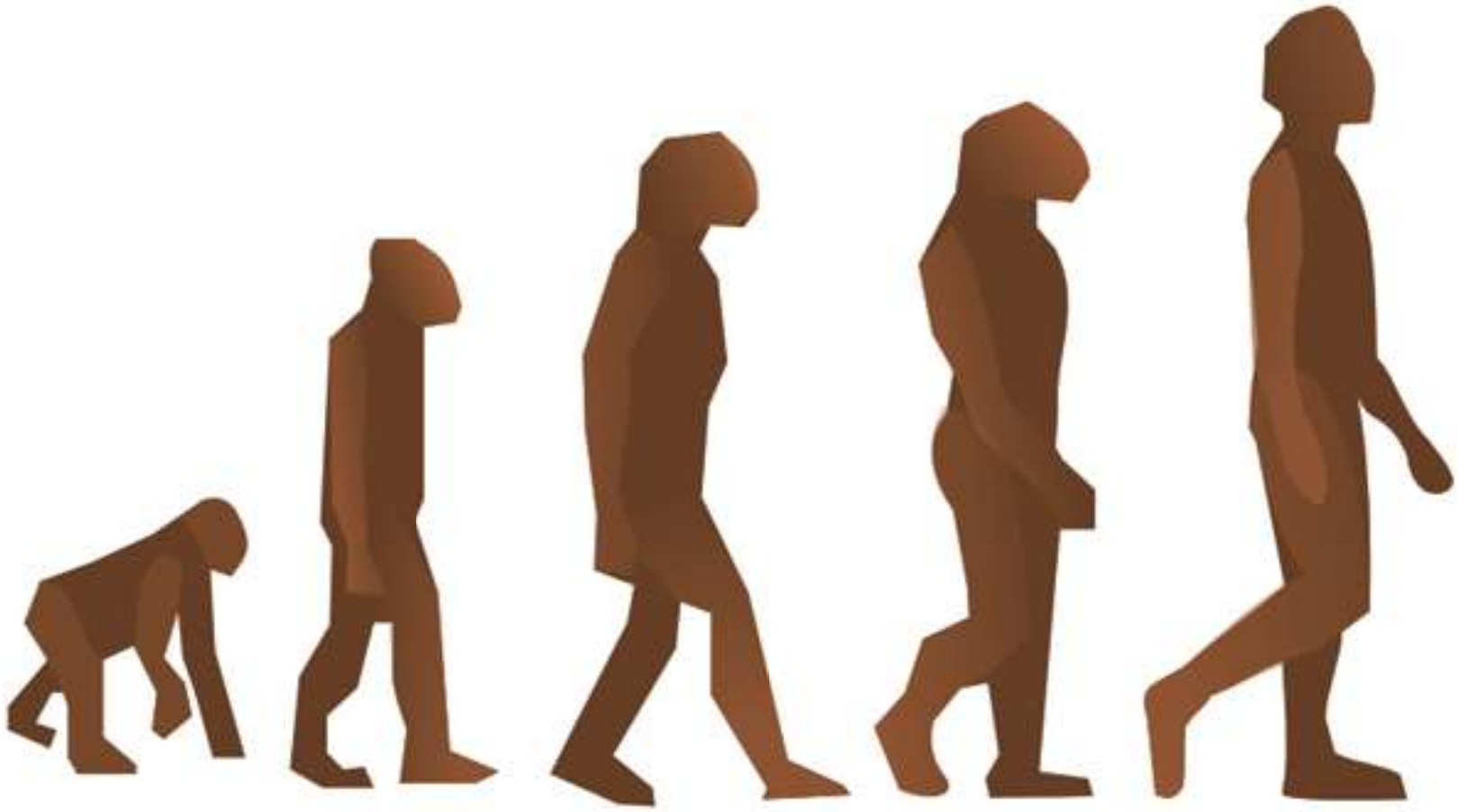
Essence, Key Concepts and Elements of Evolutionary Biology Field



Lakia Schweitzer
Suanne Yarbrough

First Edition, 2012

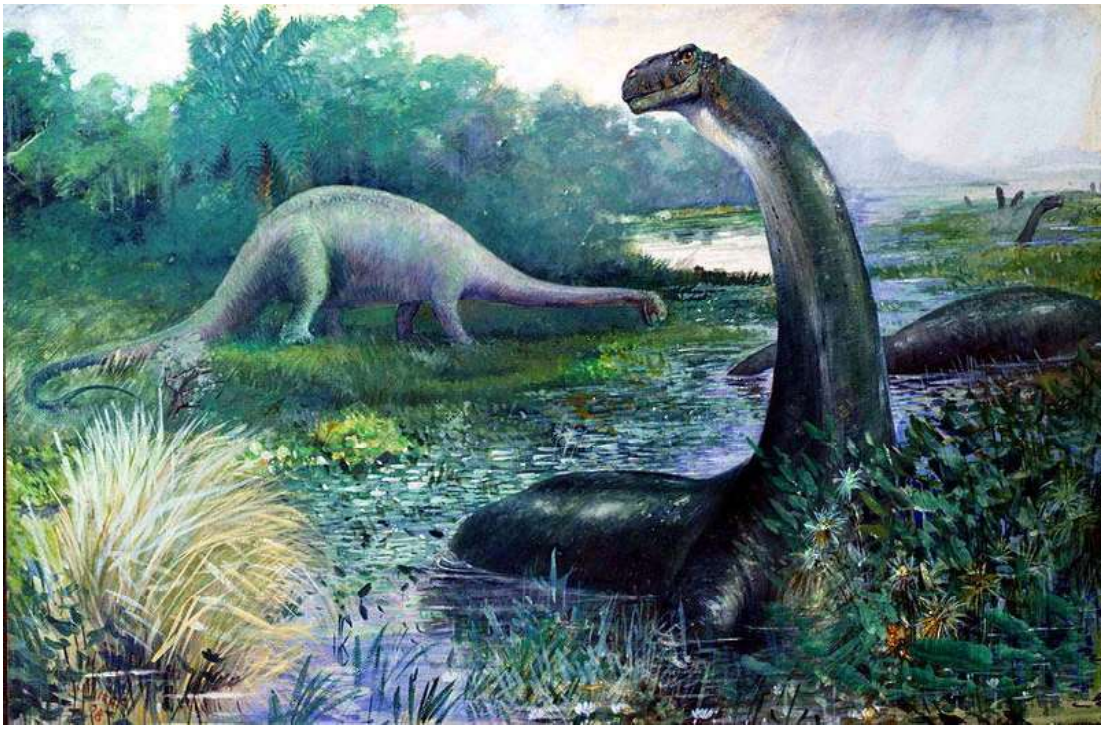
ISBN 978-81-323-0704-4

Published by:
Academic Studio
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter- 1

Introduction to Evolution



Natural selection does not lead to perfection; dramatic changes in the environment often lead to mass extinctions, as in the case of the dinosaurs nearly 65 million years ago.

Evolution is the process of change in all forms of life over generations, and evolutionary biology is the study of how evolution occurs.

