	High	10–15	< 1.0	< 0.1
24	Ichthyosis	No reduction detected	Very many	200	~ 1.0
25	Anhidrotic ectodermal dysplasia	0.6	Many	10	1–5
26 and 27	Amyelogenesis imperfecta	No reduction detected	Many	10	< 1.0
29 and 30	Deafness	0.8	Many	15–25	3–6
31	Mental defect (severe not specific)	0.0	Many	40–80	10–20
32	Borjeson's syndrome	?	2	< 0.1	< 0.1

[a] The figures here given are those of the survey by Stevenson and Kerr. Those cited in the text are from a later survey and analysis.

(From Cavalli-Sforza and Bodmer, 1971)

Let selection act against the mutant *a* allele, with the relative fitnesses of diploid individuals equaling:

AA	**Aa**	**aa**
1	**1-hs**	**1-s**

s measures the *selection coefficient* against the mutation and h measures the *dominance* of the mutation.

- Mutant allele (*a*) is dominant when h=1.
- Mutant allele (*a*) is recessive when h=0.

We will assume that s is larger than the mutation rate, μ, to *a*.

If we add mutation to the model of diploid selection, we find that the population rapidly evolves towards fixation on *A*, but rather than *a* being lost entirely, it reaches a *mutation-selection* balance:

$$\hat{q} = \frac{\mu}{h\,s}$$

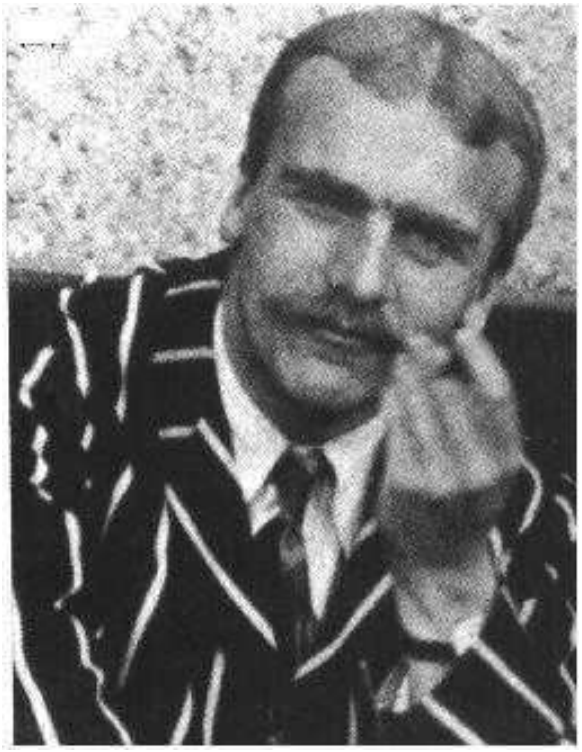
$$\hat{p} = 1 - \frac{\mu}{h\,s}$$

Mutant alleles are generally so rare that the mutant allele is almost always found in heterozygotes.

Only if the mutation is completely recessive (h=0) will homozygous mutants be common. In this special case (where only the *aa* genotype is selected against), the frequency of the mutant allele tends toward:

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

$$\hat{p} = 1 - \sqrt{\frac{\mu}{s}}$$



Much that we know about the dynamics of mutations comes from the pioneering work in population genetics done by J.B.S. Haldane in the 1920's.

"J.B.S. Haldane is reported to have said that his great pleasure was to see his ideas widely used even though he was not credited with their discovery."

-- Lewis Wolpert

For example, Haldane was the first to estimate mutation rates using the above equations.

EXAMPLE: Albinism is a recessive condition that occurs in humans at a frequency of 1/20,000 ($=q^2$). If albinos have a relative fitness of 0.9 ($s = 0.1$), then the mutation rate to the albino allele would be $\mu = 5\ 10^{-6}$.

EXAMPLE: Achondroplasia is a dominant condition causing dwarfing, which occurs at a frequency of 1/10,000.

Since it is dominant, most carriers are heterozygotes.

Therefore 2pq = 1/10,000 and q is approximately 1/20,000.

Affected individuals have an estimated relative fitness of 0.2 (s = 0.8).

The mutation rate to achondroplasia can then be estimated as $\mu = 0.8\ 1/20{,}000 = 4\ 10^{-5}$, which is consistent with the rate at which achondroplasia spontaneously appears within a population.

Mean Fitness

The mean fitness in the one-locus model with mutation and selection equals:

$$\overline{W} = p[t]^2 W_{AA} + 2\, p[t]\, q[t]\, W_{Aa} + q[t]^2 W_{aa}$$

$$= 1 - 2\, p[t]\, q[t]\, hs - q[t]^2 s$$

Mean relative fitness is therefore highest when q=0 and no mutant alleles exist in the population.

However, the population doesn't go towards q=0 (where the mean relative fitness equals 1), but rather to $q=\mu/hs$ (assuming that *a* is not completely recessive). At the mutation-selection balance, the mean fitness equals:

$$\overline{W} = 1 - 2\hat{p}\hat{q}\, hs - \hat{q}^2 s$$

$$= 1 - 2\left(1 - \frac{\mu}{hs}\right)\left(\frac{\mu}{hs}\right) hs - \left(\frac{\mu}{hs}\right)^2 s$$

$$\approx 1 - 2\mu$$

➡ Mutations reduce the mean fitness below the maximum possible by an amount equal to the mutation rate per diploid set of genes.

➡ Oddly, the decrease in mean fitness caused by deleterious mutations does not depend on the fitness effects of the mutations.

More severe mutations will exist at a lower frequency, while less severe mutations will reach a higher frequency, but the fraction of deaths caused by mutation will be the same in both cases.

Point 1 The mean fitness at equilibrium is not the maximum possible.

Point 2 The reduction in mean fitness due to mutation is not that great at one locus alone, but it may be substantial with mutations occurring throughout the genome (= "Mutation Load").

Point 3 Mean fitness can decrease over time. Whenever the population starts nearer q=0 than q=μ/hs, the mean fitness will decline towards 1-2μ.

While deleterious mutations decrease the mean fitness of a population, they are potentially an important source of genetic variation in the face of environmental change.

Genome-Wide Deleterious Mutations

If deleterious mutations at a particular gene reduce fitness by an amount 2μ, what are the fitness effects of deleterious mutations throughout the genome?

If there are no fitness interactions among genes (no epistasis), then the average fitness of a population will be $(1-2\mu)^{\#\ loci} \approx e^{-2\mu(\#\ loci)} = e^{-\mathbf{U}}$, where **U** is the sum total deleterious mutation rate in a diploid genome.

What is **U**?

Current estimates of **U** for multicellular animals and plants are roughly 0.2-2.0 (Lynch and Walsh 1998), but more data is sorely needed.

These estimates suggest a major fitness cost:

U	Average fitness = e^{-U}	Mutation load (= 1-fitness)
0.2	0.82	0.18
0.6	0.55	0.45
2.0	0.14	0.86

[If assumptions are accurate.]

Genome-Wide Mutations in a Quantitative Trait

For a quantitative trait, such as height or milk yield, knowing the mutation rate per locus or the genome wide-mutation rate is not particularly helpful, since the number of loci affecting the trait may not be known.

For certain quantitative traits, researchers have estimated the amount of new additive genetic variance arising by mutation each generation.

A common method to estimate variance due to new mutations is to take a genetically uniform population (where $V_G = 0$ and $V_P = V_E$) and subject it to selection.

The response of the population to selection can be used to determine h^2_m, the "mutational heritability" which equals the amount of new V_A caused by one generation of mutation divided by V_E.

(h^2_m will give the narrow sense heritability after one generation of mutation in a population with no genetic variance.)

Species and trait	h^2_m
Drosophila melanogaster	
Abdominal bristle number	0.0035
Ethanol resistance	0.0009
Viability	0.0003
Mouse	
Length of limb bones	0.0234
6-week weight	0.0034
Rice	
Plant size	0.0112
Reproductive traits	0.0073

(From Lynch and Walsh, 1998.)

If the genetic variability in a population is exhausted (e.g. by inbreeding or following a period of strong selection), which of the above traits would you expect to show the highest longer-term response to selection?

Why might some traits have a higher mutational heritability than others?

SOURCES:

- Li (1997) Molecular Evolution. Sinauer Associates, MA.
- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Lynch and Walsh (1998) Genetics and Analyis of Quantitative Traits. Sinauer Associates, MA.
- Cavalli-Sforza and Bodmer (1971) The Genetics of Human Populations. W. H. Freeman and Co., CA.
- Mukai et al. (1972) Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. Genetics 72: 335-355.

Evolution in Finite Populations

Up until now, we have treated populations as being infinitely large.

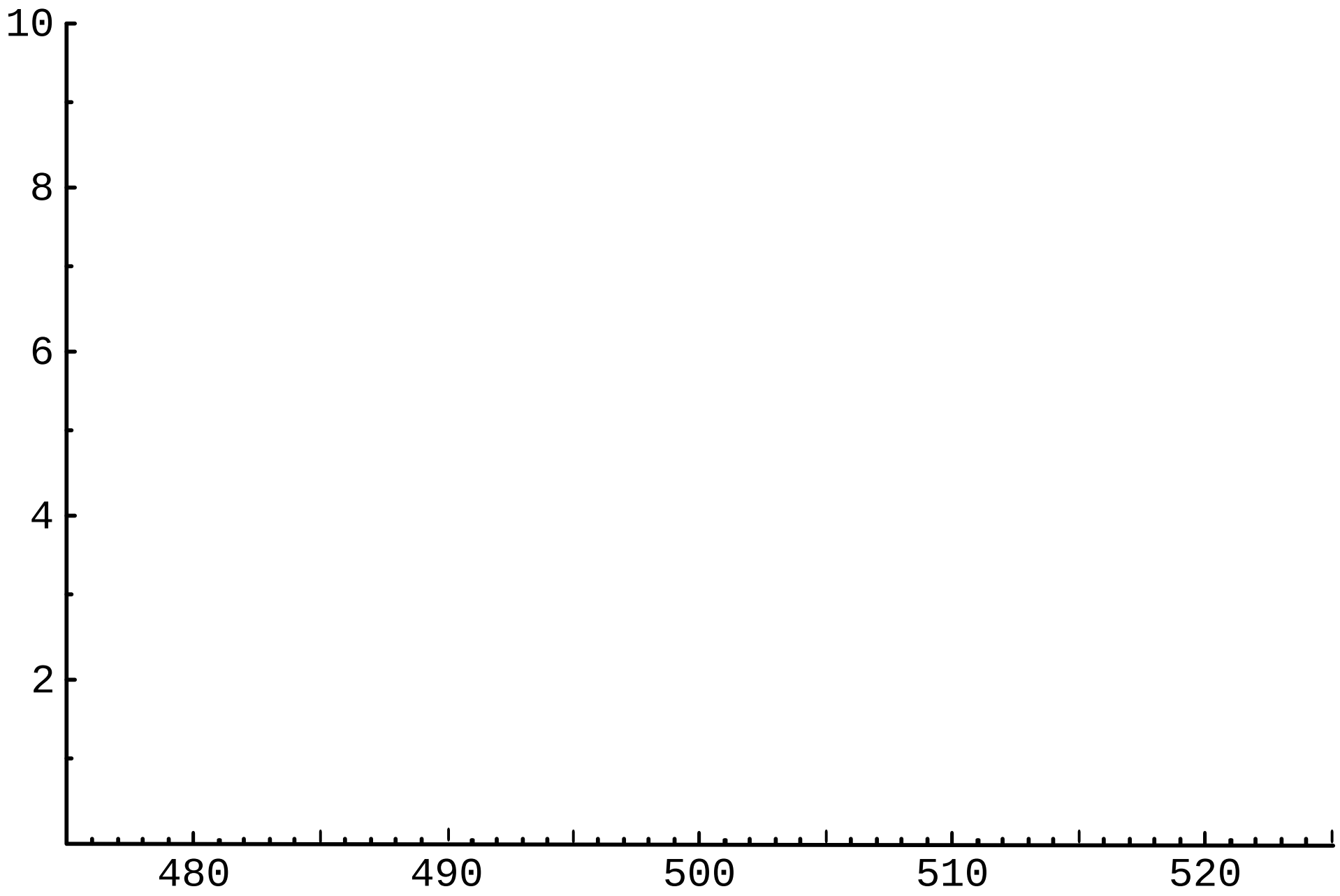
For example, we have used equations to determine p[t+1] assuming that each type produces EXACTLY as many offspring as predicted by their fitnesses.

In any finite population, however, individuals may have more or fewer offspring than expected, *simply by chance.*

The chance increases or decreases in the frequency of an allele in a finite population are called **random genetic drift**.

How does random genetic drift affect evolution in the absence of selection?

How does random genetic drift affect evolution in the presence of selection?


10
8
6
4
2
480
490
500
510
520

Random Genetic Drift in the Absence of Selection

Example: A Birth-Death Model.

Imagine a population of 1000 haploid individuals, half of which are *A* and half *a*.

At any point in time, any individual may die. It is then replaced by the reproduction of another individual chosen at random from the population.

At first, when the allele frequency remains near p=1/2:

- There is a 50% chance that *A* dies and
  - a 50% chance that it is replaced by *A*
    → No change in number of *A* alleles.
  - a 50% chance that it is replaced by *a*
    → Number of *A* alleles decreases by one.
- There is a 50% chance that *a* dies and
  - a 50% chance that it is replaced by *A*
    → Number of *A* alleles increases by one.
  - a 50% chance that it is replaced by *a*
    → No change in number of *A* alleles.

In this example, *A* and *a* are equally fit; they have an equal chance of dying and an equal chance of reproducing.

The class has simulated this model.

> Birth-death events where no change in number of *A* occurred were ignored.
>
> Of the birth-death events leading to a change in the number of *A* individuals, 50% (heads) led to an increase by one and 50% (tails) led to a decrease by one.

Consequently, the frequency of *A* varied over time from 1/2 by random drift.

(What would soon go wrong with our "simulation"?)

A more commonly used model of random genetic drift is called the **Wright-Fisher Model**, which assumes:

- A population of constant size where every individual reproduces at the same time.
- Each offspring allele is descended from a parent allele chosen at random from the previous generation.

Imagine labelling each allele in a population at some point in time. These "alleles" will drift up and down in frequency, until eventually only one remains.

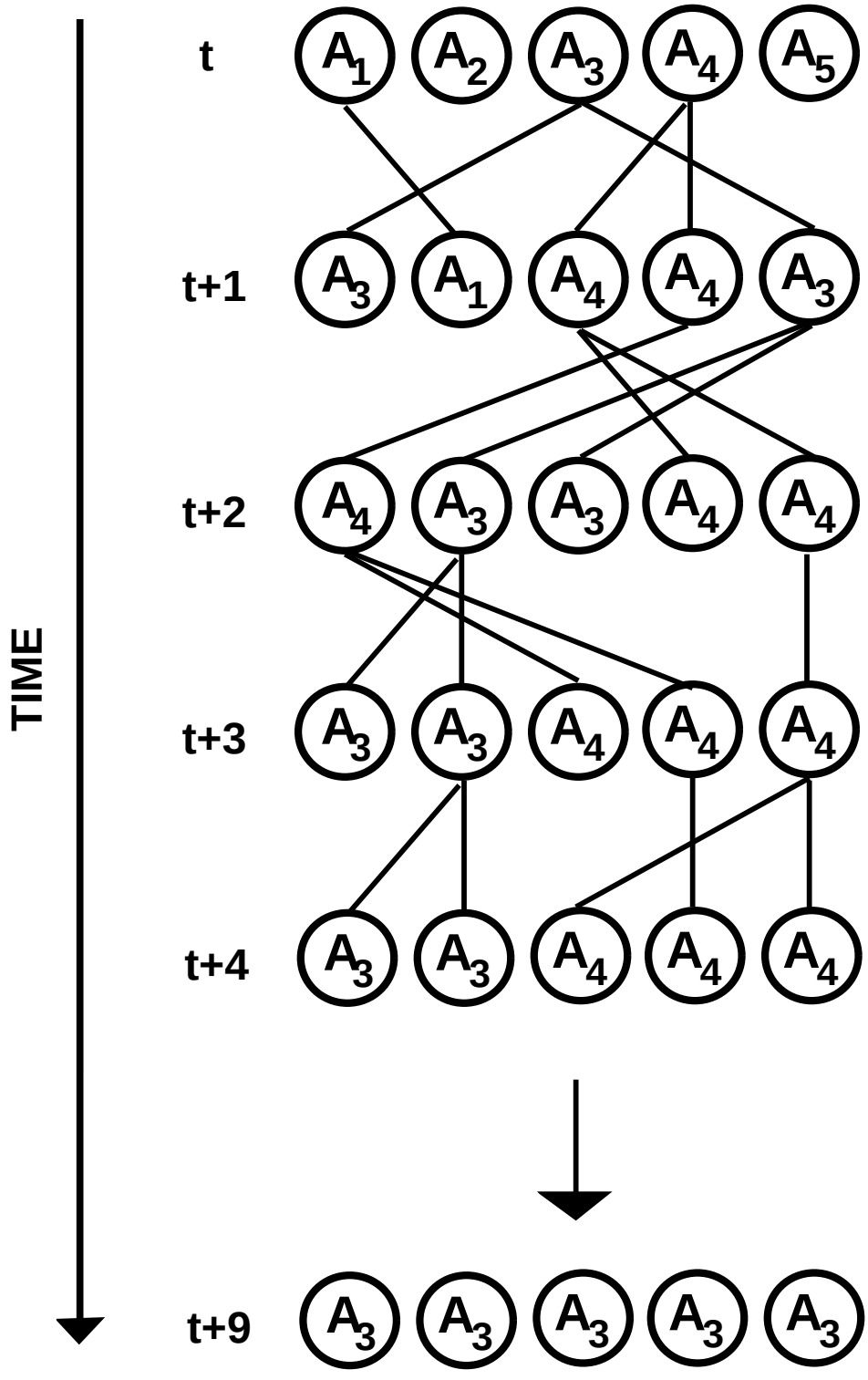
Any one of these alleles has an equal chance of being the "lucky" allele that fixes.

In a haploid population of size 5, what is the chance that allele #1 fixes?

In a haploid population of size N, what is the chance that any particular allele copy fixes?

If there are n copies of allele *A* and N-n copies of allele *a* in a haploid population, what is the probability that allele *A* eventually fixes?

Simulation with 5 alleles and parents chosen at random



In the absence of mutation and selection, allele frequencies drift up and down in frequency until, eventually, one allele becomes fixed.

We have been discussing the Wright-Fisher process for a haploid population, but the same method works for a diploid population as long as alleles within an individual are randomly sampled from a gamete pool.

In a diploid population with N individuals, there will be 2N alleles.

What is the chance that any particular allele copy fixes?

If there are n copies of allele *A* and 2N-n copies of allele *a* in a diploid population, what is the probability that allele *A* eventually fixes?

In the example shown above, it took nine generations for the third allele to fix.

In general, the AVERAGE amount of time it will take for a single allele to fix within a population is twice the number of alleles within a population:

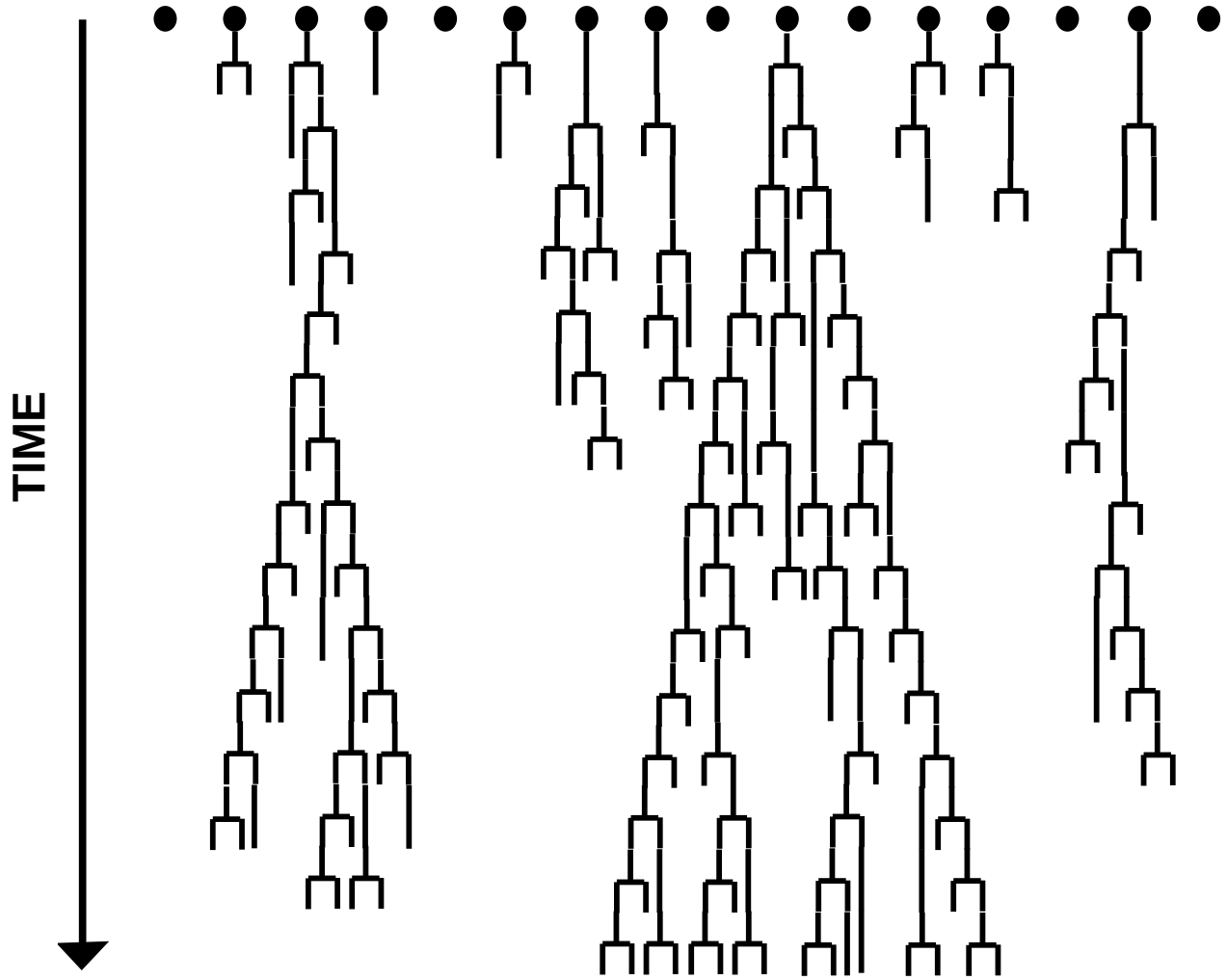
- 2N generations with N haploid individuals
- 4N generations with N diploid individuals

ASIDE: The time to fixation of an allele *A* is approximately the same whenever *A* is initially rare (small p[0]).

For example, in your PopBio assignment, p=0.1 and N=10 (diploid population), what fraction of the time should *A* fix?

How long should it take for *A* to fix when it does fix?

16 alleles (8 diploid individuals)



If it takes, on average, 4N generations for a single allele to spread to fixation within a diploid population (forwards in time), how long ago in the past, on average, must we look before all the alleles currently present in the population shared a common ancestor?

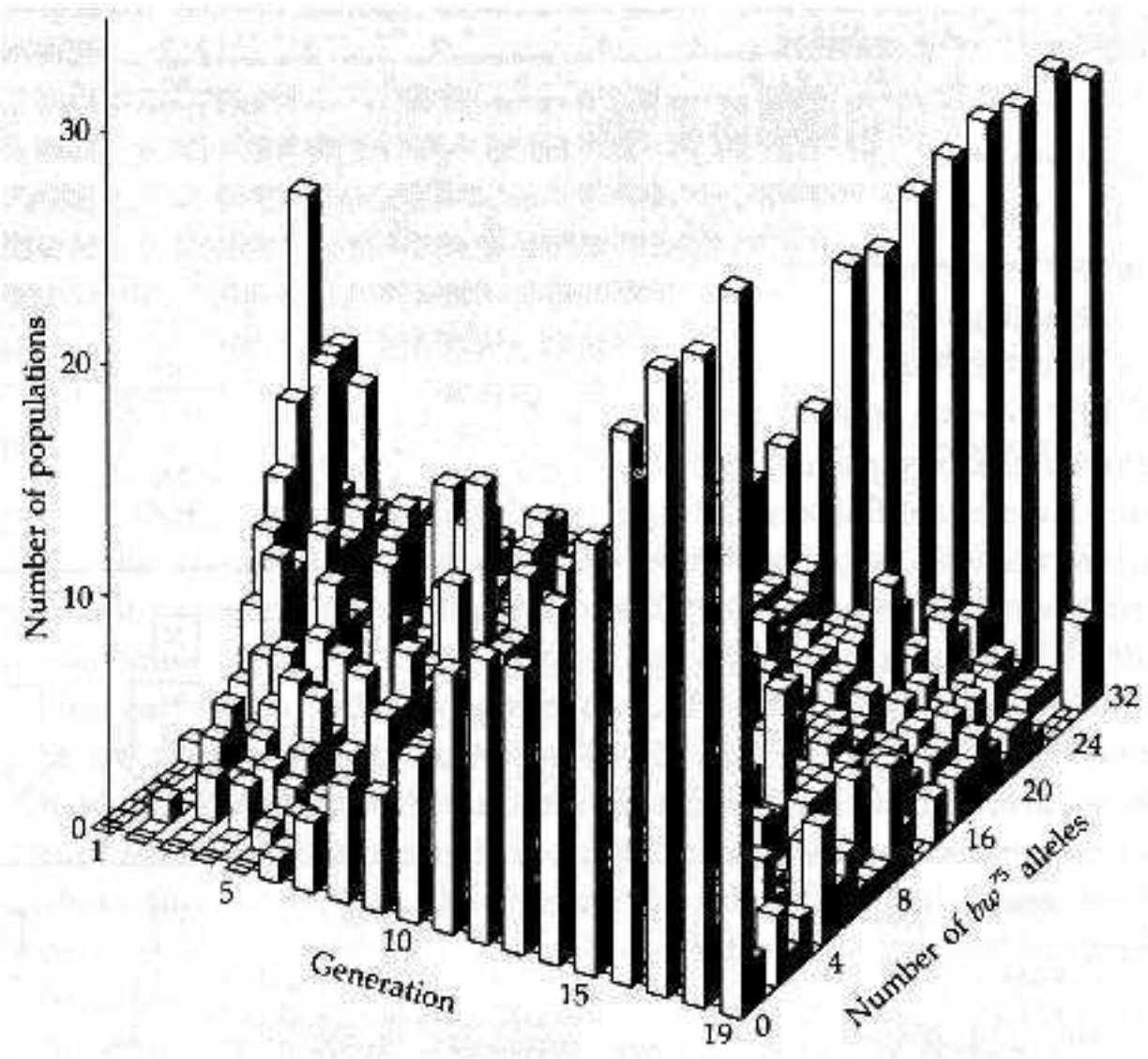
[This is called the coalescence time.]

We can look at this process forward or backwards in time!

Random genetic drift causes, on average, a loss of genetic variation within a population.

This is obviously true in the long-term, since a population eventually drifts to fixation on one allele.

For example, Buri (1956) examined 107 Drosophila populations that each started with 16 heterozygotes for a brown eye mutation (*bw*):



(Reported in Hartl and Clark, 1989.)

The loss of variability can be measured by changes in the "Expected Homozygosity" (f) of a population.

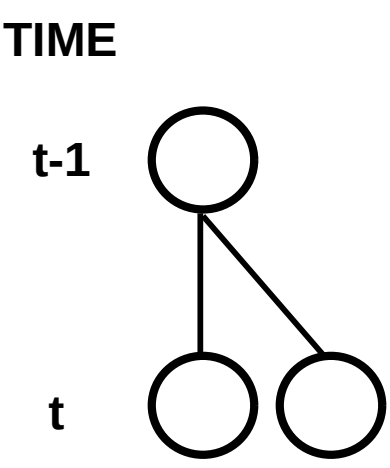
Expected Homozygosity (f) = The probability that two alleles drawn at random are the same allele.

For instance, with two alleles at a locus, the expected homozygosity will initially equal:

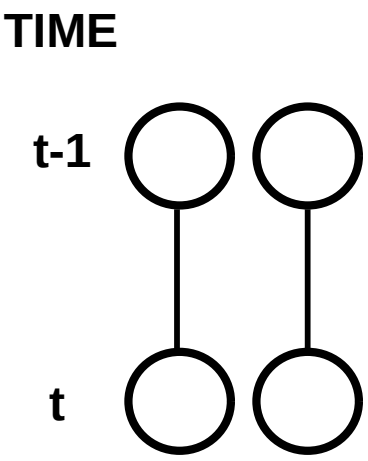
$$f[0] = p^2[0] + q^2[0],$$

since $p^2[0]$ is the probability that both the first and the second allele chosen at random are *A* and $q^2[0]$ is the probability that both the first and the second allele are *a*.

In generation t, when two offspring alleles are chosen from the population at random, there are two possibilities:



Offspring alleles are descended from the same parental allele



Offspring alleles are descended from different parental alleles

(1) What is the probability that both offspring alleles come from the same parental allele in a diploid population with 2N alleles?

(2) What is the probability that both offspring alleles come from different parental allele in a diploid population with 2N alleles?

In the first case, homozygosity becomes one (the two alleles ARE same allele in the previous generation).

In the second case, the two offspring alleles will be the same allele if they are descended from parents with the same allele, which is given by the expected homozygosity in the previous generation, f[t-1].

$$f[t] = 1/(2N) + (1-1/(2N))\, f[t-1]$$

The smaller the population, the more likely that two alleles will be the same just because they had the same parent.

➡ The expected homozygosity rises in all finite populations, but rises fastest in small populations.

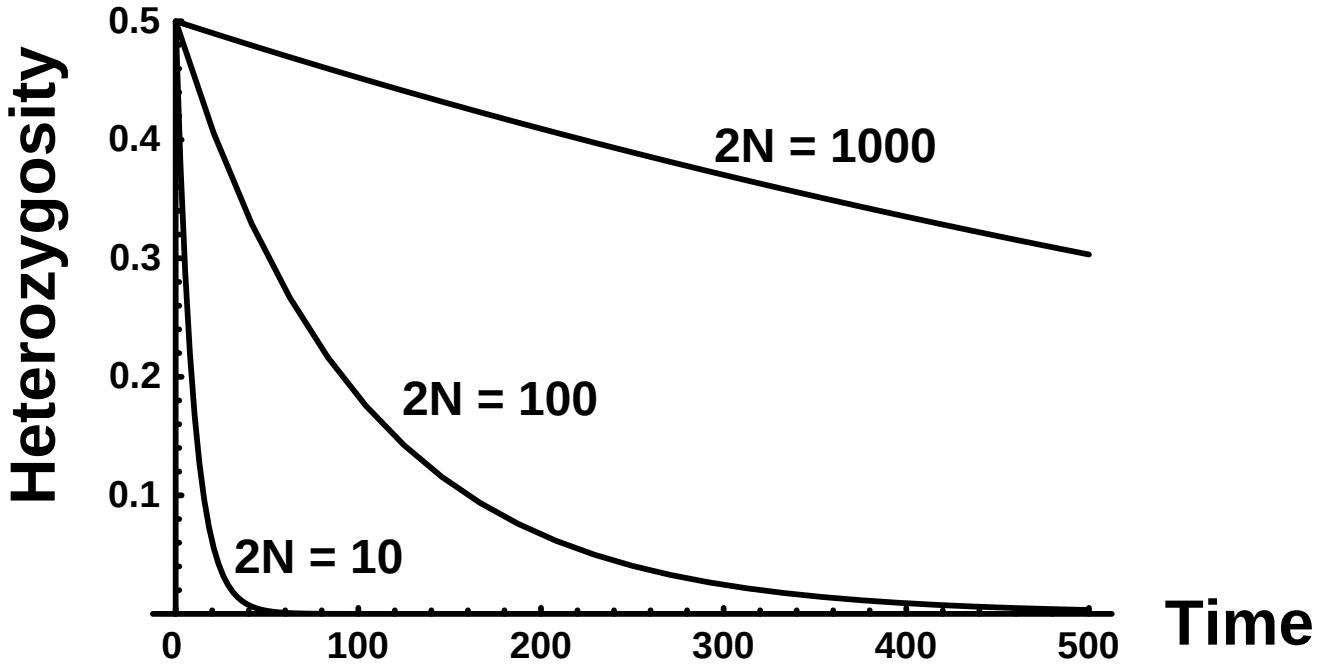
As expected homozygosity rises, the amount of genetic variation within the population declines. The amount of genetic variation is measured by:

Expected Heterozygosity (H) = The probability that two alleles drawn at random are different alleles: H=1-f.

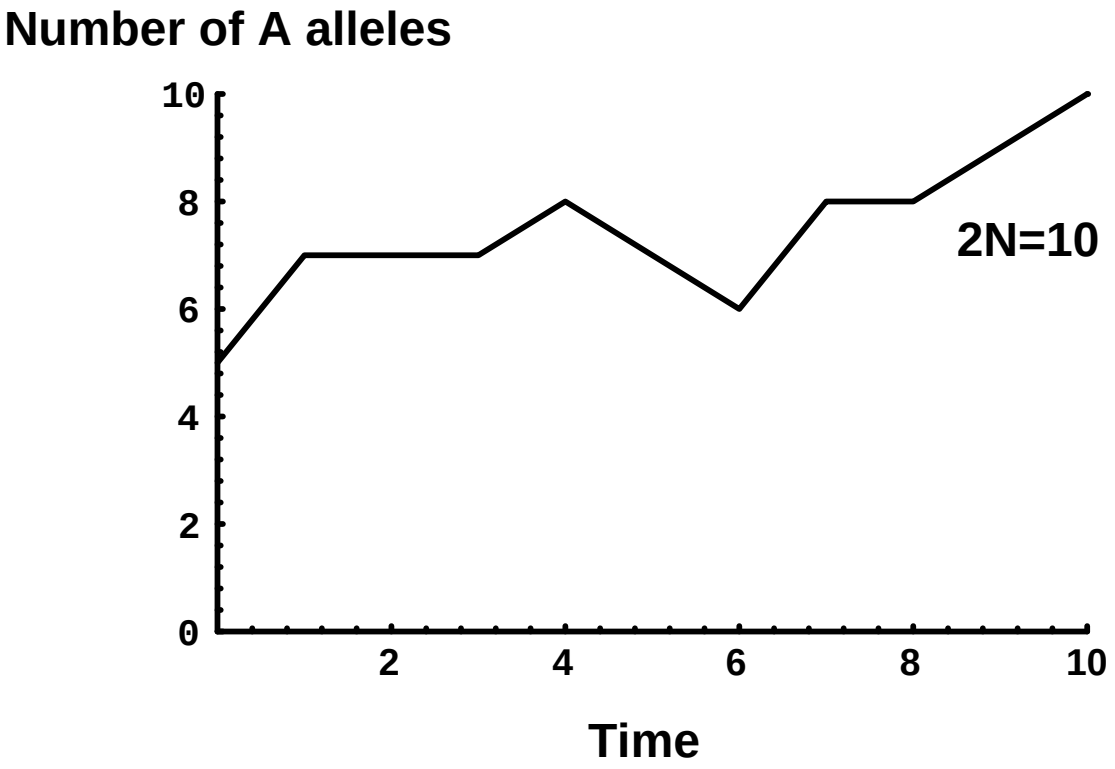
In generation t,

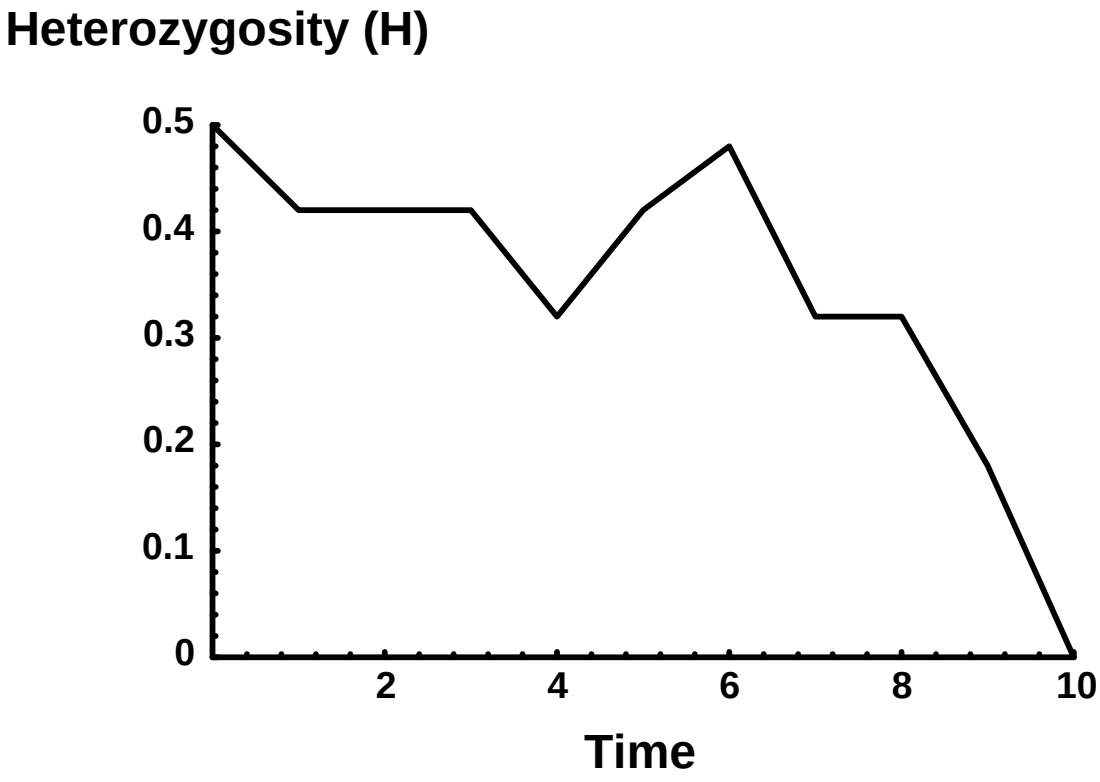
$$H[t] = 1 - f[t]$$
$$= 1 - 1/(2N) - (1-1/(2N))\, f[t-1]$$
$$= (1-1/(2N))\, H[t-1]$$

➡ The expected heterozygosity within a diploid population declines at a rate 1/(2N) each generation.



REMEMBER: These are only averages.
In any particular population, heterozygosity will rise and fall over time, eventually reaching zero.





Random Genetic Drift with Mutation

Even in small populations, heterozygosity does not disappear forever, since mutations continually arise.

➡ Two alleles drawn at random are NOT the same if either mutates to a new allele.

Therefore, for two alleles drawn at random to be the same, they must both be non-mutant:

$$f[t] = \left(1/(2N) + (1-1/(2N))\, f[t-1]\right)(1-\mu)^2$$

where μ is the mutation rate to new neutral alleles.

At equilibrium, the loss of genetic variability by drift and the gain by mutation counter-balance, so that on average:

$$f^* = 1/(4\,N\,\mu + 1)$$

$$H^* = 4\,N\,\mu/(4\,N\,\mu + 1)$$

[CHALLENGE: Derive these]

These are only averages. Any particular locus may or may not be fixed.

Not only do new mutations contribute to the amount of heterozygosity within a population, but occasionally these new mutations rise to fix within a population.

This creates a constant turn-over in the alleles carried by a population, *even in the absence of selection*.

In a diploid population, how many new mutations appear each generation?

For any one of these mutations, what is the probability that it will be the "lucky" allele from which the entire population will eventually descend?

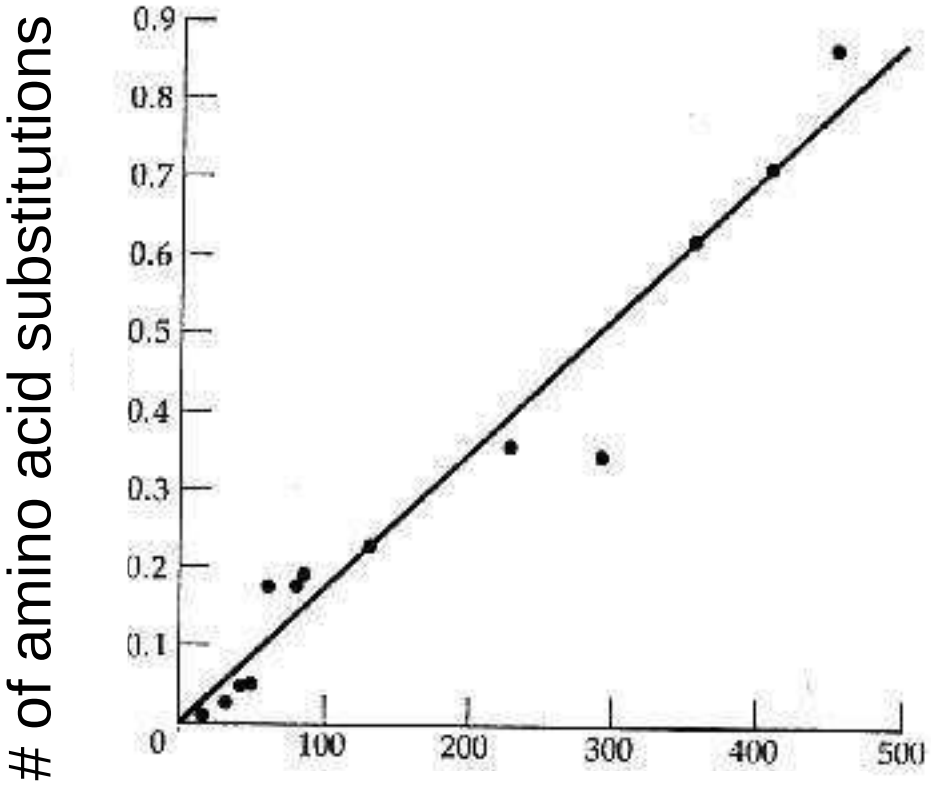
➡ The turn-over of neutral alleles will occur at a rate equal to $2N\mu/(2N) = \mu$!

This result does not depend on the population size.

➡ The number of substitutions can be used as a **MOLECULAR CLOCK**, indicating how much time has passed.

Motoo Kimura derived this result and used it to predict that the number of DNA substitutions that occur within the genome should rise as a linear function of time.

Comparing α-globin genes from various vertebrates, Kimura (1983) showed that this prediction matched the inferred numbers of amino acid substitutions:



Random Genetic Drift with Selection

Would we expect a similar rate of evolution if selection were also acting?

Haldane (1927) showed that, in a large population, the probability of a new mutant allele fixing is 2s, where s is its selective advantage.

Kimura noted that the rate of fixation of new adaptive mutants that arise at rate μ would therefore equal $2N\mu*(2s)$ in a diploid population.

This leads to a prediction: A gene sequence should evolve at a constant rate over a range of population sizes ONLY if very little selection is acting on that sequence.

When would you expect the fate of an allele to depend more on random genetic drift than on selection?

When would you expect the fate of an allele to depend more on selection than on random genetic drift?

➡ 4Ns determines the relative roles of selection (important when $4Ns >> 1$) and random genetic drift (important when $4Ns << 1$).

SOURCES:

- Hartl and Clark (1989) Principles of Population Genetics. 2nd Edition. Sinauer Associates, MA.
- For more information: Random Genetic Drift by Laurence Moran

Adaptation

Adaptation — A trait, or integrated set of traits, that increases the fitness of an organism.

More specifically --

"…a phenotypic variant that results in the highest fitness among a specified set of variants in a given environment." - Reeve & Sherman 1993

Generally speaking, adaptations are traits or characters that appear to be too well-fitted to their purpose to have arisen by chance. That is, they must be the result of selection.

Adaptations may involve morphological, physiological or behavioural traits. They arise through the accumulation of a series of small improvements over time.

"If it could be demonstrated that any complex organ existed which could not possibly have been formed by numerous successive slight modifications, my theory would absolutely break down." — Darwin.

Examples of Complex Adaptations: the eye; bird wings; the human brain; homeothermic temperature regulation; human language.

But simple traits can also be adaptations.

Example: A single change in allele for wing color in forest-dwelling moths (from black to white). This renders individuals cryptic in brightly lit environments and on pale barked trees. They survive better than black-winged moths in a novel environment -- the forest edges -- and move into this new habitat.

Temperature Regulation in a Bee Hive

The reproductive success of every bee (genus: *Apis*) in a hive hinges upon the survival of the brood, which are very sensitive to temperature during development. Larvae develop properly between the narrow temperature range of 32-36_ C. Despite this, bee colonies often survive in extremely adverse environments.

5500 m in Himalayas, at low temperatures -50_ C

60_ C heat, direct summer sunlight, over a lava field

<u>Mechanisms for Temperature Regulation</u>

(behavioural adaptations)

1. Hive Location — aerial in warm climates, within shelters in cold climates
2. Hive Orientation — entrance always faces south in northern climates
3. Heat Production — metabolic heat produced by flight muscles as bees rapidly vibrate wings inside the hive
4. Insulation — "clustering" — a living blanket surrounding the hive (begins at ambient temperatures of ~18_ C)

5. Cooling Regimen -

adults spread out in the hive

ventilation - bees open holes in the hive during the day

adults line up in rows, facing away from entrance, and fan air out of the hive with their wings

evaporative cooling — water spread about hive and over cells containing larvae; increased foraging for water

partial evacuation of hive if necessary

Adaptations are not always obvious, or easy to identify.

In order to identify a trait as an adaptation, we must first determine (or suggest) its use or function, and then **show** that the trait is advantageous. This can be difficult to do.

Example: The White Coat of Polar Bears

Hypothesis — white coat is an adaptation for camouflage

Test — observe hunting behaviour and assay use of camouflage

Result — camouflage not usually important in hunting

New Hypothesis — white coat is an adaptation for trapping solar heat

Test — hairs are actually clear and translucent, and trap 16% of incident light energy

Compare heat-trapping ability of multiple hair types. Is polar bear fur better than others? YES

Although at first it might have seemed obvious that a white coat is an adaptation for camouflage, it seems polar bears would often be just as successful at hunting if their coat were not white. Their coat does keep them warmer than other coat types, however.

We can suggest adaptive reasons for virtually any trait. The challenge is to show that the trait actually confers the advantage that we've suggested.

Studying Adaptation

First, develop alternative hypotheses about a trait's function.

Then, test these alternatives. There are several ways to do this:

1. Observational studies
2. Experiments
3. Theoretical models
4. Comparative method

Each of these methods has advantages and limitations.

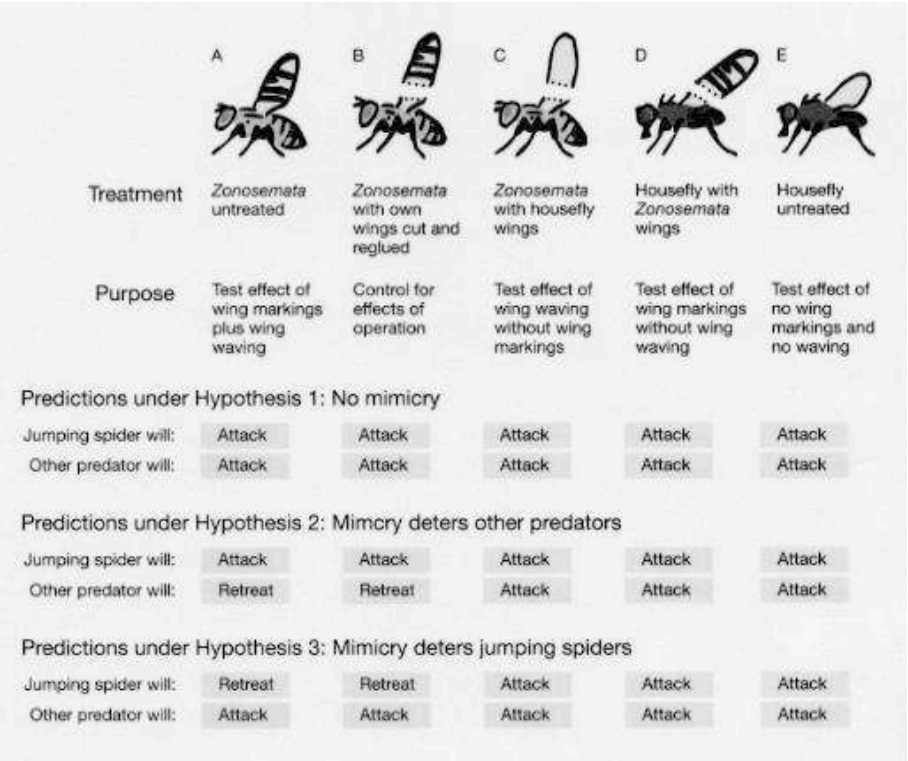
Experimental studies

Ex. Wing marks & wing waving in Tephritid flies.