The biodiversity of life evolves by means of natural selection, mutations and genetic drift. The process of natural selection is based on three conditions. First, every individual is supplied with hereditary material in the form of genes that are received from their parents then passed on to their offspring. Second, organisms tend to produce more offspring than the environment can support. Third, there are variations among offspring as a consequence of either the random introduction of new genes via mutations or reshuffling of existing genes during sexual reproduction.

Natural selection will occur when these facts of nature (heredity, overproduction of offspring, and variation) hold true. Natural selection means individuals do not have equal chances of reproductive success. As a consequence, some individuals produce more offspring and thus have a higher degree of fitness. Traits that ensure organisms are better adapted to their living conditions become more common in descendant populations. For this reason, populations will never remain exactly the same over successive generations. The forces of evolution are most evident when populations become isolated, either through geographic distance or by mechanisms that prevent genetic exchange. Over time, isolated populations can branch off into new species.

Random genetic drift describes another natural process that regulates the evolution of minor mutations in the genes, leading to changes in allele frequencies over time. These smaller mutations occur with regular frequency in and among populations. The vast majority of genetic mutations neither assist, change the appearance of, nor bring harm to individuals. These mutated genes are neutrally sorted among populations and survive across generations by chance alone. When species migrate they carry different genetic varieties to different places. When organisms mate they exchange genetic material and new individuals are born.

The outward expression of each unique genetic mixture in different environments, the phenotype, comprises a diversity of traits that can be measured and observed as individuals grow and develop. In contrast to genetic drift, natural selection is not a random process because it acts on traits that become adaptive for their functional utilities that are necessary for survival. Natural selection and random genetic drift are forever constant and dynamic parts of life. More than 99.9% of all species have become extinct since life began over 3.5 billion years ago. Evolution is more death than survival and over time this has shaped the branching structure in the tree of life.

The modern understanding of evolution began with the 1859 publication of Charles Darwin's *On the Origin of Species*. In addition, Gregor Mendel's work with plants helped to explain the hereditary patterns of genetics. Fossil discoveries in paleontology, advances in population genetics and a global network of scientific research have provided further details into the mechanisms of evolution. Scientists now have a good understanding of the origin of new species (speciation) and have observed the speciation process in the laboratory and in the wild. Evolution is the principal theory that biologists use to understand life and is used in many disciplines, including medicine, psychology, conservation biology, anthropology, forensics, agriculture and other social-cultural applications.

Darwin's idea: evolution by natural selection

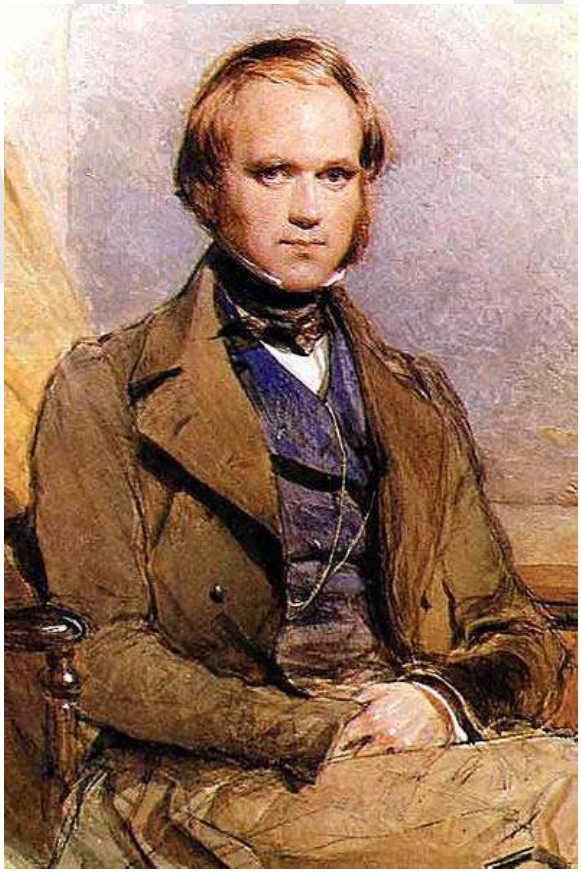
In the 19th century, natural history collections and museums were a popular pastime. The European expansion and naval expeditions employed naturalists and curators of grand museums showcasing preserved and live specimens of the varieties of life. Charles Darwin was an English graduate who was educated and trained in the disciplines of natural history science. Such natural historians would collect, catalogue, describe and

study the vast collections of specimens stored and managed by curators at these museums. Charles Darwin served as a ship's naturalist on board the HMS Beagle, assigned to a five-year research expedition around the world. During his voyage, Darwin observed and collected an abundance of organisms, being very interested in the diverse forms of life along the coasts of South America and the neighboring Galapagos Islands.

Charles Darwin gained extensive experience as he collected and studied the natural history of life forms from distant places. Through his studies, Darwin formulated the idea that each species had developed from ancestors with similar features. In 1838, he described how a process he called natural selection would make this happen.

Darwin's idea of how evolution works relied on the following observations:

- If all the individuals of a species reproduced successfully, the population of that species would increase uncontrollably.
- Populations tend to remain about the same size from year to year.
- Environmental resources are limited.
- No two individuals in a given species are exactly alike.
- Much of this variation in a population can be passed on to offspring.



Charles Darwin proposed the theory of evolution by natural selection

Darwin noted that orchids exhibited a variety of complex adaptations to ensure pollination; all derived from basic floral parts.

Darwin deduced that since organisms produce more offspring than their environment could possibly support, there must be a competitive struggle for survival—only a few individuals can survive out of each generation. Darwin realized that it was not chance alone that determined survival. Instead, survival depends on the traits of each individual and if these traits aid or hinder survival and reproduction. Well-adapted, or "fit", individuals are likely to leave more offspring than their less well-adapted competitors. Darwin realized that the unequal ability of individuals to survive and reproduce could cause gradual changes in the population. Traits that help an organism survive and reproduce would accumulate over generations. On the other hand, traits that hinder survival and reproduction would disappear. Darwin used the term *natural selection* to describe this process.

Natural selection is commonly equated with *survival of the fittest*, but this expression originated in Herbert Spencer's *Principles of Biology* in 1864, after Charles Darwin published his original works. *Survival of the fittest* describes the process of natural selection incorrectly, because natural selection is not only about survival and it is not always the fittest that survives.

Observations of variations in animals and plants formed the basis of the theory of natural selection. For example, Darwin observed that orchids and insects have a close relationship that allows the pollination of the plants. He noted that orchids have a variety of structures that attract insects, so that pollen from the flowers gets stuck to the insects' bodies. In this way, insects transport the pollen from a male to a female orchid. In spite of the elaborate appearance of orchids, these specialized parts are made from the same basic structures that make up other flowers. In *Fertilisation of Orchids* Darwin proposed that the orchid flowers did not represent the work of an ideal engineer, but were adapted from pre-existing parts, through natural selection.

Darwin was still researching and experimenting with his ideas on natural selection when he received a letter from Alfred Wallace describing a theory very similar to his own. This led to an immediate joint publication of both theories. Both Wallace and Darwin saw the history of life like a family tree, with each fork in the tree's limbs being a common ancestor. The tips of the limbs represented modern species and the branches represented the common ancestors that are shared amongst many different species. To explain these relationships, Darwin said that all living things were related, and this meant that all life must be descended from a few forms, or even from a single common ancestor. He called this process *descent with modification*.

Darwin published his theory of evolution by natural selection in *On the Origin of Species* in 1859. His theory means that all life, including humanity, is a product of continuing natural processes. The implication that all life on Earth has a common ancestor has met with objections from some religious groups who believe even today that the different types of life are due to special creation. Their objections are in contrast to the level of support for the theory by more than 99 percent of those within the scientific community today.

Source of variation

Darwin's theory of natural selection laid the groundwork for modern evolutionary theory, and his experiments and observations showed that the organisms in populations varied from each other, that some of these variations were inherited, and that these differences could be acted on by natural selection. However, he could not explain the source of these variations. Like many of his predecessors, Darwin mistakenly thought that heritable traits were a product of use and disuse, and that features acquired during an organism's lifetime could be passed on to its offspring. He looked for examples, such as large ground feeding birds getting stronger legs through exercise, and weaker wings from not flying until, like the ostrich, they could not fly at all. This misunderstanding was called the inheritance of acquired characters and was part of the theory of transmutation of species put forward in

1809 by Jean-Baptiste Lamarck. In the late 19th century this theory became known as Lamarckism. Darwin produced an unsuccessful theory he called pangenesis to try to explain how acquired characteristics could be inherited. In the 1880s August Weismann's experiments indicated that changes from use and disuse could not be inherited, and Lamarckism gradually fell from favor.

The missing information needed to help explain how new features could pass from a parent to its offspring was provided by the pioneering genetics work of Gregor Mendel. Mendel's experiments with several generations of pea plants demonstrated that inheritance works by separating and reshuffling hereditary information during the formation of sex cells and recombining that information during fertilization. This is like mixing different hands of cards, with an organism getting a random mix of half of the cards from one parent, and half of the cards from the other. Mendel called the information *factors*; however, they later became known as genes. Genes are the basic units of heredity in living organisms. They contain the information that directs the physical development and behavior of organisms.

Genes are made of DNA, a long molecule that carries information. This information is encoded in the sequence of nucleotides in the DNA, just as the sequence of the letters in words carries information on a page. The genes are like short instructions built up of the "letters" of the DNA alphabet. Put together, the entire set of these genes gives enough information to serve as an "instruction manual" of how to build and run an organism. The instructions spelled out by this DNA alphabet can be changed, however, by mutations, and this may alter the instructions carried within the genes. Within the cell, the genes are carried in chromosomes, which are packages for carrying the DNA, with the genes arranged along them like beads on a string. It is the reshuffling of the chromosomes that results in unique combinations of genes in offspring.

Although such mutations in DNA are random, natural selection is not a process of chance: the environment determines the probability of reproductive success. The end products of natural selection are organisms that are adapted to their present environments. Natural selection does not involve progress towards an ultimate goal. Evolution does not necessarily strive for more advanced, more intelligent, or more sophisticated life forms. For example, fleas (wingless parasites) are descended from a winged, ancestral scorpionfly, and snakes are lizards that no longer require limbs - although pythons still grow tiny structures that are the remains of their ancestor's hind legs. Organisms are merely the outcome of variations that succeed or fail, dependent upon the environmental conditions at the time.

Rapid environmental changes typically cause extinctions. Of all species that have existed on Earth, 99.9 percent are now extinct. Since life began on Earth, five major mass extinctions have led to large and sudden drops in the variety of species. The most recent, the Cretaceous–Tertiary extinction event, occurred 65 million years ago, and has attracted more attention than all others because it killed the dinosaurs.

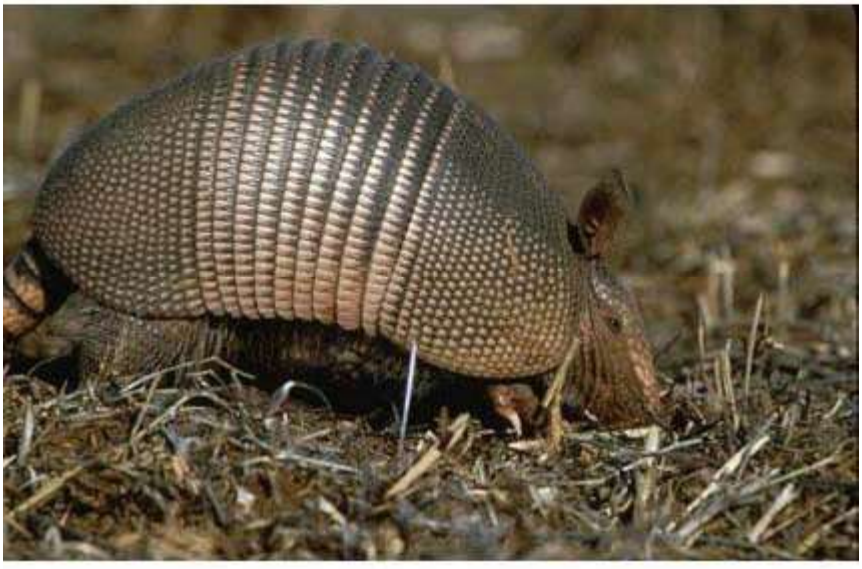
Modern synthesis

The modern evolutionary synthesis was the outcome of a merger of several different scientific fields into a cohesive understanding of evolutionary theory. In the 1930s and 1940s, efforts were made to merge Darwin's theory of natural selection, research in heredity, and understandings of the fossil records into a unified explanatory model. The application of the principles of genetics to naturally occurring populations, by scientists such as Theodosius Dobzhansky and Ernst Mayr, advanced understanding of the processes of evolution. Dobzhansky's 1937 work *Genetics and the Origin of Species* was an important step in bridging the gap between genetics and field biology. Mayr, on the basis of an understanding of genes and direct observations of evolutionary processes from field research, introduced the biological species concept, which defined a species as a group of interbreeding or potentially interbreeding populations that are reproductively isolated from all other populations. The paleontologist George Gaylord Simpson helped to incorporate fossil research, which showed a pattern consistent with the branching and non-directional pathway of evolution of organisms predicted by the modern synthesis.

The modern synthesis emphasizes the importance of populations as the unit of evolution, the central role of natural selection as the most important mechanism of evolution, and the idea of gradualism to explain how large changes evolve as an accumulation of small changes over long periods of time.

Evidence for evolution

Glyptodon - lived 2,500,000 to 10,000 years ago



Armadillo - lives presently in the Americas

During the voyage of the *Beagle*, naturalist Charles Darwin collected fossils in South America, and found fragments of armor which he thought were like giant versions of the scales on the modern armadillos living nearby. On his return, the anatomist Richard Owen showed him that the fragments were from gigantic extinct glyptodons, related to the armadillos. This was one of the patterns of distribution that helped Darwin to develop his theory.

Scientific evidence for evolution comes from many aspects of biology, and includes fossils, homologous structures, and molecular similarities between species' DNA.

Fossil record

Research in the field of paleontology, the study of fossils, supports the idea that all living organisms are related. Fossils provide evidence that accumulated changes in organisms over long periods of time have led to the diverse forms of life we see today. A fossil itself reveals the organism's structure and the relationships between present and extinct species, allowing paleontologists to construct a family tree for all of the life forms on Earth.