Tephritid flies have dark bands on their wings and wave their wings when disturbed in a manner that is reminiscent of their major predator's territorial display -- e.g., jumping spiders' leg waving.

Do flies mimic their predators?

If so, is this mimicry to deter any predator, or is it specifically to deter jumping spiders?



	A	B	C	D	E
Treatment	*Zonosemata* untreated	*Zonosemata* with own wings cut and reglued	*Zonosemata* with housefly wings	Housefly with *Zonosemata* wings	Housefly untreated
Purpose	Test effect of wing markings plus wing waving	Control for effects of operation	Test effect of wing waving without wing markings	Test effect of wing markings without wing waving	Test effect of no wing markings and no waving
Predictions under Hypothesis 1: No mimicry					
Jumping spider will:	Attack	Attack	Attack	Attack	Attack
Other predator will:	Attack	Attack	Attack	Attack	Attack
Predictions under Hypothesis 2: Mimcry deters other predators					
Jumping spider will:	Attack	Attack	Attack	Attack	Attack
Other predator will:	Retreat	Retreat	Attack	Attack	Attack
Predictions under Hypothesis 3: Mimicry deters jumping spiders					
Jumping spider will:	Retreat	Retreat	Attack	Attack	Attack
Other predator will:	Attack	Attack	Attack	Attack	Attack

Theoretical models of adaptation

Theoretical models generally fall into two classes. These differ primarily in the form of selection they assume.

1. Optimality models -- the fitness of a phenotype is independent of the phenotype of others in the population (frequency independent).

   Exs. Temperature regulation, egg size

2. Evolutionarily Stable Strategies (ESS) models -- the fitness of a phenotype is affected by the phenotype of others in the population (frequency dependent).

   Exs. Animal fighting, parental care

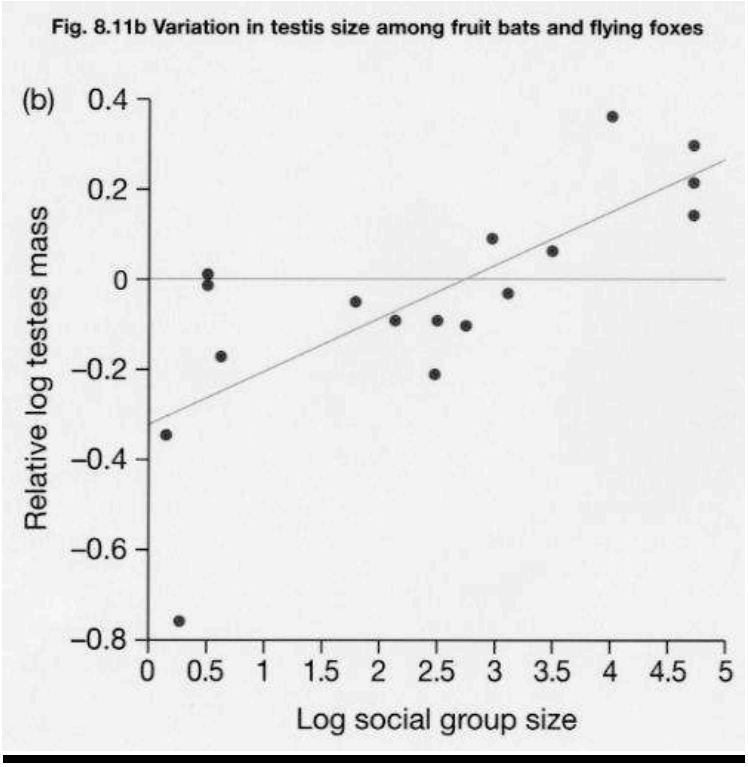
The Comparative Method

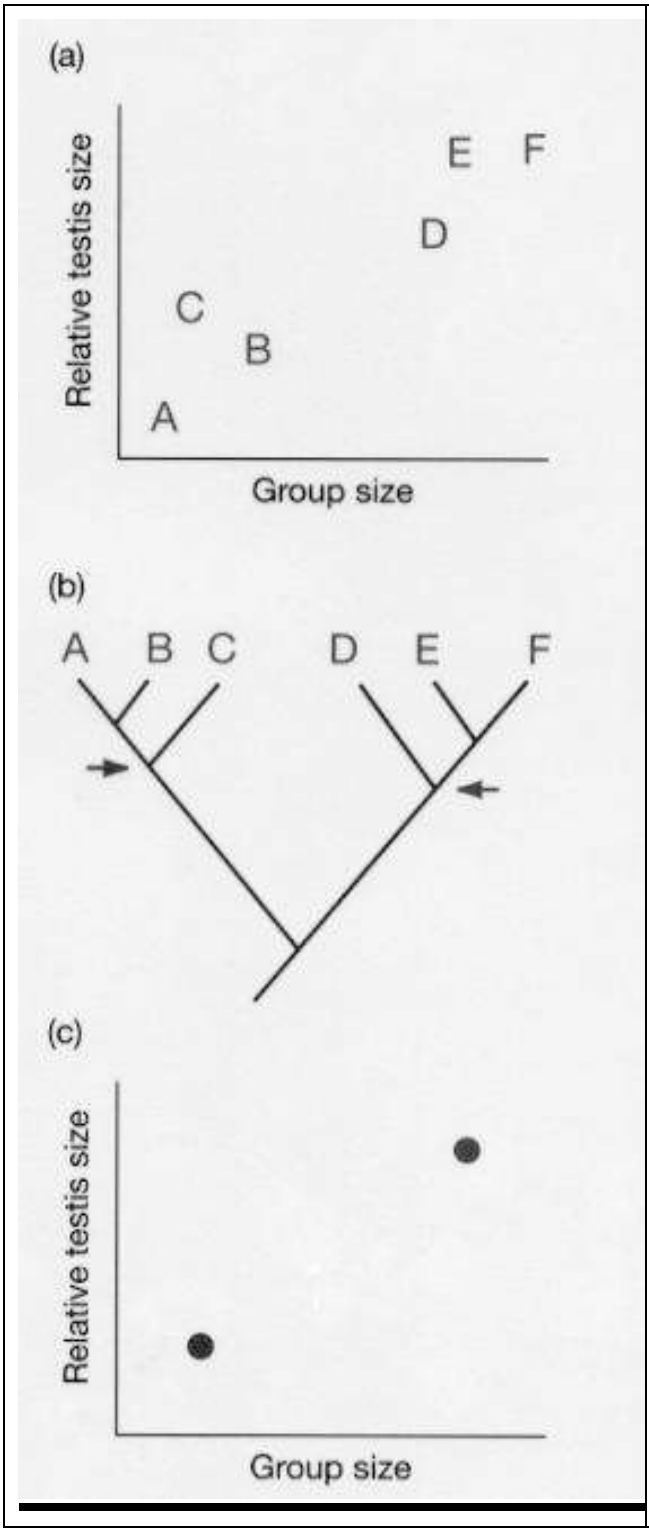
Ex. Testis size in bat species.

Male bats vary in the size of testes. One hypothesis for this variation is that large testes produce more sperm, an advantage in sperm competition.

Sperm competition might be more intense in larger social groups, where more males compete for reproductive access to females.

Hypothesis: Testis size is larger in species with larger social groups.





But not all data points are independent.

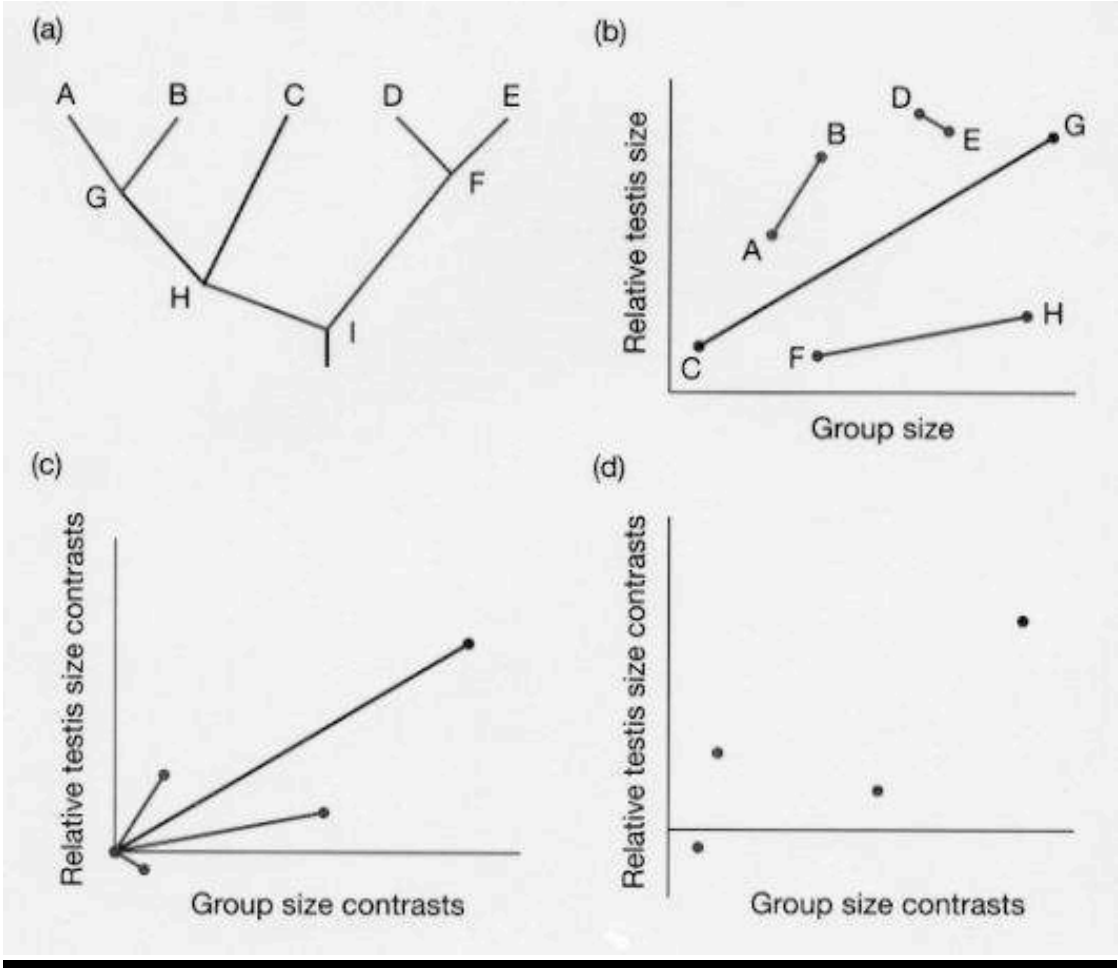
Closely related species might have similar group size and testis size NOT because of sperm competition, but because they share an ancestor who had large testes and lived in large groups.

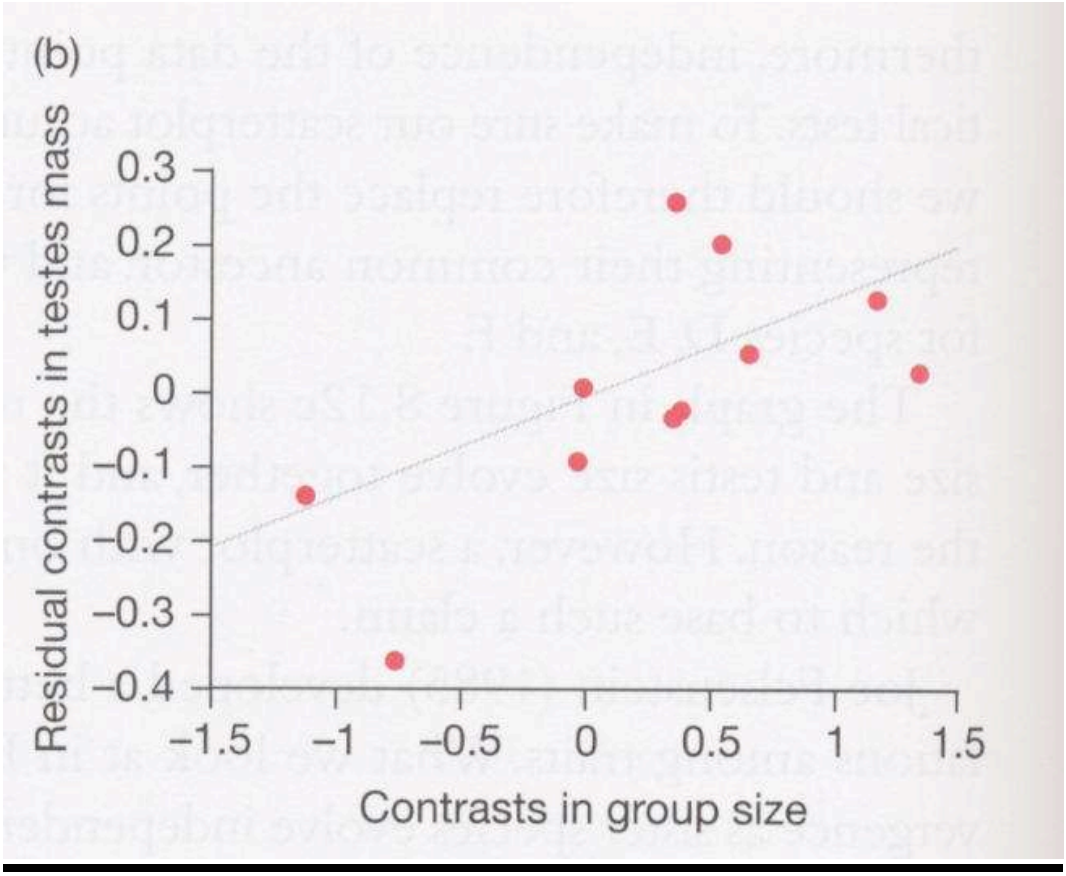
How do we control for the effects of shared history?

Independent contrasts

Compare sister taxa: when they diverge from a common ancestor, do species that evolve larger group size also evolve larger testis size?

1. Need phylogeny
2. Calculate contrasts between sister taxa
3. Evaluate relationship with phylogenetically-corrected values.





Yes, males in species with larger group sizes also have larger testes.

Adaptations are not perfect -- Constraints to adaptive evolution

Species are constantly evolving and refining their adaptations, as natural selection favours individuals best suited to their environment in each generation. A number of **selective constraints** prevent adaptations from every being perfect.

Ex. Vertebrate mouths are used for both feeding and breathing. Vertebrates can't breathe and swallow at the same time, or they choke.

Time Constraints

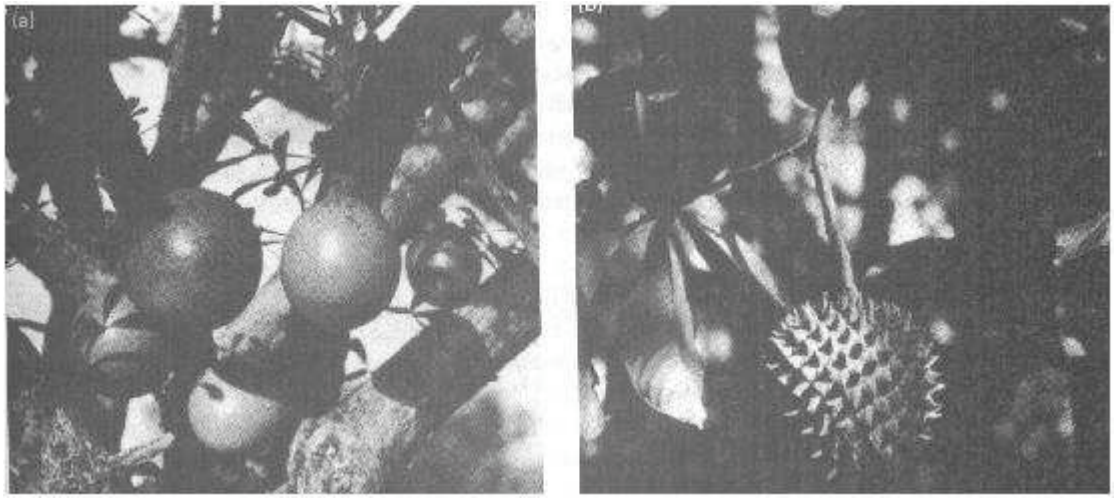
Evolution takes time (sometimes a lot of it!). Adaptations can be considered "works in progress". Often there simply hasn't been enough time for evolution to fine-tune adaptations.

Example: Neotropical Anachronisms

Many flowering plants produce fruit to entice animals to disperse their seeds. Fruits need to be attractive

to the animals, protect the seeds, and remain in the gut long enough to allow for dispersal of the seeds.

Some trees in the tropical forests of Costa Rica produce very large fruits in great numbers, which are largely inedible to local fauna.



Janzen and Martin suggest these fruits were adaptations to now extinct large herbivores that lived in the area up until about 10,000 years ago. Large fruits would have been attractive to large herbivores, and their hard exteriors would have helped to protect the seeds from being crushed by the herbivores' teeth.

10,000 years hasn't allowed enough time for the trees to evolve fruits better suited to the smaller mammals remaining today.

Genetic Constraints

We have seen that when a locus is under selection involving heterozygote advantage, the population evolves to an equilibrium where all three genotypes (AA, Aa, aa) are present. This represents a genetic constraint on adaptive evolution, since homozygotes will always be imperfectly adapted, yet remain in the population.

Example: Sickle Cell Anemia

Homozygotes can have very poor fitness, but remain at appreciable frequencies in populations where malaria acts as a selective agent.

If a new mutation were to arise which allowed malaria-resistance without sickling of the cells, it would likely spread through the population.

Developmental Constraints

Developmental Constraint — "a bias on the production of variant phenotypes (i.e. a limitation on phenotypic variability) caused by the structure, character, composition or dynamics of the developmental system." — Maynard Smith et. al.

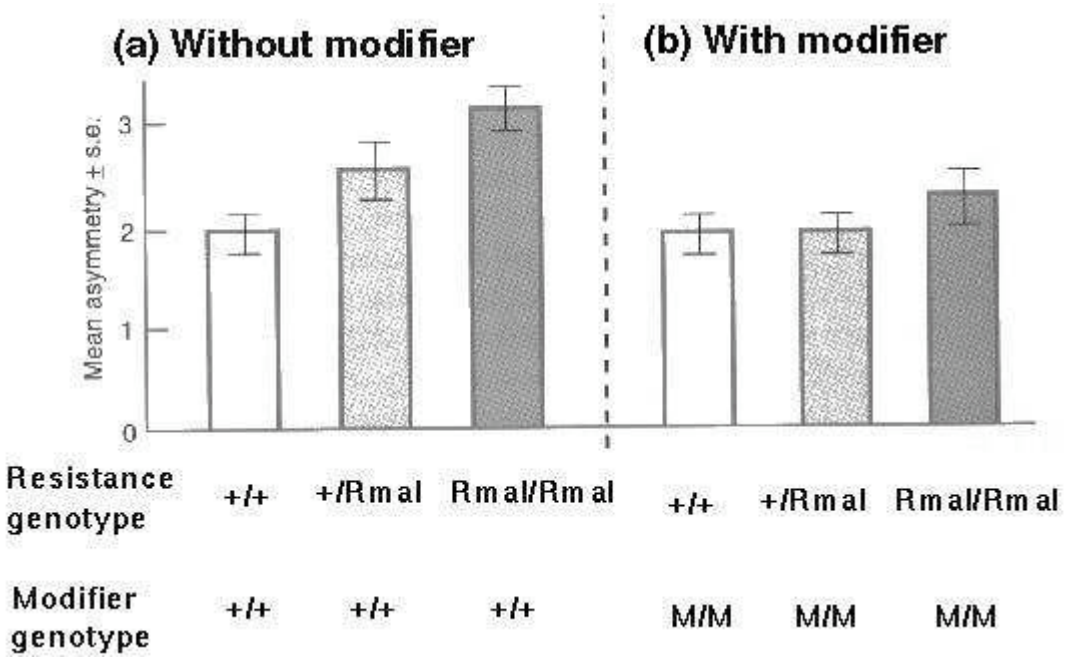
The developmental system influences the types of mutations that can occur. Pleiotropy (genes having multiple effects) may require that adaptation of one trait occurs along with changes in another trait.

NOTE: The overall effect a mutation has on fitness is what selection operates on.

Example: Australian Sheep Blowfly (*Lucilia cuprina*)

Farmers sprayed insecticide, and blowflies soon evolved resistance to it. The mutation conferring insecticide resistance also disrupted the developmental system, producing asymmetry which is maladaptive. Strong selection for insecticide resistance lead to increasing developmental

asymmetry.



Over time, mutations at other genes ("modifier loci") evolved to restore symmetry, while maintaining insecticide resistance. The developmental system adjusted to the necessity of carrying the mutation conferring insecticide resistance.

Note that in cases where selection for one trait is not this strong, adaptation of one trait may be difficult if it requires a maladaptation of another trait. (More on trade-offs in a moment)

Historical Constraints

"You can only work with what you've got."

For historical reasons, organisms may lack genetic variation that would be adaptive.

In the case of heterozygote disadvantage, populations will fix for one allele (**A** or **a**). Which allele fixes depends on :

1. The relative fitnesses of the genotypes.
2. The initial allele frequencies.

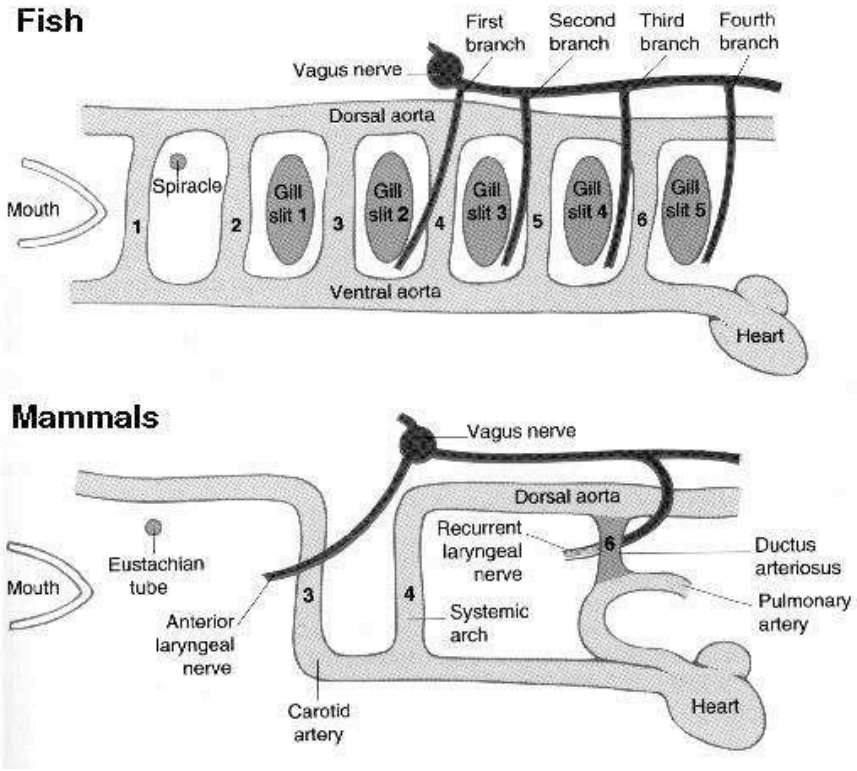
The initial allele frequencies represent a historical constraint. Even if fixation for **A** would lead to higher fitness, **a** may rise to fixation if the initial allele frequency of **A** is below the unstable equilibrium point.

Example: The Recurrent Laryngeal Nerve

4th vagus nerve first evolved in fish-like ancestors

successive branches pass behind arterial arches between the gills of fish

in modern mammals, the nerve passes from the brain, down the neck, around the dorsal aorta and then back up the neck to the larynx



this detour seems silly in humans, and ridiculous in the giraffe

mutations generating a more direct route might be beneficial, but are unlikely to evolve since they would involve extensive changes in embryological development

Trade-offs

Traits often serve multiple functions, and genes often have multiple effects. Often the value of one trait in not independent of the value of another trait.

Trade-offs arise when a limited resource (like energy or time) is divided among multiple functions.

> Ex 1. Allocating energy to growth or reproduction.

> Ex 2. Allocating time to foraging or avoiding predators.

Trade-offs can also arise when becoming better at one thing precludes being good at something else.

> Ex. Plankton feeders cannot simultaneously be good at feeding on large aquatic insects.

Traits evolve in response to selection on all of their functions, achieving the best possible overall fitness.

Example: Threespine Sticklebacks (*Gasterosteus aculeatus*)

Male sticklebacks establish territories in the spring, defend them from intrusions by other males, and court females. Male sticklebacks develop either black or red bellies during the mating season.

Males with red bellies are more successful in securing and defending territories.

> Females prefer to mate with males with red bellies.

Breeding experiments show that the colour of male bellies is controlled by a single locus with two alleles (black and red). Why then does the black allele remain in the population?

> Sticklebacks sometimes share habitats with mudminnows, which secure territories earlier in the season.
>
> Mudminnows have black bellies.

Male sticklebacks with black bellies are more successful in acquiring territories from mudminnows. Female sticklebacks will only mate with males that have nest territories.

A trade-off exists between success at intraspecific competition and interspecific competition.

Summary

Adaptations (especially complex ones) evolve through a series of small changes over time.

Studying adaptations requires posing and testing alternative hypotheses about trait function and its effects on fitness.

Adaptations need not be perfect. The are continually evolving through natural selection to better fit the current environment.

Constraints limit the types of adaptations that evolve, and include time constraints, genetic constraints, developmental constraints and historical constraints.

Trade-offs between positive and negative implications of modifying a trait can constrain adaptive evolution.

Molecular Evolution

In 1952, Frederick Sanger and coworkers determined the complete amino acid sequence of insulin.

Since that time, the amount of sequence information has grown exponentially.

For example, Genbank contains all publicly available DNA sequences, which amounts to more than 3.8 billion basepairs from 4.8 million sequences!

In addition, the entire genomes of over thirty organisms have been sequenced, including two eukaryotes (the fungus, *Saccharomyces cerevisiae*, and the nematode, *Caenorhabditis elegans*).

The human genome is also well on its way to being sequenced, with an expected date of completion in the year 2003.

Molecular evolution is a new field born from this explosion of molecular information and from our desire to understand how and why molecular sequences have evolved to be the way they are.

Topics in Molecular Evolution

In the next two lectures, we will discuss a few examples of research in the field of molecular evolution:

- Detecting selection by examining

  (1) Substitution rates

  (2) Variability within a population

  (3) Replacement versus silent changes

- Detecting historical events by examining

  (1) Number of differences between sequences

- Detecting relatedness by examining

  (1) Similarity among sequences within a population

  (2) Similarity among sequences from different species

Evidence in the DNA: Selection

1. Substitution Rates

An early theme developed within this course is that, while it is easy to establish that evolutionary change has happened, it is difficult to establish whether selection has played a role in this change.

DNA can provide a record of selection.

If we compare DNA sequences from different organisms, we can estimate the rate at which mutations appear and fix, causing basepair substitutions.

Substitution rate = the rate at which mutant alleles rise to fix within a lineage.

(For neutral mutations, the substitution rate within a population equals the mutation rate, since $2N\mu$ mutations appear, each with a 1/(2N) chance of fixation.)

Variation in the rate of substitutions among regions of the genome is due, in part, to variation in the form of selection.

Recall that:

- A new beneficial mutation has a chance of fixing within a diploid population of ~ 2s.
- A new neutral mutation has a chance of fixing within a diploid population of ~ 1/(2N).
- A new deleterious mutation has almost no chance of fixing within a large population.

Therefore, the nucleotide substitution rate is expected to be:

- highest when mutations are beneficial,
- intermediate (μ) when mutations are neutral,
- lowest when mutations are deleterious.

➡ Silent (or synonymous) mutations are more likely to be neutral.

➡ Replacement (or non-synonymous) mutations are more likely to experience selection, but the form and strength of selection depends on the gene and its function.

Gene	Non-synonymous Substitution Rate ($\times 10^{9}$)	Synonymous Substitution Rate ($\times 10^{9}$)
Histone 3	0.00	3.94
Histone 4	0.00	4.52
Actin α	0.01	2.92
Myosin β	0.10	2.15
Insulin	0.20	3.03
Growth hormone	1.34	3.79
Immunoglobulin k	2.03	5.56

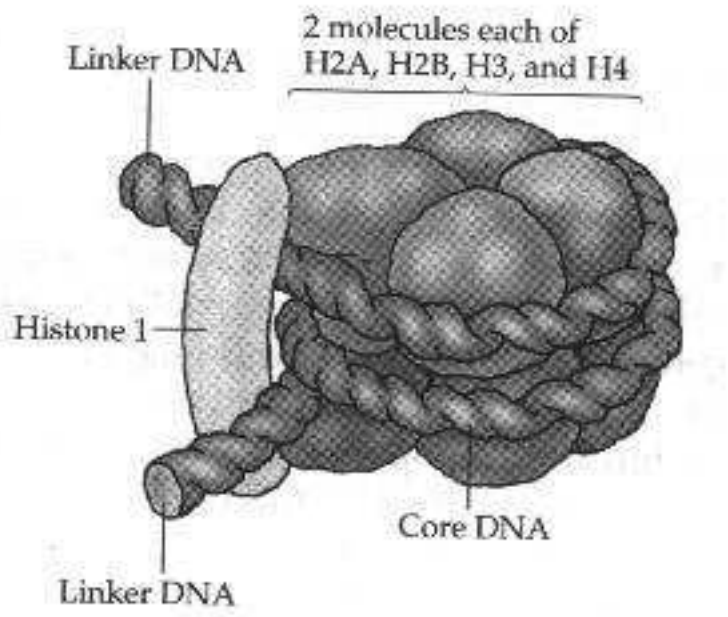
(From Li, 1997. Based on sequence differences between humans and rodents, estimated to have diverged 80MYA.)

Histones, for example, appear to have a very low rate of replacement substitutions.

This suggests that mutations causing basepair changes in histone genes may be deleterious.

Why?

Histones are DNA-binding proteins around which DNA is coiled to form chromatin. Many positions within the protein interact with the DNA or with other histones. In addition, histones are highly compact and alkaline.



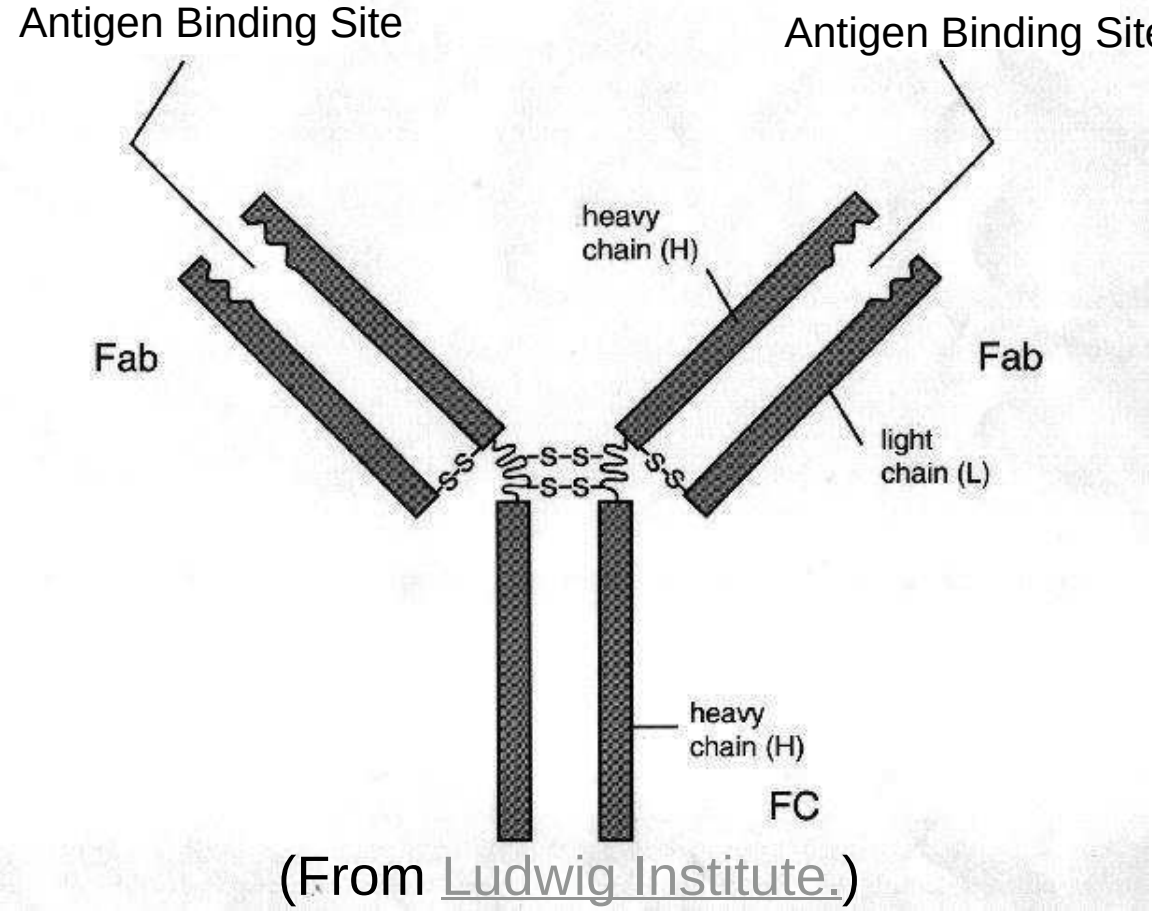
(From Li 1997)

Most amino acid changes in histone proteins may have negative or even disastrous consequences.

➡ Histone proteins have strong *functional constaints.*

Conversely, the amino acid sequences of immunoglobulins (= antibody protein) evolve at a much higher rate.

In particular, the active sites (the complementarity-determining regions) of many immunoglobulins actually have higher rates of replacement changes than silent ones!



(From Ludwig Institute.)

It is thought that selection favors mutations in these regions, thereby increasing the diversity among antibodies produced by the body and improving the immune response.

2. Levels of variability within a population

Another method to detect selective events is to examine the level of variability currently present within a population.

> If a beneficial mutation appears and sweeps through a population, what will happen to the level of polymorphism present at neighboring DNA sites?

For example, Berry et al (1991) sequenced 1.1 kilobases of the cubitus interruptus Dominant (ciD) locus in *Drosophila*.

ciD is located on a tiny fourth chromosome, which undergoes no recombination.

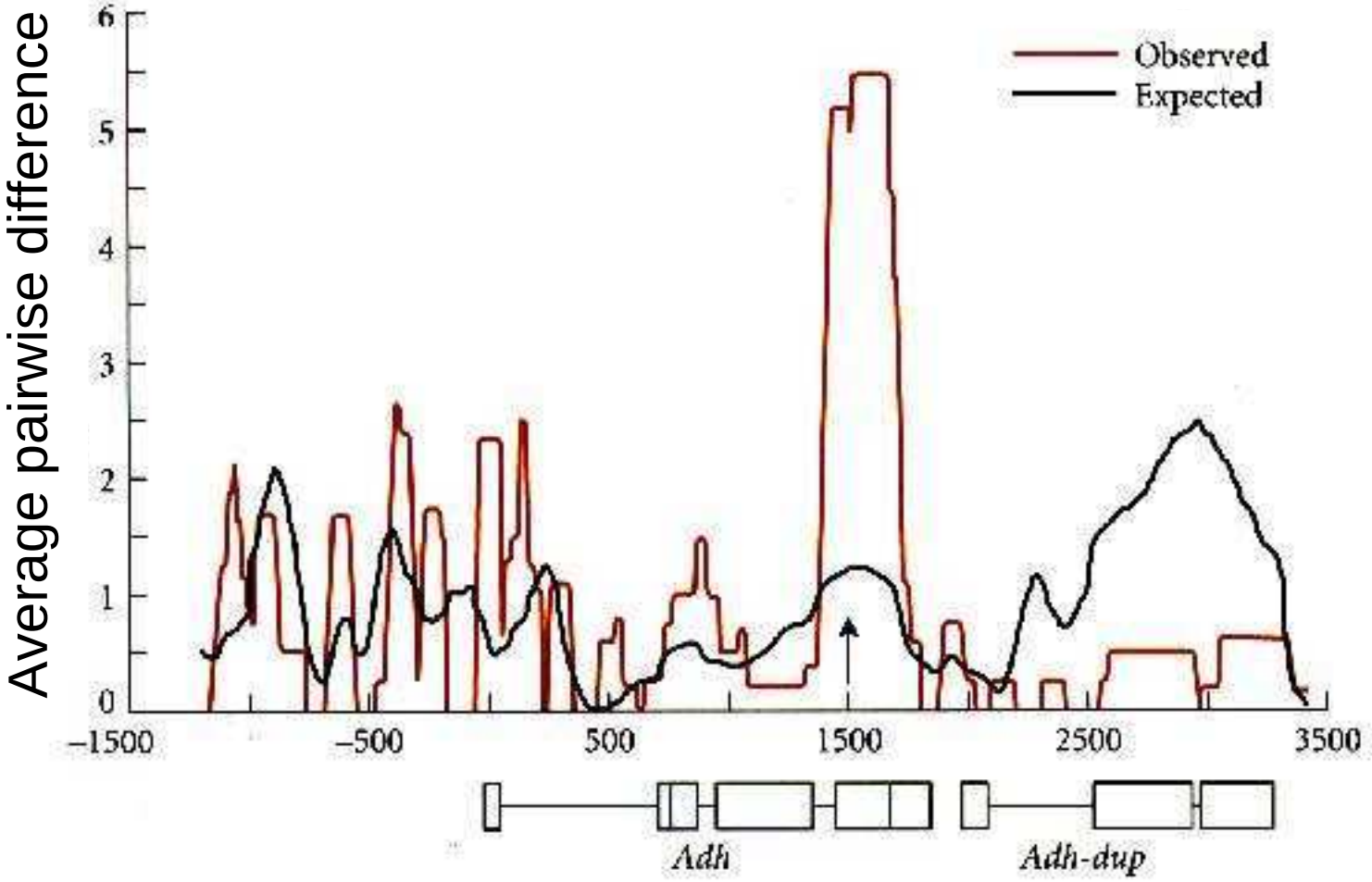
They found NO VARIATION among ten *D. melanogaster* sequences and only one basepair difference among nine *D. simulans* sequences, even though there were 54 differences between the species.

By contrast, other genes from the same individuals showed normal levels of polymorphism.

Berry et al (1991) argued that recent selective sweeps in both species may have eliminated most of the polymorphism on the fourth chromosome.

If there is overdominance at a site, what will happen to the level of polymorphism present at neighboring DNA sites?

Kreitman and Hudson (1991) sequenced a 4750 basepair region near the alcohol dehydrogenase (ADH) gene from 11 individuals of *D. melanogaster* and found higher than expected levels of polymorphism:



(From Futuyma 1998)

There is only one amino acid polymorphism (Adh^F/Adh^S) within this region, which occurs at site 1490 (see arrow).

Kreitman and Hudson (1991) hypothesize that there is selection maintaining a polymophism at or near this site.

ADH is an enzyme that breaks down ethanol.

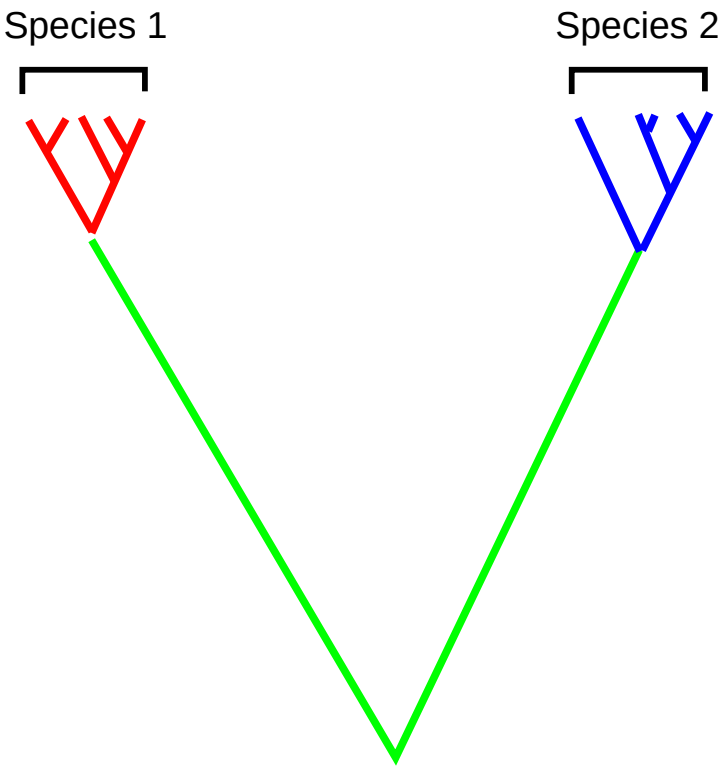
Flies carrying the Adh^F allele survive better when their food is spiked with ethanol than do flies carrying the Adh^S allele (Cavener and Clegg 1981).

Nevertheless, the factors maintaining the Adh^F/Adh^S polymorphism remain unknown.

3. Replacement versus silent changes

McDonald and Kreitman (1991) compared substitutions between species and polymorphisms within a species to construct a test to detect selection.

Imagine that five sequences are obtained from each of two species and that the tree relating these sequences is:



Any mutation that happens on a red branch will appear as a polymorphism within species 1.

Any mutation that happens on a blue branch will appear as a polymorphism within species 2.

Any mutation that happens on a green branch will appear as a fixed difference between species 1 and 2.

If mutations occur randomly over time and if the chance that a mutation does or does not cause an amino acid change remains constant, then the ratio of replacement to silent changes should be the same along any of these branches.

If mutations are neutral, any of these mutations has an equal chance of being preserved.

McDonald-Kreitman test

H_0: The ratio of replacement to silent changes among polymorphic sites (within a species) should equal the ratio among fixed differences (between species) in the absence of selection.

If new mutations are advantageous, they will fix rapidly and cause more fixed differences between species.

If new mutations are deleterious, they will rarely fix, but they will temporarily create polymorphisms.

These effects of selection should be stronger on mutations that change the amino acid sequence (replacement) than ones that don't (silent).

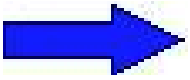
→ An excess of amino acid differences between species should be seen when replacement mutations have been beneficial and fixed by selection.

→ A lack of amino acid differences between species should be seen when replacement mutations have been deleterious and eliminated by selection.

	Fixed differences	Polymorphic sites
ADH gene		
Replacement	**7**	**2**
Silent	**17**	**42**
G6PD gene		
Replacement	**21**	**2**
Silent	**26**	**36**

(From Li, 1997. ADH study by McDonald and Kreitman, 1991: 12 *D. melanogaster,* 6 *D. simulans,* and 24 *D. yakuba* sequences. G6PD study by Eanes et al, 1993: 32 *D. melanogaster* and 12 *D. simulans*.)

For both genes, the ratio of replacement to silent substitutions is significantly lower among polymorphic sites within species (2 : 42 for ADH and 2 : 36 for G6PD) than among fixed differences between species (7 : 17 for ADH and 21 : 26 for G6PD).



The null hypothesis that selection is absent is rejected in both cases.

The excess of replacement differences between species suggests that mutations have been positively favored.

Indeed, G6PD is an important enzyme in the metabolism of pentose sugars, and it has been argued that amino acid changes may have been selectively favored in changing environments.

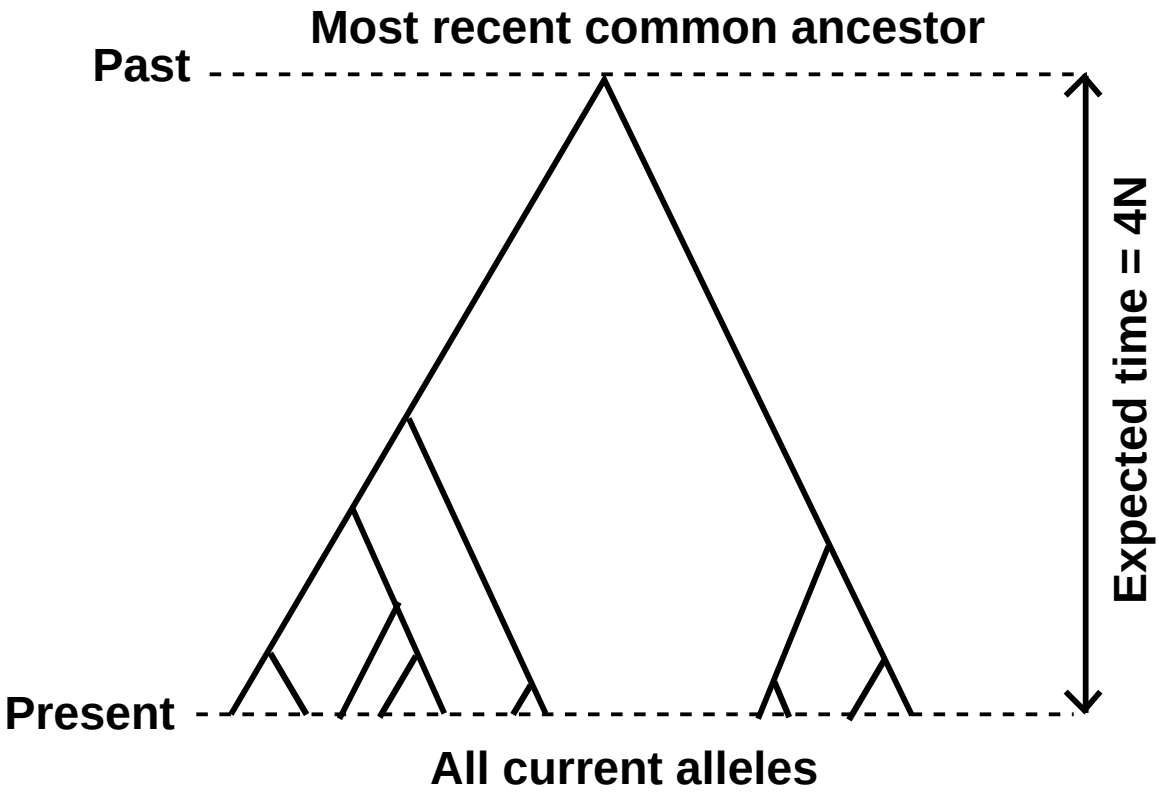
Evidence in the DNA: History

DNA sequences record other historical events besides selection.

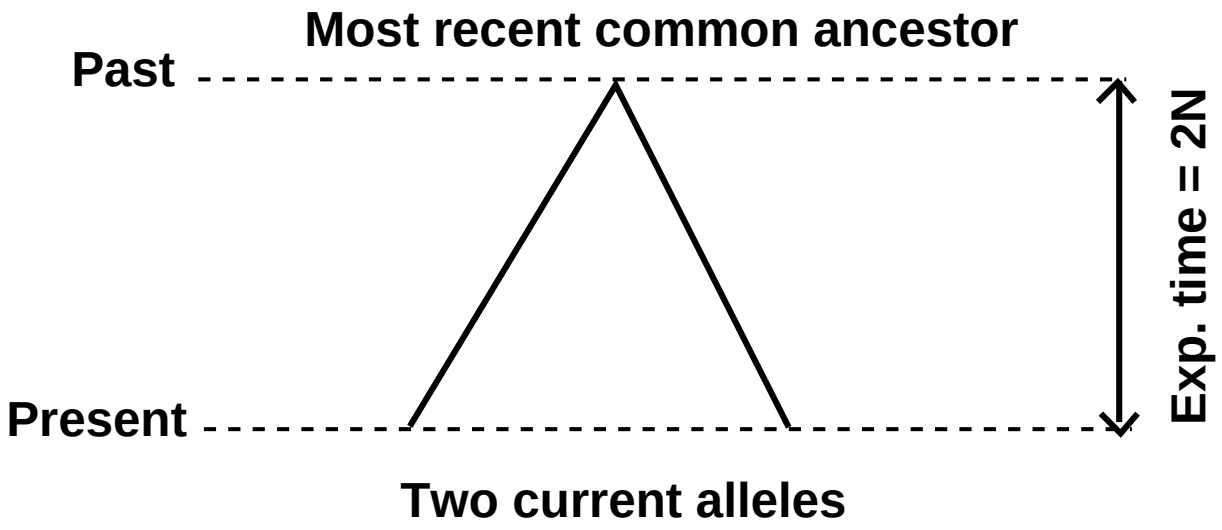
Sequence comparisons have been used to trace past migration events and also past changes in population size.