Modern paleontology began with the work of Georges Cuvier (1769–1832). Cuvier noted that, in sedimentary rock, each layer contained a specific group of fossils. The deeper layers, which he proposed to be older, contained simpler life forms. He noted that many forms of life from the past are no longer present today. One of Cuvier's successful contributions to the understanding of the fossil record was establishing extinction as a fact. In an attempt to explain extinction, Cuvier proposed the idea of "revolutions" or catastrophism in which he speculated that geological catastrophes had occurred throughout the Earth's history, wiping out large numbers of species. Cuvier's theory of revolutions was later replaced by uniformitarian theories, notably those of James Hutton and Charles Lyell who proposed that the Earth's geological changes were gradual and consistent. However, current evidence in the fossil record supports the concept of mass extinctions. As a result, the general idea of catastrophism has re-emerged as a valid hypothesis for at least some of the rapid changes in life forms that appear in the fossil records.

A very large number of fossils have now been discovered and identified. These fossils serve as a chronological record of evolution. The fossil record provides examples of transitional species that demonstrate ancestral links between past and present life forms. One such transitional fossil is *Archaeopteryx*, an ancient organism that had the distinct characteristics of a reptile (such as a long, bony tail and conical teeth) yet also had characteristics of birds (such as feathers and a wishbone). The implication from such a find is that modern reptiles and birds arose from a common ancestor.

Comparative anatomy

The comparison of similarities between organisms of their form or appearance of parts, called their morphology, has long been a way to classify life into closely related groups. This can be done by comparing the structure of adult organisms in different species or by comparing the patterns of how cells grow, divide and even migrate during an organism's development.

Taxonomy

Taxonomy is the branch of biology that names and classifies all living things. Scientists use morphological and genetic similarities to assist them in categorizing life forms based on ancestral relationships. For example, orangutans, gorillas, chimpanzees, and humans all belong to the same taxonomic grouping referred to as a family – in this case the family

called *Hominidae*. These animals are grouped together because of similarities in morphology that come from common ancestry (called *homology*).

Strong evidence for evolution comes from the analysis of *homologous* structures: structures in different species that no longer perform the same task but which share a similar structure. Such is the case of the forelimbs of mammals. The forelimbs of a human, cat, whale, and bat all have strikingly similar bone structures. However, each of these four species' forelimbs performs a different task. The same bones that construct a bat's wings, which are used for flight, also construct a whale's flippers, which are used for swimming. Such a "design" makes little sense if they are unrelated and uniquely constructed for their particular tasks. The theory of evolution explains these homologous structures: all four animals shared a common ancestor, and each has undergone change over many generations. These changes in structure have produced forelimbs adapted for different tasks.

Embryology

In some cases, anatomical comparison of structures in the embryos of two or more species provides evidence for a shared ancestor that may not be obvious in the adult forms. As the embryo develops, these homologies can be lost to view, and the structures can take on different functions. Part of the basis of classifying the vertebrate group (which includes humans), is the presence of a tail (extending beyond the anus) and pharyngeal slits. Both structures appear during some stage of embryonic development but are not always obvious in the adult form.

Because of the morphological similarities present in embryos of different species during development, it was once assumed that organisms re-enact their evolutionary history as an embryo. It was thought that human embryos passed through an amphibian then a reptilian stage before completing their development as mammals. Such a re-enactment, (often called *Recapitulation theory*), is not supported by scientific evidence. What does occur, however, is that the first stages of development are similar in broad groups of organisms. At very early stages, for instance, all vertebrates appear extremely similar, but do not exactly resemble any ancestral species. As development continues, specific features emerge from this basic pattern.

Vestigial structures

Homology includes a unique group of shared structures referred to as *vestigial structures*. *Vestigial* refers to anatomical parts that are of minimal, if any, value to the organism that possesses them. These apparently illogical structures are remnants of organs that played an important role in ancestral forms. Such is the case in whales, which have small vestigial bones that appear to be remnants of the leg bones of their ancestors which walked on land. Humans also have vestigial structures, including the ear muscles, the wisdom teeth, the appendix, the tail bone, body hair (including goose bumps), and the semilunar fold in the corner of the eye.

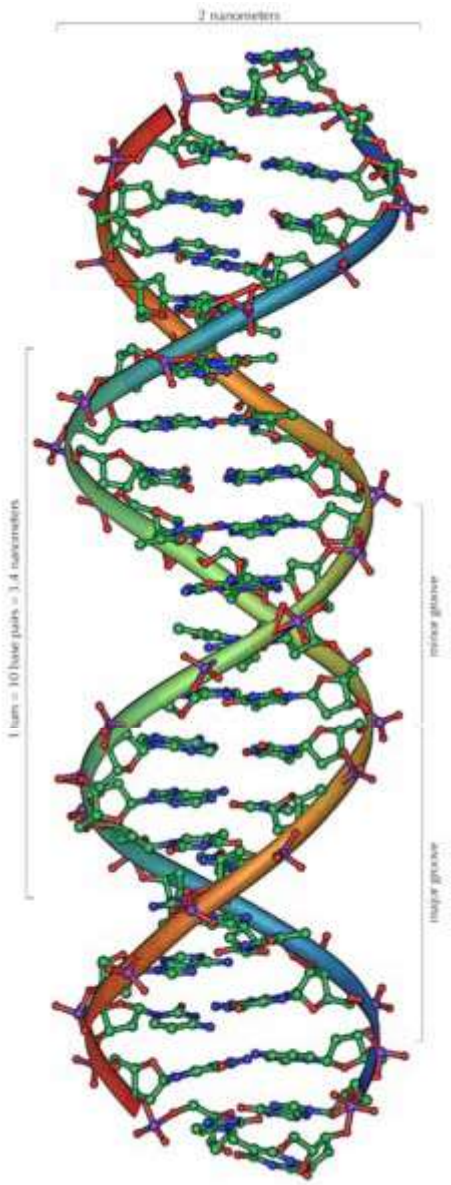
Convergent evolution

The bird and the bat wing are examples of convergent evolution

A bat is a mammal and its forearm bones have been adapted for flight

Anatomical comparisons can be misleading, as not all anatomical similarities indicate a close relationship. Organisms that share similar environments will often develop similar physical features, a process known as *convergent evolution*. Both sharks and dolphins have similar body forms, yet are only distantly related – sharks are fish and dolphins are mammals. Such similarities are a result of both populations being exposed to the same selective pressures. Within both groups, changes that aid swimming have been favored. Thus, over time, they developed similar appearances (morphology), even though they are not closely related.

Molecular biology



A section of DNA

Every living organism (with the possible exception of RNA viruses) contains molecules of DNA, which carries genetic information. Genes are the pieces of DNA that carry this information, and they influence the properties of an organism. Genes determine an individual's general appearance and to some extent their behavior. If two organisms are closely related, their DNA will be very similar. On the other hand, the more distantly related two organisms are, the more differences they will have. For example, brothers are closely related and have very similar DNA, while cousins share a more distant relationship and have far more differences in their DNA. Similarities in DNA are used to determine the relationships between species in much the same manner as they are used to show relationships between individuals. For example, comparing chimpanzees with gorillas and humans shows that there is as much as a 96 percent similarity between the DNA of humans and chimps. Comparisons of DNA indicate that humans and chimpanzees are more closely related to each other than either species is to gorillas.