For example, Takahata et al (1995) estimated the population size of early humans, using coalescence theory.

Recall that all the alleles currently present within a population are descended from a common ancestor that lived, on average, 4N generations ago.



Any two alleles chosen at random from the current population will share a common ancestor 2N generations ago, on average:



These predictions also hold for two sequences of DNA.

Therefore, if mutations occur at a rate μ per generation per basepair on each of the two branches leading to the two current sequences, the proportion of sites that differ between the two sequences is expected to be $4N\mu$.

Takahata et al (1995) used a more sophisticated version of this idea to estimate N from human gene sequences.

They studied 49 different loci from human populations.

The total number of differences between two randomly chosen sequences varied from zero at 37 loci to five at one locus.

Using an estimated mutation rate of $2\ 10^{-8}$ per basepair, Takahata et al estimated an effective population size for humans of 10,000!

[NOTE: There may have been more individuals alive. 10,000 represents the "effective" population size -- the size of an ideal population of constant size that would have led to the observed amount of sequence divergence.]

Even though the current population size of humans is nearly 6 billion, the molecular sequence divergence among humans reflects a much smaller historical population size (~ 400,000 to 50,000 years ago).

➡ Humans are genetically very similar, due in part to a recent population explosion from a relatively small number of individuals within the last few hundreds of thousands of years.

Evidence in the DNA: Relatedness

Beyond looking for clues to our past, molecular data can be used to tell us about the diversity within and among species currently alive.

1. Human genetic diversity

> "Accustomed as we are to noticing variations in skin color or facial structure, we tend to assume that the differences between Europeans, Africans, Asians, and so on must by large...This simply is not so: the remainder of our genetic makeup hardly differs at all."
>
> -- Cavalli-Sforza and Cavalli-Sforza (1995) p. 124

In a major study of human polymorphisms, Cavalli-Sforza and collaborators studied the allele frequencies of different populations at 110 genes.

In all cases, the differences between "races" were quantitative *not* qualitative.

That is, there was not a single gene for which two races were totally different.

Instead, slight differences in allele frequencies were observed at most loci, e.g.:

Allele Frequency	European	African	Asian
GC-1	**72%**	**88%**	**76%**
HP-1	**38%**	**57%**	**23%**

In a similar study, Nei and Roychoudhury (1982) found that 85% of the genetic variation in the human species exists *within* populations and that only 8% is among the major "races".

> "If everyone on earth became extinct except for the Kikiyu of East Africa, about 85% of all human variation would still be present in the reconstituted species."
>
> -- Lewontin et al. (1984)

Similarly, Brown (1980) studied the mitochondrial DNA from 21 humans of diverse origin.

868 nucleotide sites were examined and only a few differences were observed between any pair of individuals.

Overall, the sequences differed from a postulated ancestral mtDNA sequence at only 0.18% of the sites.

Using a substitution rate estimate of 10^{-8} per basepair per year, Brown concluded that humans passed through a severe population bottleneck ~180,000 years ago.

2. Mammalian genetic diversity

How much do we differ genetically from other mammals?

Species comparison	% sequence difference
Human-Chimp	**1.45**
Human-Gorilla	**1.51**
Human-Orangutan	**2.98**
Human-Rhesus Monkey	**7.51**

(From Li 1997. Based on 5.3 kb of non-coding DNA.)

Surprisingly little!

Tables, such as the above, provide information about the relatedness of different species.

This data can be used to reconstruct the phylogenetic relationships among the species involved.

In the next few lectures, we'll discuss how phylogenetic reconstruction is accomplished and look at some interesting phylogenies.

SOURCES:

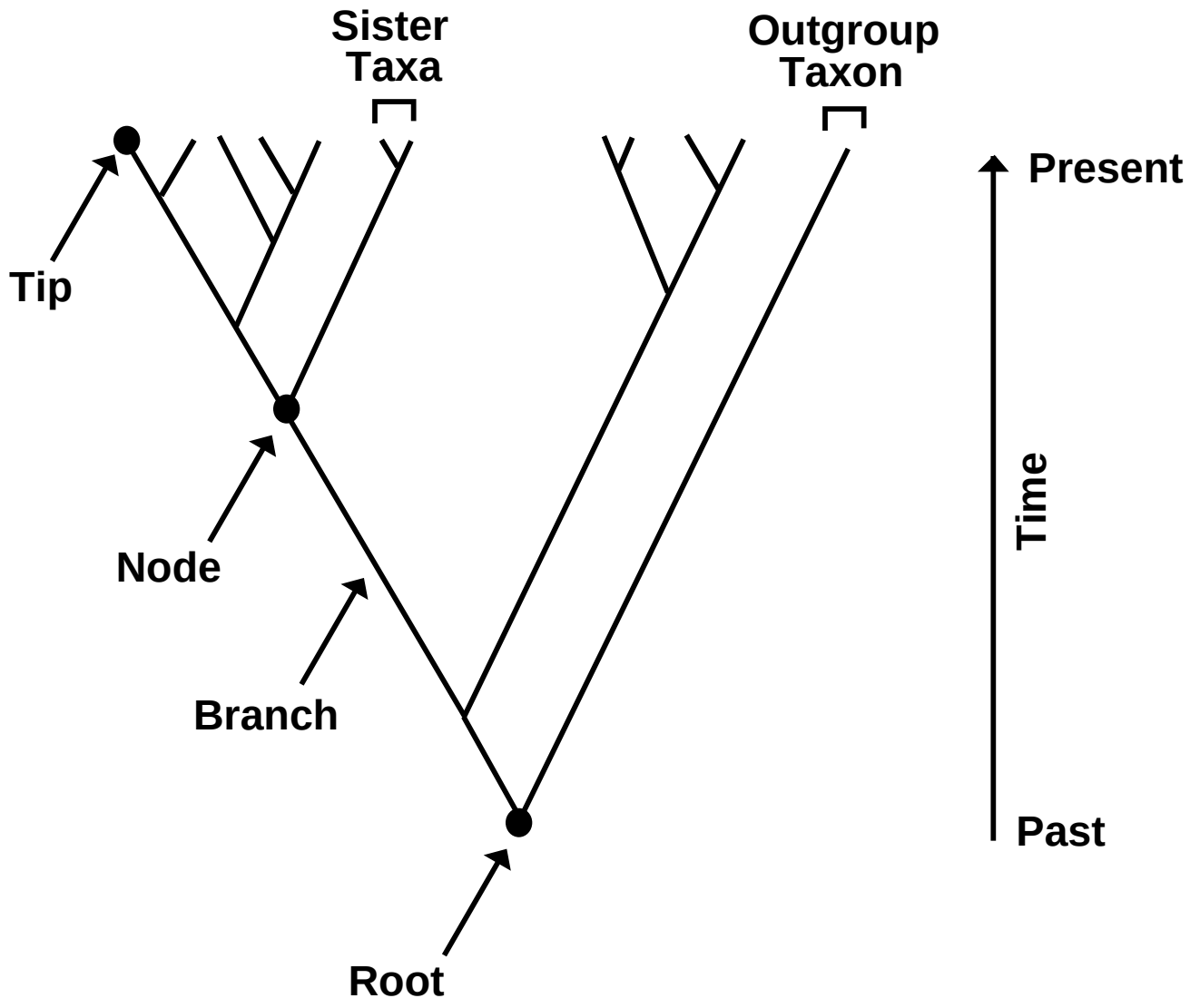
- **Li (1997) Molecular Evolution. Sinauer Associates, MA.**
- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Cavalli-Sforza and Cavalli-Sforza (1995) The Great Human Diasporas. Addison-Wesley, NY.

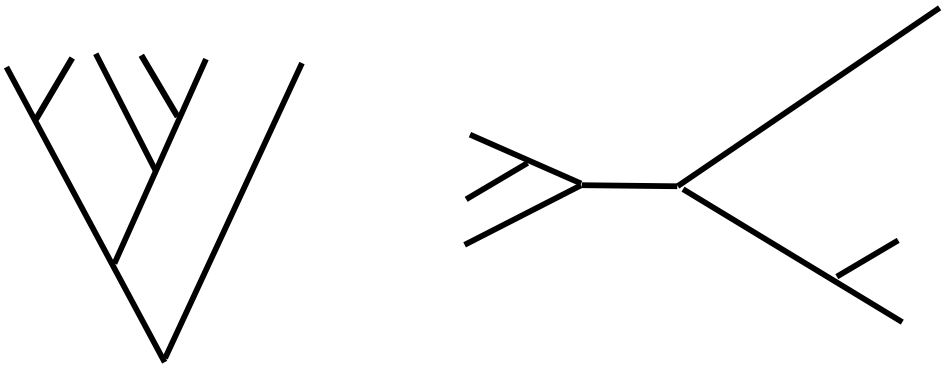
Phylogenetic Reconstruction

Phylogeny: The history of descent from a group of taxa such as species from their common ancestors, including the order of branching and sometimes absolute ages of divergence; also applied to the genealogy of genes derived from a common ancestral gene.

-- Futuyma 1998

PHYLOGENETIC TREE:





Rooted Tree **Unrooted Tree**

"Rooted" trees make a statement about the passage of time.

Nodes near the bottom of a rooted tree represent older divergences between two lineages.

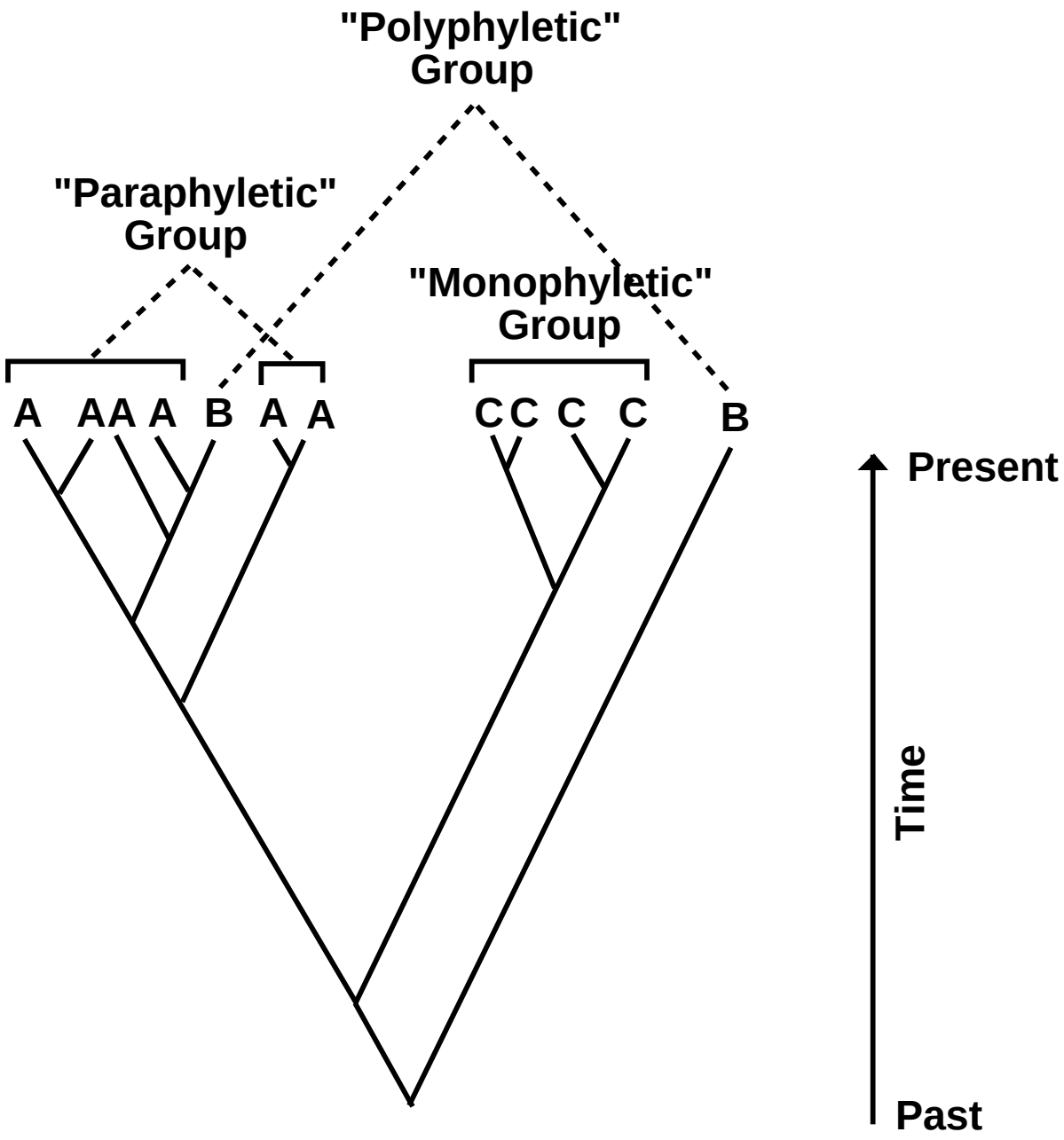
Nodes near the top of the tree represent recent divergences between two closely related lineages.

The root of a tree is often determined by an "outgroup".

An outgroup is *presumed* to be outside of the group of interest (i.e., it diverged prior to the taxa in a phylogenetic analysis).

An unrooted tree makes no claim about which of the divergences is oldest.

Phylogenetic trees sometimes do and sometimes do not correspond to the Linnean classification system.

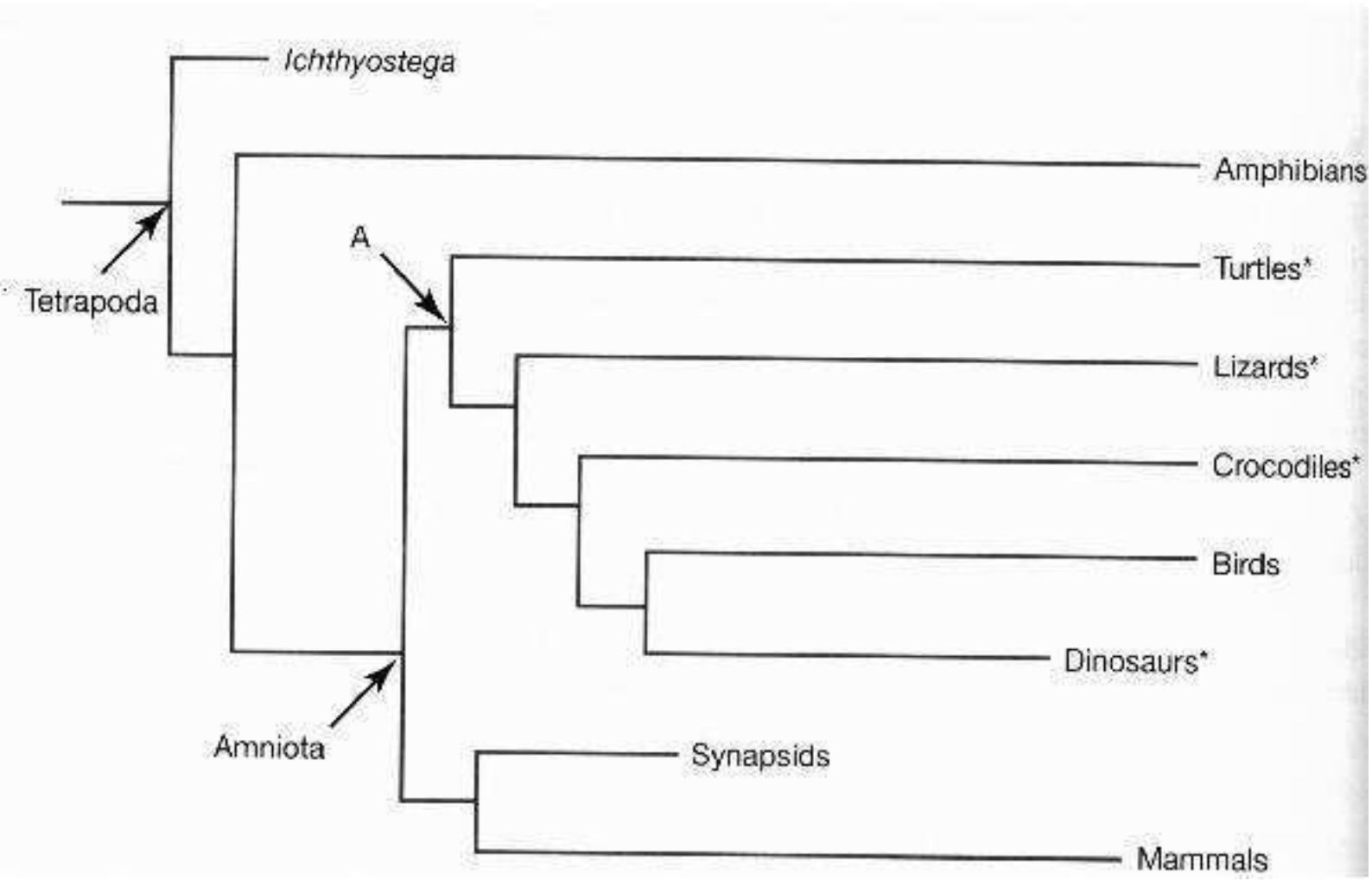


For instance, mammals make a good phylogenetic group (or clade), because all mammals are more closely related to each other than they are to any other taxon.

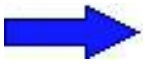
➡ Monophyletic group

Conversely, reptiles do not represent all the descendants of their common ancestor. Birds and mammals are also descended from the common ancestor of all reptiles (living and extinct). This means that some reptiles (e.g. crocodiles) may be more closely related to a non-reptile (e.g. birds) than they are to other reptiles.

➡ Paraphyletic group



Finally, some systematic groupings are completely artificial and based only on superficial resemblance and convergent evolution rather than true relatedness. For example, Linnaeus grouped together several unrelated worms into the artificial group "Vermes".

 Polyphyletic group

Should the classification scheme we use be based purely on monophyletic groups?

Choice of Characters

Phylogenetic trees may be based on many different forms of data: morphological, physiological, biochemical, molecular.

For any type of character, there are four attributes that are key to a successful phylogenetic analysis:

- **Numbers:** There should be a large number of characters.
- **Independence:** The characters should evolve independently of one another.
- **Homologous:** The characters must be derived from the same character in a common ancestor.
- **Low risk of convergence:** The characters should reflect common descent not "homoplasy".

Homoplasy: Similarity in the characters found in different species that is due to convergent evolution, parallelism, or reversal -- not common descent.

-- Freeman and Herron 1998

We are going to focus on reconstructing phylogenies from molecular data, specifically from DNA sequences.

Attributes of molecular data:

- **Numbers:** Large numbers of characters can be generated.
- **Independence:** Basepairs *largely* evolve independently of one another.
- **Homologous:** Sequences can be aligned using many different taxa to attempt to place basepairs in homologous positions.
- **Low risk of convergence:** No!!

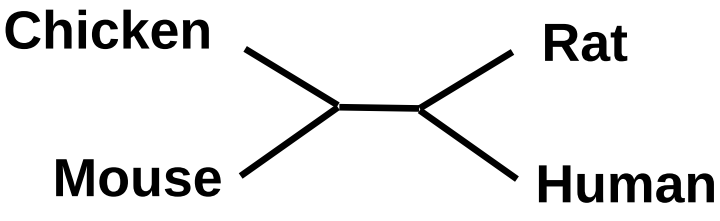
The main problem with using molecular data is that there is a high risk of homoplasy.

That is, if two sequences both have an adenine at a particular site, we do not know if this is because both descended from a common ancestor that had an adenine or because adenine happened to arise independently in both lineages.

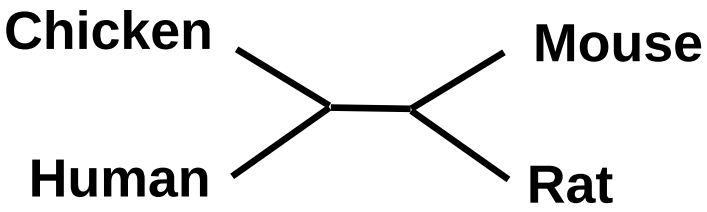
EXAMPLE OF HOMOPLASY

Sites 1220-1225 in cytochrome oxidase 1 (mtDNA)

Chicken: AATAGT
Mouse: AATAGT
Rat: TATGGT
Human: TATGGT



Tree based on these sites



True tree

➡ The risk of homoplasy is greatest if the DNA sequence evolves rapidly relative to the species divergences being examined.

Choice of Trees

There are several different *criteria* and *algorithms* used to reconstruct phylogenetic trees.

We'll focus on the conceptual criteria used in three different methods:

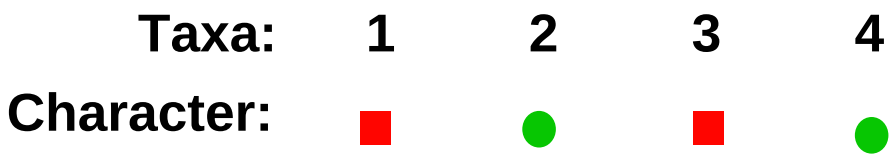
- Parsimony analysis
- Distance analysis
- Maximum Likelihood analysis

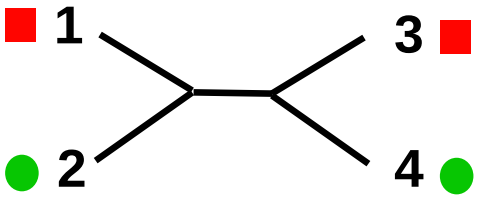
1. Parsimony analysis

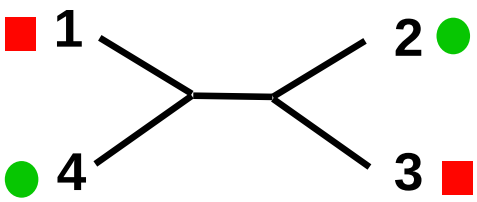
> "The principle of this method is to infer the amino acid or nucleotide sequences of the ancestral species and choose a tree that requires the *minimum number of mutational changes*".
>
> -- Nei (1987)

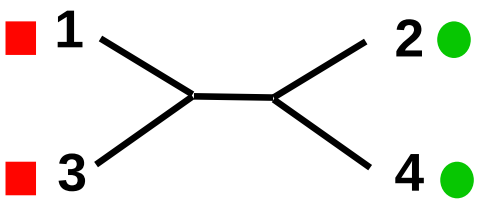
Parsimony's guiding principle is Occam's razor, the philosophical principle that it is preferable to choose the simplest of alternative explanations.

In practise, this means determining the tree (or trees) that require the *fewest number of mutations* in order to explain the data that you have.









With multiple characters, the minimum number of mutations on each possible tree has to be determined.

In Figure 17.13, Ridley provides an *algorithm* for determining some (but not all) possible ancestral sequences and for finding the smallest number of mutations required by a tree.

More sophisticated algorithms exist for searching all possible trees and all possible ancestral states.

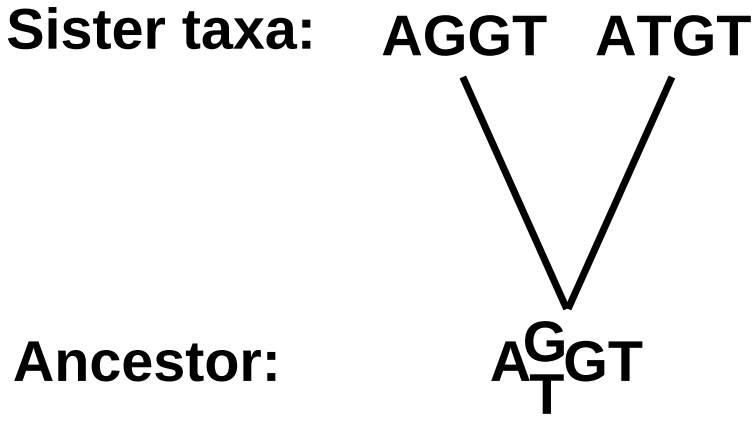
An algorithm for finding the minimum number of mutations on a tree

Part A: Determining ancestral states

(1) Pick a pair of sister taxa.

(2) Write an inferred sequence for the most recent common ancestor of these two taxa at the node connecting them. Site by site, determine:

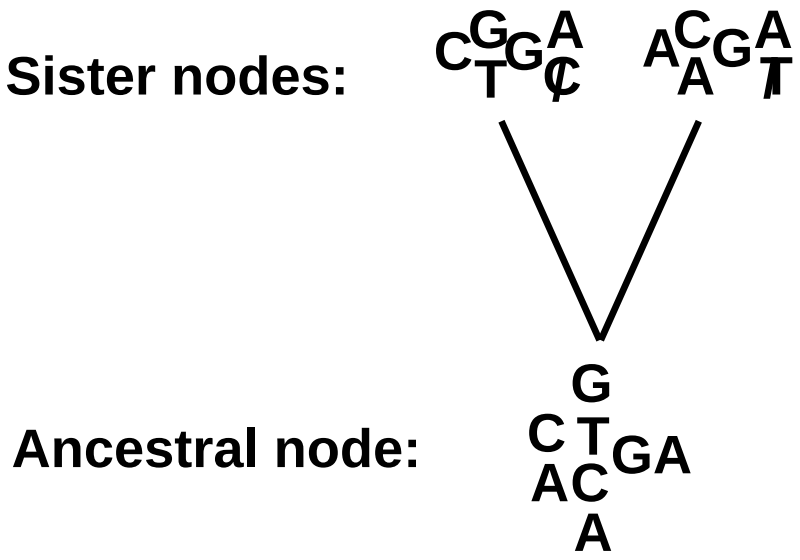
- if the basepairs are the same in both sister taxa. Add this basepair to the ancestral sequence.
- if the basepairs differ between the sister taxa. Add both basepairs to the ancestral sequence (stacked on top of each other).



(3) Now ignore the original sister taxa, and treat their ancestral node as a new taxon.

(4) Repeat steps (1)-(3), with the following additional instruction for step (2). When comparing sites that have stacks of possible basepairs:

- If the stacks have no basepairs in common, add all basepairs to the ancestral sequence (stacked on top of each other).
- If the stacks have a basepair in common, strike out all other members of the two stacks. Then move back up the tree to resolve the stacks in previous parts of the tree if necessary.



(5) Once all ancestral nodes have been determined, resolve any remaining stacks, being careful to choose the same basepair at a site on both sides of a branch whenever possible.

NOTE: There are often several different possible sets of ancestral states that would give the same minimum number of mutations.

Part B: Counting the minimum number of mutations

(1) Along each branch, make a mark for each difference between the two sequences at either end of the branch.

(2) Count the total number of marks on the tree.

➡ Minimum number of mutations required to explain this sequence data with this tree.

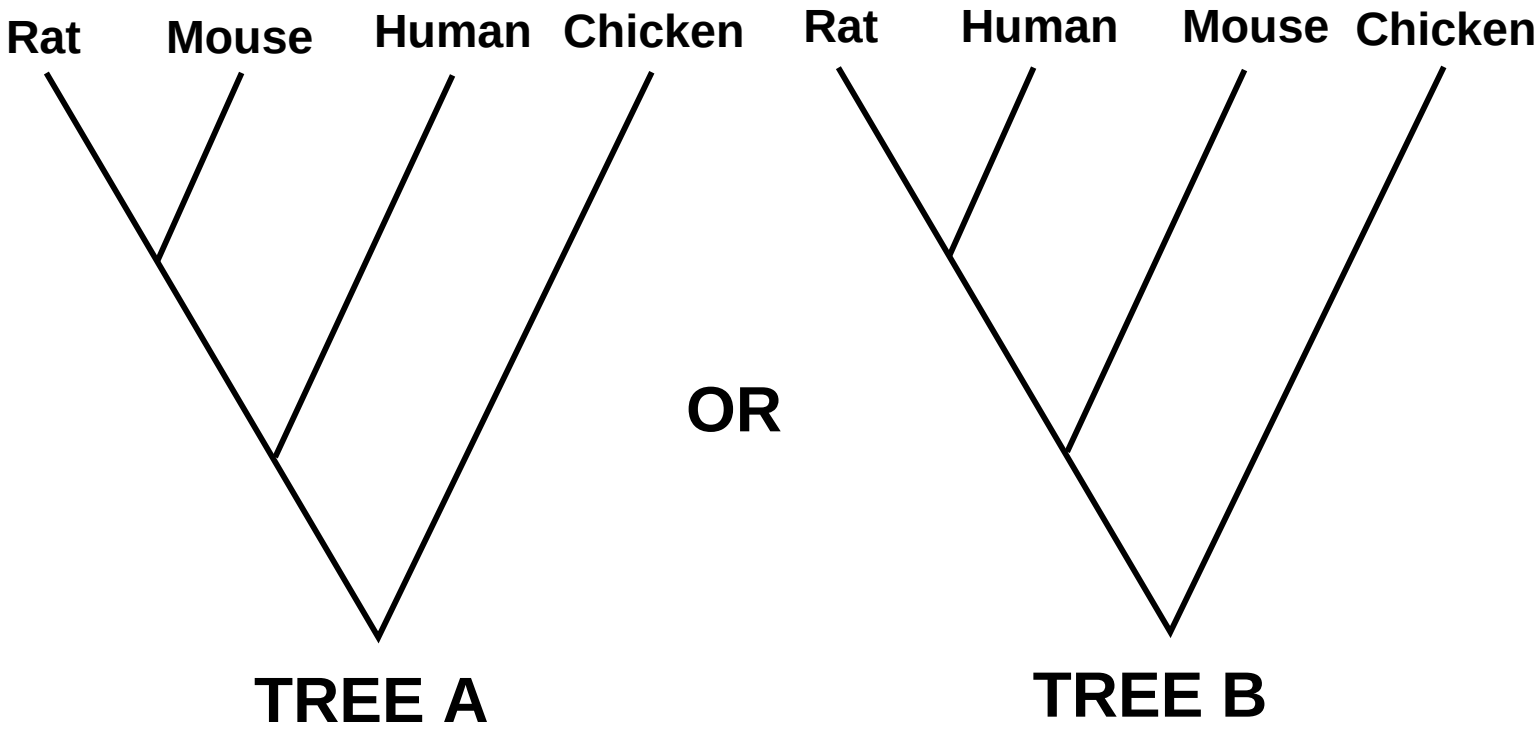
(Figure 17.13 describes this method Ridley.)

Example

For the following data set, which tree is most parsimonious?

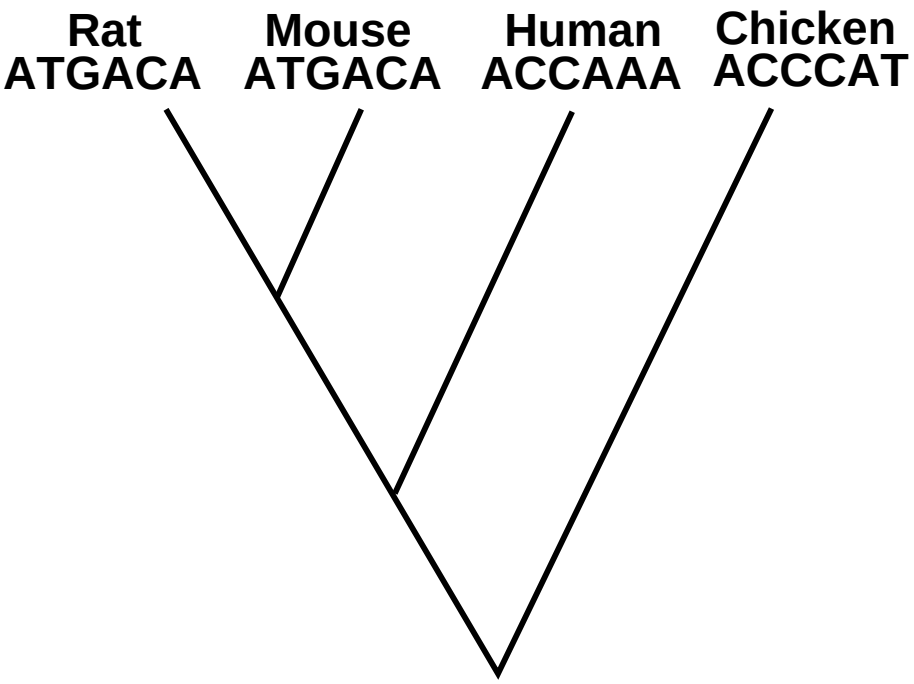
Sites 819-824 in cytochrome oxidase 1 (mtDNA)

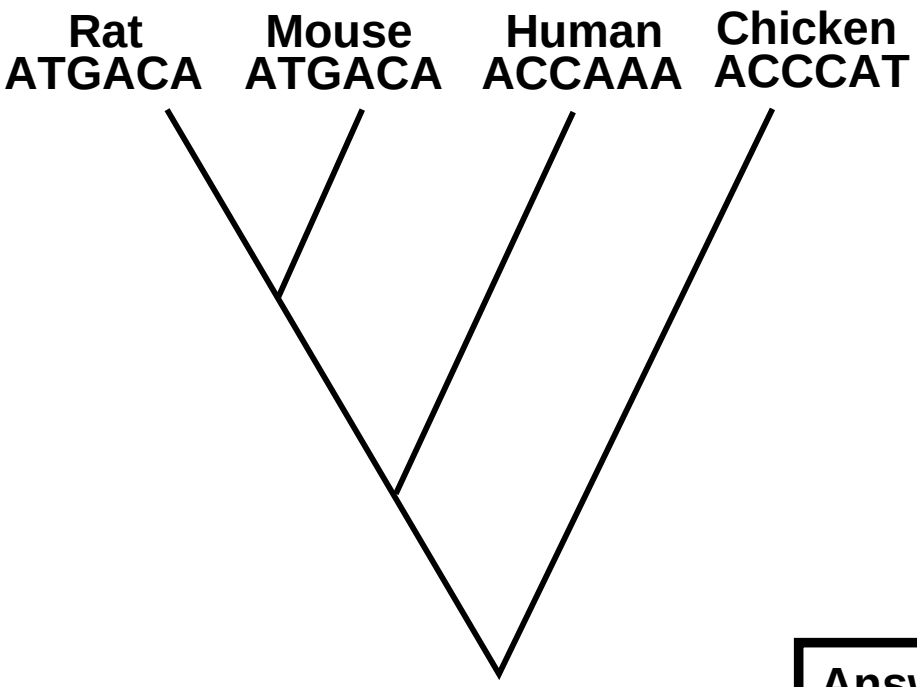
Chicken:	**ACCCAT**
Mouse:	**ATGACA**
Rat:	**ATGACA**
Human:	**ACCAAA**



TREE A **TREE B**

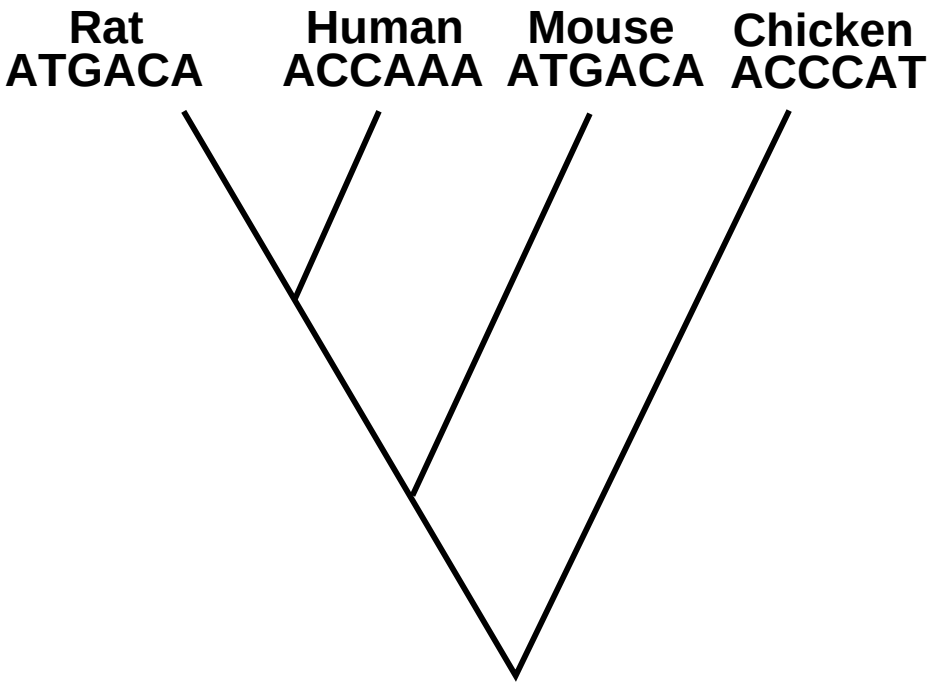
Worksheet for tree A

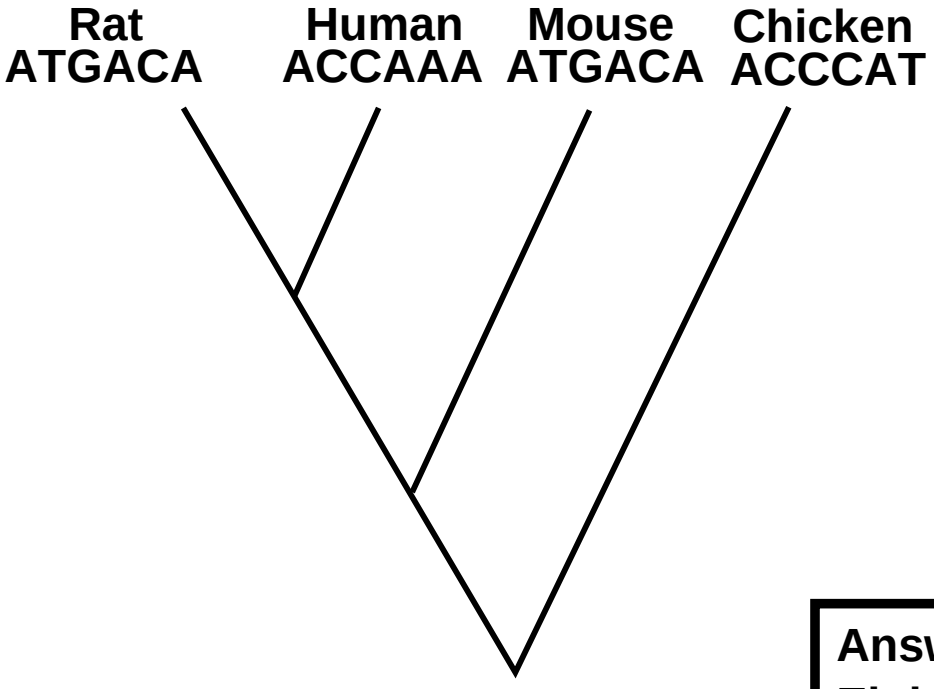




Answer:
Five mutations

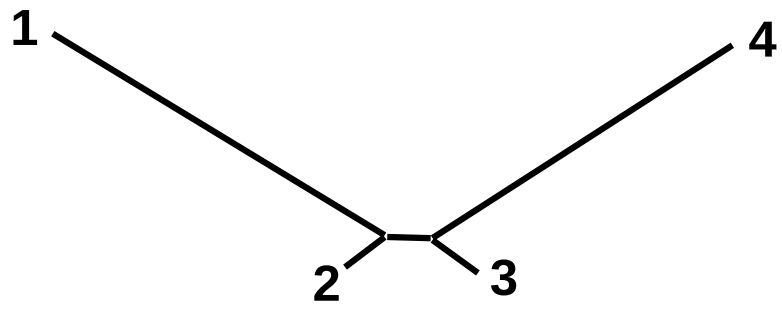
Worksheet for tree B





Answer:
Eight mutations

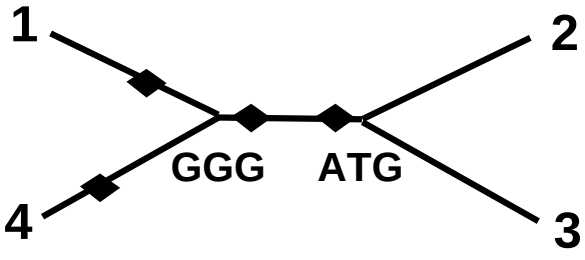
Interestingly, parsimony can fail as a method, because evolution may take more steps than absolutely necessary to get from the ancestral sequence to the current sequences.



True Tree

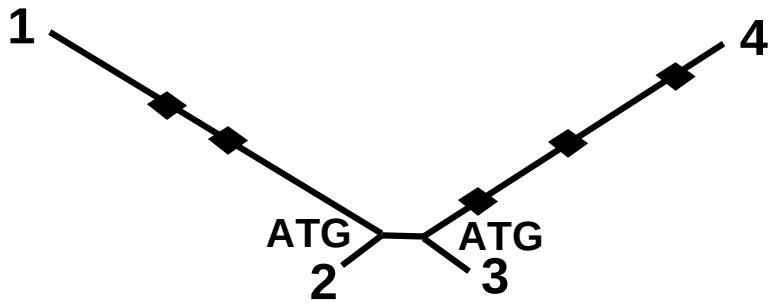
Mutations on both long branches may be common.
Mutations on internal branch may be rare.

Species 1: CGG
Species 2: ATG
Species 3: ATG
Species 4: GGT



Tree requires:
Four mutations

Most Parsimonious Tree



Tree requires:
Five mutations

True Tree

Advantages of Parsimony Analysis

- Conceptually easy to understand
- Straightforward to calculate the length of a tree
- Accurate if few evolutionary changes have occurred (homoplasy unlikely)

Disdvantages of Parsimony Analysis

- Underestimates the true amount of evolutionary change
- Can strongly favor the wrong tree ("positively misleading")

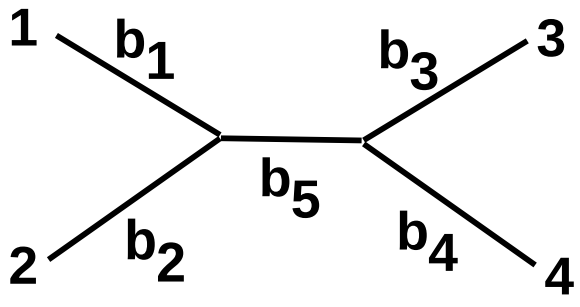
2. Distance analyses

Species comparisons are often presented as distances between each pair of species (e.g. the number of sequence differences).

Sometimes only distance data are available, such as the strength of DNA-DNA hybridization.

Distance methods choose a tree on the basis of how well it coincides with the observed distances between every pair of species.

For any particular tree, the expected distance between two taxa can be found by summing the branch lengths separating the two taxa:



For example, the expected distance between species 1 and species 3 is $d_{13} = b_1 + b_3 + b_5$.

Distance methods attempt to minimize the discrepancy between the observed distances (D_{ij}) and the expected distances (d_{ij}), e.g. by minimizing:

$$\Sigma\ w_{ij}(D_{ij} - d_{ij})^2,$$

where w_{ij} is a weighting term that can be used, for example, to diminish the importance of distantly related taxa.

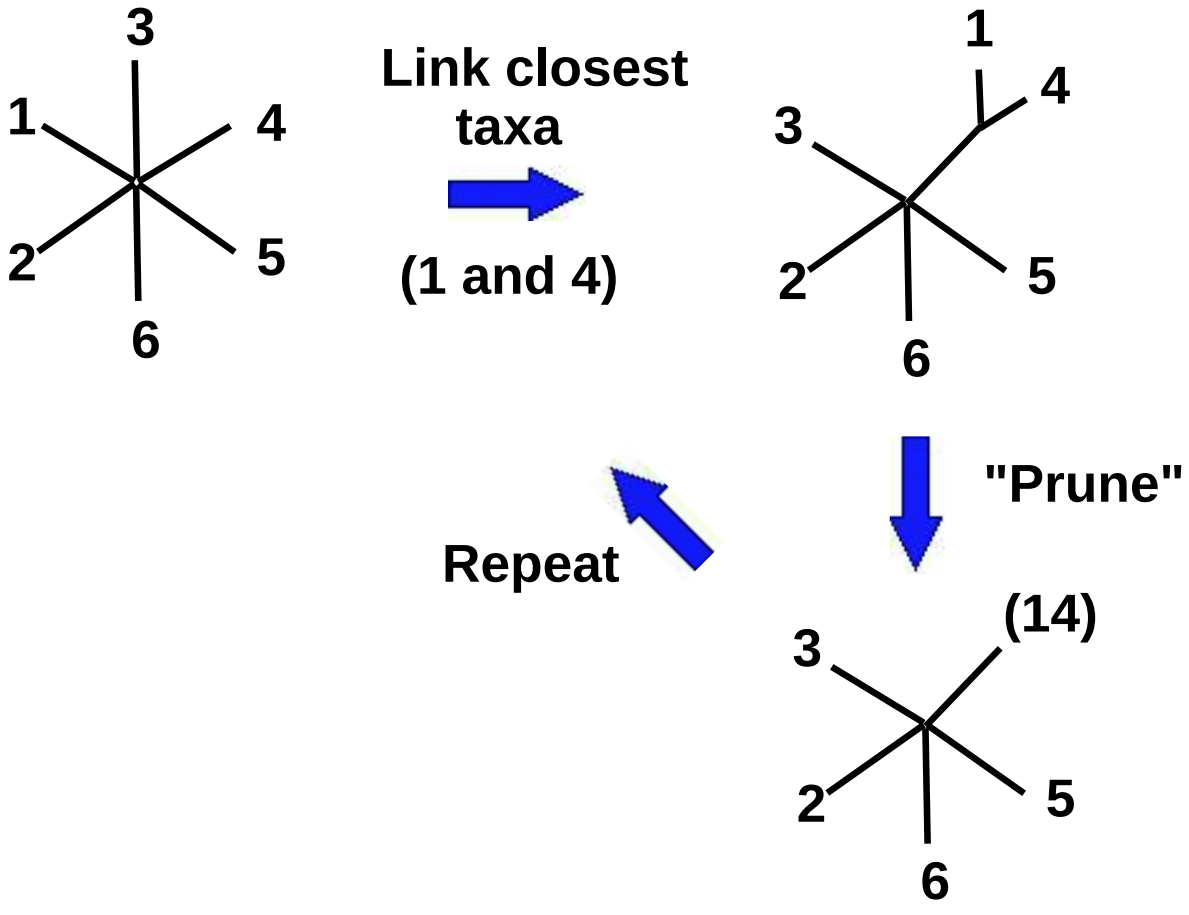
A particularly common distance method is neighbor-joining.

Neighbor-joining is an algorithm, meaning that one follows a recipe to get the tree rather than figuring out how to mimimize functions like the one above.

Neighbor-joining: Starting from a star-like tree, the two closest taxa are placed together as neighbors. [Aside: The distances are first corrected to take into account potential differences in rates between the taxa.]

These two taxa are then represented by their common ancestral node and removed from the analysis.

The procedure is repeated until the full tree is resolved.



Advantages of Distance Analysis

- The only method available for distance data
- Fast (especially neighbor-joining)
- Better able to handle large data sets

Disdvantages of Distance Analysis

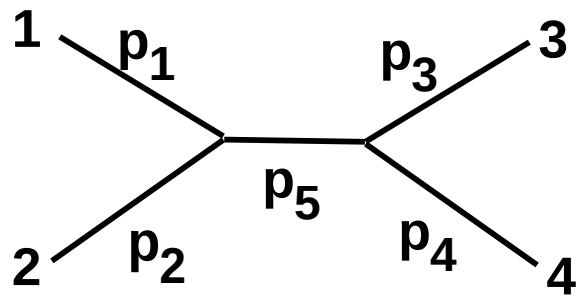
- Distances can hide convergent evolution (homoplasy)
- Distance methods can generate incorrect trees (when distances do not scale with time)

3. Maximum likelihood analysis

In a maximum likelihood analysis, a specific model is used to determine the probability that a given base substitution will occur along a given branch on a tree.

The maximum likelihood tree is the one that can generate the observed data with the highest probability.

For any one site, the likelihood of observing the data given a particular tree and a particular model of sequence evolution is calculated:



Likelihood of this tree:

The total probability of making each transition on the tree (= $p_1 p_2 p_3 p_4 p_5$), summed over all possible internal nodes.

The likelihood for the whole sequence is then calculated as the product of the likelihoods for each site.