The field of molecular systematics focuses on measuring the similarities in these molecules and using this information to work out how different types of organisms are related through evolution. These comparisons have allowed biologists to build a *relationship tree* of the evolution of life on Earth. They have even allowed scientists to unravel the relationships between organisms whose common ancestors lived such a long time ago that no real similarities remain in the appearance of the organisms.

Co-evolution

Co-evolution is a process in which two or more species influence the evolution of each other. All organisms are influenced by life around them; however, in co-evolution there is evidence that genetically determined traits in each species directly resulted from the interaction between the two organisms.

An extensively documented case of co-evolution is the relationship between *Pseudomyrmex*, a type of ant, and the acacia, a plant that the ant uses for food and shelter. The relationship between the two is so intimate that it has led to the evolution of special structures and behaviors in both organisms. The ant defends the acacia against herbivores and clears the forest floor of the seeds from competing plants. In response, the plant has evolved swollen thorns that the ants use as shelter and special flower parts that the ants eat. Such co-evolution does not imply that the ants and the tree choose to behave in an altruistic manner. Rather, across a population small genetic changes in both ant and tree benefited each. The benefit gave a slightly higher chance of the characteristic being passed on to the next generation. Over time, successive mutations created the relationship we observe today.

Artificial selection

The results of artificial selection: a Chihuahua mix and a Great Dane

Artificial selection is the controlled breeding of domestic plants and animals. Humans determine which animal or plant will reproduce and which of the offspring will survive; thus, they determine which genes will be passed on to future generations. The process of artificial selection has had a significant impact on the evolution of domestic animals. For example, people have produced different types of dogs by controlled breeding. The differences in size between the Chihuahua and the Great Dane are the result of artificial selection. Despite their dramatically different physical appearance, they and all other dogs evolved from a few wolves domesticated by humans in what is now China less than 15,000 years ago.

Artificial selection has produced a wide variety of plants. In the case of maize (corn), recent genetic evidence suggests that domestication occurred 10,000 years ago in central Mexico. Prior to domestication, the edible portion of the wild form was small and difficult to collect. Today *The Maize Genetics Cooperation • Stock Center* maintains a

collection of more than 10,000 genetic variations of maize that have arisen by random mutations and chromosomal variations from the original wild type.

In artificial selection the new breed or variety that emerges is the one with random mutations attractive to humans, while in natural selection the surviving species is the one with random mutations useful to it in its non-human environment. In both natural and artificial selection the variations are a result of random mutations, and the underlying genetic processes are essentially the same. Darwin carefully observed the outcomes of artificial selection in animals and plants to form many of his arguments in support of natural selection. Much of his book *On the Origin of Species* was based on these observations of the many varieties of domestic pigeons arising from artificial selection. Darwin proposed that if humans could achieve dramatic changes in domestic animals in short periods, then natural selection, given millions of years, could produce the differences seen in living things today.

Species

There are numerous species of cichlids that demonstrate dramatic variations in morphology.

Given the right circumstances, and enough time, evolution leads to the emergence of new species. Scientists have struggled to find a precise and all-inclusive definition of *species*. Ernst Mayr (1904–2005) defined a species as a population or group of populations whose members have the potential to interbreed naturally with one another to produce viable, fertile offspring. (The members of a species cannot produce viable, fertile offspring with members of *other* species). Mayr's definition has gained wide acceptance among biologists, but does not apply to organisms such as bacteria, which reproduce asexually.

Speciation is the lineage-splitting event that results in two separate species forming from a single common ancestral population. A widely accepted method of speciation is called *allopatric speciation*. Allopatric speciation begins when a population becomes geographically separated. Geological processes, such as the emergence of mountain ranges, the formation of canyons, or the flooding of land bridges by changes in sea level may result in separate populations. For speciation to occur, separation must be substantial, so that genetic exchange between the two populations is completely disrupted. In their separate environments, the genetically isolated groups follow their own unique evolutionary pathways. Each group will accumulate different mutations as well as be subjected to different selective pressures. The accumulated genetic changes may result in separated populations that can no longer interbreed if they are reunited. Barriers that prevent interbreeding are either *prezygotic* (prevent mating or fertilization) or *postzygotic* (barriers that occur after fertilization). If interbreeding is no longer possible, then they will be considered different species.

Usually the process of speciation is slow, occurring over very long time spans; thus direct observations within human life-spans are rare. However speciation has been observed in present day organisms, and past speciation events are recorded in fossils. Scientists have documented the formation of five new species of cichlid fishes from a single common ancestor that was isolated fewer than 5000 years ago from the parent stock in Lake Nagubago. The evidence for speciation in this case was morphology (physical appearance) and lack of natural interbreeding. These fish have complex mating rituals and a variety of colorations; the slight modifications introduced in the new species have changed the mate selection process and the five forms that arose could not be convinced to interbreed.

Different views on the mechanism of evolution

James Hutton

Stephen Jay Gould

Richard Dawkins

The theory of evolution is widely accepted among the scientific community, serving to link the diverse specialty areas of biology. Evolution provides the field of biology with a solid scientific base. The significance of evolutionary theory is best described by the title of a paper by Theodosius Dobzhansky (1900–1975), published in *American Biology Teacher*; "Nothing in Biology Makes Sense Except in the Light of Evolution". Nevertheless, the theory of evolution is not static. There is much discussion within the scientific community concerning the mechanisms behind the evolutionary process. For example, the rate at which evolution occurs is still under discussion. In addition, there are conflicting opinions as to which is the primary unit of evolutionary change – the organism or the gene.

Rate of change

Two views exist concerning the rate of evolutionary change. Darwin and his contemporaries viewed evolution as a slow and gradual process. Evolutionary trees are based on the idea that profound differences in species are the result of many small changes that accumulate over long periods.

The view that evolution is gradual had its basis in the works of the geologist James Hutton (1726–1797) and his theory called "gradualism". Hutton's theory suggests that profound geological change was the cumulative product of a relatively slow continuing operation of processes which can still be seen in operation today, as opposed to catastrophism which promoted the idea that sudden changes had causes which can no longer be seen at work. A uniformitarian perspective was adopted for biological changes. Such a view can seem to contradict the fossil record, which shows evidence of new species appearing suddenly, then persisting in that form for long periods. The paleontologist Stephen Jay Gould (1941–2002) developed a model that suggests that evolution, although a slow process in human terms, undergoes periods of relatively rapid change over only a few thousand or million years, alternating with long periods of relative stability, a model called "punctuated equilibrium" which explains the fossil record without contradicting Darwin's ideas.

Unit of change

A common unit of selection in evolution is the organism. Natural selection occurs when the reproductive success of an individual is improved or reduced by an inherited characteristic, and reproductive success is measured by the number of an individual's surviving offspring. The organism view has been challenged by a variety of biologists as well as philosophers. Richard Dawkins (born 1941) proposes that much insight can be gained if we look at evolution from the gene's point of view; that is, that natural selection operates as an evolutionary mechanism on genes as well as organisms. In his 1976 book *The Selfish Gene*, he explains:

> “Individuals are not stable things, they are fleeting. Chromosomes too are shuffled to oblivion, like hands of cards soon after they are dealt. But the cards themselves survive the shuffling. The cards are the genes. The genes are not destroyed by crossing-over; they merely change partners and march on. Of course they march on. That is their business. They are the replicators and we are their survival machines. When we have served our purpose we are cast aside. But genes are denizens of geological time: genes are forever.”

Others view selection working on many levels, not just at a single level of organism or gene; for example, Stephen Jay Gould called for a hierarchical perspective on selection.

Summary

Several basic observations establish the theory of evolution, which explains the variety and relationship of all living things. There are genetic variations within a population of individuals. Some individuals, by chance, have features that allow them to survive and thrive better than their kind. The individuals that survive will be more likely to have offspring of their own. The offspring might inherit the useful feature.

Evolution is not a random process. While mutations are random, natural selection is not. Evolution is an inevitable result of imperfectly copying, self-replicating organisms reproducing over billions of years under the selective pressure of the environment. The outcome of evolution is not a perfectly designed organism. The outcome is simply an individual that can survive better and reproduce more successfully than its neighbors in a particular environment. Fossils, the genetic code, and the peculiar distribution of life on Earth provide a record of evolution and demonstrate the common ancestry of all organisms, both living and long dead. Evolution can be directly observed in artificial selection, the selective breeding for certain traits of domestic animals and plants. The diverse breeds of cats, dogs, horses, and agricultural plants serve as examples of evolution.

Although some groups raise objections to the theory of evolution, the evidence of observation and experiments that has been gathered for over a hundred years by thousands of scientists supports evolution. The result of four billion years of evolution is the diversity of life around us, with an estimated 1.75 million different species in existence today.

Chapter- 2

Evolution

Evolution (also known as **biological**, **genetic** or **organic evolution**) is the change in the inherited traits of a population of organisms through successive generations. This change results from interactions between processes that introduce variation into a population, and other processes that remove it. As a result, variants with particular traits become more, or less, common. A trait is a particular characteristic—anatomical, biochemical or behavioural—that is the result of gene–environment interaction.

The main source of variation is mutation, which introduces genetic changes. These changes are heritable (can be passed on through reproduction), and may give rise to alternative traits in organisms. Another source of variation is genetic recombination, which shuffles the genes into new combinations which can result in organisms exhibiting different traits. Under certain circumstances, variation can also be increased by the transfer of genes between species, and by the extremely rare, but significant, wholesale incorporation of genomes through endosymbiosis.

Two main processes cause variants to become more common or rarer in a population. One is natural selection, through which traits that aid survival and reproduction become more common, while traits that hinder survival and reproduction become rarer. Natural selection occurs because only a small proportion of individuals in each generation will survive and reproduce, since resources are limited and organisms produce many more offspring than their environment can support. Over many generations, heritable variation in traits is filtered by natural selection and the beneficial changes are successively retained through differential survival and reproduction. This iterative process adjusts traits so they become better suited to an organism's environment: these adjustments are called adaptations.

However, not all change is adaptive. Another cause of evolution is genetic drift, which leads to random changes in how common traits are in a population. Genetic drift is most important when traits do not strongly influence survival—particularly so in small populations, in which chance plays a disproportionate role in the frequency of traits passed on to the next generation. Genetic drift is important in the neutral theory of molecular evolution, and plays a role in the molecular clocks that are used in phylogenetic studies.

A key process in evolution is speciation, in which a single ancestral species splits and diversifies into multiple new species. There are several modes through which this occurs. Ultimately, all living (and extinct) species are descended from a common ancestor via a long series of speciation events. These events stretch back in a diverse "tree of life" which has grown over the 3.5 billion years during which life has existed on Earth. This is visible in anatomical, genetic and other similarities between groups of organisms, geographical distribution of related species, the fossil record and the recorded genetic changes in living organisms over many generations.

Evolutionary biologists document the fact that evolution occurs, and also develop and test theories which explain its causes. The study of evolutionary biology began in the mid-nineteenth century, when research into the fossil record and the diversity of living organisms convinced most scientists that species changed over time. The mechanism driving these changes remained unclear until the theory of natural selection was independently proposed by Charles Darwin and Alfred Wallace. In 1859, Darwin's seminal work *On the Origin of Species* brought the new theory of evolution by natural selection to a wide audience, leading to the overwhelming acceptance of evolution among scientists.

In the 1930s, Darwinian natural selection became understood in combination with Mendelian inheritance, forming the modern evolutionary synthesis, which connected the substrate of evolution (inherited genetics) and the mechanism of evolution (natural selection). This powerful explanatory and predictive theory has become the central organizing principle of modern biology, directing research and providing a unifying explanation for the history and diversity of life on Earth. Evolution is applied and studied in fields as diverse as agriculture, anthropology, conservation biology, ecology, medicine, paleontology, philosophy, and psychology along with other specific topics in the previous listed fields.

History of evolutionary thought

Around 1854 Charles Darwin began writing out what became *On the Origin of Species*

The roots of naturalistic thinking on biology can be dated to at least the 6th century BCE, with the Greek philosopher Anaximander.

Early Christian Church Fathers and Medieval European scholars treated the Genesis creation myth as allegory and believed that natural organisms were unstable and capricious, but the Protestant Reformation inspired Biblical literalism and a natural theology in which species were static and fixed.

As emerging science explored mechanical philosophy in the 18th century, proto-evolutionary ideas were set out by a few natural philosophers such as Pierre Maupertuis in 1745 and Erasmus Darwin in 1796.

The word *evolution* itself (from the Latin *evolutio*, meaning "to unroll like a scroll") was initially used to refer to embryological development; its first use in relation to development of species came in 1762, when Charles Bonnet used it for his concept of "pre-formation", in which females carried a miniature form of all future generations. The term gradually gained a more general meaning of growth or progressive development. The first published modern use of the word has been attributed to the *Edinburgh New Philosophical Journal* in 1826, edited by Robert Jameson but arguably authored by Robert Edmond Grant.

The Bible-based Ussher chronology of the 1650s had calculated creation at 4004 BC, but by the 1780s geologists assumed a much older world. Wernerians thought strata were deposits from shrinking seas, but James Hutton proposed a self-maintaining infinite cycle. Georges Cuvier's paleontological work in the 1790s established the reality of extinction, which he explained by local catastrophes, followed by repopulation of the affected areas by other species. He held that species were fixed, and marginalised the ideas of the biologist Jean-Baptiste Lamarck about transmutation of species which were only taken up by radicals.

Geologists such as Adam Sedgwick adapted Cuvier's catastrophism to show repeated worldwide annihilation and creation of new fixed species adapted to a changed environment, initially identifying the most recent catastrophe as the biblical flood. In opposition to this view, Charles Lyell adapted Hutton's concept into a stricter uniformitarianism which strongly influenced the young geologist Charles Darwin during the *Beagle* expedition. Darwin initially followed Lyell's idea of repeated "centres of creation" of fixed species, but questioned Lyell's views and in 1836, near the end of the voyage, he expressed doubts that species were fixed. Darwin formulated his idea of natural selection in 1838 and was still developing his theory in 1858 when Alfred Russel Wallace sent him a similar theory, and both were presented to the Linnean Society of London in separate papers. At the end of 1859, Darwin's publication of *On the Origin of Species* explained natural selection in detail and presented evidence leading to increasingly wide acceptance of the occurrence of evolution.