Advantages of Maximum Likelihood Analysis

- Extremely flexible (any model can be used)
- Statistically justifiable
- Will always infer the right tree given enough data (if the model is correct)

Disdvantages of Maximum Likelihood Analysis

- Impossible to know if the model is correct
- Computer intensive
- Practically impossible with many taxa

Evaluating a Tree

Frequently, many trees are optimal or near optimal on the basis of a criterion. Generating a "best" tree does not say how much better it is than other trees.

One of the most common methods used to evaluate the support in the data for the phylogenetic relationships shown on a tree is the *bootstrap* resampling procedure.

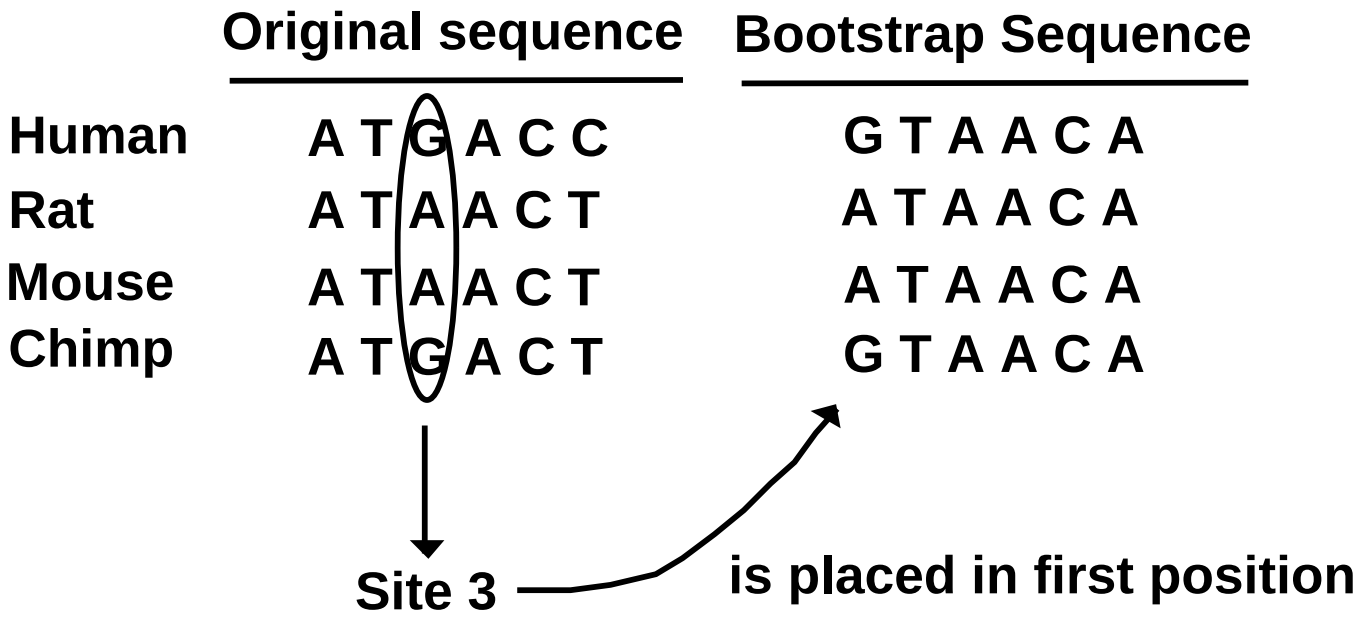
The bootstrap technique involves generating artificial sequences by randomly sampling sites from the original sequences with replacement.

This randomly generated data set has the same sequence length but a slightly different composition (i.e some sites will be oversampled and others not).

For example, consider a simple sequence with 6 sites.

Say that the first site chosen randomly is 3. For each species, site 3 is placed in the first position of the bootstrap sequence.

This is repeated until the bootstrap sequence is also 6 bp long.



(Then the next five randomly chosen sites: 2, 1, 1, 5, 4, are placed in the next five positions.)

The "best" tree is then determined from the bootstrap sequences, using the same method as used with the original data set.

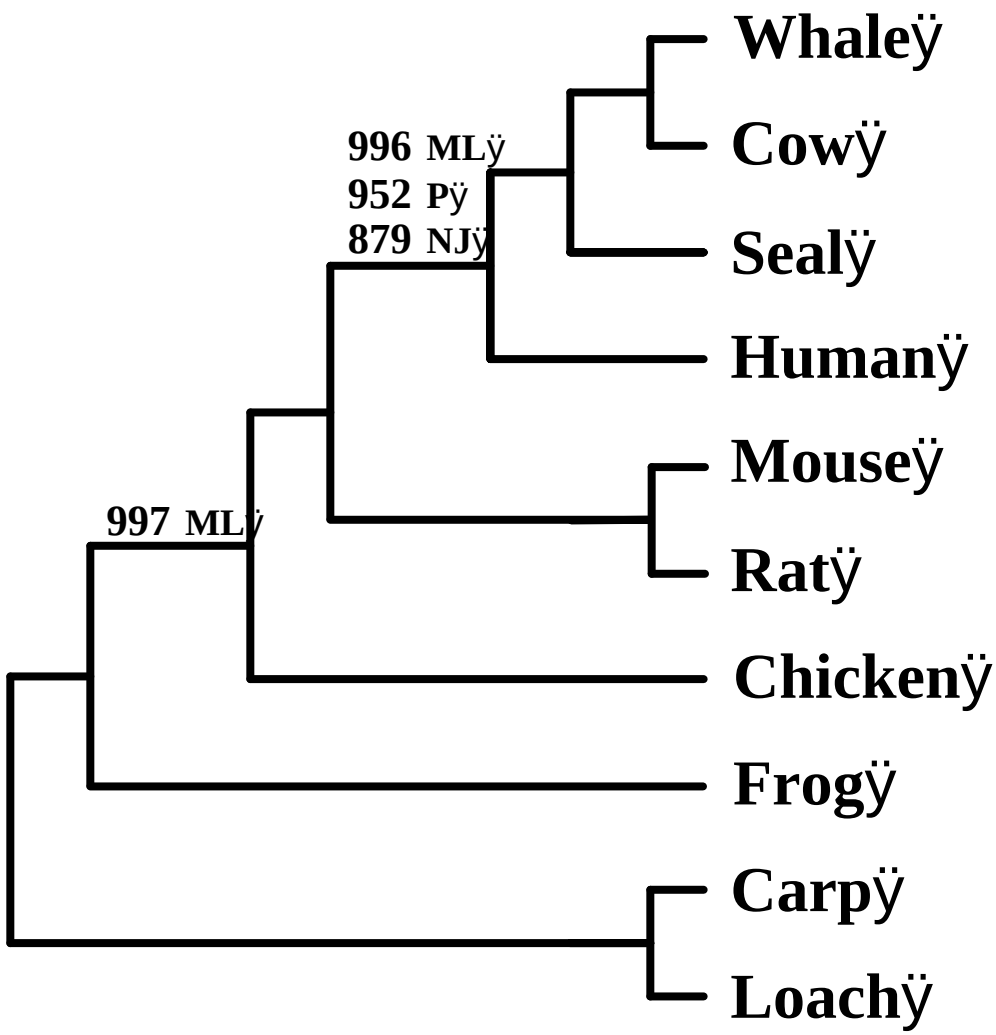
This whole process is repeated at least 100 times.

The number of times that a clade is seen among the bootstrap trees is reported.

The more often a clade is present among the bootstrap trees, the more strongly the data support that clade, because the result is insensitive to which basepairs happen to be sampled.

EXAMPLE:

Cummings et al (1995) used the entire mitochondrial genomes of ten vertebrates and obtained the following tree using parsimony (P), neighbor joining (NJ), and maximum likelihood (ML) methods:



All clades were supported in 1000/1000 bootstrap data sets, with the exception of the two clades shown, which still had strong support.

Conclusion

We have focused on the criteria used to build trees.

The criteria have been refined to take into account several other factors including:

- transition/transversion bias
- mutation rate heterogeneity across a sequence
- rate variation along a tree.

In practise, efficient algorithms have to be used in order to evaluate all the possible arrangements of taxa on trees.

Several computer programs are available that implement these phylogenetic methods (the commonly used, general-purpose programs are Phylip, PAUP, and MEGA).

For more information, check out this web-site of phylogenetic resources.

SOURCES:

- **Hillis, Moritz, and Mable (1996) Molecular Systematics. Sinauer Associates, MA.**
- **Felsenstein (1988) Phylogenies from molecular sequences: Inference and reliability. Annual Review of Genetics 22:521-565.**
- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Freeman and Herron (1998) Evolutionary Analysis. Prentise Hall.

Examples of Phylogenetic Reconstruction

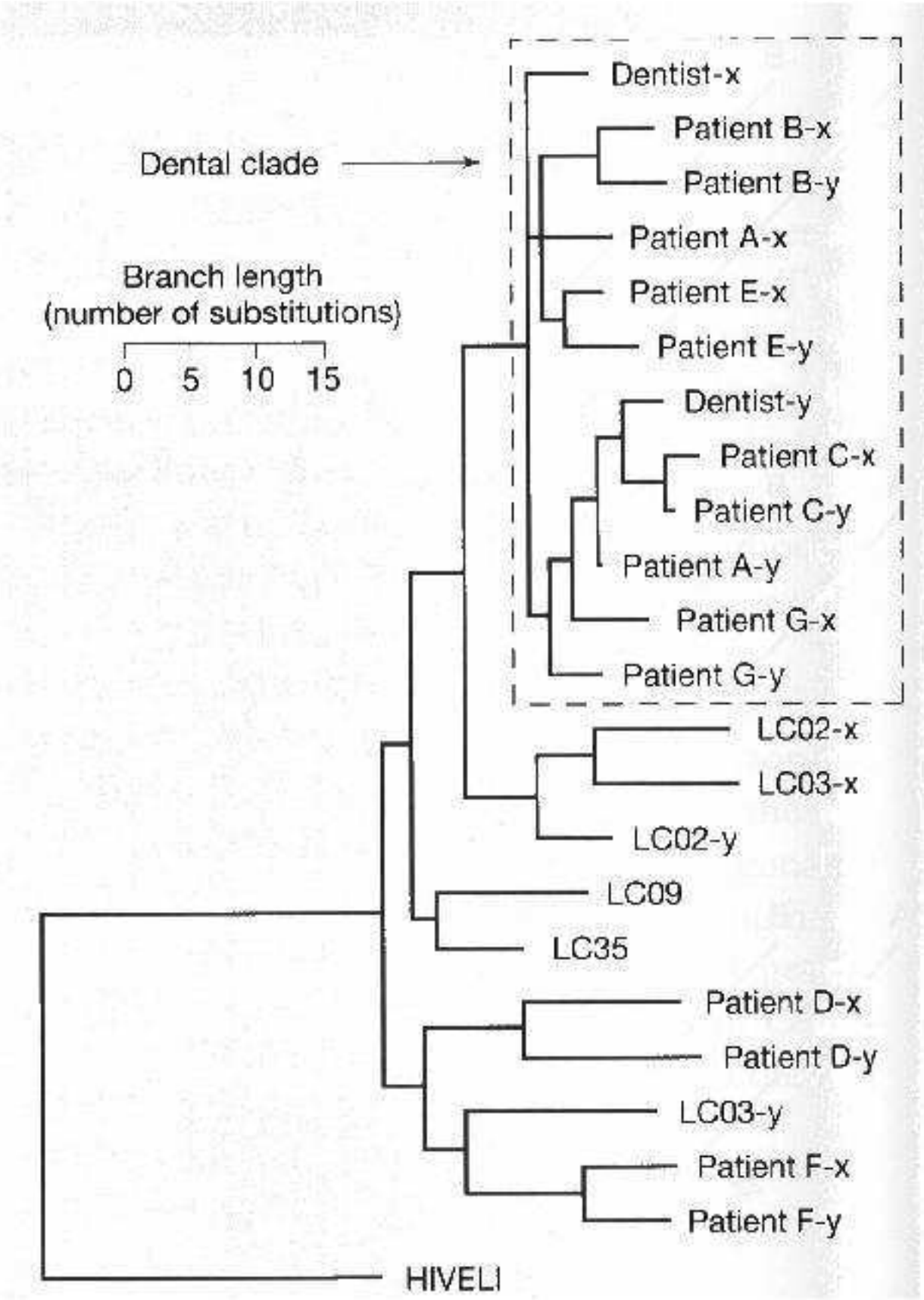
1. HIV transmission

Recently, an HIV-positive Florida dentist was suspected of having transmitted the HIV virus to his dental patients.

Although a number of his patients were HIV-positive, it was unclear whether they had been infected by visiting the dentist.

The Centers for Disease Control sequenced the gp120 gene from viruses in the dentist, his HIV-positive patients, and a number of HIV-positive people from the same community.

This data was analysed by Ou et al (1992) and reanalysed using a number of methods (parsimony, distance, maximum likelihood) by Hillis et al (1994):



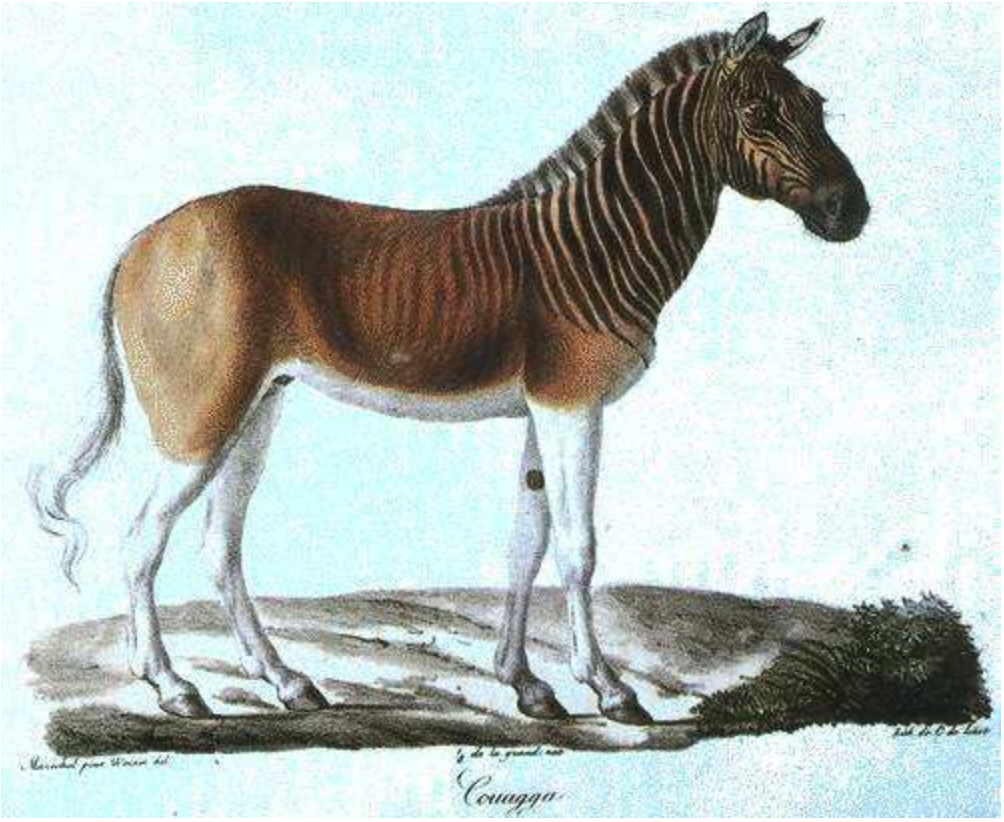
(From Freeman and Herron, 1998; x and y represent different viruses sampled from the same individual)

All methods supported the existence of a "dental clade" (dashed box), which suggests that the dentist’s HIV strain is ancestral to those found in patients A, B, C, E, and G.

This case was very important in publicizing the need for careful, sterile techniques in dental and medical practise.

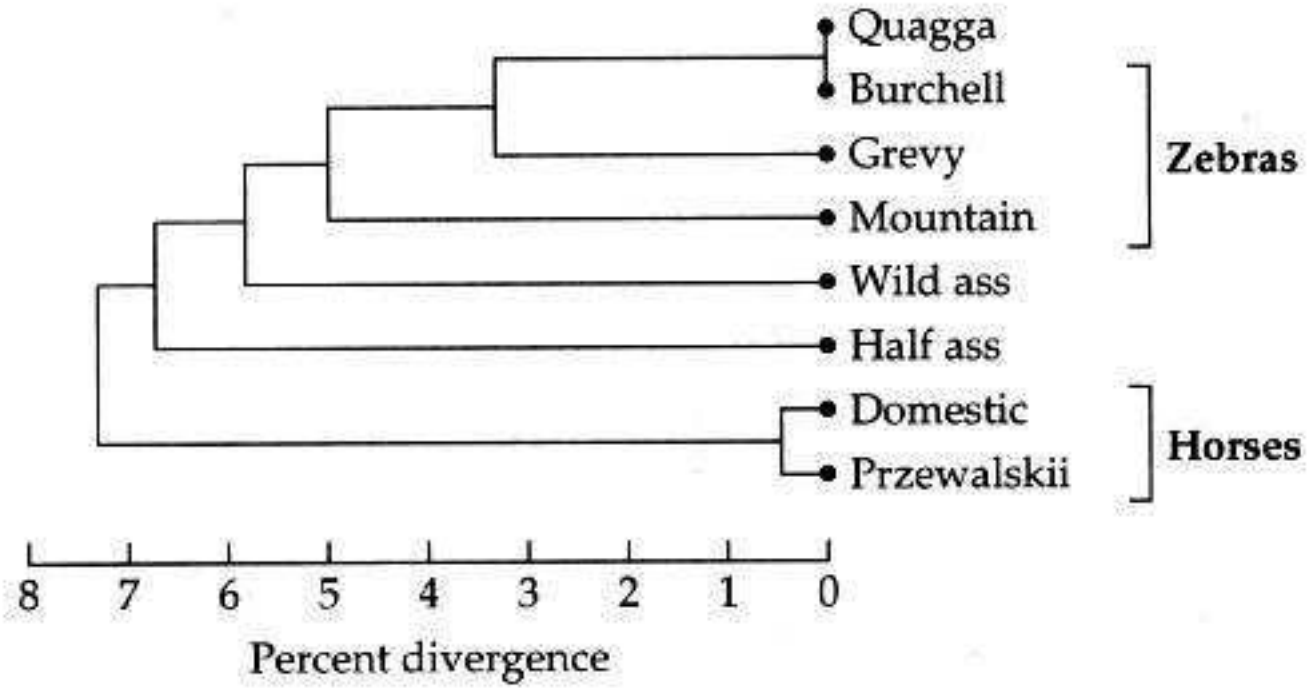
2. Ancient DNA

In 1984, Higuchi et al succeeded in isolating DNA from the 140-year old skin of a quagga, a species in the horse genus which had gone extinct with the death of the last quaaga in the Amsterdam Zoo on August 12 1883.



Numerous changes happen to ancient DNA, including pyrimidine modifications, missing bases, and intermolecular cross-links.

Although these changes make it difficult to obtain DNA for analysis, enough mitochondrial DNA was obtained to reconstruct the following phylogeny (Paabo et al, 1989):



(From Li, 1997)

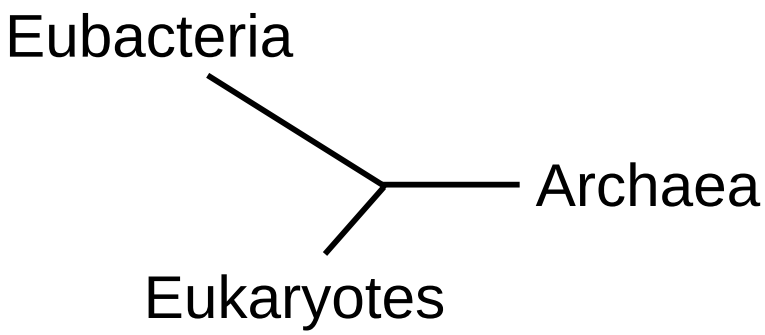
Rather than being closely related to the domestic horse, this phylogeny supports the view that the quagga is closely related to the Burchell's zebra (and is perhaps even a subspecies).

3. The Tree of Life

Recently, our view of the major groupings of living organisms has undergone a major revision, following the revelation that three major branches of life exist.

The traditional view divided living organisms into two main domains: eukaryotes and bacteria.

Molecular phylogenies of eukaryotes and bacteria suggested instead that a third group, the Archaea, exist as well.

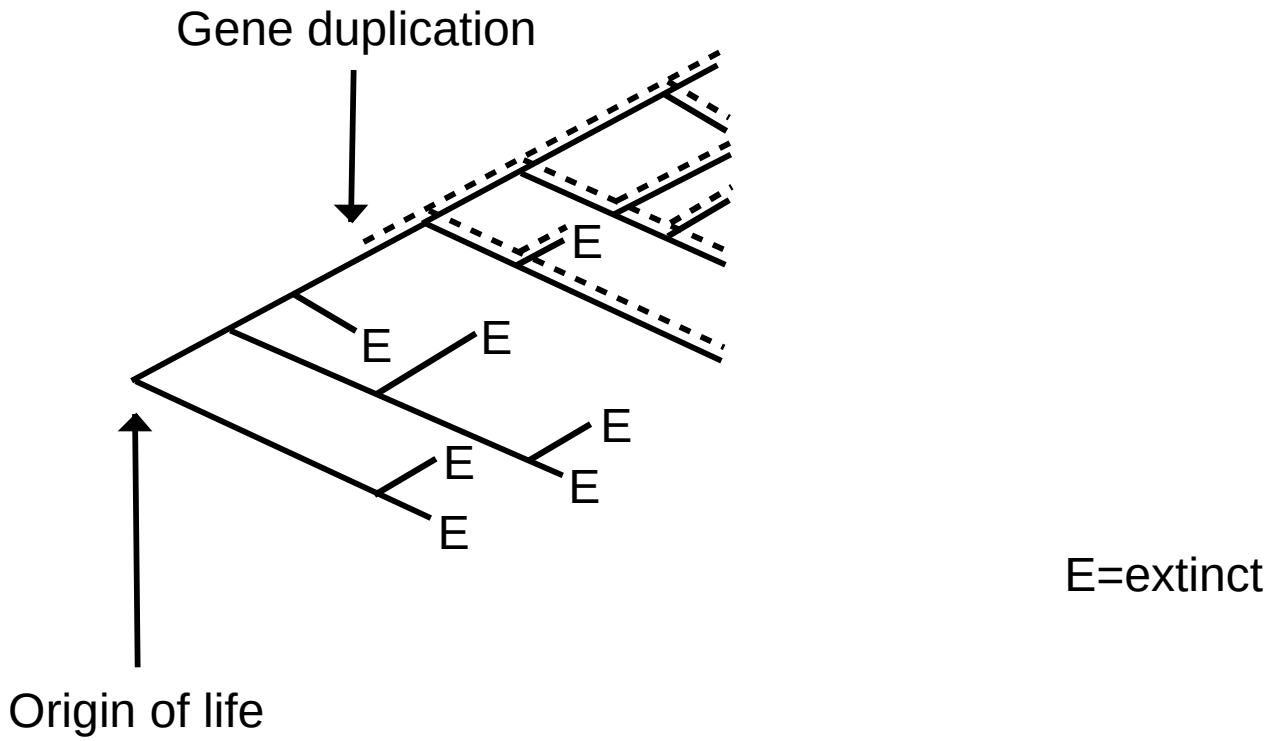


The Archaea include a number of prokaryotes living in harsh environments, including thermophiles ("heat-loving") and halophiles ("salt-loving") prokaryotes.

But how can we root the universal tree of life without an outgroup?

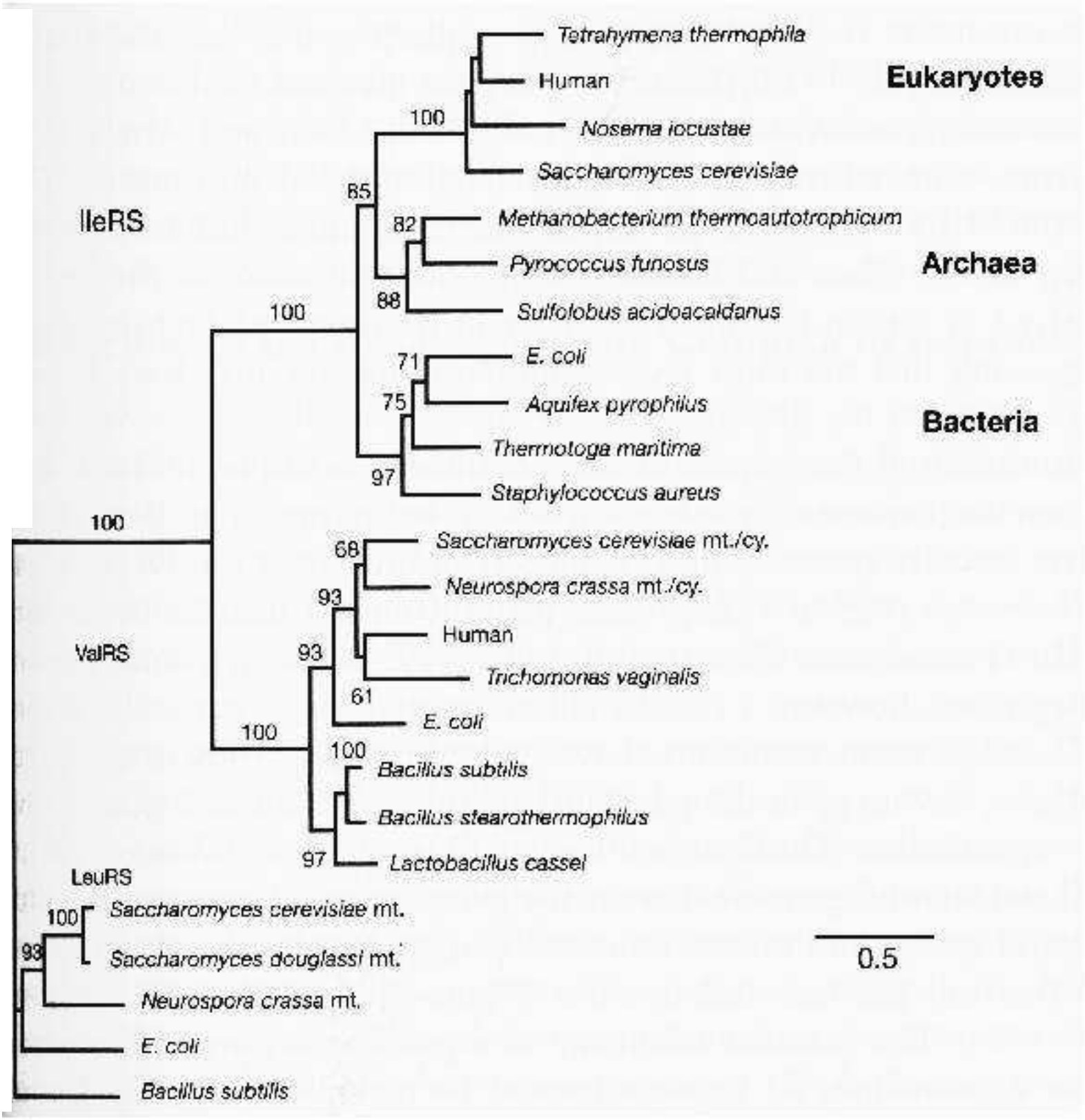
Ingeniously, Schwartz and Dayhoff suggested that a pseudo-outgroup exists in the form of duplicated genes.

If a gene duplicated before the diversification of all the branches of organisms currently alive, then the two genes will share a common ancestor before this diversification and one gene can be used to root the other.



To root the tree of life, Brown and Doolittle (1995) used aminoacyl-tRNA synthetase genes, which add amino acids to tRNAs.

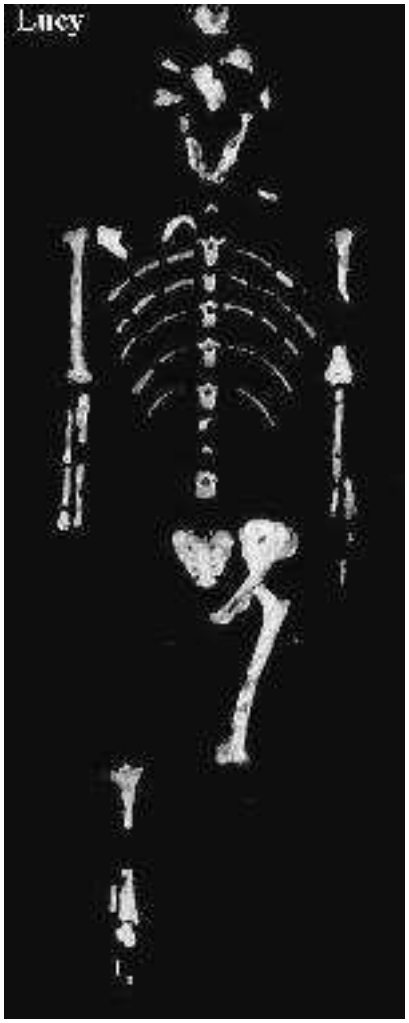
The aminoacyl-tRNA synthetase genes that add isoleucine (IleRS), valine (ValRS), and leucine (LeuRS) are structurally similar and are thought to represent very early gene duplication events.



(From Freeman and Herron, 1998)

Similar studies in other laboratories using different genes confirms this phylogeny, suggesting that the closest relatives to eukaryotes are the Archaea.

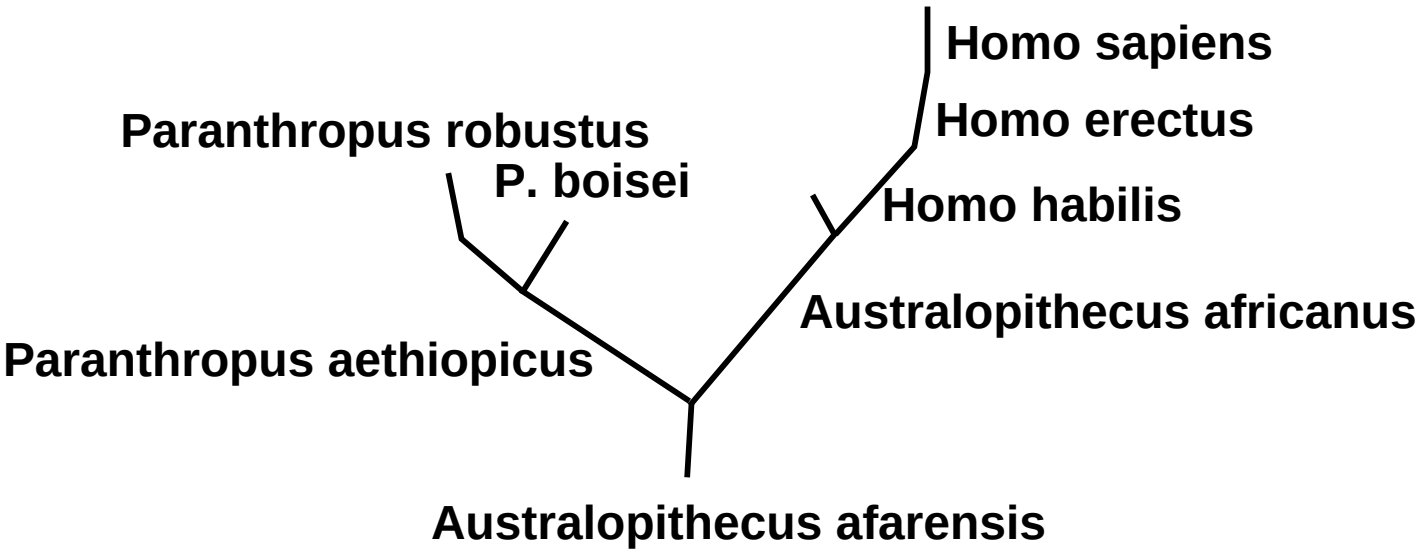
4. Mitochondrial Eve



The earliest fossils in the genus *Homo* have been found in African deposits nearly two million years old.

These early hominids, *Homo habilis*, form a link between older, more ape-like *Australopithecus* species (e.g. "Lucy") and more modern precursors to humans, *Homo erectus*.

Fossil *Homo erectus* specimens are found during a period from 1.6 MYA to 0.3 MYA, with the later specimens appearing more and more similar to *Homo sapiens* fossils which date back to 0.4 MYA.

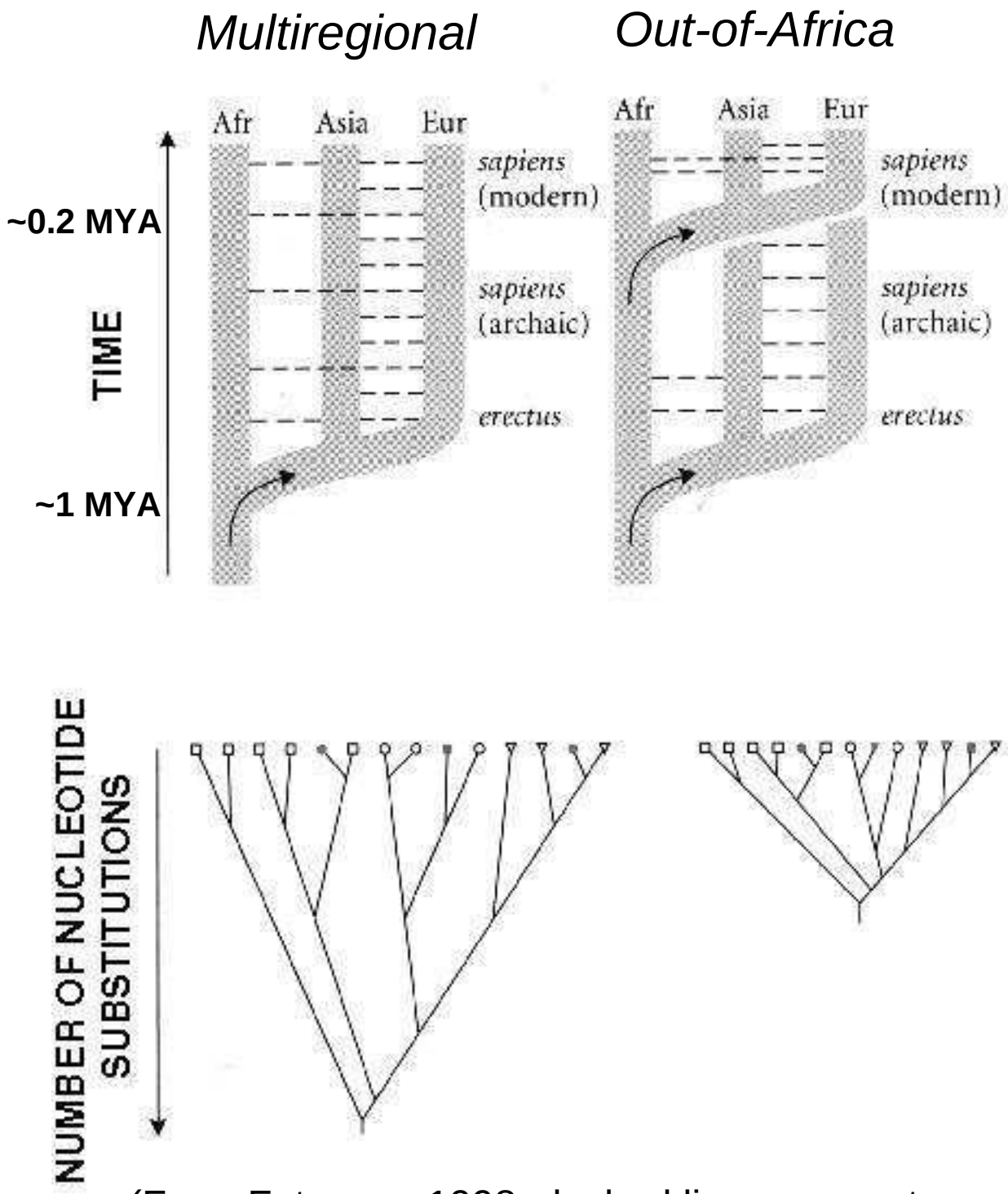


"On average, brain size (cranial capacity) increases throughout hominid history, although not at a constant rate, and there are progressive changes, from [*Australopithecus*] *afarensis* to [*Australopithecus*] *africanus* to [*Homo*] *erectus* to [*Homo*] *sapiens*, in many other features, such as the teeth, face, pelvis, hands, and feet...Although many issues remain unresolved, the most important point is fully documented: *modern humans evolved from an ape-like ancestor*"

-- Futuyma (1998), p. 733

How modern *Homo sapiens* are related to the *Homo erectus* populations that existed throughout Africa and Asia has been a subject of much debate.

The two primary views are the *multiregional* hypothesis and the *out-of-Africa* hypothesis:



(From Futuyma, 1998; dashed lines suggest gene flow)

The multiregional hypothesis claims that modern *Homo sapiens* evolved from precursors throughout Africa and Asia, with gene flow ensuring that modern traits were common to all populations.

The out-of-Africa hypothesis claims that modern *Homo sapiens* evolved fairly recently from a population within Africa and then migrated out of Africa to form modern *Homo sapiens*.

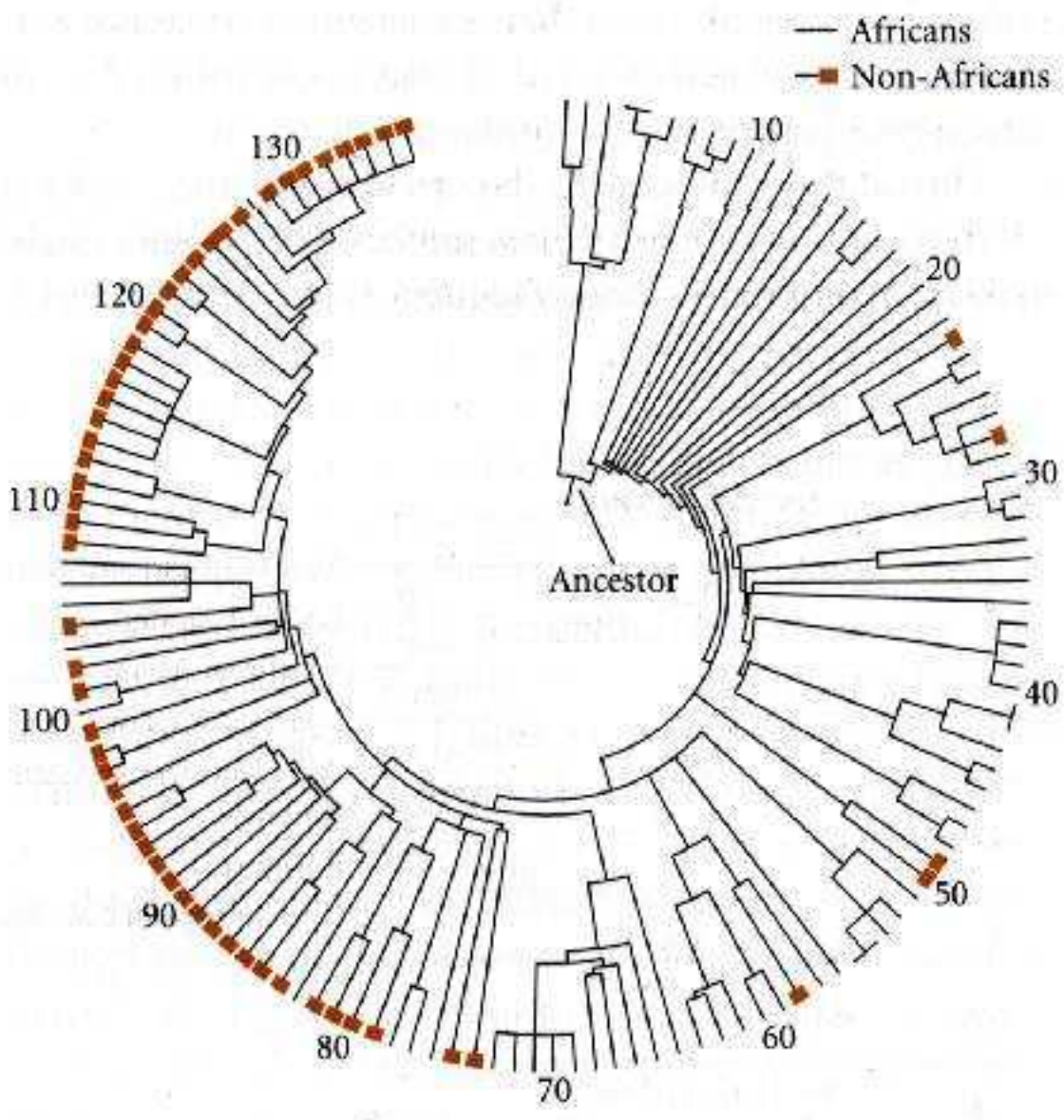
The two hypotheses lead to very different predictions:

	Multiregional	Out-of-Africa
Most recent common ancestor:	>1,000,000 YA	~ 200,000 YA
Genetic diversity:	High	Low

Cann et al (1987) and Vigilant et al (1991) used phylogenies estimated from mitochondrial DNA to test the above hypotheses.

For example, Vigilant et al (1991) sequenced two rapidly evolving segments of the mitochondrial genome from 189 individuals.

From this data, they constructed a tree using parsimony:



The time until the most recent common ancestor of these sequences (the "Mitochondrial Eve") was estimated to have lived 166,000 - 249,000 years ago, consistent with the out-of-Africa hypothesis.

In addition, the greatest genetic diversity was found in Africa, suggesting that modern *Homo sapiens* did evolve in Africa, with other geographical regions containing only some of the mitochondrial genomes present in Africa.

This study was critisized for a number of reasons, most importantly that only one tree was presented within the paper. Many trees were equally parsimonious and some did not support an African origin for humans.

Other studies soon followed:

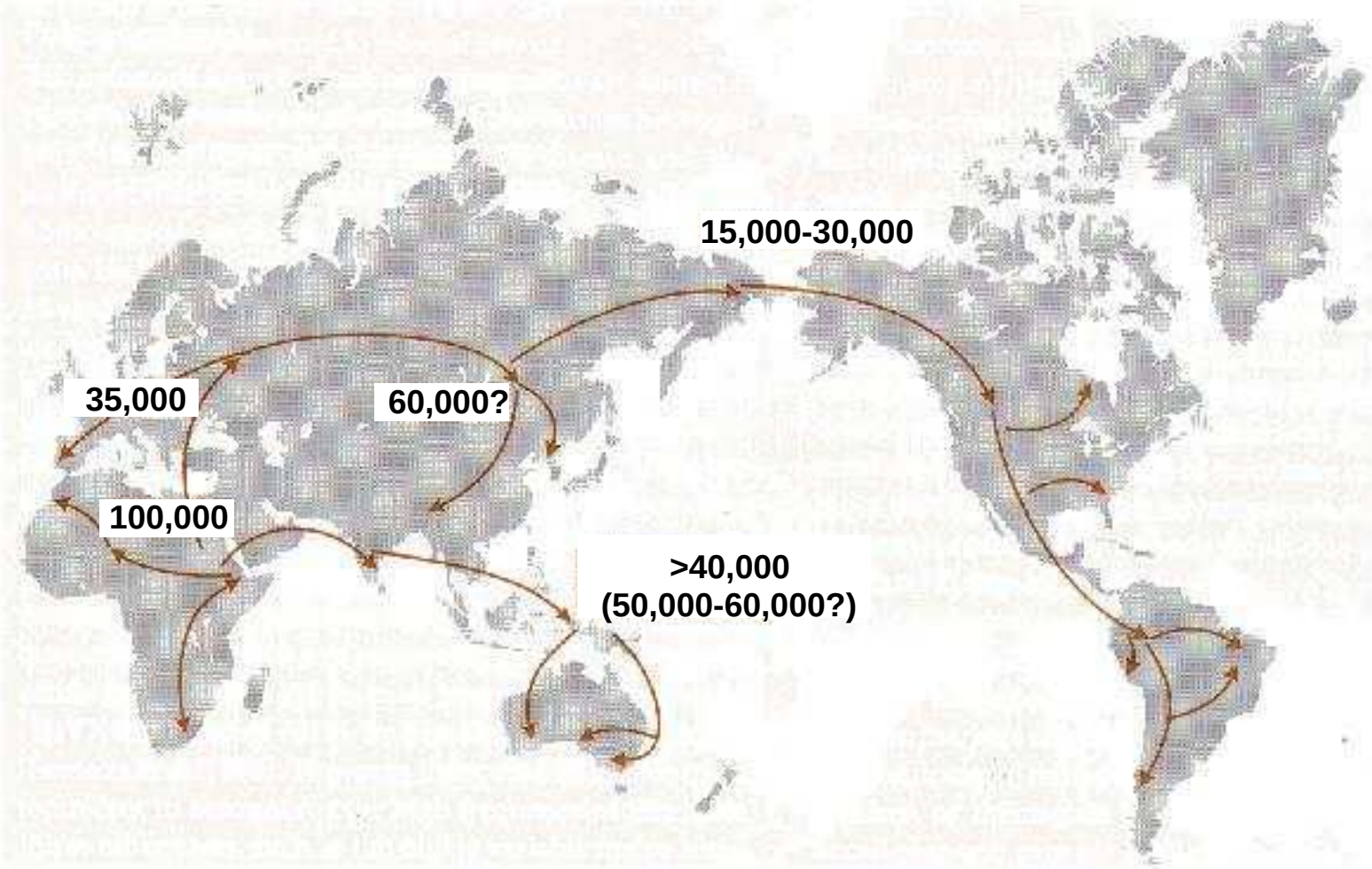
- Ruvolo et al (1993) used similar data and dated Eve to 129,000 - 536,000 YA
- Horai et al (1995) used entire mitochondrial genomes (!) to date Eve to 125,000 - 161,000 YA
- Bowcock et al (1994) used microsatellite data to confirm that the greatest genetic diversity occurs in Africa
- Goldstein et al (1995) used this microsatellite data to date Eve to 75,000 - 287,000 YA

These last studies are particularly important.

It is possible for a single gene (or a completely linked segment like the mitochondrial genome) to provide a biased historical picture. [For example, a beneficial mutation may have recently appeared and fixed.]

Several unlinked genes are needed to get a reliable picture of the phylogeny of a species.

Although the debate continues and has not been definitively settled, the balance of evidence suggests that modern *Homo sapiens* did evolve relatively recently (100,000-300,000 YA) from populations of archaic *Homo sapiens* within Africa and then migrated throughout the world.



SOURCES:

- HIV transmission and the Tree of Life figures and discussion: Freeman and Herron (1998) Evolutionary Analysis. Prentise Hall.
- Ancient DNA figures and discussion: Li (1997) Molecular Evolution. Sinauer Associates, MA.
- Mitochondrial Eve figures and discussion: Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Take a walk through human prehistory.

Species Concepts

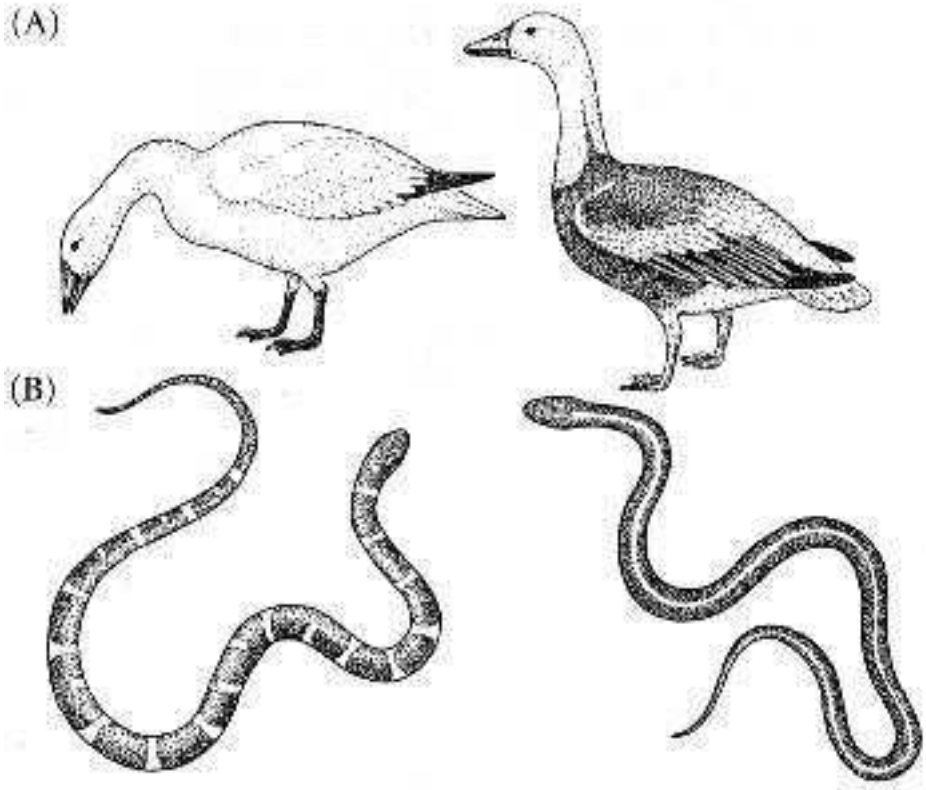
What are species?

Species concepts and definitions continue to be one of the most controversial topics in biology.

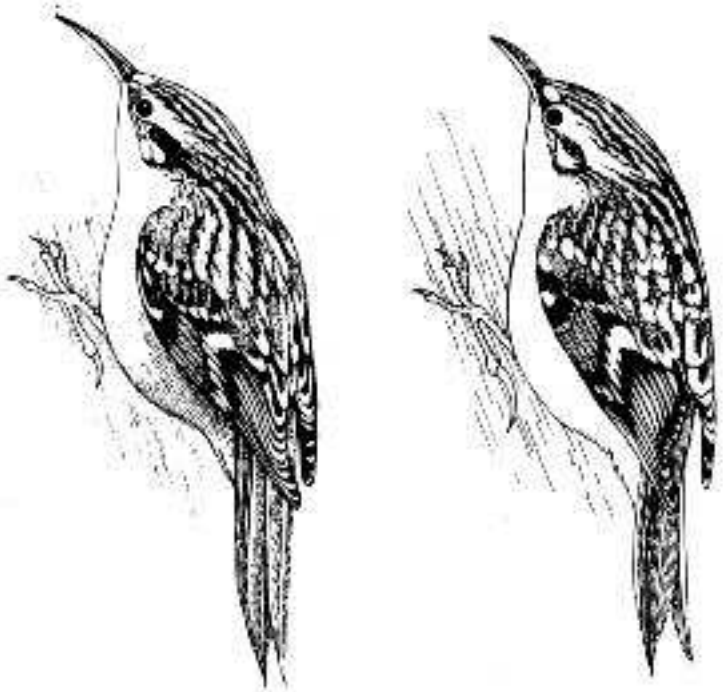
A large number of alternative species concepts exist, each with its own strengths and limitations.

1. The typological species concept: A species is a set of organisms that resemble one another and is distinct from other sets (Linnaeus). But:

- Do large differences in phenotype always reflect large differences in relatedness among organisms?
- How well are we able to discern small, but significant, differences?



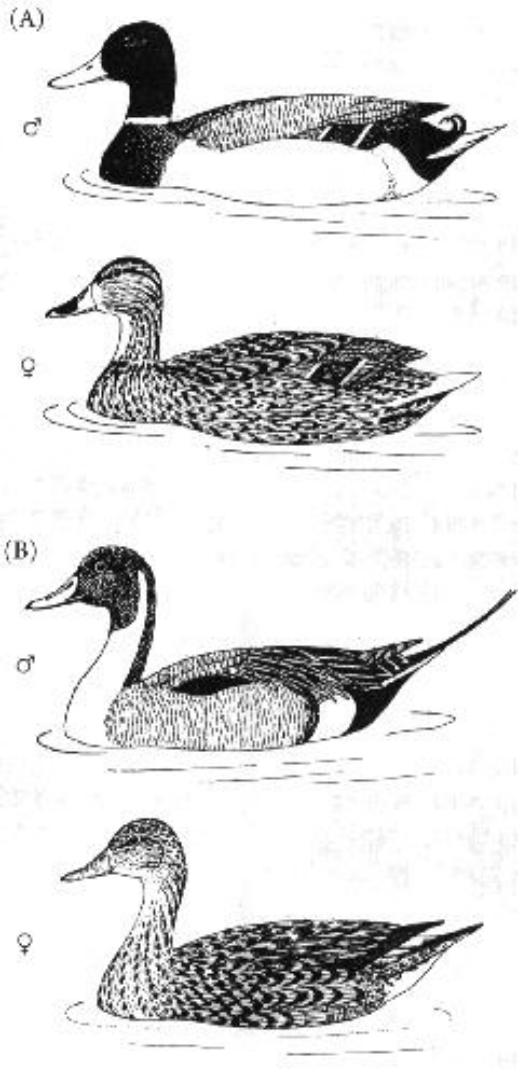
Large differences in morphology may be due to single-locus color polymorphisms, as seen in snow geese and king snakes.



Small differences in morphology may be difficult to detect, as with the slight brown coloration on the flank and shorter claw on the hind toe of the short-toed treecreeper.

2. The biological species concept (BSC): "Species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups" (from Mayr 1942; Dobzhansky 1935). But:

- How do we evaluate "potentially interbreeding" for populations that do not live in the same location?
- How much reproductive isolation is needed? 100%?
- How does the BSC apply to asexuals?



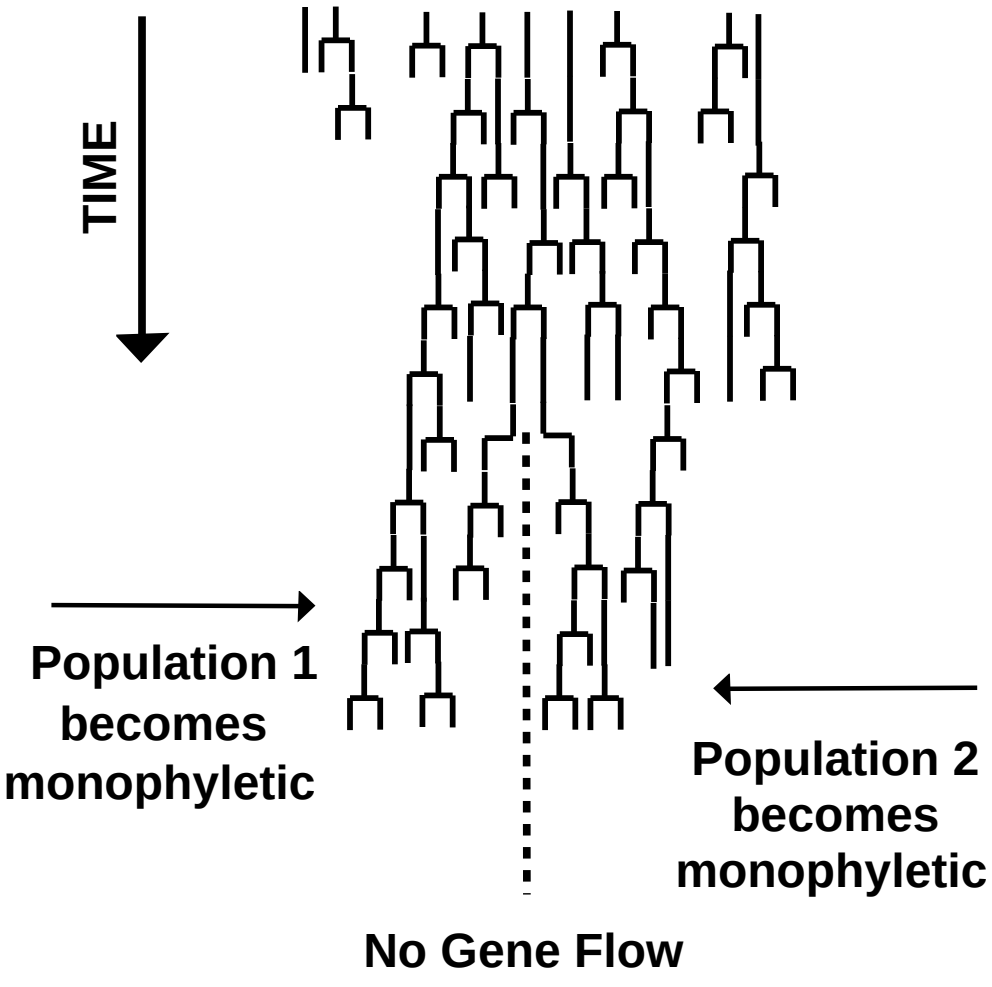
Mallards can hybridize with the northern pintail, but they rarely do so in nature.

3. The evolutionary species concept: A species is a single lineage of populations or organisms that maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate (Wiley 1978). But:

- How much "identity" is needed?
- Will we ever know the "evolutionary tendency" of a population?
- How can we assess the "historical fate" of a population?

4. The genealogical species concept: A species is the smallest monophyletic group of common ancestry (de Queiroz and Donoghue 1990). (Meaning: The smallest grouping of organisms for which individuals within the group are more closely related to each other than to any other organism.) But:

- Can we know that a currently monophyletic population will remain monophyletic?
- Certain genetic polymorphisms can be shared (i.e. not become monophyletic) for a very long period of time, does this alone prevent speciation?



5. The ecological species concept: A species is a lineage that occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range (Van Valen 1976). But:

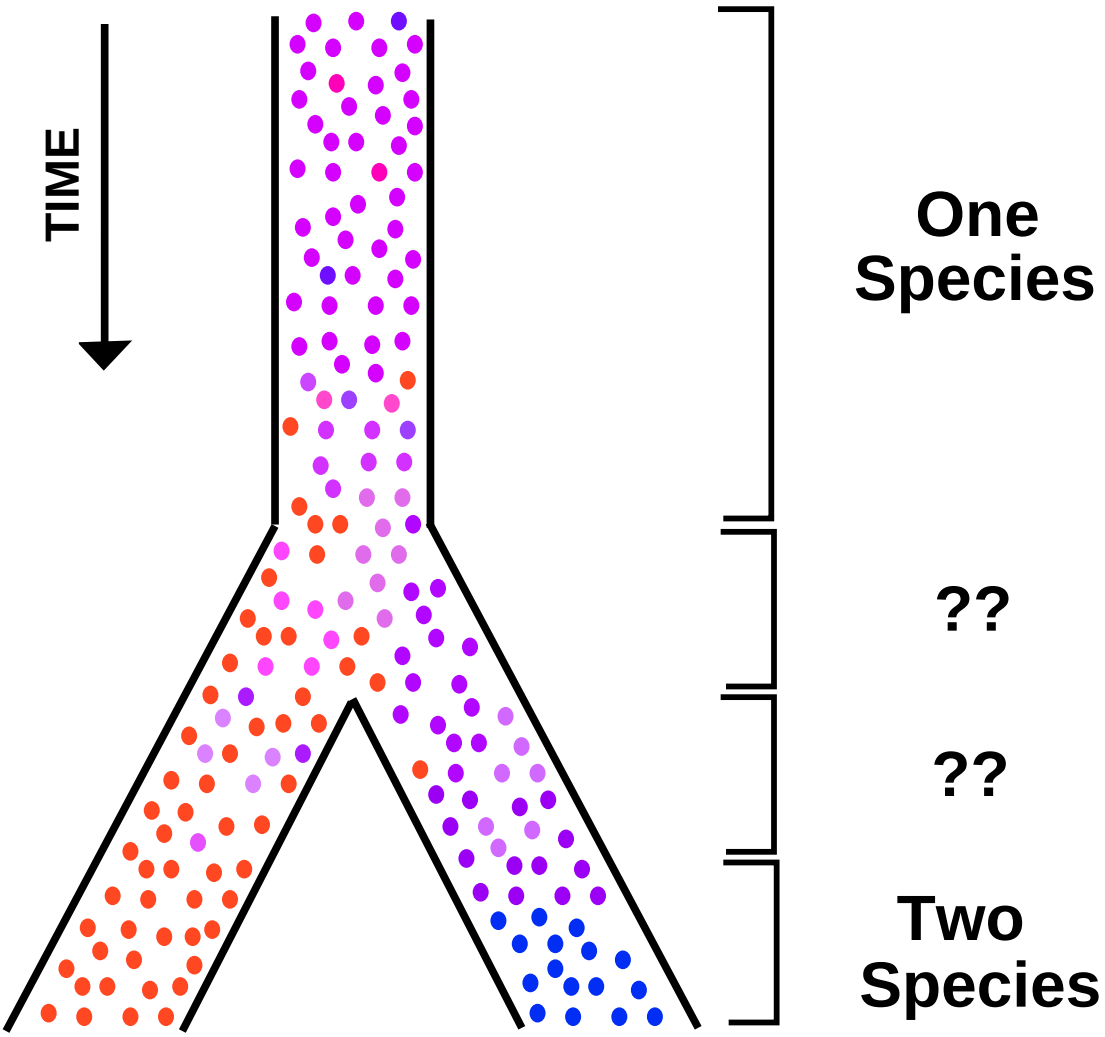
- Might many different genotypes converge upon the phenotype allowing survival in this adaptive zone?
- How can we know that a lineage will, in the future, evolve separately?

And these aren't even all of the species concepts!!

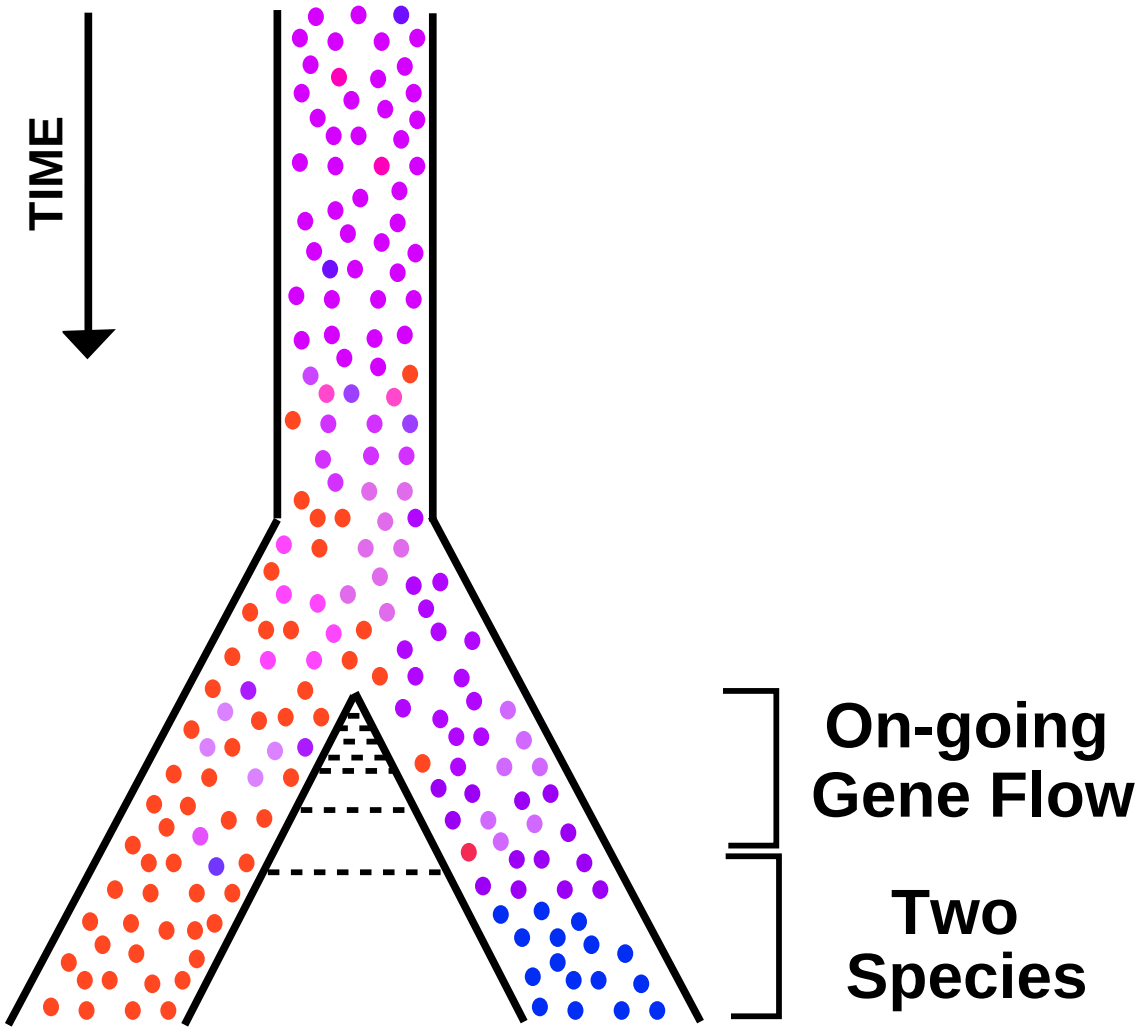
Phenetic species concept, recognition species concept, cohesion species concept, genotypic-cluster species concept, internodal species concept, phylogenetic species concept...

Why is defining "species" such a controversial headache?

For sexual organisms, the evolutionary transition from a single group of interbreeding individuals to two distinct groups of interbreeding individuals will almost always take a substantial period of time, during which it may be difficult determine whether one or two species exist.



Furthermore, hybridization often occurs among recently diverged (and even not so recently diverged) species.

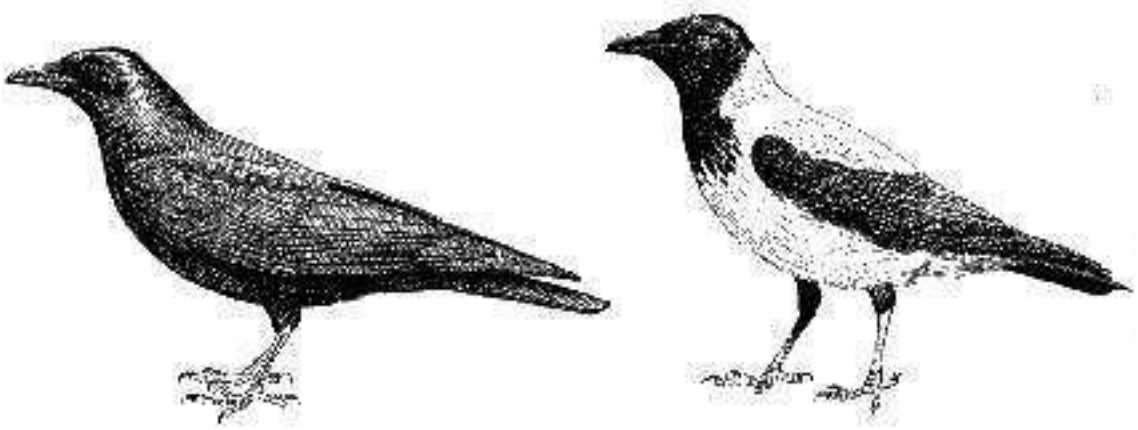


This gene flow may prevent further differentiation of the groups, but it may just slow down the accumulation of differences.

Hybridization and the possibility of limited gene flow is common among closely related species:

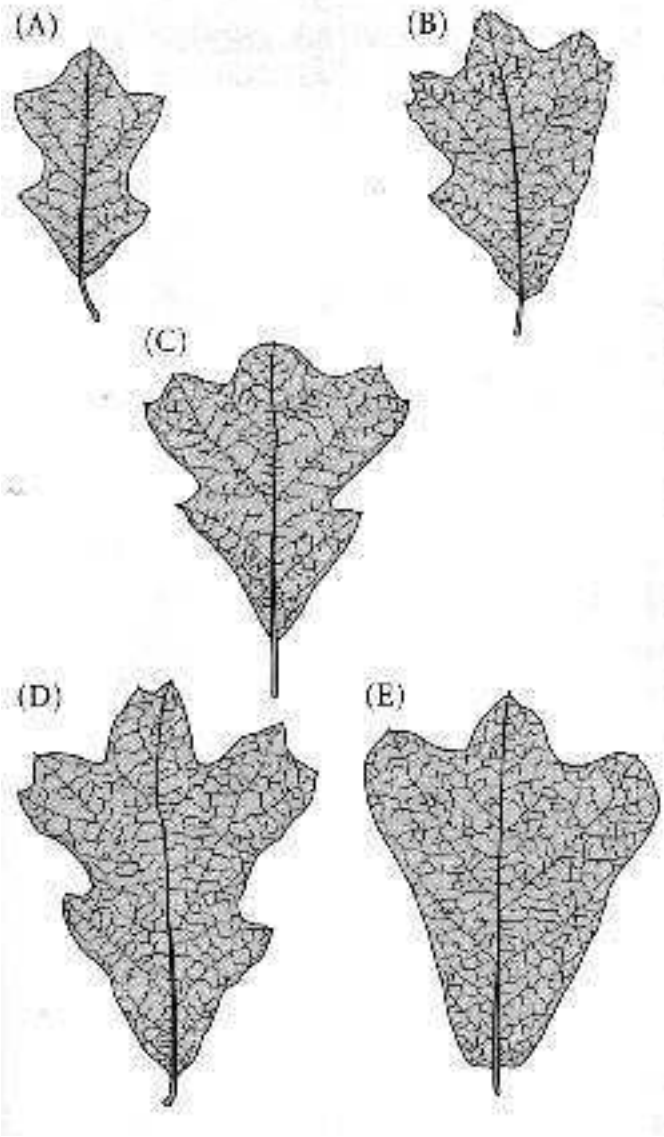
1/10000 wild-caught, pregnant females of the species *Drosophila pseudoobscura* and *Drosophila persimilis* had mated with the opposite species.

5/2000 fish in a collection of *Catostomus catostomus* and *C. commersoni* (sucker fish) were hybrids.



Carrion crows and hooded crows hybridize along a narrow hybrid zone in central Europe.

Oaks frequently hybridize with one another, despite clear and persistent phenotypic differences between the species.

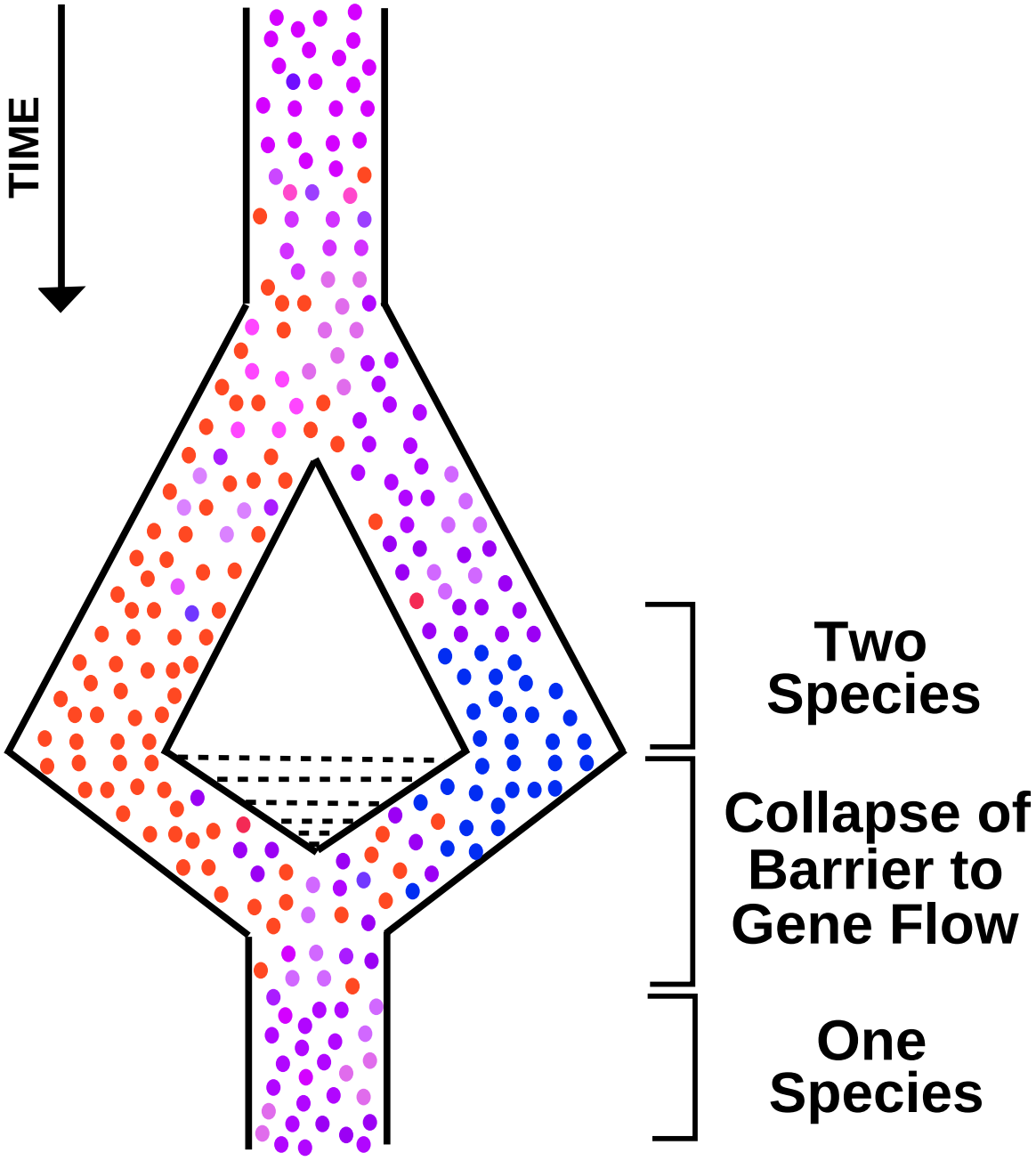


(A): Bear oak, (E) Blackjack oak, (B)-(D) hybrids found in several locations along the Atlantic coast of the US.

In any particular case, it is difficult to know whether current phenotypic differences will be maintained in the face of on-going hybridization or may disappear.

This points to the fundamental problem in identifying species: we do not have a crystal ball and cannot always correctly predict the evolutionary fate of a group of organisms.

It is difficult or perhaps impossible to know whether the factors maintaining differentiation between two closely related groups will continue to act or whether they might, in the future, collapse.



Murky Species Definitions

Example of collapsing species boundaries:
Study of cichlid fish diversity by Seehausen et al (1997).

Cichlid fishes have undergone dramatic speciation in African rift lakes.

Mylochromis lateristriga male, photo © by M. K. Oliver
http://www.connix.com/~mko/mw08024.htm

http://www.geocities.com/Heartland/Plains/8115/afcichlids.html
http://www.geocities.com/Heartland/Plains/8115/

At least 500 species of cichlids evolved in Lake Victoria (the largest of these lakes), utilizing "almost all resources available to freshwater fishes in general, despite having evolved in perhaps as little as 12,400 years and from a single ancestral species."

Many of these species have vanished in recent years, partly due to the introduction of the predatory Nile perch.

However, even those species not eaten by Nile perch have been going extinct.

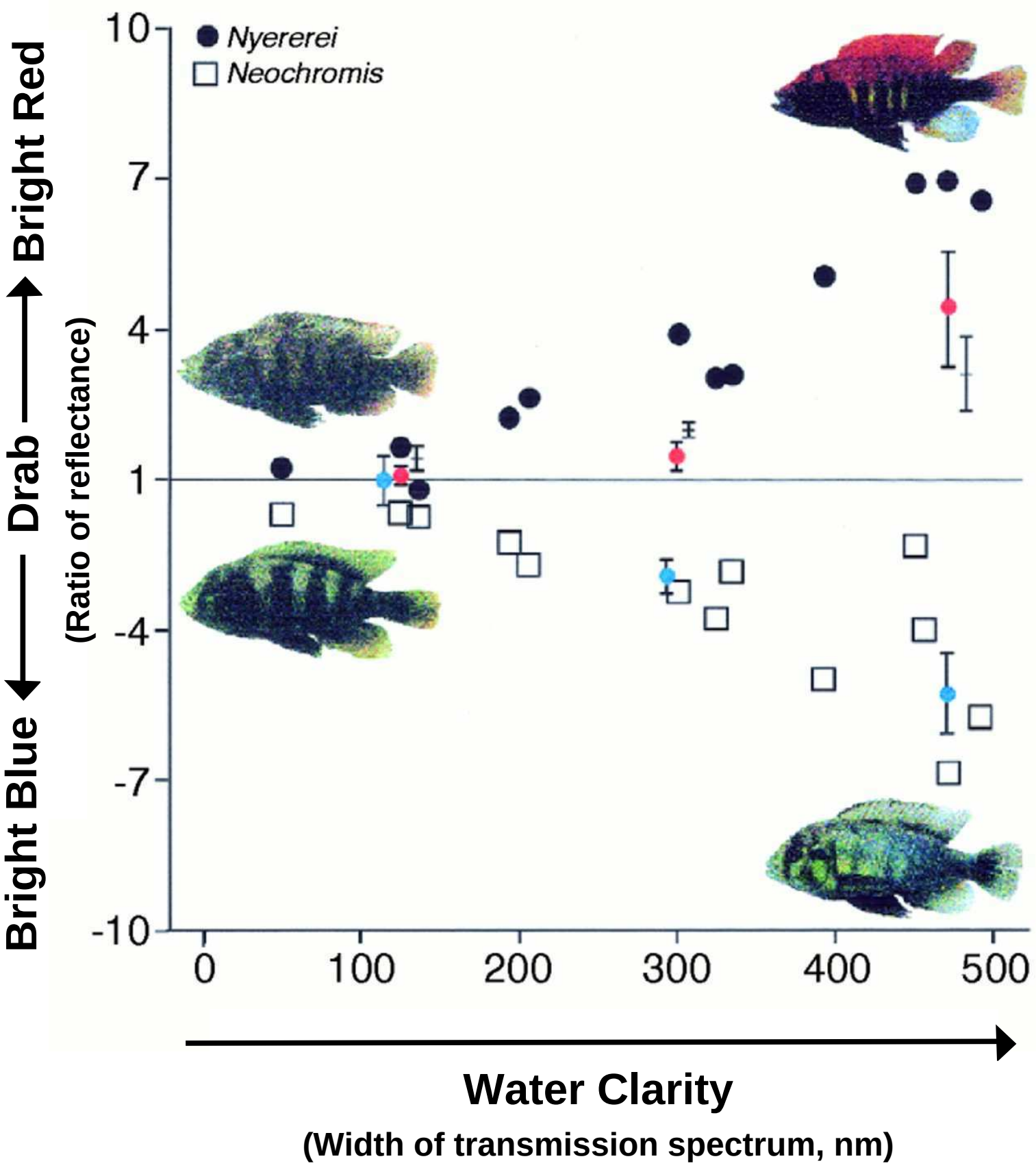
Cichlids species are isolated from one another by mate choice; cichlids can interbreed and produce fertile offspring but prefer to mate with members of their own species.

Since the 1920's, Lake Victoria has become eutrophic (rich in nutrients due to agricultural run-off and deforestation), leading to increased turbidity of the water.

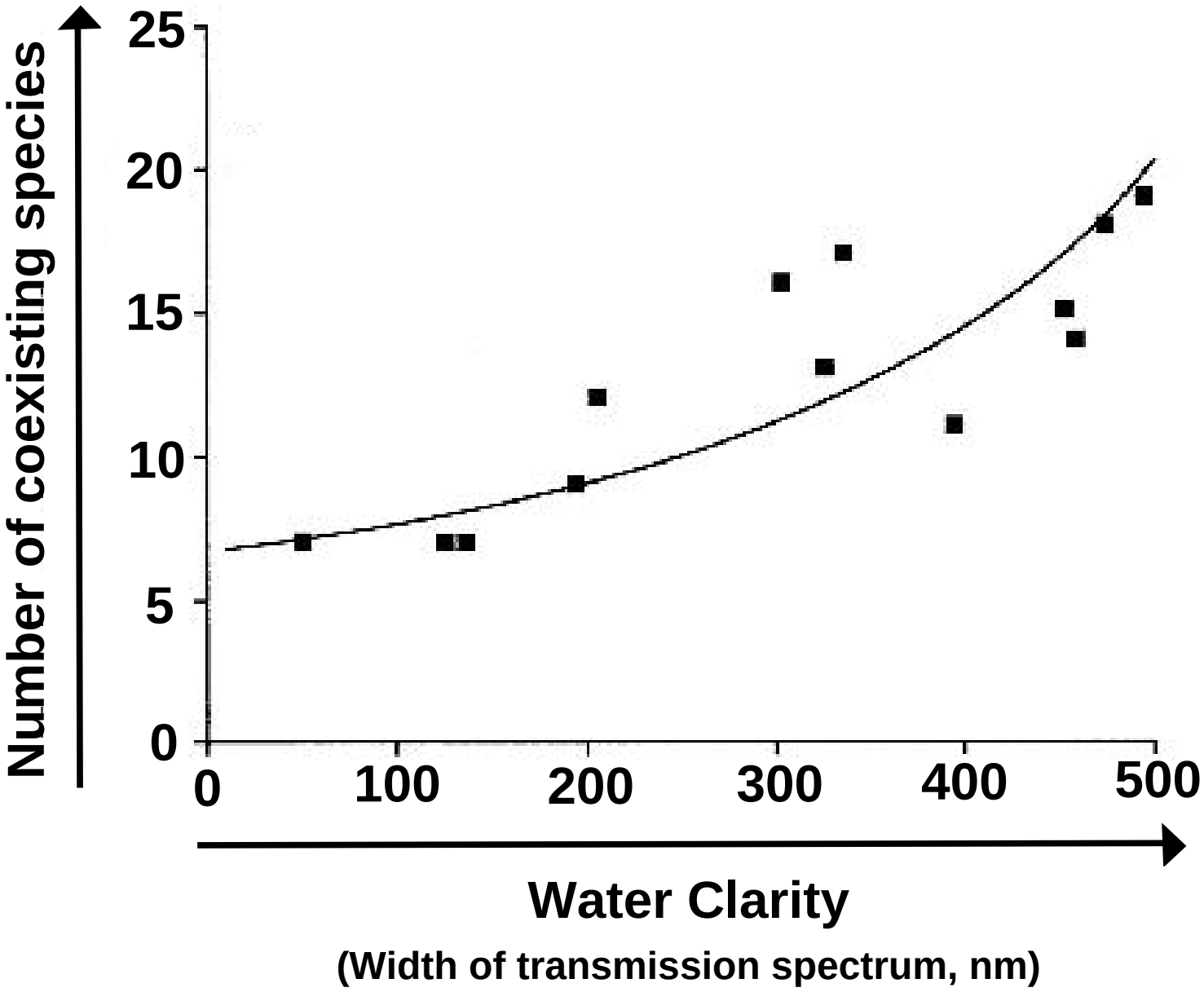
Seehausen et al (1997) observed:

1. In the lab, mating preferences for individuals of the same species disappeared under monochromatic light.

2. Along a transect in Lake Victoria, an increased brightness in red (*Haplochromis nyererei*) and in blue (*Neochromis* "velvet black"/"blue scraper") species in clearer waters.



3. Along a transect in Lake Victoria, an increased number of co-existing species in clearer waters.



➡ Some cichlid species may have been reproductively isolated from one another before a change in environment (eutrophication of Lake Victoria due to human activities) that led to a breakdown in the ability to distinguish and prefer mates of the same species.

➡ Loss of species diversity due to breakdown of reproductive isolation rather than extinction.

A clear definition of species is biologically and legally important.

Many species concepts exist, each with its strengths and limitations.

In classifying a group of organisms, gray areas necessarily exist during the early evolutionary diversification of a pair of species.

Although the different species concepts may be equally able to recognise clear-cut species, each makes different decisions in these gray areas.

SOURCES:

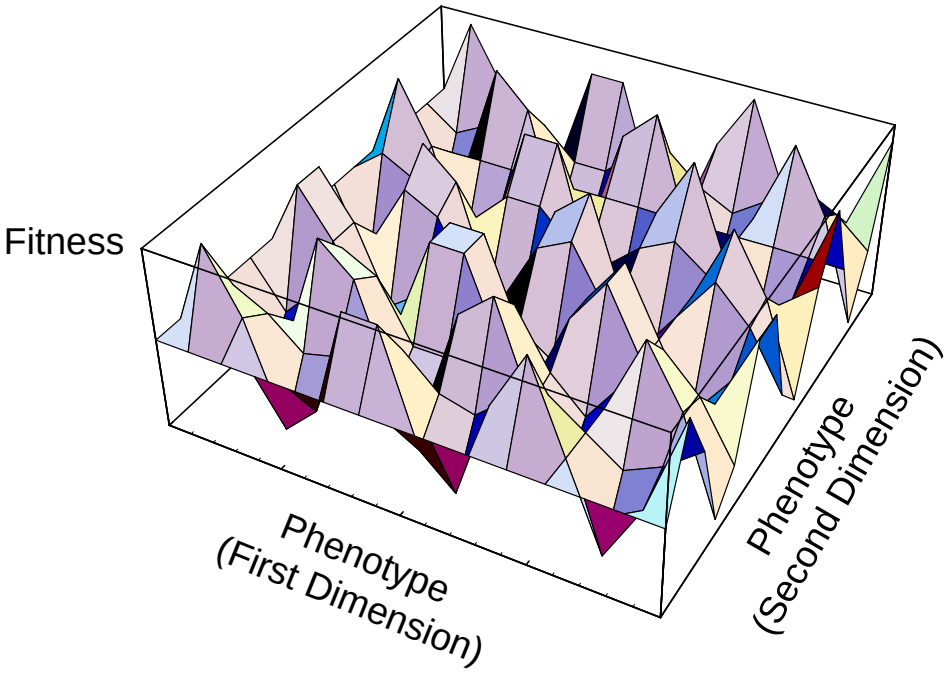
- Species concepts and figures: Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Cichlid story and figures: O. Seehausen, J. J. M. van Alphen, F. Witte (1997) Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. Science 277: 1808 - 1811

The Origin of Species

"...the living world is not a formless mass of randomly combining genes and traits, but a great array...of gene combinations, which are clustered on a large but finite number of adaptive peaks."

Theodosius Dobzhansky (1951)

Adaptive Landscape



Why do living organisms cluster into discrete species?

What are the processes at work that lead to the origin of discrete species?

(The question that Darwin (1859) called "that mystery of mysteries".)

In this lecture, we will examine a number of different modes by which speciation may happen in nature.

These have been classified by Mayr (1963) as:

1. Gradual speciation
   - 1A. Allopatric speciation (with geographic isolation)
     - ¤ Vicariant speciation (following the appearance of a barrier)
     - ¤ Peripatric speciation (by evolution in an isolated colony)
   - 1B. Parapatric speciation (with contiguous but non-overlapping distributions)
   - 1C. Sympatric speciation (with overlapping distributions)
2. Instantaneous speciation
   - 2A. Single genetic mutations
   - 2B. Cytological changes
     - ¤ Chromosome rearrangements
     - ¤ Polyploidy
3. Hybridization (producing new reproductively isolated species)

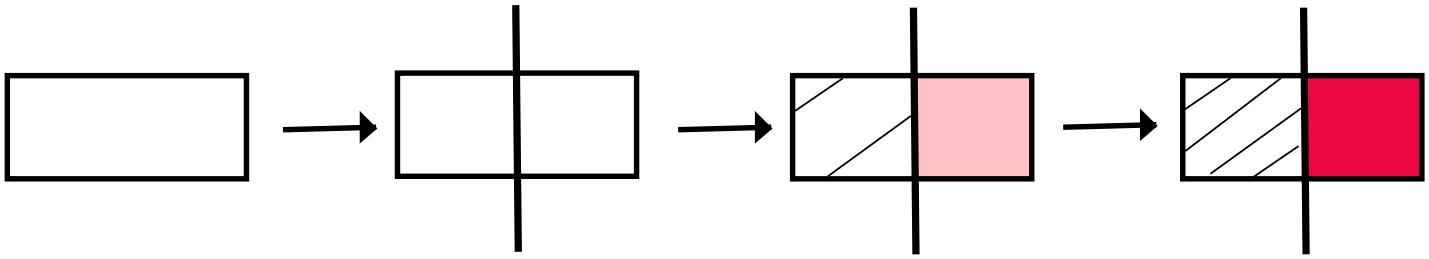
Our primary focus will be on how reproductive isolation can evolve under the gradual modes of speciation.

1A. Allopatric speciation

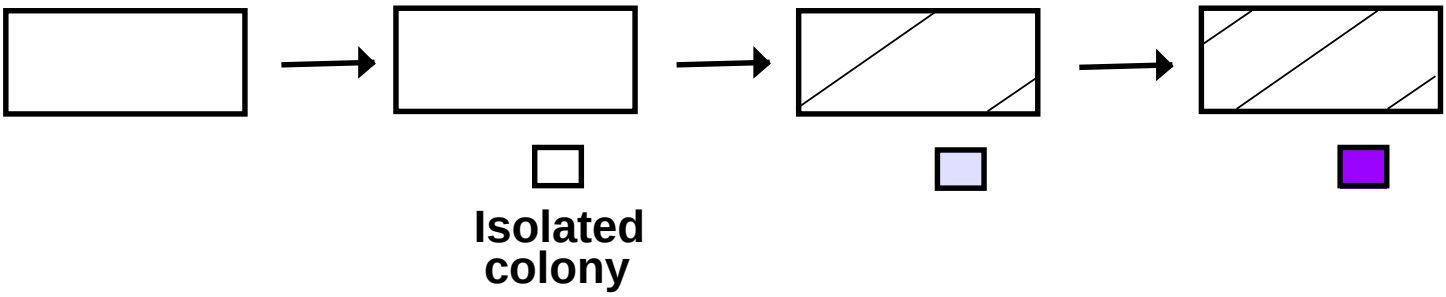
"Allopatric speciation is the evolution of genetic reproductive barriers between populations that are geographically separated by an extrinsic, physical barrier such as topography, water (or land) or unfavorable habitat."

Futuyma (1998) p. 482

Vicariant speciation (Allopatric)

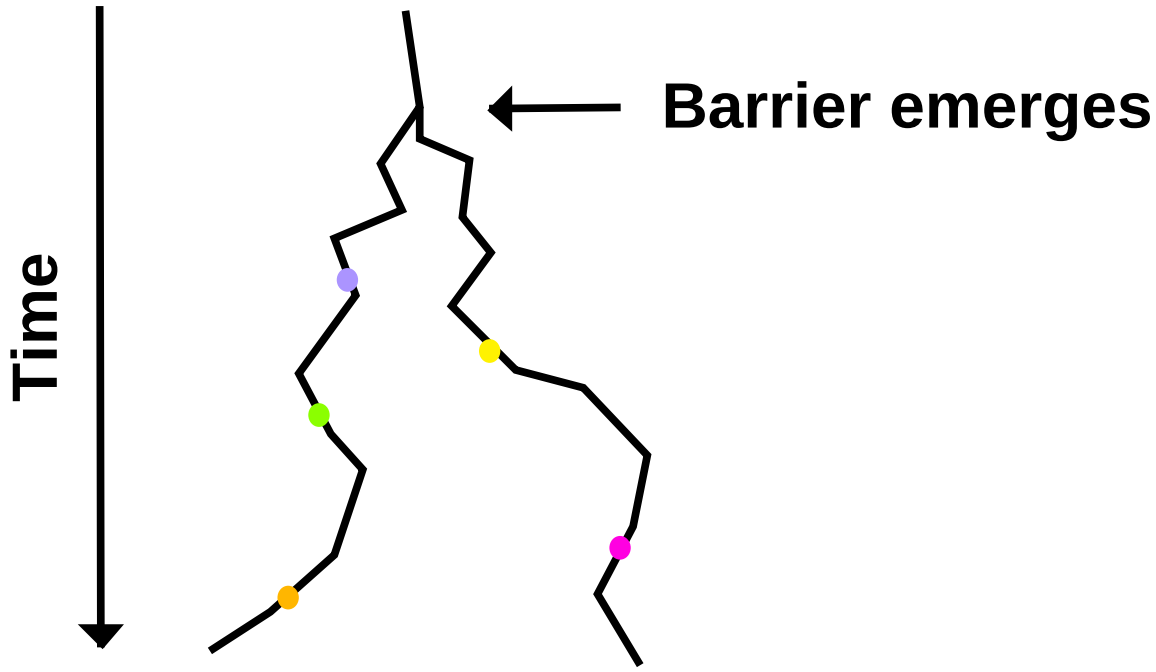


Peripatric speciation (Allopatric)



The extrinsic barrier halts gene flow between the isolated populations.

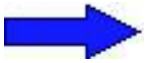
Without gene flow, each population follows an independent evolutionary trajectory, accumulating genetic changes through drift and/or selection.



For these changes to lead to speciation, the genetic changes in one population must be incompatible with the genetic changes in the other population.

Allen Orr (1995) noted that as populations accumulate changes over time (R), the number of possible ways in which an incompatibility can occur rises at least as fast as R^2: the more changes occur in one population, the more possible ways in which a genetic incompatibility could occur with the second population.

Once genetic incompatibilities have arisen between two separately evolving populations, the populations are no longer able to cross and produce fertile offspring.

 Speciation.

The genetic changes that accumulate in the two populations may be caused by natural selection, sexual selection, or by random genetic drift.

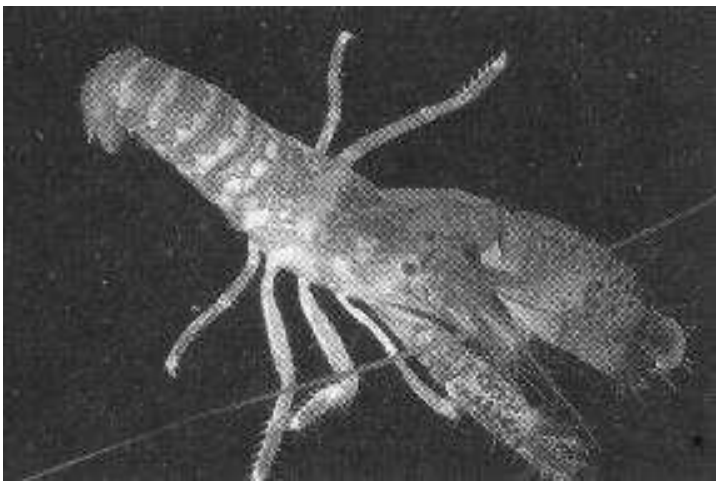
Which forces would you expect to be faster?

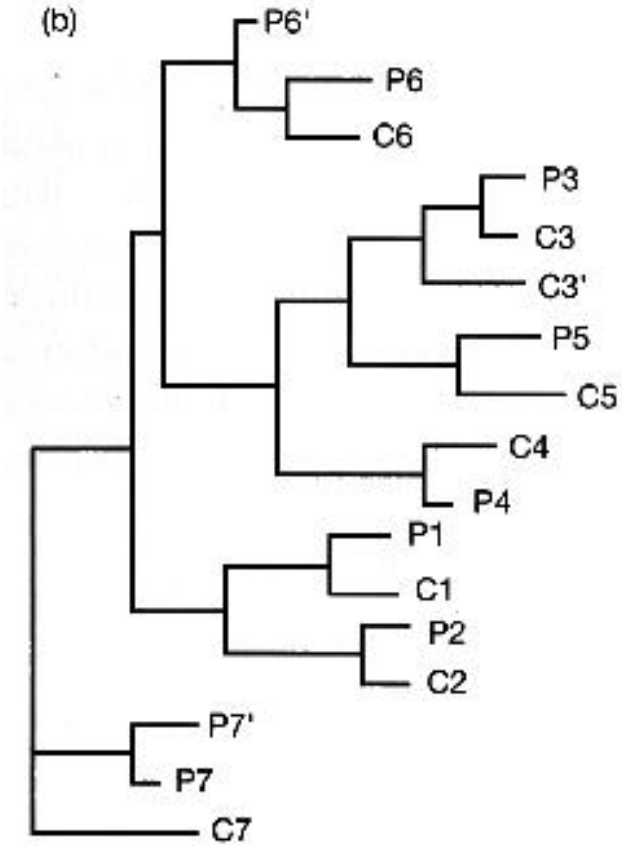
Which forces would you expect to be relatively more important in peripatric speciation?

Example: Isthmus of Panama

About 3 million years ago, North and South America merged, forming a land bridge (the Isthmus of Panama) and isolating the marine communities of the Carribean and Pacific.

Knowlton et al (1993) studied snapping shrimp from seven pairs of similar species on either side of the isthmus.





A phylogeny of the seven pairs indicated that each of the species was most closely related to a similar looking species from the other coast.

Mating experiments in the lab indicated that, in each case, the sister species from the Carribean and from the Pacific failed to produce viable offspring.