Debate about the mechanisms of evolution continued, and Darwin could not explain the source of the heritable variations which would be acted on by natural selection. Like Lamarck, he still thought that parents passed on adaptations acquired during their lifetimes, a theory which was subsequently dubbed Lamarckism. In the 1880s, August Weismann's experiments indicated that changes from use and disuse were not heritable, and Lamarckism gradually fell from favour. More significantly, Darwin could not account for how traits were passed down from generation to generation. In 1865 Gregor Mendel found that traits were inherited in a predictable manner. When Mendel's work was rediscovered in the 1900s, disagreements over the rate of evolution predicted by early geneticists and biometricians led to a rift between the Mendelian and Darwinian models of evolution.

Yet it was the rediscovery of Gregor Mendel's pioneering work on the fundamentals of genetics (of which Darwin and Wallace were unaware) by Hugo de Vries and others in the early 1900s that provided the impetus for a better understanding of how variation occurs in plant and animal traits. That variation is the main fuel used by natural selection to shape the wide variety of adaptive traits observed in organic life. Even though Hugo de Vries and other early geneticists rejected gradual natural selection, their rediscovery of and subsequent work on genetics eventually provided a solid basis on which the theory of evolution stood even more convincingly than when it was originally proposed.

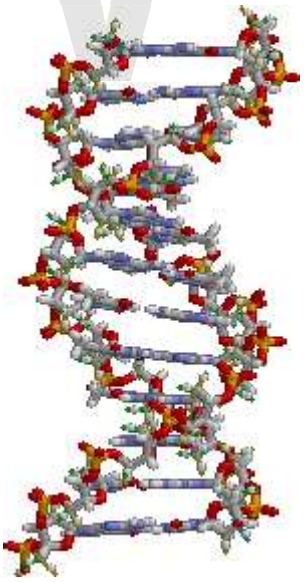
The apparent contradiction between Darwin's theory of evolution by natural selection and Mendel's work was reconciled in the 1920s and 1930s by evolutionary biologists such as

J.B.S. Haldane, Sewall Wright, and particularly Ronald Fisher, who set the foundations for the establishment of the field of population genetics. The end result was a combination of evolution by natural selection and Mendelian inheritance, the modern evolutionary synthesis. In the 1940s, the identification of DNA as the genetic material by Oswald Avery and colleagues and the subsequent publication of the structure of DNA by James Watson and Francis Crick in 1953, demonstrated the physical basis for inheritance. Since then, genetics and molecular biology have become core parts of evolutionary biology and have revolutionised the field of phylogenetics.

In its early history, evolutionary biology primarily drew in scientists from traditional taxonomically oriented disciplines, whose specialist training in particular organisms addressed general questions in evolution. As evolutionary biology expanded as an academic discipline, particularly after the development of the modern evolutionary synthesis, it began to draw more widely from the biological sciences. Currently the study of evolutionary biology involves scientists from fields as diverse as biochemistry, ecology, genetics and physiology, and evolutionary concepts are used in even more distant disciplines such as psychology, medicine, philosophy and computer science.

In the 21st century, current research in evolutionary biology deals with several areas where the modern evolutionary synthesis may need modification or extension, such as assessing the relative importance of various ideas on the unit of selection and evolvability and how to fully incorporate the findings of evolutionary developmental biology.

Heredity



DNA structure. Bases are in the center, surrounded by phosphate–sugar chains in a double helix

Evolution in organisms occurs through changes in heritable traits – particular characteristics of an organism. In humans, for example, eye colour is an inherited characteristic and an individual might inherit the "brown-eye trait" from one of their parents. Inherited traits are controlled by genes and the complete set of genes within an organism's genome is called its genotype.

The complete set of observable traits that make up the structure and behaviour of an organism is called its phenotype. These traits come from the interaction of its genotype with the environment. As a result, many aspects of an organism's phenotype are not inherited. For example, suntanned skin comes from the interaction between a person's genotype and sunlight; thus, suntans are not passed on to people's children. However, some people tan more easily than others, due to differences in their genotype; a striking example are people with the inherited trait of albinism, who do not tan at all and are very sensitive to sunburn.

Heritable traits are passed from one generation to the next via DNA, a molecule that encodes genetic information. DNA is a long polymer composed of four types of bases. The sequence of bases along a particular DNA molecule specify the genetic information, in a manner similar to a sequence of letters spelling out a sentence. Before a cell divides, the DNA is copied, so that each of the resulting two cells will inherit the DNA sequence.

Portions of a DNA molecule that specify a single functional unit are called genes; different genes have different sequences of bases. Within cells, the long strands of DNA form condensed structures called chromosomes. The specific location of a DNA sequence within a chromosome is known as a locus. If the DNA sequence at a locus varies between individuals, the different forms of this sequence are called alleles. DNA sequences can change through mutations, producing new alleles. If a mutation occurs within a gene, the new allele may affect the trait that the gene controls, altering the phenotype of the organism.

However, while this simple correspondence between an allele and a trait works in some cases, most traits are more complex and are controlled by multiple interacting genes. The study of such complex traits is a major area of current genetic research. Another unsolved question in genetics is whether or not epigenetics is important in evolution. Epigenetics is when a trait is inherited without there being any change in gene sequences.

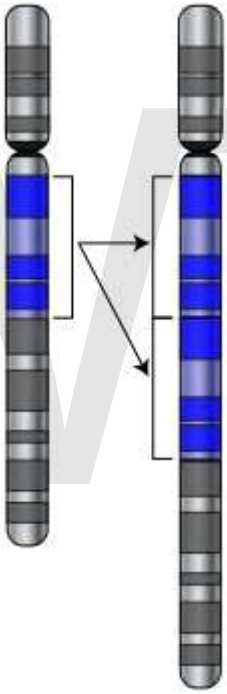
Variation

An individual organism's phenotype results from both its genotype and the influence from the environment it has lived in. A substantial part of the variation in phenotypes in a population is caused by the differences between their genotypes. The modern evolutionary synthesis defines evolution as the change over time in this genetic variation. The frequency of one particular allele will fluctuate, becoming more or less prevalent relative to other forms of that gene. Evolutionary forces act by driving these changes in allele frequency in one direction or another. Variation disappears when a new allele

reaches the point of fixation — when it either disappears from the population or replaces the ancestral allele entirely.

Variation comes from mutations in genetic material, migration between populations (gene flow), and the reshuffling of genes through sexual reproduction. Variation also comes from exchanges of genes between different species; for example, through horizontal gene transfer in bacteria, and hybridisation in plants. Despite the constant introduction of variation through these processes, most of the genome of a species is identical in all individuals of that species. However, even relatively small changes in genotype can lead to dramatic changes in phenotype: for example, chimpanzees and humans differ in only about 5% of their genomes.