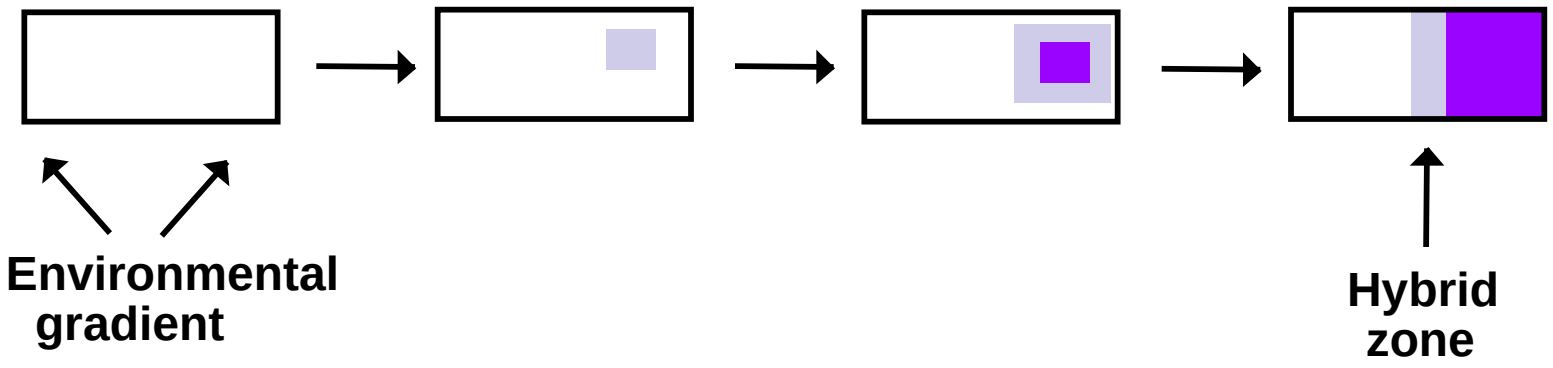
The formation of the land bridge between North and South America led to the isolation of populations on either side. With gene flow cut off, the populations diverged, leading to new species in 7/7 cases.

1B. Parapatric speciation

"Parapatric speciation is the evolution of reproductive isolation between populations that are continuously distributed in space, so that there is substantial movement of individuals, and hence gene flow, between them."

Futuyma (1998) p. 498

Parapatric speciation



Without an extrinsic barrier to gene flow, genetic differences that arise within a population may be swamped by genetic exchange.

If, however, the population is distributed over space and:

- little gene flow occurs between distant locations and
- there is an environmental gradient favoring different genetic combinations in different places,

selection may be strong enough to maintain genetic differences, creating a *genetic cline*.

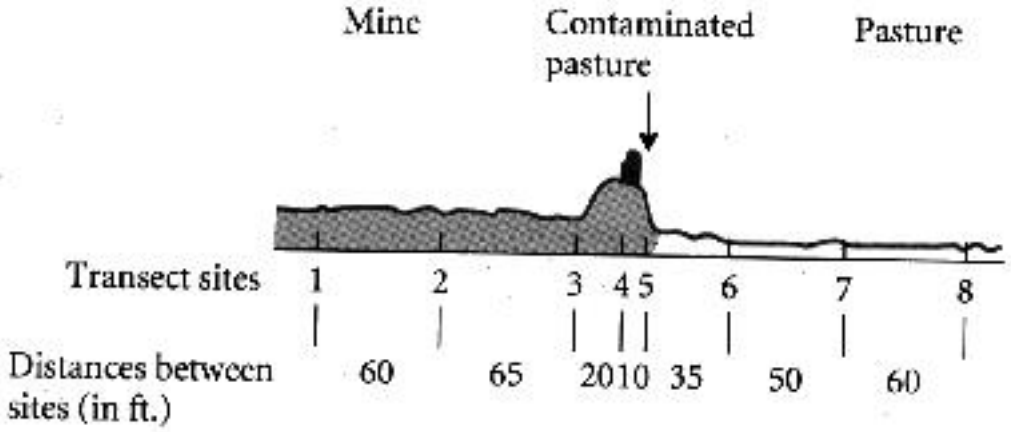
If the habitat changes fairly abruptly, there may be a sharp border between different types: *a hybrid zone*

Over time, the hybrid zone may shrink or remain stable.

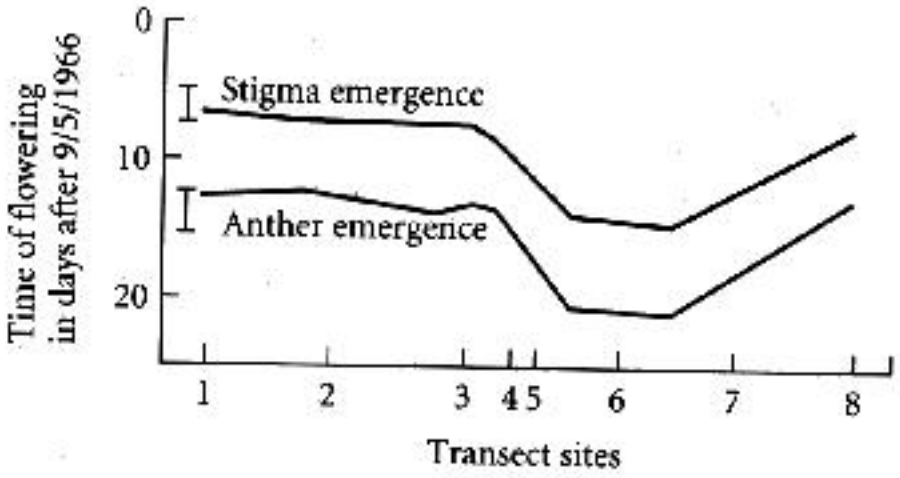
Example: Heavy metal tolerance

Antonovics and co-workers studied heavy metal tolerance in grasses growing near mines on land contaminated with lead and zinc.

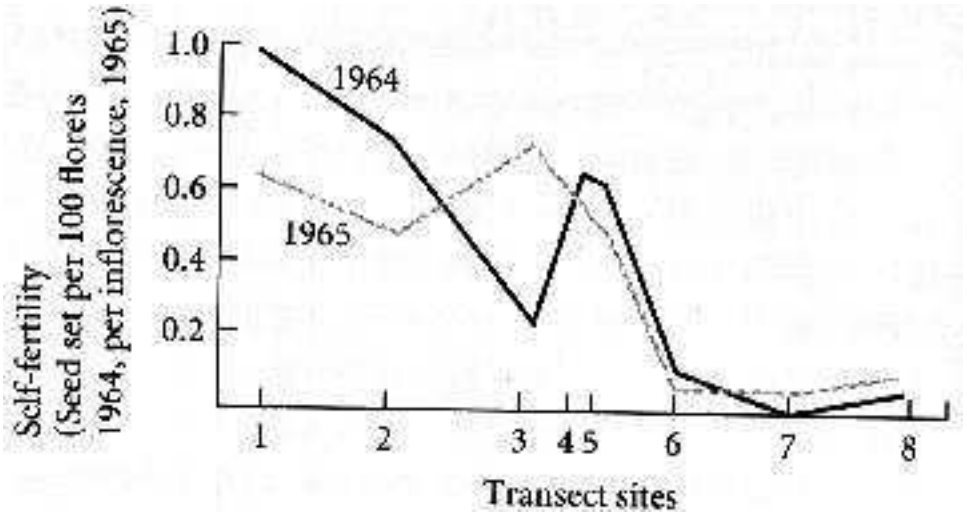
Plants from the grass species *Anthoxanthum odoratum* were tolerant near the mine but remained intolerant at distant sites.



Heavy metal tolerance was not the only difference observed; flowering time also differed along the transect.



Selfing rates were also higher near the mine.



The divergence in flowering time and increased selfing rates have reduced gene flow between tolerant and non-tolerant grasses and increased the reproductive isolation of grasses near the mine.

In this case, heavy metal tolerance has evolved recently, during which time the species has always had a continuous distribution.

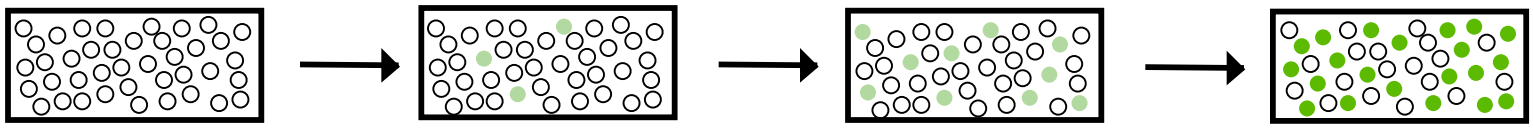
Therefore, parapatric speciation processes can be inferred. In most other cases, however, it is difficult to know whether a species that is currently parapatric may have been previously allopatric.

1C. Sympatric speciation

"Sympatric speciation is the evolution of reproductive isolation within a randomly mating population."

Futuyma (1998) p. 498

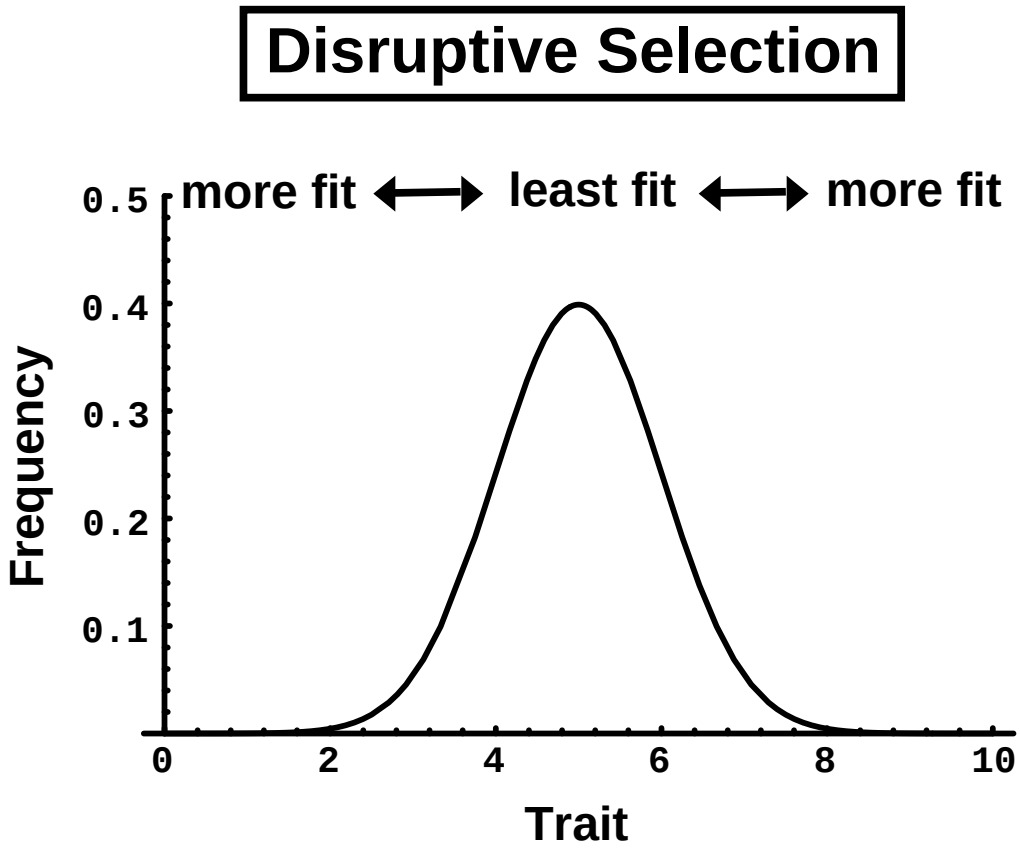
Sympatric speciation



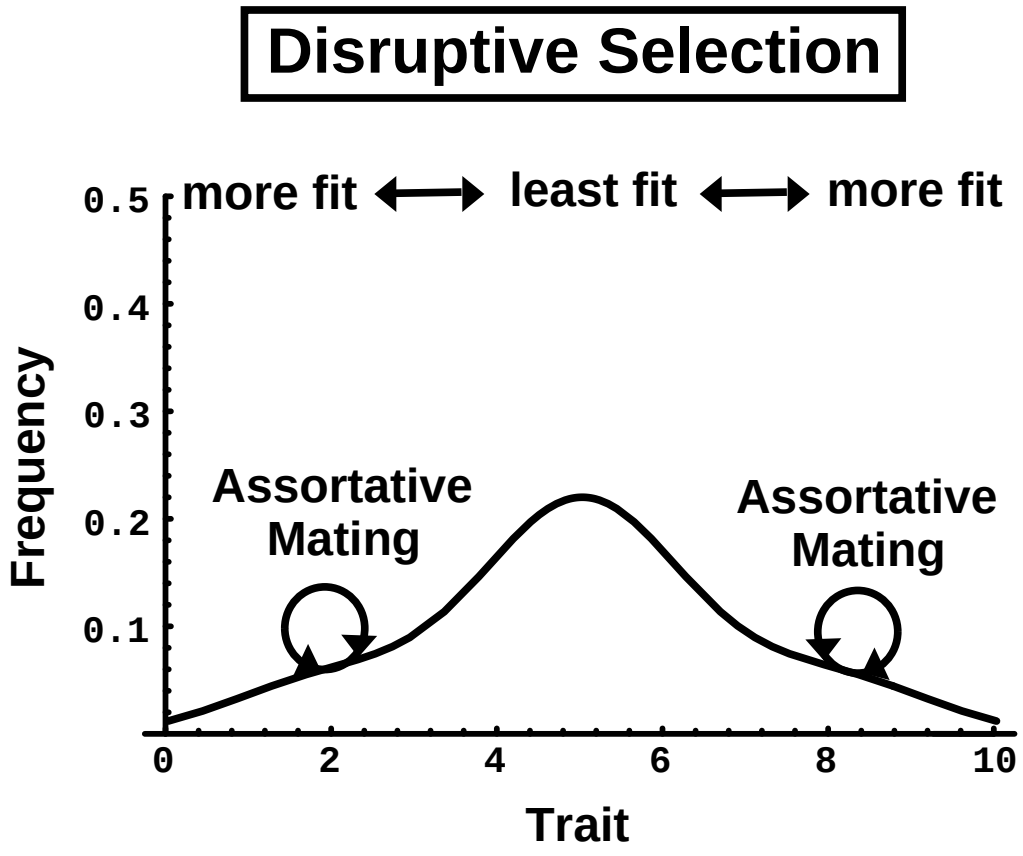
Sympatric speciation is the most controversial of all the modes of speciation, since divergence must occur in the face of high levels of genetic exchange.

Imagine a single species utilizing a resource that could support a wider variety of individuals.

If the extreme phenotypes in a population have the highest fitness (disruptive selection), the population will not split apart if mating is random.



If, however, genes arise that cause extreme individuals to *assortatively mate* (= mate with similar types), then more extreme offspring can be produced:



The problem with this idea is that it is difficult for assortative mating to evolve in a way that is coupled with the trait subject to disruptive selection.

One way for this coupling to be achieved is if there is tight linkage (low recombination) between the genes causing the trait and the genes causing assortative mating.

A second, more likely way is if the traits themselves lead to assortative mating.

For example, if the traits involve utilizing different resources and mating tends to happen around resources, then individuals with extreme traits will naturally mate with similar individuals.

Over time, if sufficiently strong assortative mating evolves, sympatric speciation will be the result.

Example: Apples and hawthorns

The classic case of sympatric speciation is that of the apple maggot fly, *Rhagoletis pomonella*.

Hawthorns were the original host plant of *Rhagoletis pomonella* until about 150 years ago, when the fly was observed on cultivated apple trees (introduced to the Americas).

The apple maggot fly is now widespread in the northeastern US and causes millions of dollars of damage each year to apple crops.

Rhagoletis courts, mates, and lays eggs on its preferred host plant.

Individuals that changed their preferred food source to apple trees (= a host shift), consequently tend to mate with other individuals who prefer apples.

Since the host shift, flies that prefer hawthorns and apples have diverged genetically.

Feder et al (1988, 1990) found statistically significant differences in allele frequencies at six loci, differences that allow individuals from the two populations to be identified.

Changes in the timing of mating have also occurred: mating on apples occurs ~ 3 weeks earlier.

Overall, matings between hawthorn and apple flies has been reduced to only ~6% and the two are now considered incipient species.

The exact mechanisms by which reproductive isolation is accomplished vary but fall into two main classes:

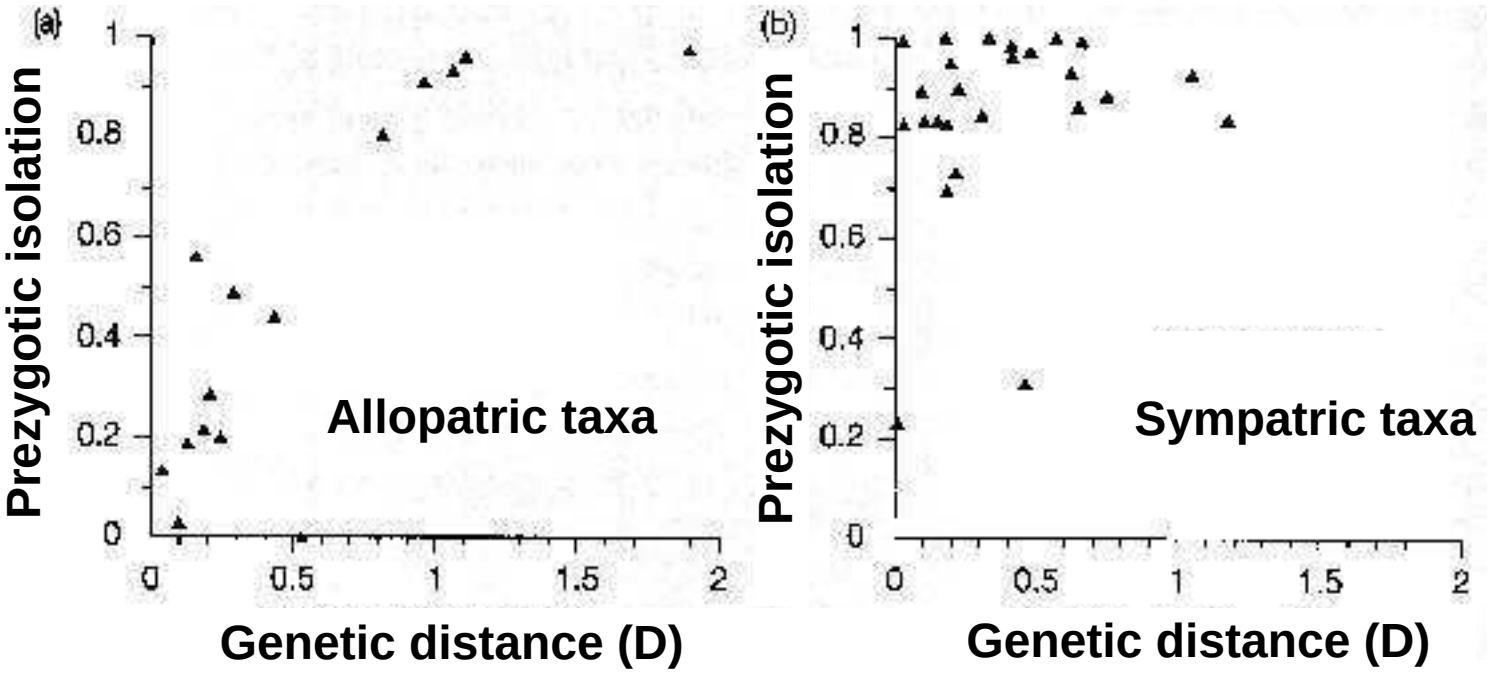
1. *Premating* (or prezygotic) isolating mechanisms: Mating and fertilization are prevented.

Premating isolation may be accomplished by: changes in habitat preferences, changes in timing of reproduction, physical incompatibilities, changes in mating preferences or mating behaviors.

2. *Postmating* (or zygotic) isolating mechanisms: Mating occurs but hybrid offspring are inviable, infertile, or produce inviable/infertile offspring.

Would you expect premating isolation to play more of a role in allopatric or sympatric speciation?

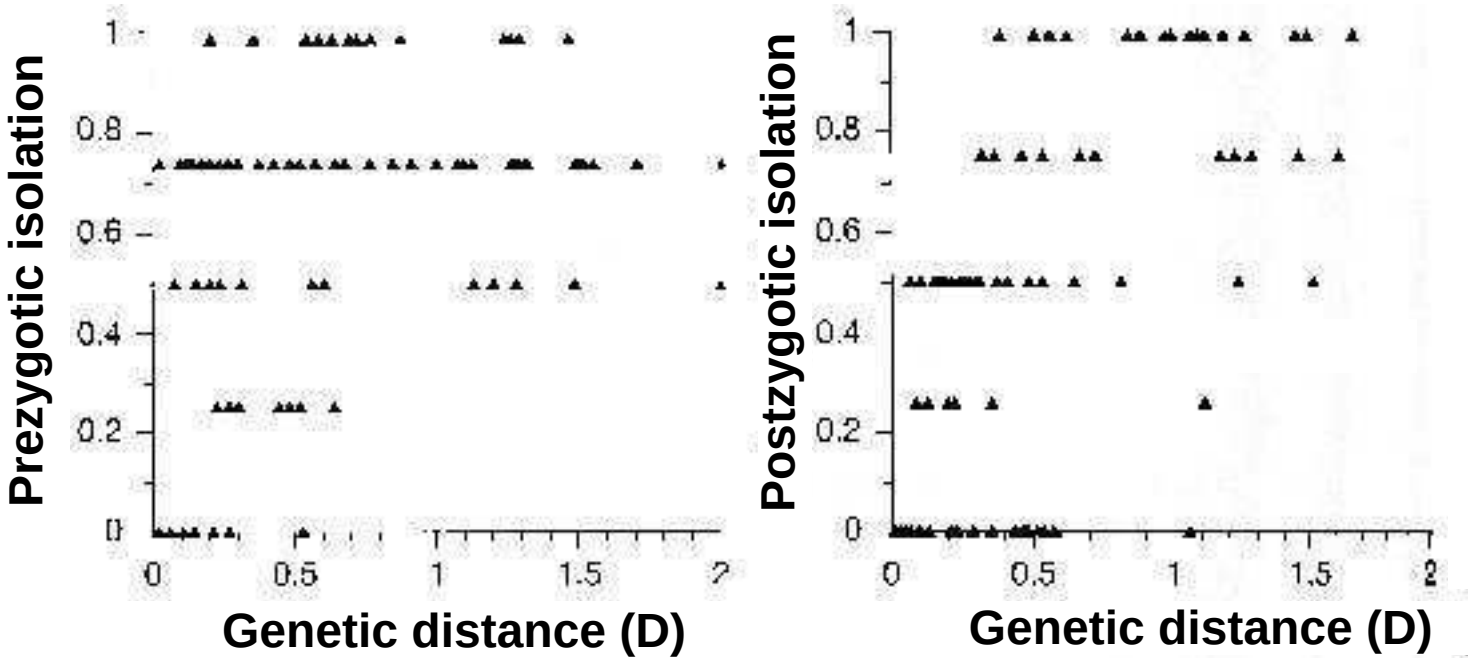
Coyne and Orr (1997) examined the causes of reproductive isolation in a number of sister species in *Drosophila*, finding:



➡ More rapid and complete prezygotic isolation between sympatric species pairs.

Would you expect premating or postmating isolation to evolve sooner between sympatric pairs?

... between allopatric pairs?



➡ Prezygotic isolation evolves earlier in the speciation process, primarily due to the rapid mating isolation that evolves between sympatric species pairs.

Although examples of each mode of speciation are known, it remains unclear how frequently each occur.

Many evolutionary biologists suspect that allopatric speciation is responsible for the vast majority of speciation events.

SOURCES:

- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Freeman and Herron (1998) Evolutionary Analysis. Prentice Hall, NJ.

Mating Preferences and Sexual Selection

Individuals rarely mate at random for a number of reasons:

- Dispersal may be limited
- Individuals may or may not be able to self
- Individuals may reproduce asexually
- Individuals may compete for mates
- Individuals may choose particular mates

Non-random mating has a number of important evolutionary consequences.

In this lecture, we will focus on the evolution of mate choice and sexual selection.

Sexual Selection

"Sexual selection depends on the success of certain individuals over others of the same sex, in relation to the propagation of the species; while natural selection depends on the success of both sexes, at all ages, in relation to the general conditions of life.

The sexual struggle is of two kinds: in the one it is between the individuals of the same sex, generally the males, in order to drive away or kill their rivals, the females remaining passive; while in the other, the struggle is likewise between the individuals of the same sex, in order to excite or charm those of the opposite sex, generally the females, which no longer remain passive, but select the more agreeable partners."

-- Darwin (1871)

The Descent of Man and Selection in Relation to Sex (p. 639)

Sexual Selection: Classification

Following Darwin, two main forms of sexual selection are recognized:

(1) Intrasexual selection
(competition within a sex for the opportunity to mate)

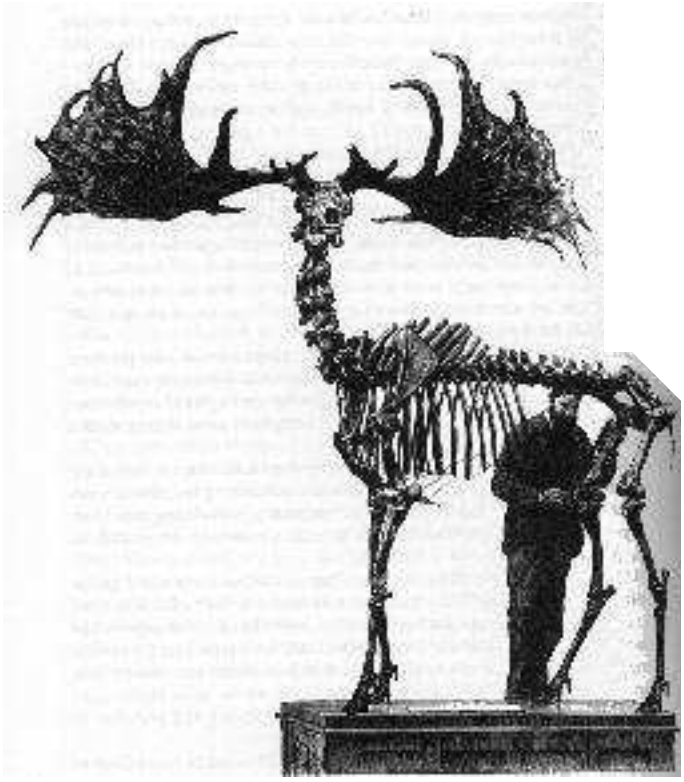
(2) Intersexual selection
(choosiness on the part of one sex for mates)

Although female-female competition and male choosiness are known, we'll focus on the more common patterns of male-male competition and female choice.

(Why might these be the more common patterns?)

(1) Intrasexual selection

Horns, antlers, tusks, spurs and other weapons provide some of the most extreme examples of sexual dimorphism.



Possible explanations:

(1) Weapons against predators

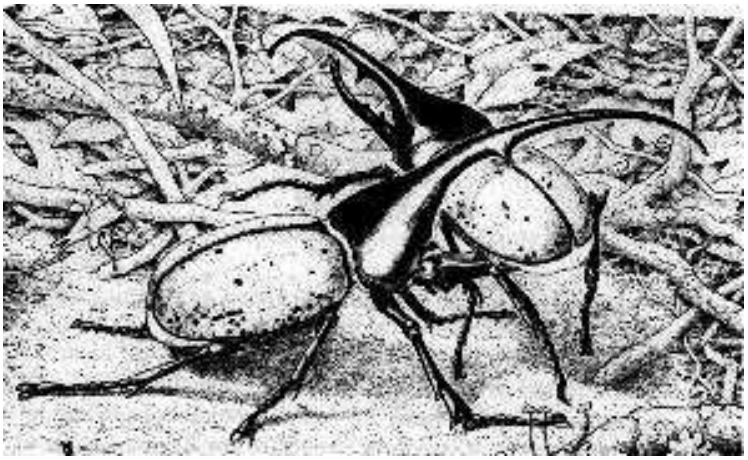
(2) Weapons against other males

(3) Indicators of male strength and fighting ability (male-male competition)

(4) Indicators of sexual vigor and quality (female choice)

Examples:

Barrette and Vandal (1990) studied sparring in caribou. Of 713 matches between males of different antler size, males with smaller antlers withdrew 90% of the time.



Eberhard (1979,1980) studied the use of horns in seventeen species of beetles, finding that they tend to be used either to pry a rival off his site or to lift and drop the rival to the ground.

(Topi)

"Among the explanations for sexually dimorphic horns, antlers, tusks, and spurs, the empirical support is strongest for the idea that they have evolved and are favored in males as weapons in contests over females."

-- Andersson (1994) *Sexual Selection* (p. 314)

(2) Intersexual selection

Some of the more profoundly beautiful traits seen in nature have evolved in response to female choice.

(Peacock)

(Sage Grouse)

(Greater Frigate)

Theories for the existence of female preferences:

(2A) Female choice and male traits co-evolve
(Fisher's *Runaway Process*)

(2B) Choosy females gain *direct benefits* from their mates

(2C) Female are choosy because of a *sensory bias*

(2A) Fisher's Runaway Process

PHASE 1: Female preferences initially evolve because they favor a trait in males that is also favored by natural selection. The offspring of choosy females are then more likely to carry the advantageous trait.

> "Whenever appreciable differences exist in a species..., there will be a tendency to select also those individuals of the opposite sex which most clearly discriminate the difference to be observed, and which most decidedly prefer the more advantageous type."
>
> -- R. A. Fisher (1930)

PHASE 2: Once female preferences exist, they can favor even more extreme traits in males. This can in turn favor the evolution of stronger female preferences, leading to a *runaway process*.

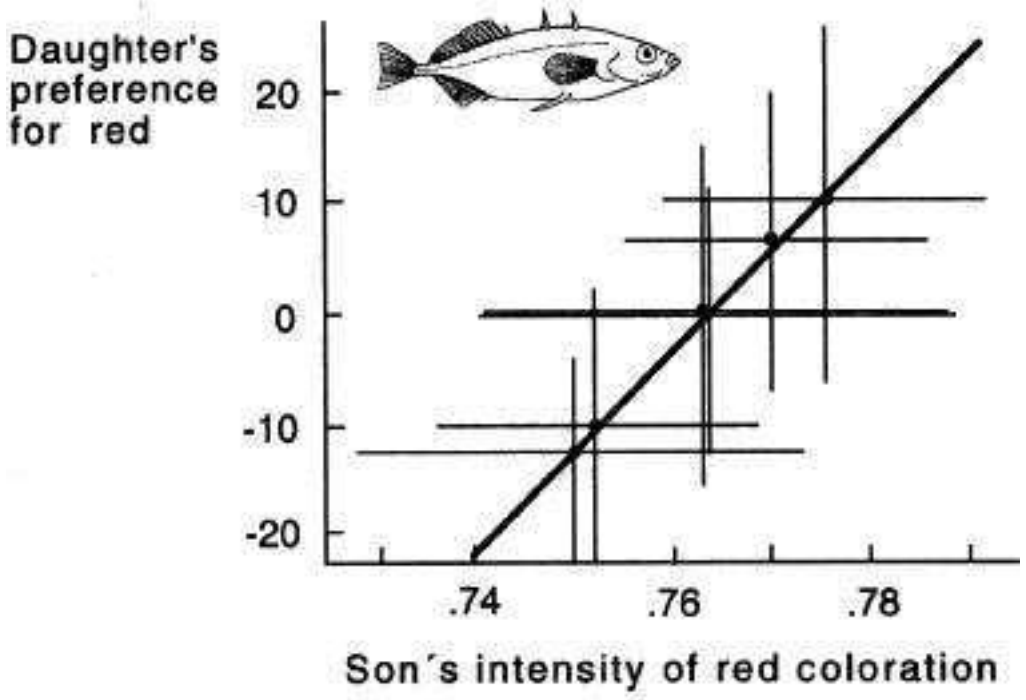
"...the further development of the plumage character will still proceed, by reason of the advantage gained in sexual selection, even after it has passed the point in development at which its advantage in Natural Selection has ceased."

-- R. A. Fisher (1930)

The runaway process will halt when genetic variation is exhausted or when the trait becomes so costly that natural selection balances sexual selection.

Example:

In a breeding experiment with the threespine stickleback, Bakker (1993) observed a genetic correlation between red coloration among sons and preferences for red coloration among daughters, as expected under the Fisherian process.



(2B) Direct Benefits to Choosiness

The Fisherian model of sexual selection is, however, ineffective in the face of costs to female choosiness, such as

- Time and energy in evaluating mates
- Risk of remaining unmated

Example: Engelhardt et al (1982) found that choosy female seaweed flies had reduced fertility.

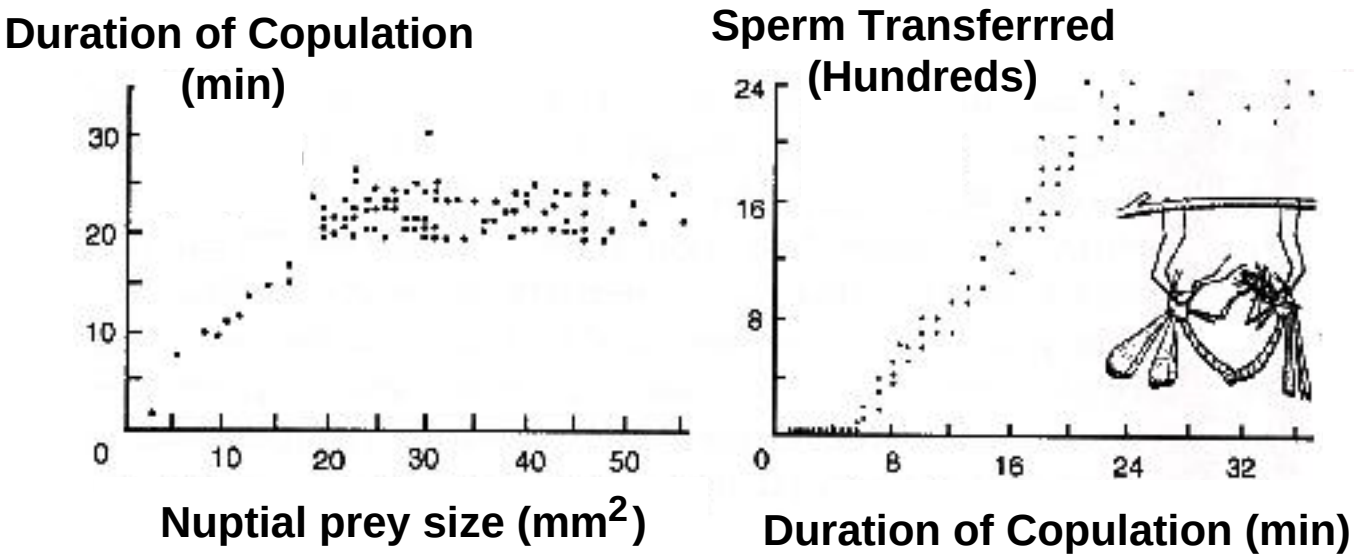
Occasionally, female preferences may be directly beneficial, such that natural selection itself favors the evolution of preferences.

Possible benefits of being choosy:

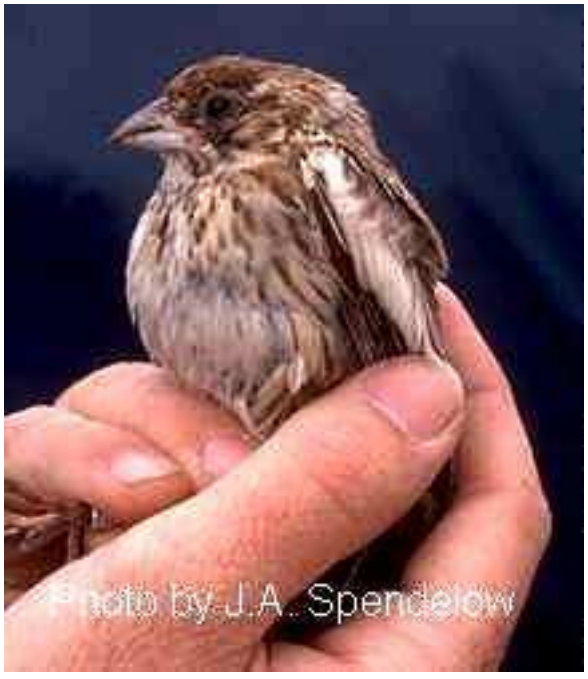
- Lower risk of mating with the wrong species
- Mate may provide paternal care
- Mate may provide food (eg nuptial package)
- Mate may be more fecund
- May avoid diseases/parasite transmission
- Offspring may be more fit (*Good genes* hypothesis)

Examples:

Thornhill (1983) showed that female hangingflies lay more eggs with males that provide larger nuptial food gifts.



Pleszczynska (1978) showed that male lark buntings with more nest cover in their territory attracted more females and these females had higher breeding success.



(2C) Sensory Bias

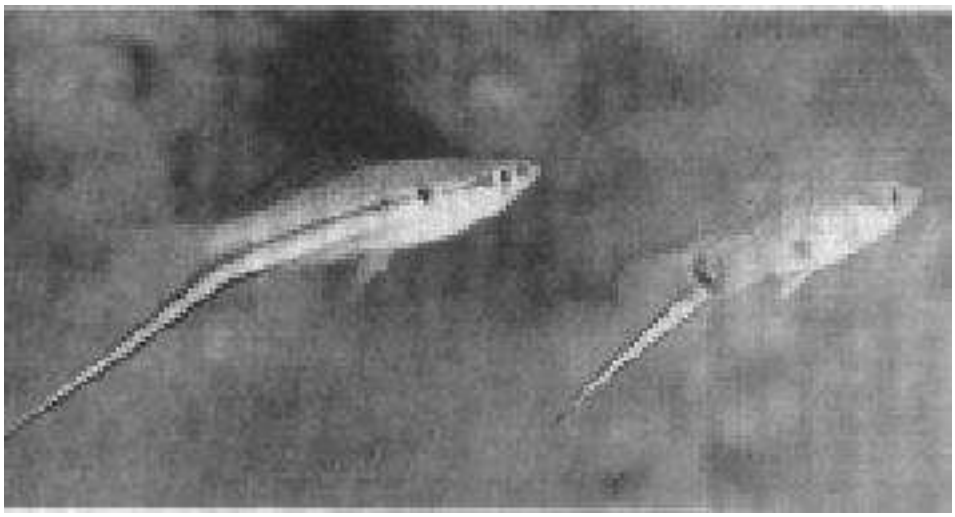
Female preferences may evolve as a correlated response to the evolution of a favored male trait or due to direct benefits of choosiness.

An alternative possibility is that female preferences are simply a side-effect (a "pleiotropic" effect) of how sensory systems have evolved.

Examples:

Searcy (1992) found that female common grackles preferred males singing an artificial repertoire with four song types even though males in this species sing only one song type.

Basolo (1990) showed that female platyfish preferred males with swords artificially attached, arguing that evolution in the sister taxa (swordtails) was shaped by this pre-existing bias for swords.



(On the left is a tropical green swordtail, on the right a southern male platyfish with a sword artificially attached.)

Mating Preferences and Sexual Selection

Regardless of how mating preferences have evolved, their presence in a population has profound influences on the evolution of morphology, behavior, and communication.

Sexual selection has undoubtedly contributed to the evolution of some of the more spectacular traits seen in the natural world (e.g. the radiant feathers of peacocks, the flashing lights of fireflies, the nightly song of crickets).

In addition, sexual selection can lead to rapid reproductive isolation of populations, thereby contributing to speciation.

For example, sexual selection has played a crucial role in the explosive radiation of Drosophila species on Hawaii (800-900 species) and of cichlids in the African Rift Lakes.

SOURCES:

Malte Andersson's (1994) book entitled Sexual Selection is a fantastic source of information about the theoretical and empirical support for various hypotheses about mate choice. Most of the examples used in this lecture are drawn from his book.

Additional pictures come from Sexual Selection (1989) by Gould and Gould.

Other web sites of interest are:

- See a cartoon about Enquist and Arak's neural network model of sensory bias.
- See a bird of paradise.
- The importance of being flashy.
- Costly signals and the handicap principle.

The Evolution of Sex

The unit of reproduction is the individual in the case of asexually reproducing organisms but the couple in sexually reproducing organisms.

Unless the sexual couple produces twice as many offspring as the asexual individual, there will be a **cost of sex** (fewer offspring per sexual parent than per asexual parent).

In other words, the "sexual female propagates her genome, or any given element of her genome, only half as efficiently as the asexual female" (Bell 1982).

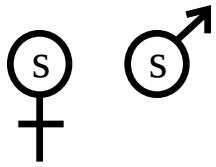
→ The Paradox of sexuality

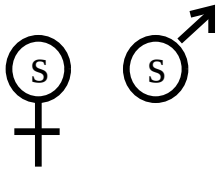
Other costs of sex:

- Cost of finding a mate
- Risk of disease transmission in mating
- Risk of not finding a mate
- Loss of heterozygosity

If every female produces exactly two offspring...

SEXUAL POPULATION



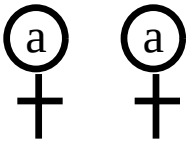


2 offspring, but only one daughter on average

Daughter produces two grandoffspring, only one of which is female, on average.

Expected number of sexual individuals remains constant.

ASEXUAL POPULATION



2 offspring, both asexual females

Both daughters produce two asexual female offspring, for a total of four female grandoffspring.

Expected number of asexual individuals doubles each generation.

The Evolution of Sex

What might offset the cost of sex?

- If males contribute to raising offspring such that couples produce more offspring than can a single female

What might pay the cost of sex?

In some fashion or other, all explanations of sex rely on the fact that sex can generate greater variability through recombination and segregation.

Fisher - Muller hypothesis

Sex allows advantageous alleles that arise in different individuals to be combined together into the same individual.

Related: Sex generates novel combinations of alleles - generating a wide variety of genotypes, some of which may have greater fitness.

Also known as the **The Vicar of Bray Hypothesis** after:

> "an English cleric noted for an ability to change his religion whenever a new monarch ascended the throne [to emphasize] that there may be great advantages of easily and gracefully adapting to changed circumstances."
>
> -- Bell 1986.

Red Queen Hypothesis

Since the world is constantly changing, organisms must, like the Red Queen in Alice in Wonderland, constantly adapt just to stay in place.

If sexual organisms adapt faster, they may be able to keep up with rapid changes in the environment, while asexual populations fall behind.

→ Asexuals may be more prone to extinction.

Parasites are an especially important part of the environment. Organisms must evolve defenses etc continuously and rapidly against viruses, bacteria, insect parasites, etc.

(When focused on the evolutionary arms race" between hosts and parasites, the hypothesis is known as the **Host-Parasite Hypothesis**.)

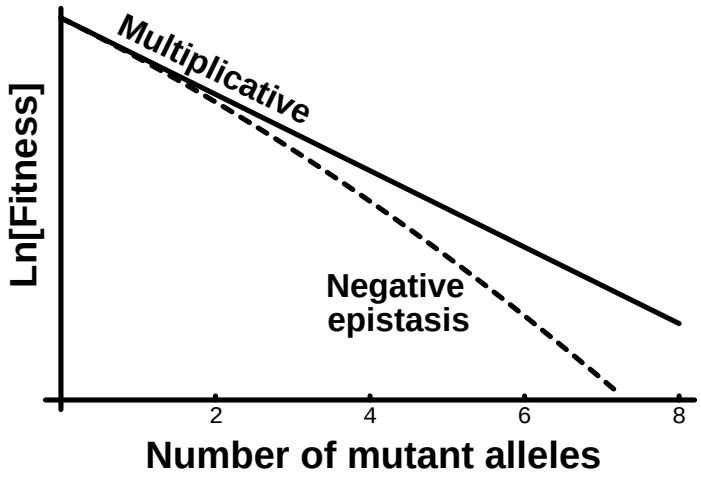
Problems with the Fisher-Muller and the Red Queen Hypotheses:

Sex and recombination can break apart new beneficial (e.g. resistant) genotypes once they arise, whereas asexuals can reproduce their genomes exactly.

Mutation and dispersal can sometimes generate and retain more beneficial genotypic combinations than sex.

Negative epistasis hypothesis

If individuals with multiple mutations are less fit than expected based on the effects of each mutation considered separately (i.e. negative epistasis is present), extreme genotypes will become underrepresented in the population.



Loss of genetic variation

Sex and recombination can regenerate the missing extreme genotypes and increase the amount of genetic variance.

With this increased variability, selection is more efficient at increasing the frequency of the fittest alleles.

Muller's Ratchet

In any population, the fittest genotype may be lost as a result of sampling fluctuations.

Sexual populations can regenerate these lost genotypes by recombination (as long as the alleles have not been lost).

Asexual populations cannot.

Once the least mutated class is lost in an asexual population, the population reequilibrates.

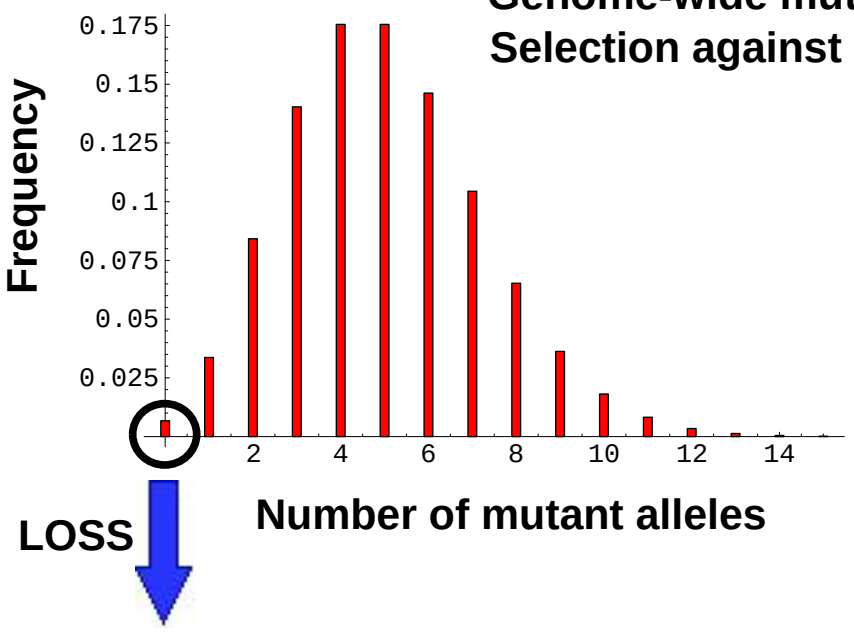
This reequilibration reduces the frequency of individuals in the next best class, which can then be lost, repeating the cycle [= the ratchet].

The ratchet will be fastest when mutation rates are high, selection is weak, and population sizes are small: Greatest advantage to sex under these circumstances.

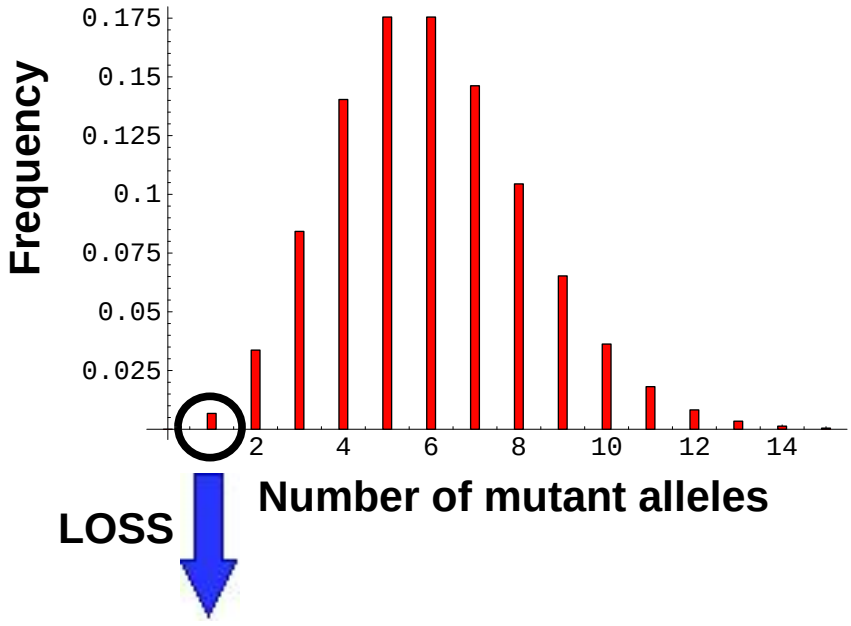
SOURCES:

- Bell (1986) The Masterpiece of Nature. California Press

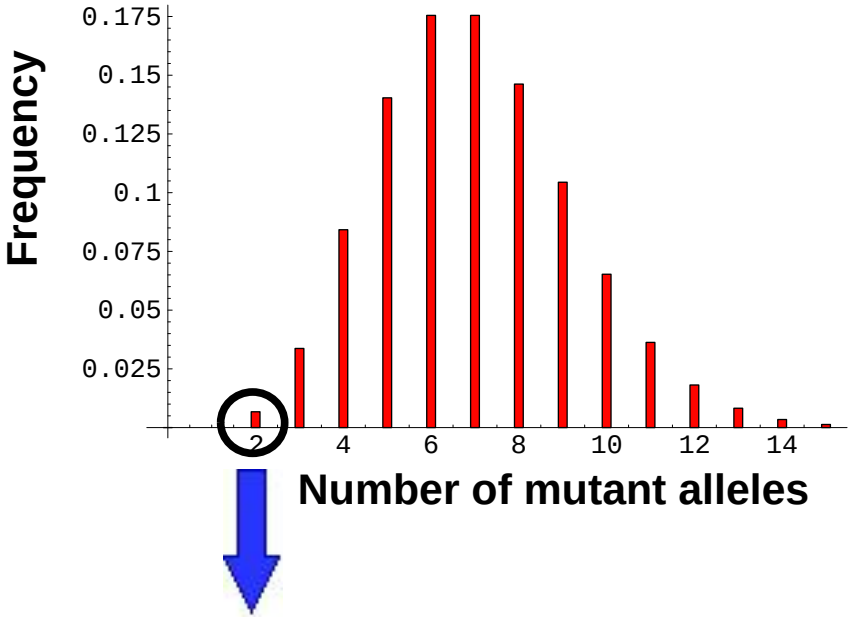
Genome-wide mutation rate: U=0.1
Selection against each mutation: s=0.02



LOSS



LOSS



802338 (R01597) 9-94

The Geological Time Scale

Geological time scaled to a cross-country tour of Canada.

The universe came into existence about 14 billion years ago, through an explosion known as the "big bang".

Our galaxy formed from clouds of dust and gas about 10 billion years ago.

We begin our cross-country tour in St. John’s Newfoundland, the eastern-most city in Canada (6240 km from Vancouver), which we will make correspond in distance to the formation of the earth around 4.6 billion years ago (BYA).

Every kilometer in our tour will therefore cover 0.737 million years (MY).

Saint John's, Newfoundland: [6240 km]
~ 4.6 billion years ago (BYA), Beginning of Hadean Eon.

➡ The formation of the earth and solar system.

Earth was molten and extremely hot at first, slowly cooling and forming land and water masses.

Initially, little atmospheric oxygen was present.

Saint John, New Brunswick: [5211 km]
~3.8 BYA, Eon Hadean.

➡ Oldest surviving rocks on earth.

Quebec City, Quebec: [4643 km]
~3.5 BYA, Near beginning of Archean Eon.

➡ First evidence of life

Bacteria-like microfossils and layered fossil mats, known as stromatolites provide the first evidence of life on earth.

Stromatolites formed by living cyanobacteria

(From Futuyma, 1998, p. 169)

Dryden, Ontario: [2628 km]
~2 BYA, Near beginning of Proterozoic Eon.

➡ First eukaryotes (single-celled algae).
The symbiotic origin of mitochondria and chloroplasts followed by 1.4 BYA.

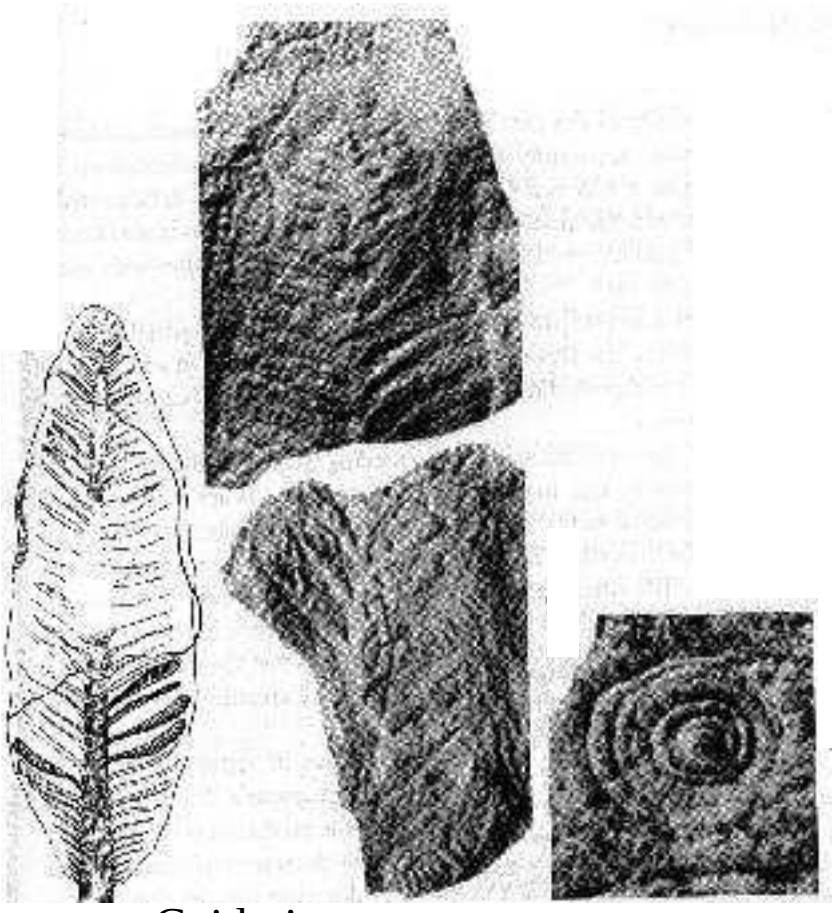
Calgary, Alberta: [962 km]
~800 MYA, Eon Proterozoic.

➡ Trace evidence, including burrows, of first metazoans (= multicellular animals with differentiated tissues)

More definite fossil evidence of metazoans comes from the late Proterozoic. In particular, the Ediacaran fauna from Australia consists of a number of soft-bodied, aquatic animals resembling jellyfish and worms.

Ediacaran Fauna (~600 MYA)

(From Ridley, 1998, p. 547)



Cnidarian

Jellyfish

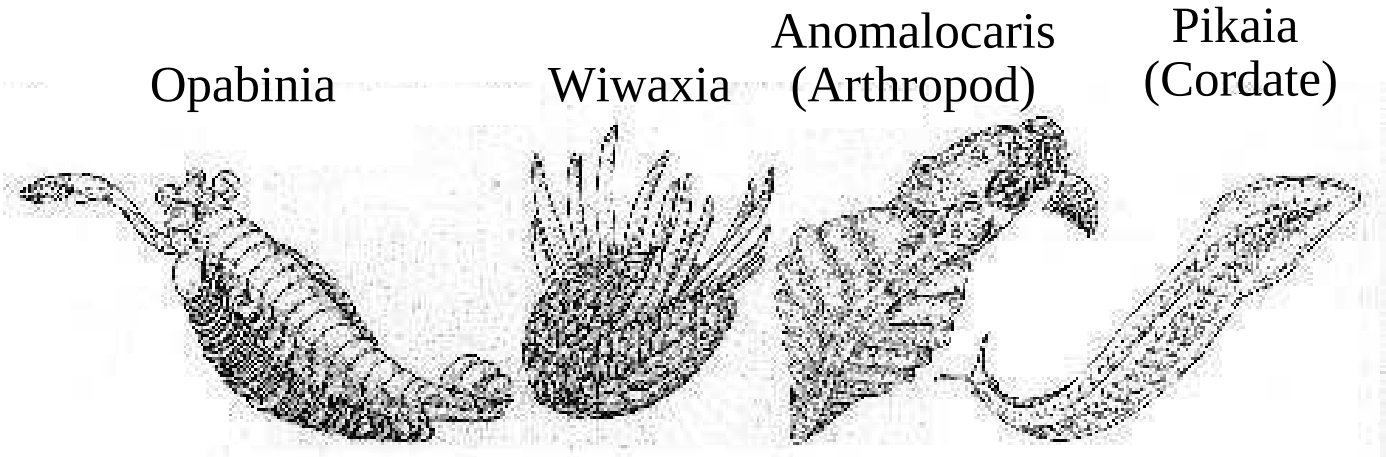
Yoho National Park, British Columbia: [740 km]
~543 MYA, Beginning of Paleozoic Era (Cambrian Period).

➡ Diversification of metazoans.
Increase in oxygen levels, approaching modern levels.

Cambrian Explosion: "Almost all the modern phyla and classes of skeletonized marine animals, as well as many groups that may represent extinct phyla and classes, suddenly appear in the fossil record...within about 30, perhaps only 5 to 10, million years."

-- Futuyma (1998) p. 172

Burgess Shale Fauna



(From Futuyma, 1998, p. 174)

Most of the fundamental body plans (baupläne) had evolved by the end of the Cambrian Period. The most spectacular collection of Cambrian fossils comes from the Burgess shale, right here in Yoho National Park, BC!

Important events in the Paleozoic Era:

- 1st shells in the Cambrian Period (543-500 MYA)
- 1st fish in the Ordovician Period (500-439 MYA)

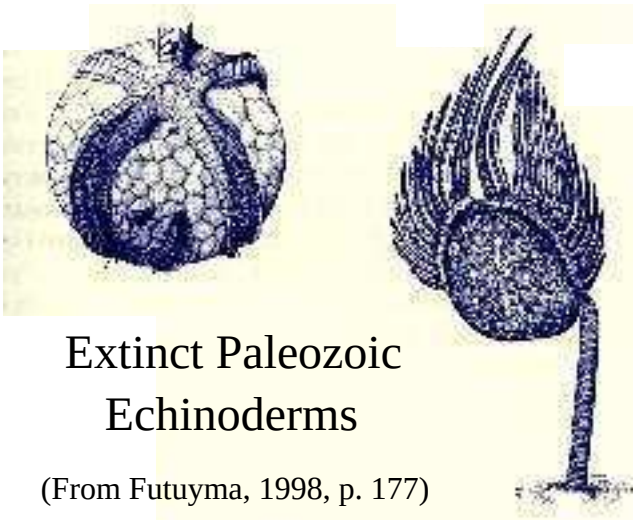
⚡ Mass extinction

- 1st land plants in the Silurian Period (439-409 MYA)
- 1st amphibians and true insects in the Devonian Period (409-354 MYA)

⚡ Mass extinction

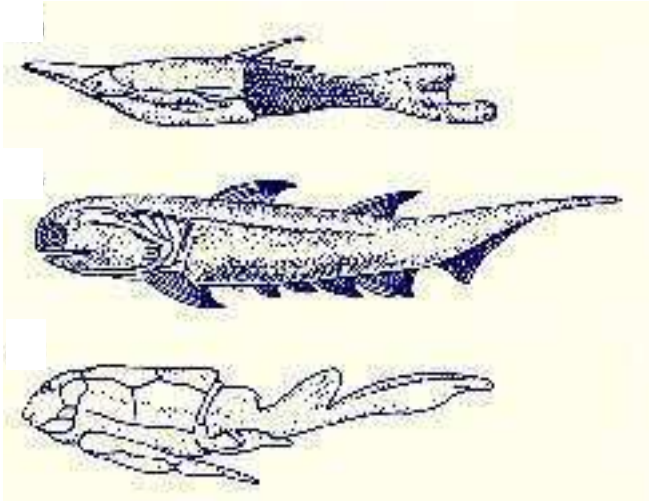
- 1st reptiles in the Carboniferous Period (354-290 MYA)
- Diversification of the reptiles in the Permian Period (290-251 MYA)

⚡ Mass extinction (P/Tr boundary)



Extinct Paleozoic Echinoderms

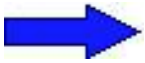
(From Futuyma, 1998, p. 177)



Extinct Paleozoic Fish

(From Futuyma, 1998, p. 178)

Kamloops, British Columbia: [348 km]
~251 MYA, Beginning of Mesozoic Era (Triassic Period).

 "Age of Reptiles". Break-up of Pangaea (single large landmass including all present-day continents). Warm climate.

Important events in the Mesozoic Era:

- 1st dinosaurs and mammals in the Triassic Period (251-206 MYA)

 Mass extinction

- 1st birds and angiosperms in the Jurassic Period (206-144 MYA)
- Diversification of mammals, birds, and angiosperms in the Cretaceous Period (144-65 MYA)

 Mass extinction (K/T boundary)

Chilliwack, British Columbia: [90 km]
~65 MYA, Beginning of Cenozoic Era (Tertiary Period).

➡ "Age of Mammals". Break-up of Gondwanaland (landmass containing southern continents and India). Cooling of climate.

Important events in the Cenozoic Era:

- Radiation of mammals, birds, snakes, angiosperms, pollinating insects, and teleost fish into their modern orders in the Tertiary Period (65-1.8 MYA)

- 1st *Homo* fossils in Quaternary Period (1.8 MYA - Present)

Broadway and Vine (Kitsilano), Vancouver: [6.8 km]
~5 MYA, Era Cenozoic (Tertiary Period).

➡ Divergence of hominid and chimpanzee lineages.

UBC Golf Course, Vancouver: [2.4 km]
~1.8 MYA, Era Cenozoic (Quaternary Period).

➡ Beginning of Pleistocene epoch. Massive fluctuations in temperature (~100,000 year period), leading to major glacial advances and retreats.

Forest Sciences building, UBC, Vancouver: [24 m]
~18,000 YA, Era Cenozoic (Quaternary Period).

➡ Last ice age at its maximum.

Room 1005 door, FSB, UBC, Vancouver: [16 m]
~12,000 YA, Era Cenozoic (Quaternary Period).

➡ Development of agriculture.

The span of my hand: [19 cm]
139 YA, Era Cenozoic (Quaternary Period).

➡ The Origin of Species is published.

MYA	EON	ERA	PERIOD	EPOCH	MAJOR EVENTS
0.01	PHANEROZOIC	CENOZOIC	Quarternary	Recent	Development of agriculture and human civilization
1.8				Pleistocene	Appearance and world-wide spread of the genus *Homo*. Repeated glaciations. Extinctions of large mammals and birds.
5.2			Tertiary	Pliocene	Continued diversification of modern birds, placental mammals, snakes, teleost fish, pollinating insects, grasses and angiosperms.
24				Miocene	
34				Oligocene	
56				Eocene	
65				Paleocene	
144		MESOZOIC	Cretaceous		Diversification of flowering plants, birds and mammals.
206			Jurassic		1st birds and angiosperms. Dinosaurs abundant.
251			Triassic		1st dinosaurs and mammals. Gymnosperms become abundant. Continents moving apart.
290			Permian		Diversification of reptiles, including mammal-like species. Land masses form single continent, Pangea.

MYA	EON	ERA	PERIOD	EPOCH	MAJOR EVENTS
354	PHANEROZOIC	PALEOZOIC	Carboniferous		1st reptiles and winged insects. Warm humid conditions result in huge forests of primitive plants, which formed extensive coal deposits.
409			Devonian		1st amphibians and true insects. Atmospheric oxygen at present levels or higher. Continents moving toward one another.
439			Silurian		1st land plants. Atmospheric oxygen about 20 percent.
500			Ordovician		1st fish.
543			Cambrian		1st shelled organisms. Trilobites abundant. Probably all metazoan phyla present, including arthropods and early chordates. Atmospheric oxygen reaches about 2%.
2,500	PROTEROZOIC				Abundant prokaryotic life. Eukaryotes may have appeared by 2,000 million years ago. Atmospheric oxygen about 0.2%.
3,600	ARCHEAN				Oldest known rocks and prokaryotes.
4,600	HADEAN				Earth forms. No geological record.

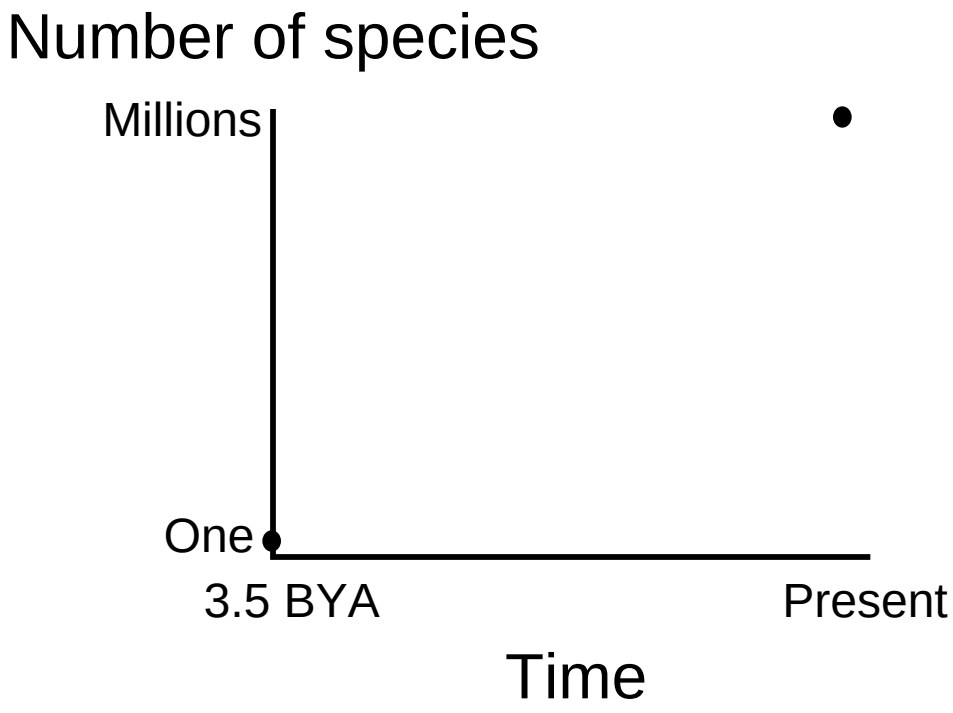
MYA	EON	ERA	PERIOD	EPOCH	MAJOR EVENTS
0.01	PHANEROZOIC	CENOZOIC	Quarternary	Recent	Development of agriculture and human civilization
1.8				Pleistocene	Appearance and world-wide spread of the genus *Homo*. Repeated glaciations. Extinctions of large mammals and birds.
5.2			Tertiary	Pliocene	Continued diversification of modern birds, placental mammals, snakes, teleost fish, pollinating insects, grasses and angiosperms.
24				Miocene	
34				Oligocene	
56				Eocene	
65				Paleocene	
144		MESOZOIC	Cretaceous		Diversification of flowering plants, birds and mammals.
206			Jurassic		1st birds and angiosperms. Dinosaurs abundant.
251			Triassic		1st dinosaurs and mammals. Gymnosperms become abundant. Continents moving apart.
290		PALEOZOIC	Permian		Diversification of reptiles, including mammal-like species. Land masses form single continent, Pangea.
354			Carboniferous		1st reptiles and winged insects. Warm humid conditions result in huge forests of primitive plants, which formed extensive coal deposits.
409			Devonian		1st amphibians and true insects. Atmospheric oxygen at present levels or higher. Continents moving toward one another.
439			Silurian		1st land plants. Atmospheric oxygen about 20 percent.
500			Ordovician		1st fish.
543			Cambrian		1st shelled organisms. Trilobites abundant. Probably all metazoan phyla present, including arthropods and early chordates. Atmospheric oxygen reaches about 2%.
2,500	PROTEROZOIC				Abundant prokaryotic life. Eukaryotes may have appeared by 2,000 million years ago. Atmospheric oxygen about 0.2%.
3,600	ARCHEAN				Oldest known rocks and prokaryotes.
4,600	HADEAN				Earth forms. No geological record.

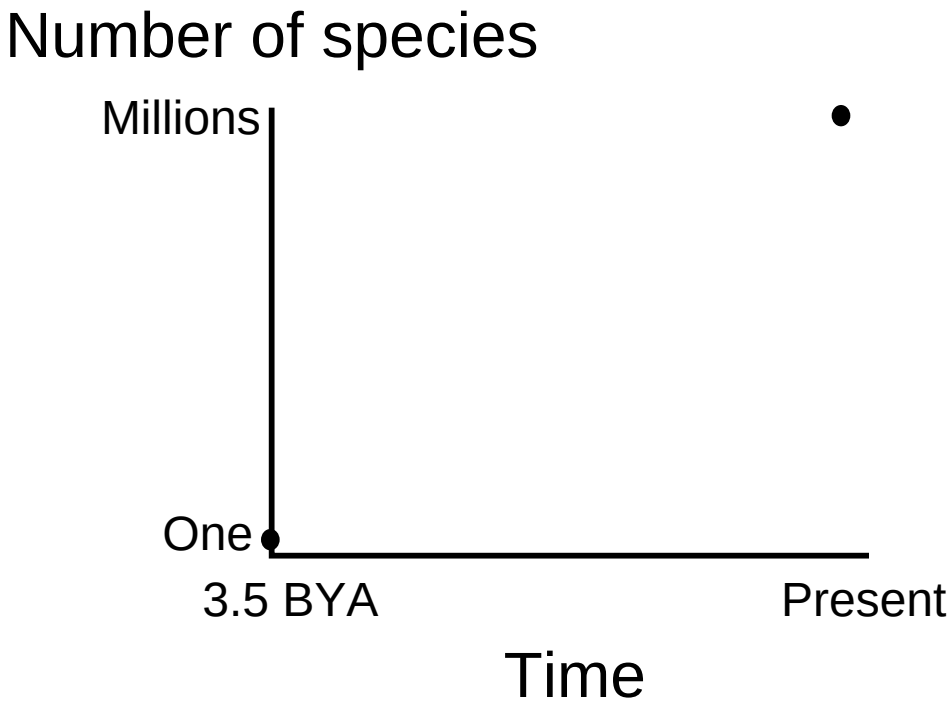
The Evolution of Biological Diversity

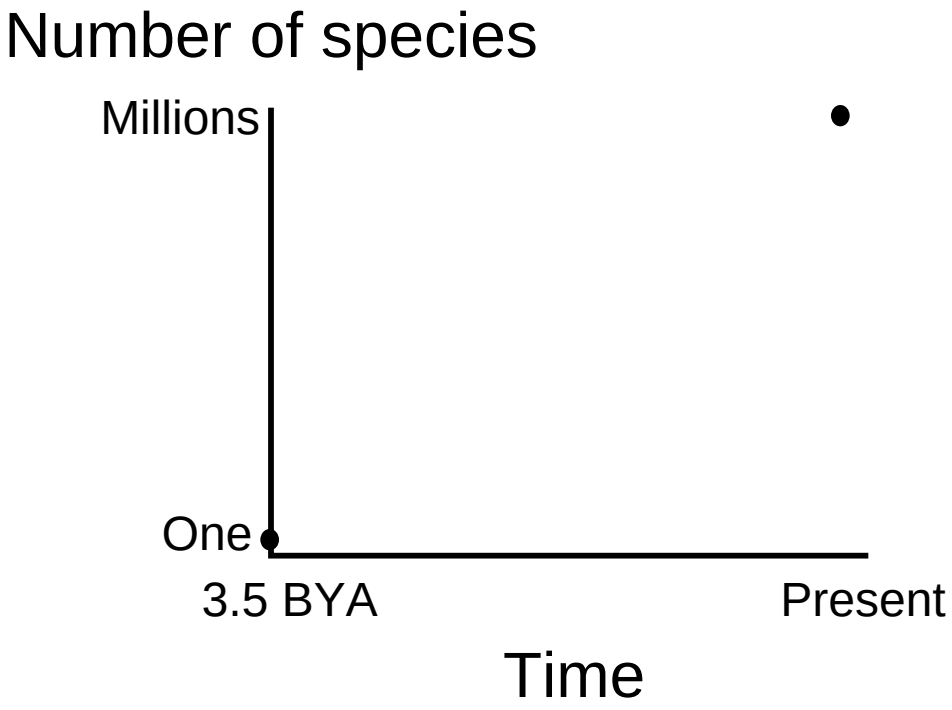
All living organisms are descended from an ancestor that arose between 3 and 4 billion years ago.

The diversity of life on earth currently includes some 5 to 50 million species!

Before discussing the observed patterns, it is worth thinking about the possible ways in which the number of species present at any point in time may have changed over the history of the earth:




Number of species
Millions
One
3.5 BYA
Present
Time


Number of species
Millions
One
3.5 BYA
Present
Time

The patterns and rates of diversification reflect the rates of *speciation* (S) and *extinction* (E) of taxa.

At the simplest conceptual level, the number of species (N) present on earth will change over time according to the formula:

$$\Delta N = (S - E)\, N\, \Delta t,$$

where ΔN is the change in number of species over an amount of time Δt and where S and E are rates of speciation and extinction per species.

If S and E remain approximately constant over time, exponential growth will result.

Both speciation and extinction rates may vary over time and from species to species depending on:

- resource and habitat availability ("niche space")
- interactions among species
- key adaptations (changing "adaptive zones")
- climactic changes
- catastrophes

Diversity-dependent growth: An alternative possibility is that either the rate of speciation or extinction (or both) depends on diversity levels (N).

Why might this be true?

If (S - E) goes down as N increases, then logistic growth in the number of species is expected, potentially resulting in a global equilibrium.

The number of species is difficult to determine from the fossil record.

- Are allopatric populations the same or different species?
- Are populations from different points in time the same or different species?
- Are there sufficient fossil remains of most species?

Consequently, paleontologists tend to study diversity by counting the number of higher taxa (e.g. families).

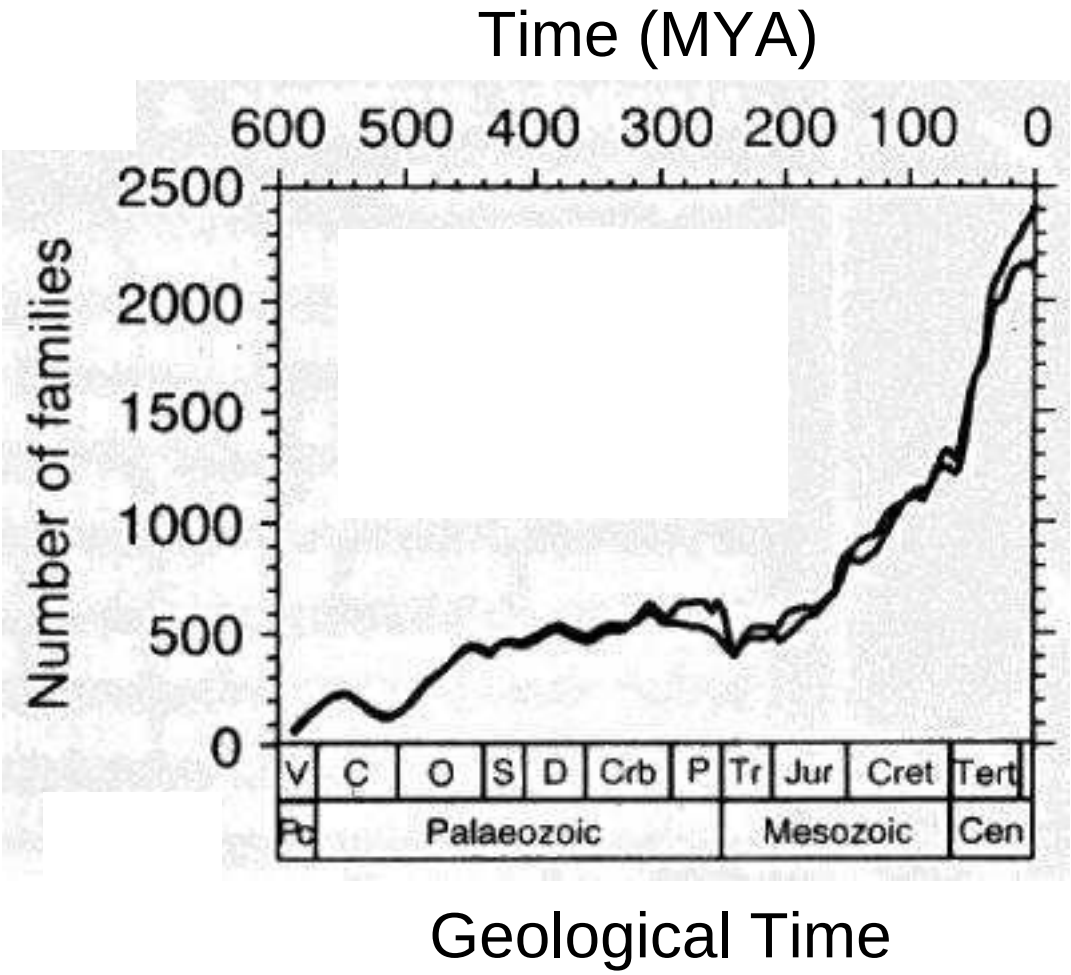
Two major databases have been compiled:

- John Sepkoski (1984) compiled data on the temporal ranges of more than 4000 marine skeletonized families throughout the 543 MY of the Phanerozoic.
- Michael Benton (1993) compiled the *Fossil Record 2*, containing 7186 families of all groups of microbes, algae, fungi, protists, plants, and animals.

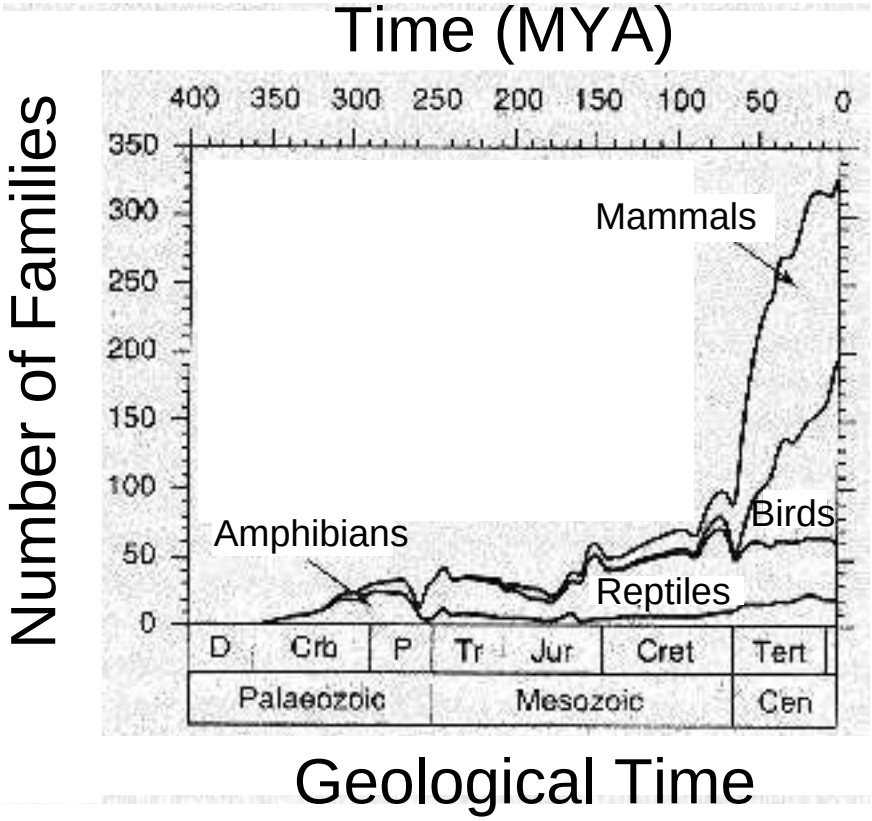
The pattern of diversification depends on the taxa examined.

For several groups, an exponential rise in the number of families best fits the data.

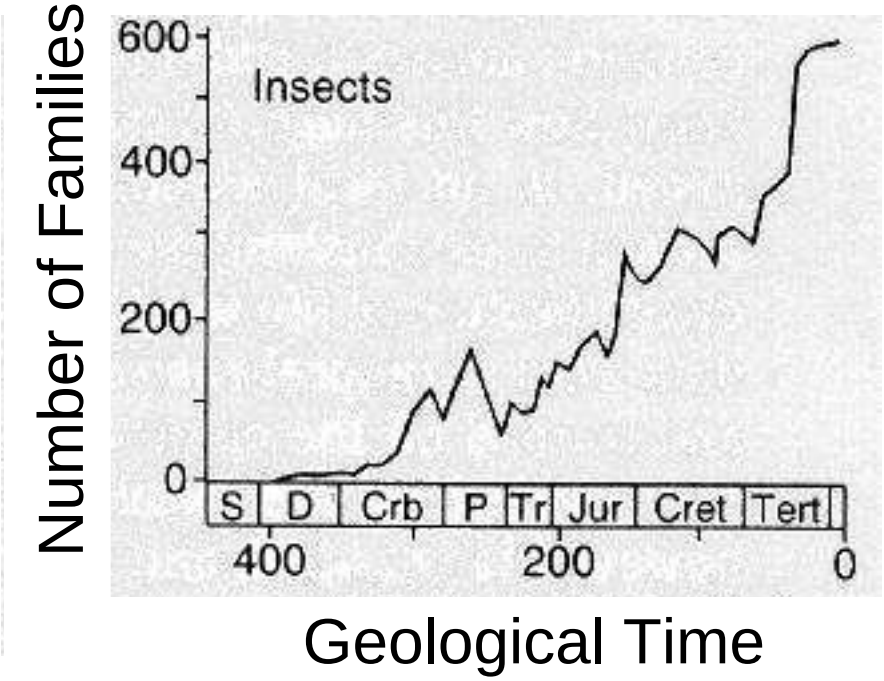
(1) All of the families in the *Fossil Record 2* database:



(2) Tetrapod families:



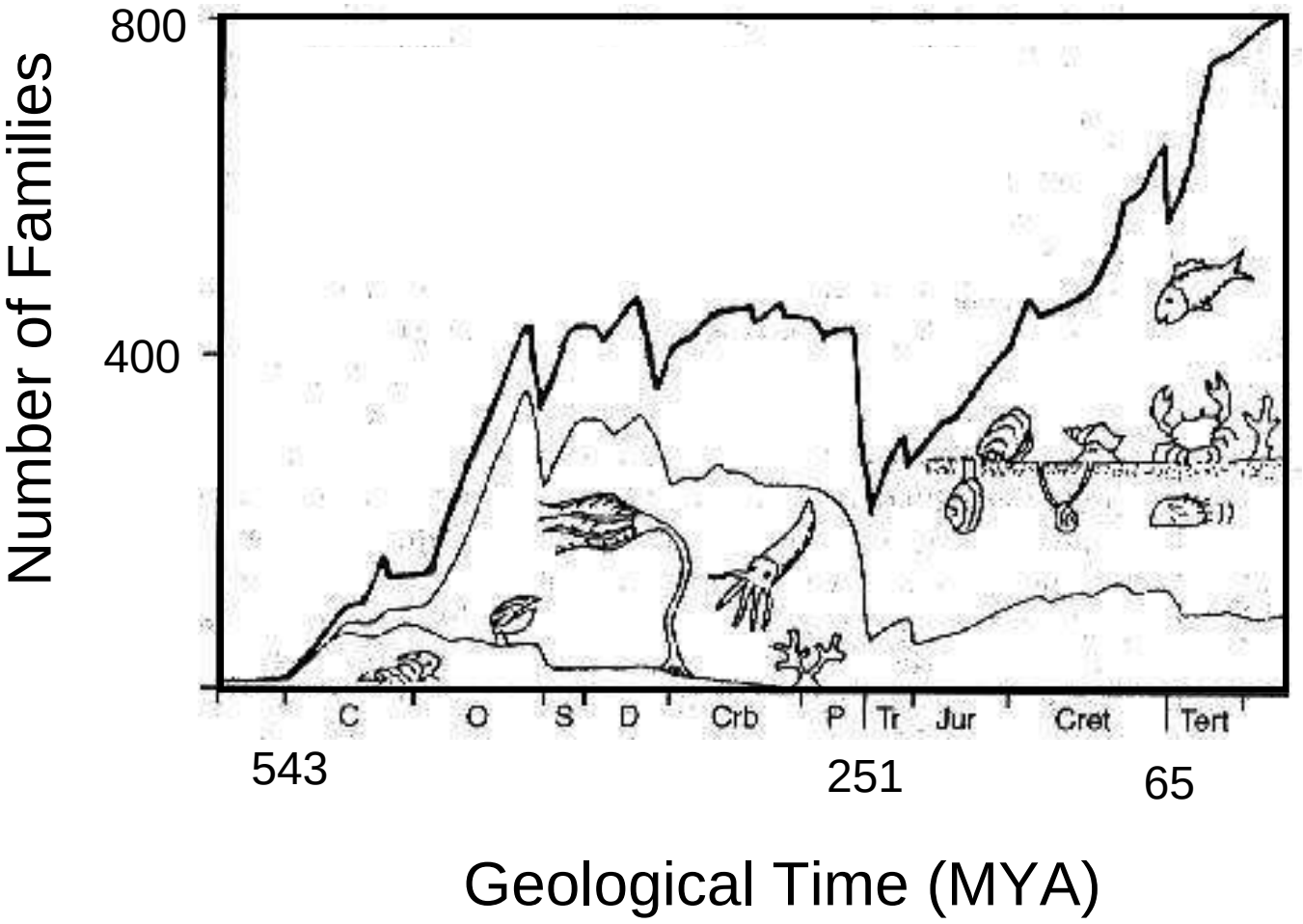
(3) Insect families:



(REMEMBER: Extinctions are happening throughout.)

For other groups, the patterns are more complicated.

For example, the growth of marine families appears to show a number of steps:

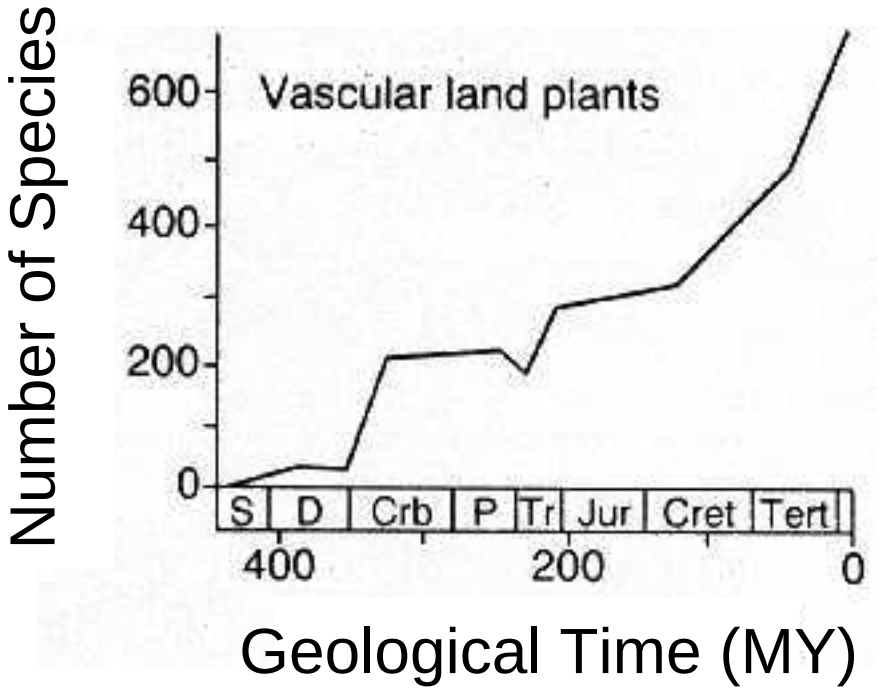


Sepkoski interpreted this as the result of the radiation of three distinct "evolutionary faunas":

- Cambrian fauna with broad habitat and trophic requirements (e.g. trilobites, inarticulate brachiopods...)
- Palaeozoic fauna with more specialized requirements (e.g. crinoids, cephalopods, soft corals...)
- Modern fauna with predatory and defensive strategies (e.g. bivalves, hard corals, malacostracan crustaceans, fish...)

Why might marine fauna have increased rapidly and then reached a plateau between 400 and 250 MYA whereas tetrapods and insects have not plateaued?

Similarly, the number of species of vascular land plants appears to have increased in steps:



Knoll and colleagues (1984, 1985) interpreted this as the appearance and radiation of new major baupläne (= body forms):

- Devonian flora of early vascular plants
- Carboniferous flora of club mosses, ferns, conifers
- Triassic flora of gymnosperms
- Cretaceous flora of angiosperms

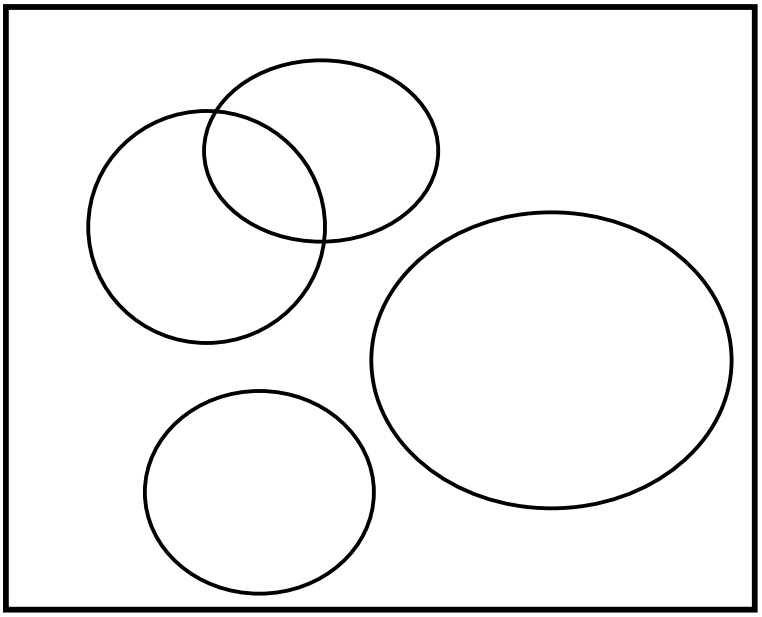
The subsequent rise of angiosperms has proceeded exponentially.

Surprisingly, these data suggest that the number of species on earth has not reached a global equilibrium.

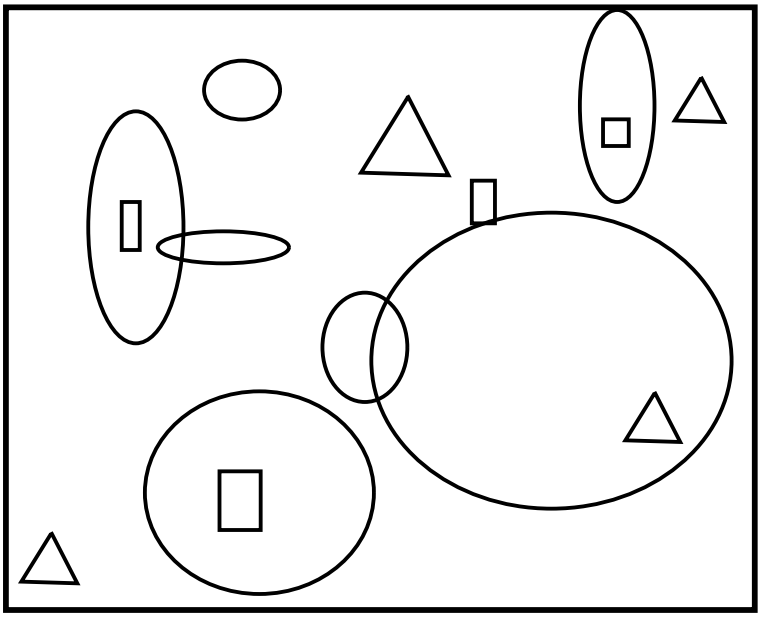
If anything, the most prevalent pattern is of exponential growth.

Can exponential growth of species be sustained?

How?



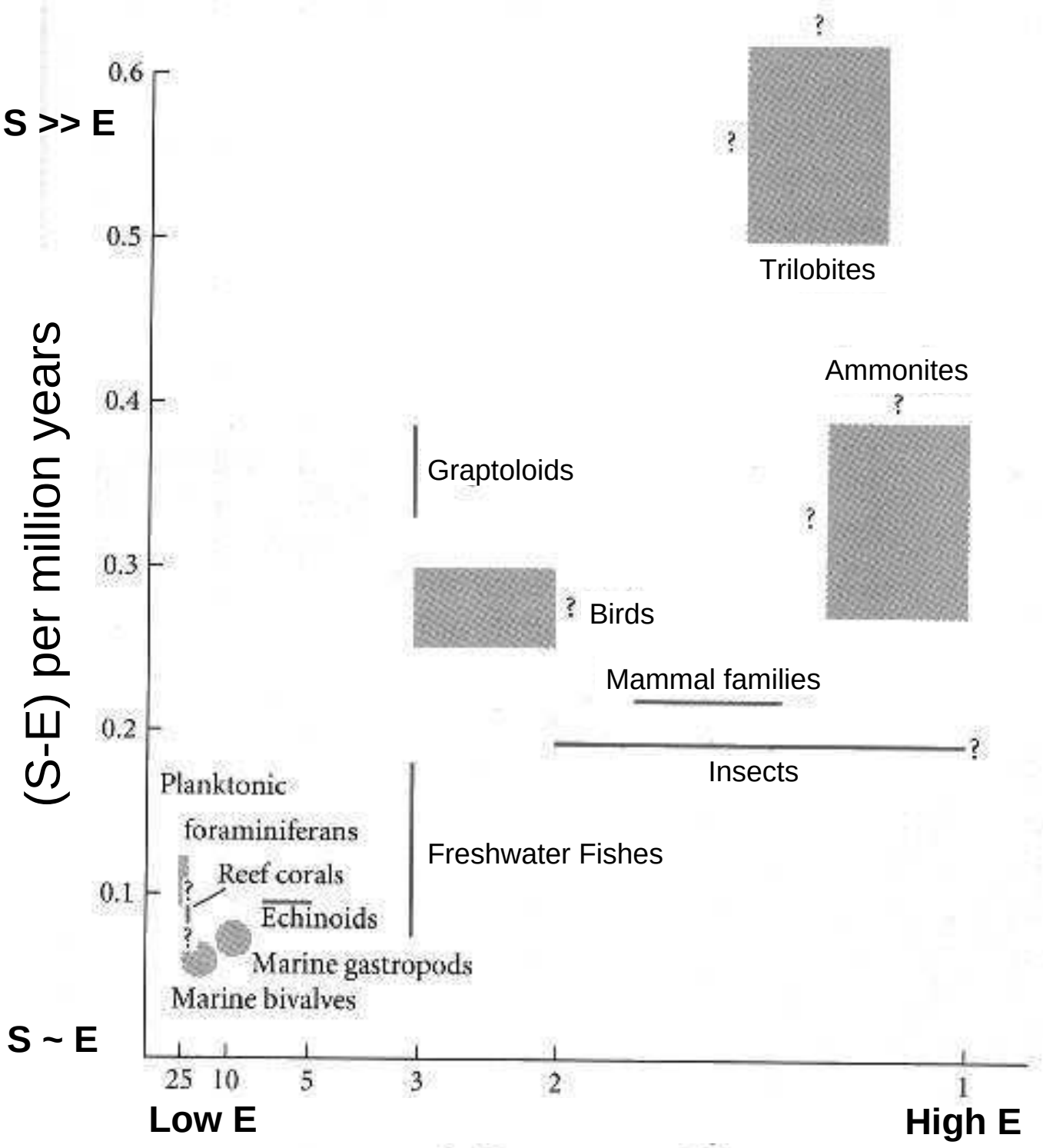
Earth



Earth

That is, exponential growth of species may occur by changing patterns of "species packing", with generalists being replaced by more specialized forms and with an increase in the complexity of the ecological community.

Are the rates of speciation and extinction the same for different organisms?



No!

Interestingly, those taxa with high rates of increase (S-E) also tend to have high rates of extinction.

What might explain this odd result?

- Specialists may be more likely to speciate because of their patchy distribution but may also be at higher risk of extinction.

- Species with small population sizes may be more likely to speciate (if drift is important) but are at higher risk of extinction.

- Species with low dispersal rates may be more likely to speciate (lower gene flow) but may be more likely to go extinct following local environmental changes.

Factors Affecting the Origin of Biological Diversity

Speciation rates are higher in some lineages than others and at certain times over others. Here we explore several possible explanations.

(1) Ecological Opportunity:

Living organisms present numerous examples of spectacular radiations following the colonization of isolated islands or bodies of water, e.g.:

- honeycreepers on Hawaii
- drosophilids on Hawaii
- cichlids in African Great Lakes
- amphipod crustaceans in Lake Baikal in Siberia

In these cases, the fauna was locally depauperate before the arrival of the original colonist.

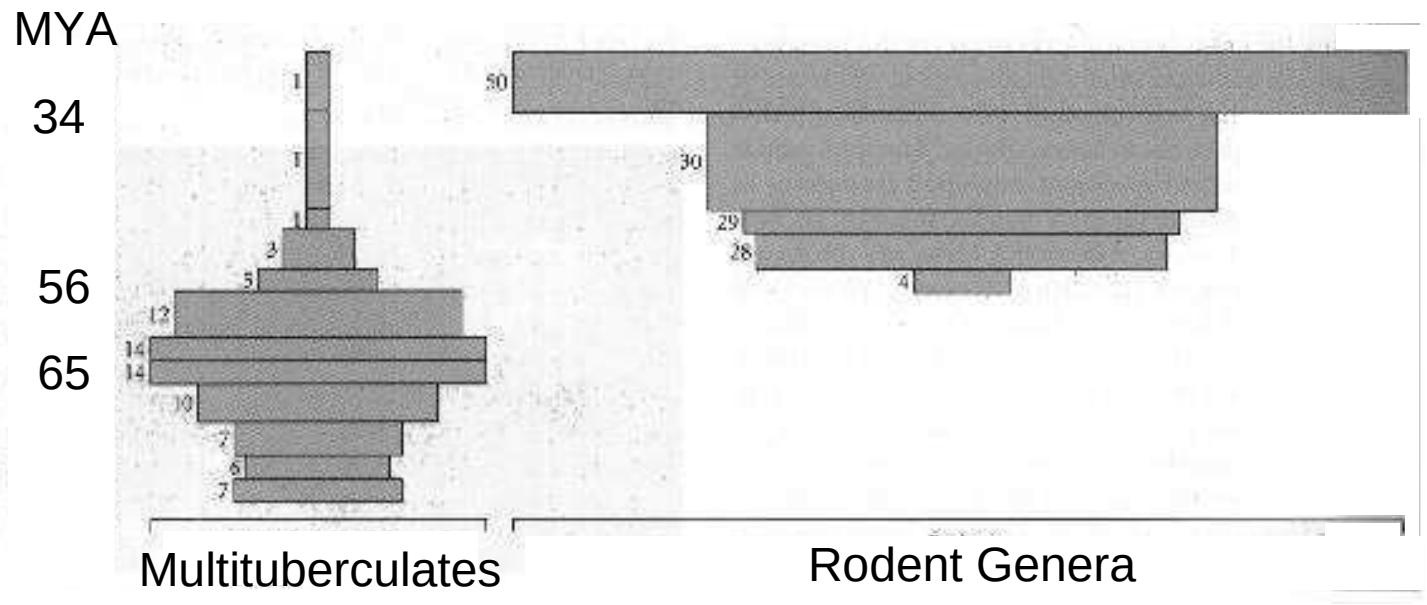
➡ "Vacant niches" existed into which the newly arrived organisms diversified.

Similarly, there are several examples in the fossil record where the decreased representation of one group is followed or accompanied by a proliferation of another group.

➡ The new group may cause the extinction of the former group (**displacement**)

➡ The new group may be released from competition by the extinction of the former group (**incumbent replacement**)

For example, rodent-like, non-placental mammals (multituberculates) decreased in diversity in North America following the appearance of placental rodents.



The correlated pattern of increase and decrease suggests displacement or incumbent replacement?

Lineages often are seen to radiate following the mass extinction of another group.

For example, mammals radiated in the Tertiary period following the mass extinction of dinosaurs near the K-T boundary (Cretaceous-Tertiary boundary ~65MYA).

Does this suggest displacement or incumbent replacement?

(2) Key Adaptations:

Speciation rates within a group may rise after the evolution of a new adaptive trait.

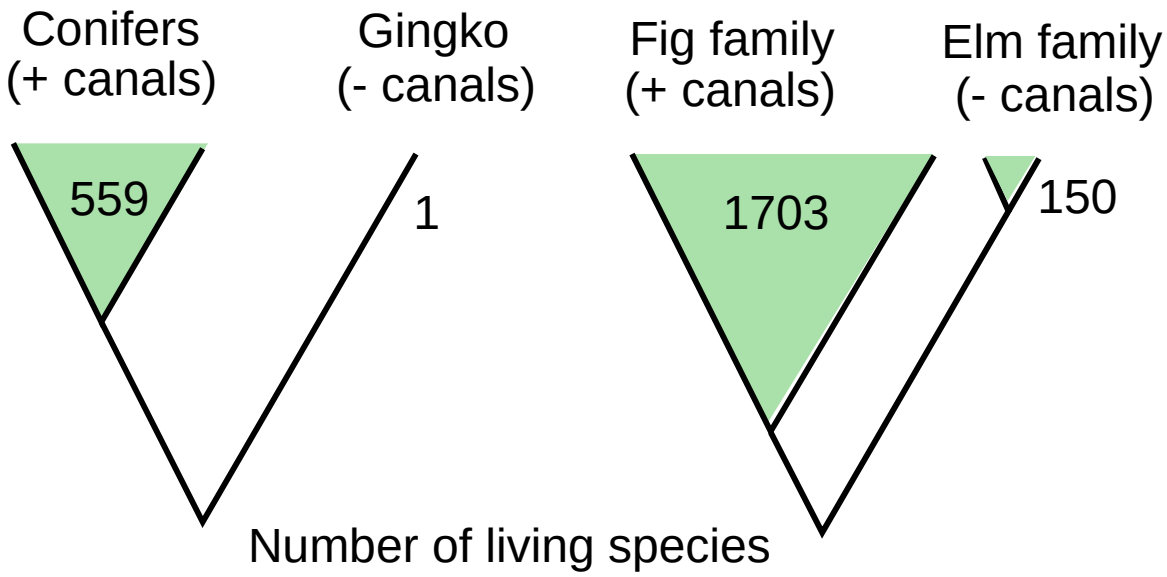
How can we tell whether a trait increases speciation rates?

Replicated Sister-Group Comparisons

For example, Ehrlich and Raven (1964) suggested that defenses against herbivores (e.g. latex and resin canals) promoted diversification in plants.

Mitter, Farrel and colleagues (1988,1991) tested this hypothesis by identifying 16 sister groups of plants with and without these canals.

In 13/16 cases, the canal-bearing groups contained more species than their sister clades.



Similarly, the fossil record suggests that key adaptations in marine organisms (specialization, predation, swimming, hard shells) promoted their diversification.

Key adaptations allow organisms to evolve into a greater variety of niches, creating a more complex and tiered community structure.

(3) Provinciality:

Speciation rates will depend on the extent to which organisms are distributed through space.

There has been a general trend over the last 250MY from wide-spread distributions to more localized distributions.

As Pangaea began to break apart during the Triassic, land and ocean masses became more **spatially separated**.

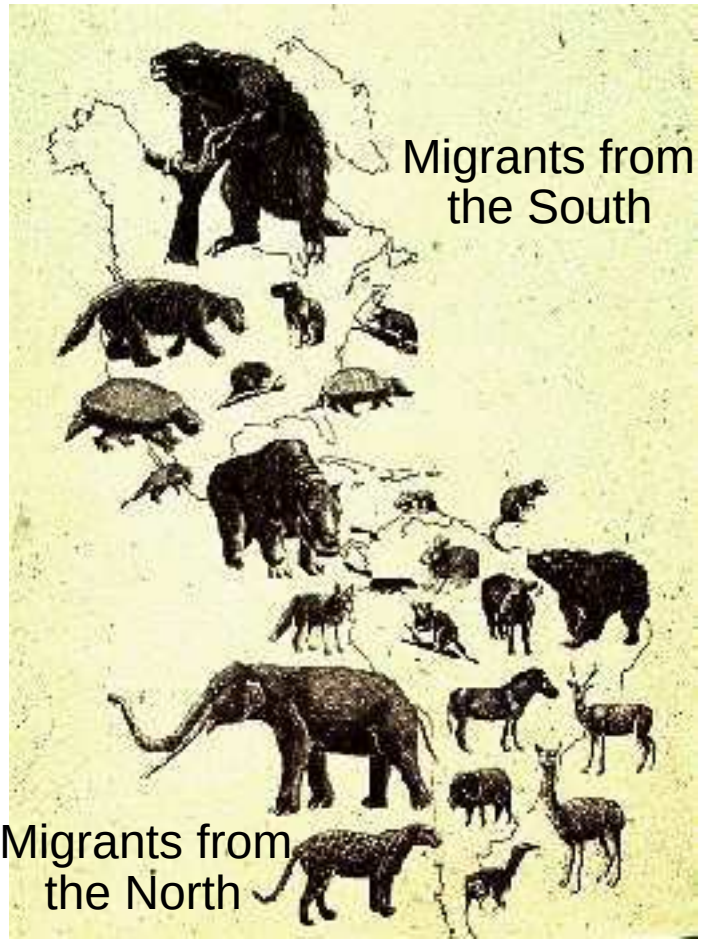
Ocean currents also changed, leading to a more pronounced **temperature gradient**.

These changes have increased the number of biological "provinces" (= a self-contained region wherein speciation rather than colonization dominates the appearance of new taxa).

An example (in reverse) is the extinction caused when the separate land provinces of North and South America became connected by the Isthmus of Panama (~2MYA).

23 families of mammals were endemic to South America (incl. sloths, armadillos, opossums, raccoons), 25 to North America (incl. mammoths, mastodons, saber-toothed cats, and camels), and 2 occurred in both.

The "Great American Interchange" describes the resulting migration of animals across the Isthmus of Panama.



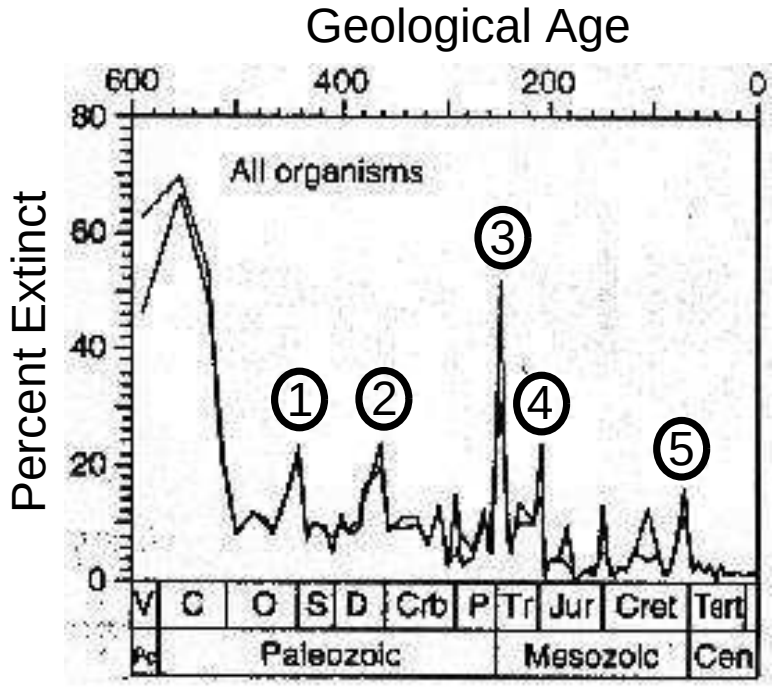
Diversity, supported by provinciality, was lost following the Interchange: only 38/50 families remained, 24% went extinct.

Extinctions and The Decline of Biological Diversity

The tree of life has been severely and often pruned.

The vast majority of species that have ever lived have gone extinct.

Extinction rates vary over time, but are dominated by the "big five" mass extinction events.



[Note: Peak in Cambrian probably reflects low diversity at that time.]

Extinction Event	MYA	Family Loss	Species Loss
End-Cretaceous	65	~14%	~76%
End-Triassic	208	~30%	~80%
End-Permian **	245	~60%	~95%
Late Devonian	367	~30%	~83%
End-Ordovician	439	~23%	~85%

[Family extinctions reflect all organisms in the *Fossil Record 2* database. Species extinctions reflect loss of marine species estimated by Jablonski (1991,1995).]

Climate change has been cited as a major factor involved in each mass extinction event.

The largest of the extinction events, at the end of the Permian, is associated with a number of catastrophic climate changes (the "world-went-to-hell" hypothesis) including:

- major sea level regression
- ocean anoxia (= decreased oxygen)
- Siberian flood basalts (=magma flows) over 1.5 million km^2
- increased CO_2
- global warming

Major climate changes also surround the end-Cretaceous extinction (K-T), possibly resulting from a massive asteroid hitting the earth.

(A buried crater has been detected in the Yucatan peninsula of Mexico with a diameter of 180 km!)

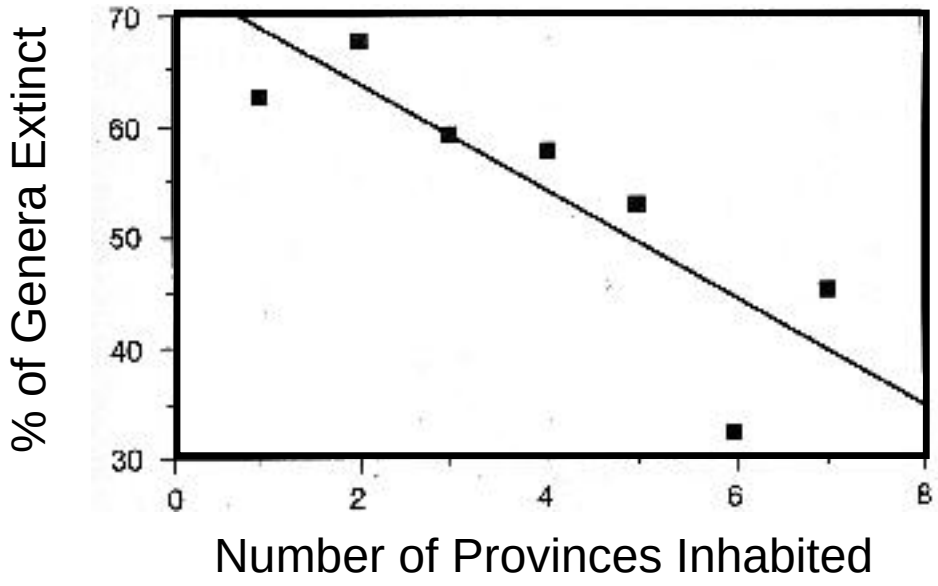
These mass extinctions have played a major role in shaping the biota we see today.

Regardless of how adapted a species may be in "normal" times, if it succumbs to extinction during a massive climate change, it will play no further role in evolution.

For instance, at the K-T boundary, insects, amphibians, crocodilians, mammals, and turtles suffered few extinctions, whereas several bird species, ammonites, dinosaurs, and other large reptiles went extinct.

Although we generally do not know why extinction events are so selective, some patterns have emerged.

Jablonski (1986) found that bivalves and gastropods with wider geographic distributions were less likely to go extinct at the end of the Cretaceous.



Without these mass extinctions, the world would be a very different place.

For one thing, the rise and diversification of mammals may never have happened.

Macroevolutionary Patterns

In this lecture we will examine the *tempo and mode of evolution*, in the words of George Gaylord Simpson (1902-1984).

Simpson devoted himself to understanding whether *macroevolutionary* patterns arise from the *microevolutionary* processes studied by population geneticists.

Simpson showed that major evolutionary developments in the fossil record took place in the irregular and undirected manner expected under Darwinian evolution.

- Does morphological evolution occur gradually or in fits and starts?

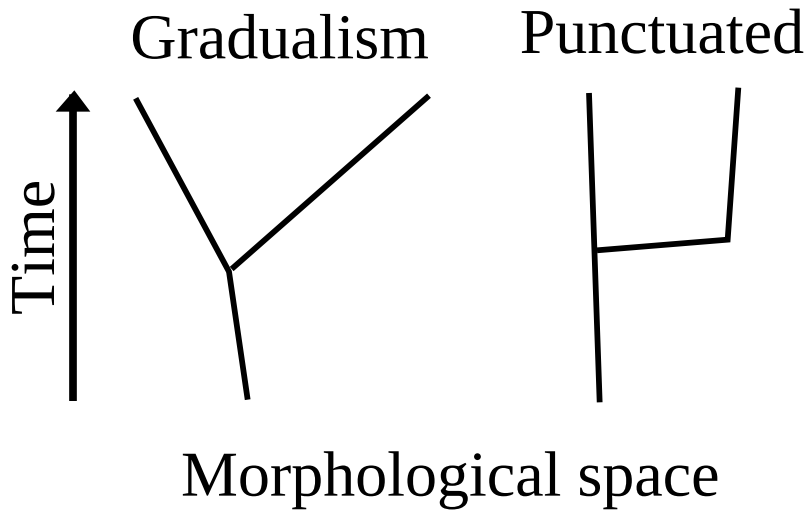
- Is there a tendency for organisms to evolve in a particular direction?
  - Towards greater size?
  - Towards greater complexity?

Punctuated Equilibria versus Gradualism

Following Darwin, the prevailing view of evolution by natural selection held that evolution is *gradual.*

➡ Expectation that macroevolutionary changes (= large changes in morphology that define higher taxonomic divisions) accumulate over long periods of time by gradual microevolutionary processes.

Nevertheless, the fossil record does not always show continuous and gradual changes.



Simpson (1944) noted that higher taxa (e.g. orders of mammals) appear suddenly in the fossil record, describing this pattern as "quantum evolution".

Major morphological innovations sometimes appear suddenly in the fossil record, often preceded and followed by periods of relative stasis.

➡ Interpreted as inaccuracy of the fossil record.

Eldredge and Gould (1972) argued otherwise.

(1) The pattern was real

(2) The pattern reflected a process whereby most evolutionary change happens around speciation events.

➡ Punctuated equilibrium model of evolution

This was an extremely controversial interpretation.

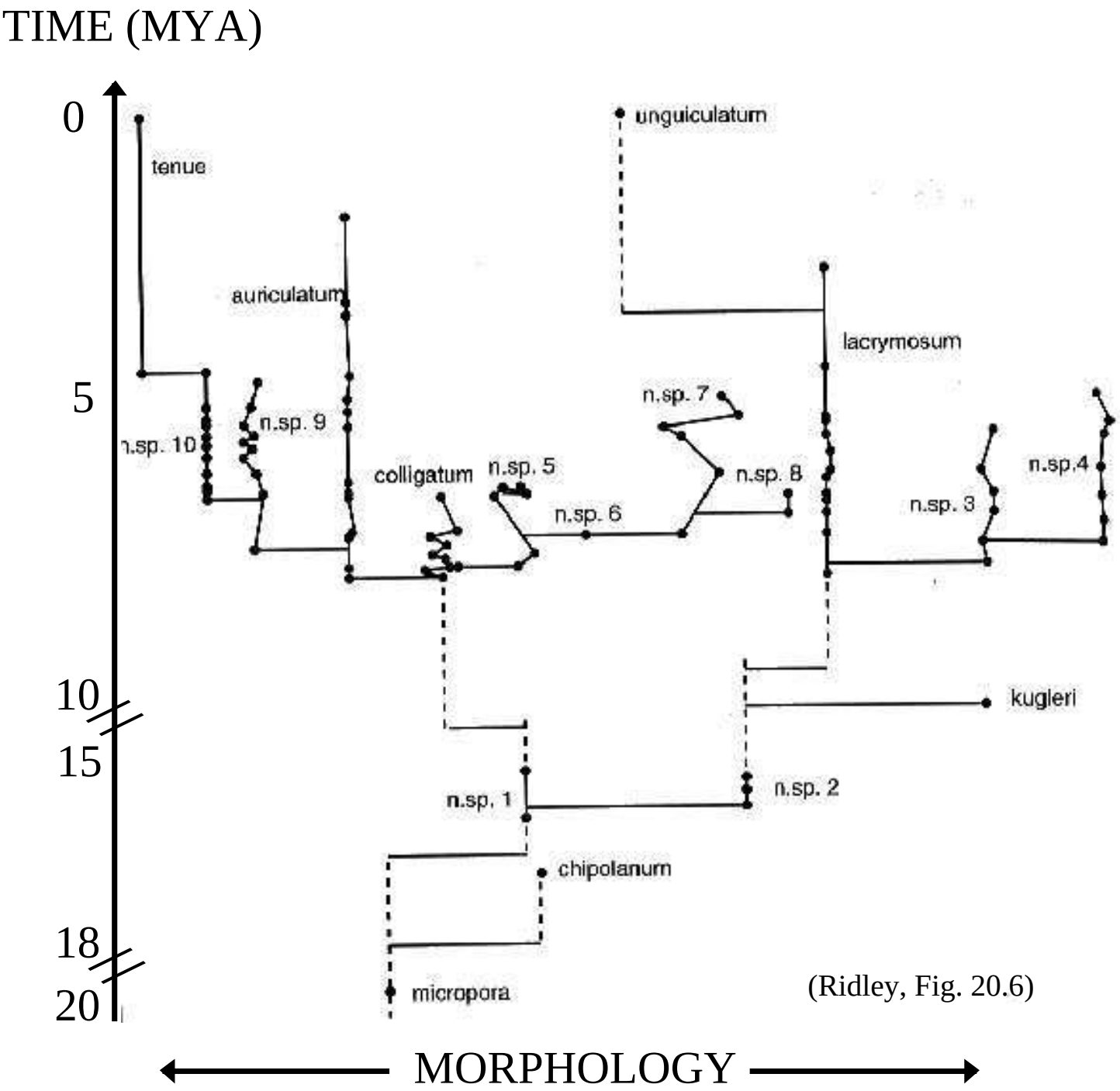
[Eldredge and Gould did not argue for instantaneous evolutionary change but rather a concentration of gradual evolutionary change near a speciation event.]

(1) Is the pattern real?

Example: Punctuated change in Bryozoans

Cheetham (1986) examined 1000 fossil specimens from the Bryozoan genus *Metrarabdotus*, an aquatic invertebrate.

Using 46 morphological characters, Cheetham drew a phylogenetic tree connecting the specimens:



(Ridley, Fig. 20.6)

Relatively little change occurred within a morphospecies, while large shifts were observed between morphospecies.

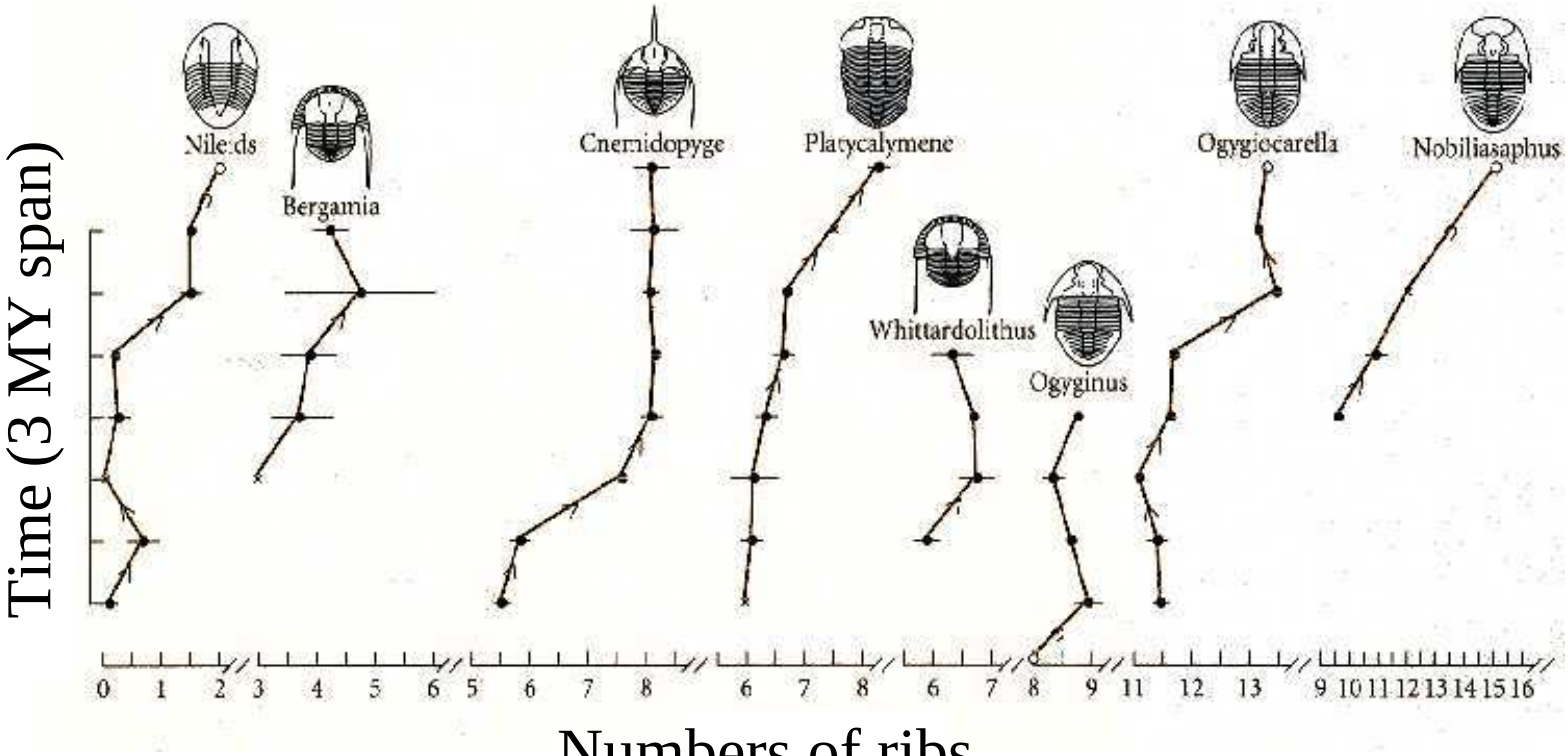
Almost no intermediates were found in the fossil record between these morphospecies.

[Interestingly, Jackson and Cheetham (1990,1994) examined 7 living Bryozoan species from this genus and confirmed that the morphospecies identified differed significantly from one another at a number of allozyme loci.]

Nevertheless, other examples exist of fairly gradual evolutionary change.

For example, Sheldon (19) studied 3458 specimens from eight trilobite lineages.

These lineages showed gradual change of a sufficiently pronounced nature that the specimens at the beginning and end of each lineage would be classified as different species (and in one case a different genus).



Such examples illustrate that punctuated and gradualist processes can both occur.

Reviewing 58 such studies, Erwin and Anstey (1995) conclude:

> "Paleontological evidence overwhelmingly supports a view that speciation is sometimes gradual and sometimes punctuated, and that no one mode characterizes this very complicated process."

Eldredge and Gould also argue that "stasis is data", which should play a more prominent role in evolutionary explanations.

(2) What explains punctuated evolution?

Why might morphological evolution be rapid around speciation events?

Why might morphological evolution be relatively static during other periods of time?

Eldredge and Gould's (1972) explanation (following Mayr): Peripatric speciation of a small isolated population might lead to rapid changes in a daughter population (drift), whereas large parental populations remain relatively unchanged.

Gould and Eldredge's (1993) explanation (following Futuyma): Populations are constantly changing, but genetic mixture across populations prevents sustained differences from accumulating. Speciation "locks up" the changes that a population has undergone.

Alternative explanations??

Directionality in Evolution

> "It is absurd to talk of one animal being higher than another...*We* consider those, where the intellectual faculties most developed, as highest. -- A bee doubtless would [use]...instincts."
>
> Charles Darwin's Notebooks 1833-1844 (B46, 74)

"Progress" is a thorny concept in evolution, since it implies that there is a goal towards which evolution proceeds.

Natural selection and mutation are "myopic" processes: they act in the present and have no foresight.

Nevertheless, change does occur and often follows a particular trend (with exceptions).

A directional trend has been argued to occur along the following axes:

- Size
- Complexity

Size

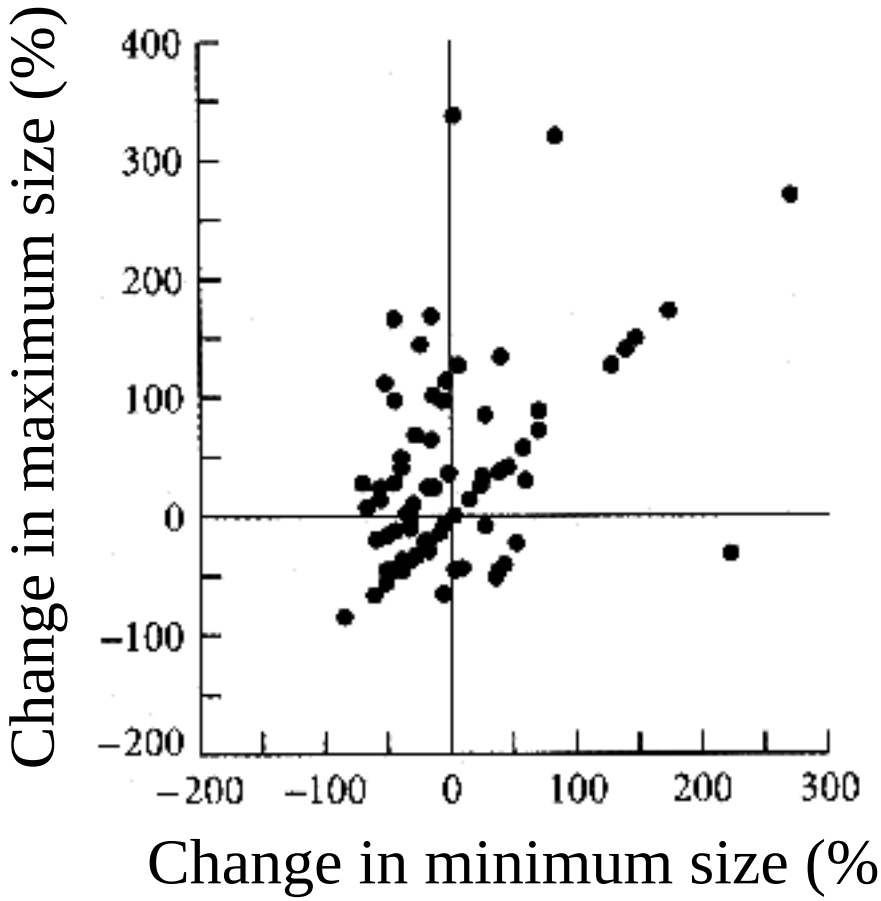
Cope's rule: Body size increases within a lineage over evolutionary time.

This rule has often been explained by the potential advantages of being large: increased defense, mating success, foraging success, improved homeostasis (= sustaining a constant state in a changing environment).

However, we tend to focus on extreme cases where increased body size has clearly increased. Is Cope's rule generally true?

Jablonski (1996) examined 191 bivalve and gastropod lineages over a 16 MY period, in the most extensive study of Cope's rule.

Bivalve genera (N=81)



[Figure from Futuyma (1998).]

The body size of the largest species within a genus often increased (top half) but also decreased 36% of the time.

Interestingly, the body size of the smallest species within a genus decreased (left half) more often than it increased (36% of the time).

This suggests that the most prevalent pattern is one of *increased variability* rather than a trend towards larger size.

Complexity

How might complexity be measured??

A cautionary note: There is a definite risk in defining complexity that we are simply seeking "most human-like".

1. Genome size

Species	Genome Size*
Escherichia coli (bacteria)	0.005
Saccharomyces cerevisiae (yeast)	0.009
Drosophila melanogaster	0.18
Arabidopsis thaliana (a weed)	0.2
Homo sapiens	3.5
Triturus cristatus (a newt)	19
Fritillaria assyriaca (a monocot plant)	127
Protopterus aethiopicus (a lungfish)	142

[*Haploid genome size, measured in picograms (1 pg ~ 10^9 base pairs) from Maynard Smith (1989)]

What may account for these differences?

Assuming that the common ancestor to all living organisms had a small genome size (~bacterial in size), it would be easier for mutations and selection to increase genome size than the reverse.

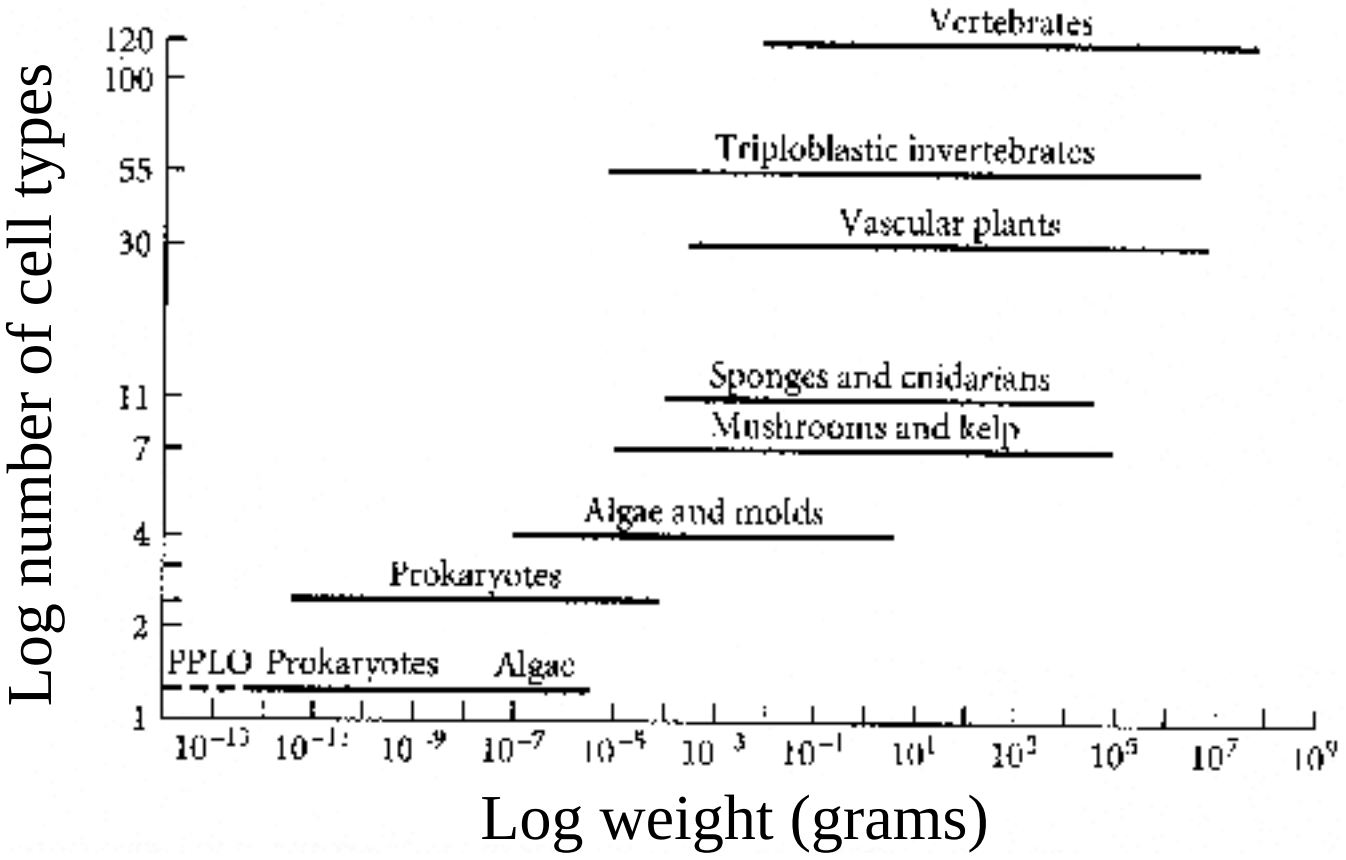
Junk DNA may accumulate as the result of transposable elements (or other repeat elements) copying themselves throughout the genome.

There are more coding sequences in some organisms than others: E. coli have ~4000 genes, yeast ~6000 genes, Drosophila ~10,000 genes, humans ~100,000 genes.

Does the complexity of an organism double if it becomes tetraploid but otherwise looks the same?

2. Number of cell types

The total number of recognizably different cell types is much larger in vertebrates than in invertebrates, plants, fungi, etc.



Nevertheless, there is no evidence that the number of cell types has increased within any of these phyla since the Cambrian (Futuyma, 1998).

Most of the net trend toward an increased number of cell types was established early in evolution (before the Cambrian).

Again, assuming that the common ancestor to all living organisms had one cell, the only direction in which evolution could proceed is up.

"Our strong and biased predilection for focusing on extremes...generates all manner of deep and stubborn errors. Most notable of these misconceptions is the false and self-serving notion that evolution displays a central and general thrust towards increasing complexity, when life, in fact, has been dominated by its persistent bacterial mode for all 3.5 billion years of its history on Earth."

-- Stephen J. Gould (1997, Nature 385: 199-200)

It may be frustrating that we cannot draw broad and sweeping generalizations about evolutionary processes.

Evolution may occur rapidly...or slowly.

Evolution may increase size...or decrease it.

Evolution may lead to greater complexity...or greater simplicity.

Yet the resulting view that evolution is a complex process leading to a richness in the forms and varieties of life is, in its own way, satisfying.

> "It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms...have all been produced by laws acting around us....There is grandeur in this view of life."
>
> -- C. Darwin (Origin of Species, 6th edition)

SOURCES OF INFORMATION AND FIGURES:

- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Freeman and Herron (1998) Evolutionary Analysis. Prentice Hall, NJ.
- Gould and Eldredge (1993) Punctuated equilibrium comes of age. Nature 366:223-227.
- Eldredge and Gould (1972) Punctuated equilibria: An alternative to phyletic gradualism, in Models in Paleobiology.
- Jablonski (1997) Body-size evolution in Cretaceous molluscs and the status of Cope's rule. Nature 385:250-252.

The Human Factor in Evolution

Nearly 6,000,000,000 humans currently inhabit the earth.

Massive human population growth has been supported by agricultural, industrial, and technological developments, which have forever altered the course of earth's evolutionary history.

Homo sapiens has been the first species to impose its own subjective assessments of the "quality" of members of other species, redirecting selection for its own purposes.

➡ Artificial selection

Example: Average egg production rose from 20 eggs/chicken to over 100 eggs/chicken between 1933 and 1965.

Example: Mean corn cob length has increased from about 1 cm (in teosinte) to more than 12 cm over a period of 7000 years.

Humans have recently caused directed mutations (via genetic engineering) in an attempt to hasten the evolutionary response.

Example: Genes for human growth hormone injected into mice zygotes resulted in giant mice that grew to twice the normal adult size. Giant mice then passed on the gene to their offspring.

In this lecture, we will examine two aspects of how humans have altered the tempo and mode of evolutionary change:

- human-induced mass extinction
- cultural evolution

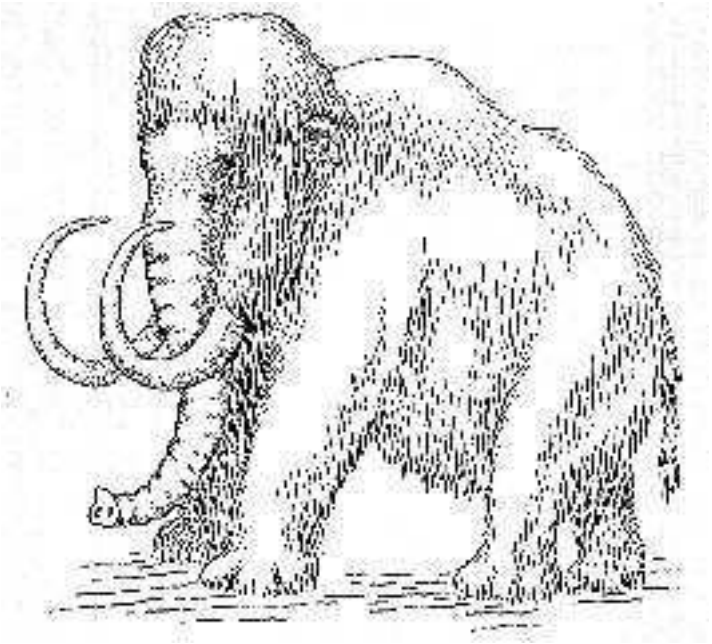
The Next Mass Extinction

From the early evolution and spread of *Homo sapiens*, our species has been associated with the disappearance of species.

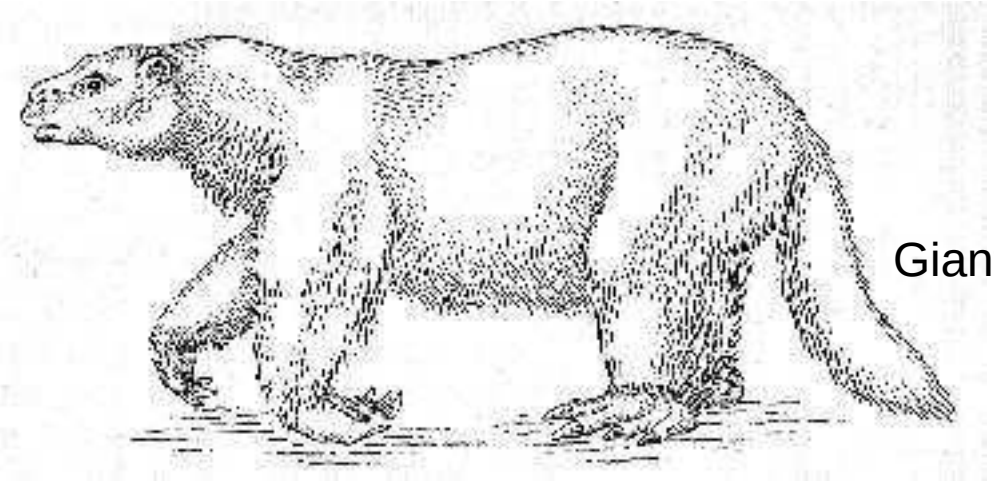
During the last 10,000 to 100,000 years, a number of large, terrestrial mammals have gone extinct.

In North America, 70% (38/54) of the large mammal species went extinct, including:

- an endemic camel
- an endemic musk ox
- the mastodon
- the wooly mammoth
- the saber-toothed cat
- a large beaver (~ 150 kg!!)
- the American cheetah
- two species of tapir
- the giant ground sloth



Wooly Mammoth



Giant Sloth

Two possible explanations have been advanced for these extinctions: climate changes over the last 100,000 years and/or hunting by early humans.

What is unusual about these extinction events is that they were so selective: small-bodied mammals, marine invertebrates, plants, insects, and birds were unaffected.

Such selectivity is consistent with the idea that large mammals went extinct from hunting, but this is still hotly debated.

A clearer example of pre-historic human-induced extinctions comes from island species of birds.

Steadman (1995) estimated that 2000 species of birds have gone extinct on Pacific islands following human colonization.

The evidence presented by Steadman and colleagues includes:

- Loss of 60 species on the Hawaiian islands following the arrival of humans 1500 ybp.
- Loss of 44 species (including moas) in New Zealand following human colonization
- Loss of 21/27 species of birds on ’Eua (Tonga) following human arrival
- Loss of 20 taxa on the Galapagos islands in the last two hundred years (compared to three in the preceding 4000-8000 years)

What caused these losses?

Evidence in early fire pits and middens suggests that humans directly consumed many of these bird species.

Humans brought to these islands additional threats to island species: mammalian predators, (especially rats and dogs) and diseases (e.g. birdpox introduced into Hawaii in the 1890's).

Finally, humans altered island habitats, via slash-and-burn agriculture, deforestation, and irrigation.

→ Extinctions have been a hallmark of human history.

The Crisis Expands

Sadly, many species have gone extinct during recent human history and many more remain greatly threatened.

Current extinction rates are estimated to be 100-1000 times above background levels (estimates from such eminent ecologists as Stuart Pimm, Paul Ehrlich, and E. O. Wilson).

	# species certified extinct since 1600
Molluscs	191
Insects	61
Vertebrates	229
Plants	584

[From Smith et al. (1993); estimates are underestimates for poorly studied groups.]

Gone forever are the dodo bird, the passenger pigeon, the great auk, the Carolina parakeet, the quagga, Stellar's sea cow, Schomburgk's deer, Antarctic wolf, the Tasmanian wolf...

But while these numbers are large, they do not (YET!) approach the levels of the "big five" mass extinction events.

The *known* losses over the last 400 years represent around 1% or less of extant species (compared to the 50%-90% losses experienced during the "big five").

But extend these losses for a few thousand years and the numbers will become comparable.

And the situation will worsen.

Increased demands on resources:

Human population size continues to grow and is expected to double within the next forty years.

As countries develop, human demand on land and fossil fuels increases faster than population growth.

Resource	US consumption (per capita)	Indian consumption (per capita)
Petroleum	2.53 metric tons	0.057 metric tons
Natural gas	81.3 metric tons	0.47 metric tons

Reduced resource base:

By 1989, tropical forests had declined to about 50% of their pre-human levels, declining at a further rate of 1.8% per year (and growing).

> "In 1989 the surviving rain forests occupied an area about that of the contiguous forty-eight states of the United States, and they were being reduced by an amount equivalent to the size of Florida each year."
>
> --E. O. Wilson (1992) *The Diversity of Life*

Land area of forests in Costa Rica

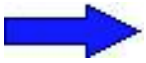
These tropical rain forests are thought to contain at least half of the world's species, 10 to 25 percent of which are expected to go extinct within the next 30 years.

Deterioration of remaining habitats:

In addition to deforestation, the physical environment of all living species has been adversely affected by

- air pollution
- water pollution
- loss of top-soil
- acid rain
- depletion of ozone layer
- rise in global temperatures

Similarly, the biotic environment has been adversely affected by the introduction of non-native competitors, predators, and diseases.

 As many as 150-200 species may be going extinct every day (estimate from Rio+5 group sponsored by the Earth Council).

As with each mass extinction, the process of evolution will forever-more be shaped by those fortunate species that survive decimation.

Unlike the former mass extinctions, however, there is one major characteristic that defines who will survive and who will perish: ability to co-exist and thrive alongside human populations.

Cockroaches, pigeons, starlings, rats, dandelions, and similar species are the likely winners in this, the sixth, mass extinction.

Cultural Evolution

Evolutionary change in a trait occurs whenever:

- individuals vary in some trait (VARIANCE)
- individuals with some trait values are more likely to live and/or reproduce than other individuals (SELECTION)
- trait values may be passed from individual to individual (HERITABILITY)

Although we have focused on the evolution of traits that are passed from parents to offspring via genetic inheritance, evolution can also act on traits that are "inherited" culturally.

Traditions, ideas, phrases, fads, or skills can be passed from individual to individual, not through genes but through communication.

Besides vertical transmission (= from parents to offspring), cultural transmission can occur horizontally (= among peers) or obliquely (= from any member of an older generation to any member of a younger generation).

Example: Cavalli-Sforza et al (1981) collected information about the cultural inheritance of a variety of traits.

Salt consumption

Father x Mother	"High" offspring
"High x High"	60%
"High x Low"	36%
"Low x High"	65%
"Low x Low"	26%

(Example of maternal transmission.)

Political Interest

Father x Mother	"High" offspring
"High x High"	72%
"High x Low"	44%
"Low x High"	40%
"Low x Low"	25%

(Example of biparental transmission.)

Richard Dawkins (1976) coined the term "meme" to refer to the unit of cultural evolution (analogous to gene).

Examples of memes: roller-blades [inventions], "hey dude" [phrases], lattes [great ideas].

Like genes, memes can "mutate" during transmission.

Like mutant alleles, mutant memes can be selectively favored (high transmission) or disfavored (low transmission).

For instance, Imo, a two-year old Japanese macaque was the first to learn to wash sand off of sweet potatoes by throwing them in water. This behavioral "mutation" then rapidly spread, via learning, to other members of her troop.

Cultural and genetic evolution are not mutually exclusive, as demonstrated by the example of lactose tolerance.

Adult mammals are generally lactose intolerant.

With the increase in dairy farming over 6000 years, however, lactose tolerance has increased from about 20% (among populations without dairy traditions) to 90%.

The change in lactose absorption is due mainly to genetic differences, but the change in farming practises that led to selection for absorption was cultural.

Cultural evolution can be rapid and can change the way we view the world.

A final thought:

➡ What sorts of "memes" may help slow the rate of extinctions caused by humankind?

What factors may favor the spread of such "memes"?

SOURCES OF INFORMATION AND FIGURES:

- Futuyma (1998) Evolutionary Biology. Sinauer Associates, MA.
- Freeman and Herron (1998) Evolutionary Analysis. Prentice Hall, NJ.
- Cavalli-Sforza and Feldman (1981) Cultural Transmission and Evolution.
- Feldman and Laland (1981) Gene-cultural coevolutionary theory. TREE 11:453-457.