Mutation



Duplication of part of a chromosome

Random mutations constantly occur in the genomes of organisms; these mutations create genetic variation. Mutations are changes in the DNA sequence of a cell's genome and are caused by radiation, viruses, transposons and mutagenic chemicals, as well as errors that occur during meiosis or DNA replication. These mutations involve several different types of change in DNA sequences; these can either have no effect, alter the product of a gene, or prevent the gene from functioning. Studies in the fly *Drosophila melanogaster* suggest that if a mutation changes a protein produced by a gene, this will probably be harmful, with about 70 percent of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial.

Due to the damaging effects that mutations can have on cells, organisms have evolved mechanisms such as DNA repair to remove mutations. Therefore, the optimal mutation rate for a species is a trade-off between costs of a high mutation rate, such as deleterious mutations, and the metabolic costs of maintaining systems to reduce the mutation rate, such as DNA repair enzymes. Viruses that use RNA as their genetic material have rapid mutation rates, which can be an advantage since these viruses will evolve constantly and rapidly, and thus evade the defensive responses of e.g. the human immune system.

Mutations can involve large sections of a chromosome becoming duplicated (usually by genetic recombination), which can introduce extra copies of a gene into a genome. Extra copies of genes are a major source of the raw material needed for new genes to evolve. This is important because most new genes evolve within gene families from pre-existing genes that share common ancestors. For example, the human eye uses four genes to make structures that sense light: three for colour vision and one for night vision; all four are descended from a single ancestral gene.

New genes can be created from an ancestral gene when a duplicate copy mutates and acquires a new function. This process is easier once a gene has been duplicated because it increases the redundancy of the system; one gene in the pair can acquire a new function while the other copy continues to perform its original function. Other types of mutation can even create entirely new genes from previously noncoding DNA.

The creation of new genes can also involve small parts of several genes being duplicated, with these fragments then recombining to form new combinations with new functions. When new genes are assembled from shuffling pre-existing parts, domains act as modules with simple independent functions, which can be mixed together creating new combinations with new and complex functions. For example, polyketide synthases are large enzymes that make antibiotics; they contain up to one hundred independent domains that each catalyze one step in the overall process, like a step in an assembly line.

Changes in chromosome number may involve even larger mutations, where segments of the DNA within chromosomes break and then rearrange. For example, two chromosomes in the *Homo* genus fused to produce human chromosome 2; this fusion did not occur in the lineage of the other apes, and they retain these separate chromosomes. In evolution, the most important role of such chromosomal rearrangements may be to accelerate the divergence of a population into new species by making populations less likely to interbreed, and thereby preserving genetic differences between these populations.

Sequences of DNA that can move about the genome, such as transposons, make up a major fraction of the genetic material of plants and animals, and may have been important in the evolution of genomes. For example, more than a million copies of the Alu sequence are present in the human genome, and these sequences have now been recruited to perform functions such as regulating gene expression. Another effect of these mobile DNA sequences is that when they move within a genome, they can mutate or delete existing genes and thereby produce genetic diversity.

Sex and recombination

In asexual organisms, genes are inherited together, or *linked*, as they cannot mix with genes of other organisms during reproduction. In contrast, the offspring of sexual organisms contain random mixtures of their parents' chromosomes that are produced through independent assortment. In a related process called homologous recombination, sexual organisms exchange DNA between two matching chromosomes. Recombination and reassortment do not alter allele frequencies, but instead change which alleles are associated with each other, producing offspring with new combinations of alleles. Sex usually increases genetic variation and may increase the rate of evolution. However, asexuality is advantageous in some environments as it can evolve in previously sexual animals. Here, asexuality might allow the two sets of alleles in their genome to diverge and gain different functions.

Recombination allows even alleles that are close together in a strand of DNA to be inherited independently. However, the rate of recombination is low (approximately two events per chromosome per generation). As a result, genes close together on a chromosome may not always be shuffled away from each other, and genes that are close together tend to be inherited together, a phenomenon known as linkage. This tendency is measured by finding how often two alleles occur together on a single chromosome, which is called their linkage disequilibrium. A set of alleles that is usually inherited in a group is called a haplotype. This can be important when one allele in a particular haplotype is strongly beneficial: natural selection can drive a selective sweep that will also cause the other alleles in the haplotype to become more common in the population; this effect is called genetic hitchhiking.

When alleles cannot be separated by recombination – such as in mammalian Y chromosomes, which pass intact from fathers to sons – harmful mutations accumulate. By breaking up allele combinations, sexual reproduction allows the removal of harmful mutations and the retention of beneficial mutations. In addition, recombination and reassortment can produce individuals with new and advantageous gene combinations. These positive effects are balanced by the fact that sex reduces an organism's reproductive rate, can cause mutations and may separate beneficial combinations of genes. The reasons for the evolution of sexual reproduction are therefore unclear and this question is still an active area of research in evolutionary biology, that has prompted ideas such as the Red Queen hypothesis.

Population genetics

White peppered moth

Black morph in peppered moth evolution

From a genetic viewpoint, evolution is a *generation-to-generation change in the frequencies of alleles within a population that shares a common gene pool*. A population is a localised group of individuals belonging to the same species. For example, all of the moths of the same species living in an isolated forest represent a population. A single gene in this population may have several alternate forms, which account for variations between the phenotypes of the organisms. An example might be a gene for colouration in moths that has two alleles: black and white.

A gene pool is the complete set of alleles for a gene in a single population; the allele frequency measures the fraction of the gene pool composed of a single allele (for example, what fraction of moth colouration genes are the black allele). Evolution occurs when there are changes in the frequencies of alleles within a population of interbreeding organisms; for example, the allele for black colour in a population of moths becoming more common.

To understand the mechanisms that cause a population to evolve, it is useful to consider what conditions are required for a population not to evolve. The *Hardy-Weinberg principle* states that the frequencies of alleles (variations in a gene) in a sufficiently large population will remain constant if the only forces acting on that population are the random reshuffling of alleles during the formation of the sperm or egg, and the random combination of the alleles in these sex cells during fertilisation. Such a population is said to be in *Hardy-Weinberg equilibrium*; it is not evolving.

Gene flow

When they mature, male lions leave the pride where they were born and take over a new pride to mate, causing gene flow between prides.

Gene flow is the exchange of genes between populations, which are usually of the same species. Examples of gene flow within a species include the migration and then breeding of organisms, or the exchange of pollen. Gene transfer between species includes the formation of hybrid organisms and horizontal gene transfer.

Migration into or out of a population can change allele frequencies, as well as introducing genetic variation into a population. Immigration may add new genetic material to the established gene pool of a population. Conversely, emigration may remove genetic material. As barriers to reproduction between two diverging populations are required for the populations to become new species, gene flow may slow this process by spreading genetic differences between the populations. Gene flow is hindered by mountain ranges, oceans and deserts or even man-made structures such as the Great Wall of China, which has hindered the flow of plant genes.

Depending on how far two species have diverged since their most recent common ancestor, it may still be possible for them to produce offspring, as with horses and donkeys mating to produce mules. Such hybrids are generally infertile, due to the two

different sets of chromosomes being unable to pair up during meiosis. In this case, closely related species may regularly interbreed, but hybrids will be selected against and the species will remain distinct. However, viable hybrids are occasionally formed and these new species can either have properties intermediate between their parent species, or possess a totally new phenotype. The importance of hybridisation in creating new species of animals is unclear, although cases have been seen in many types of animals, with the gray tree frog being a particularly well-studied example.

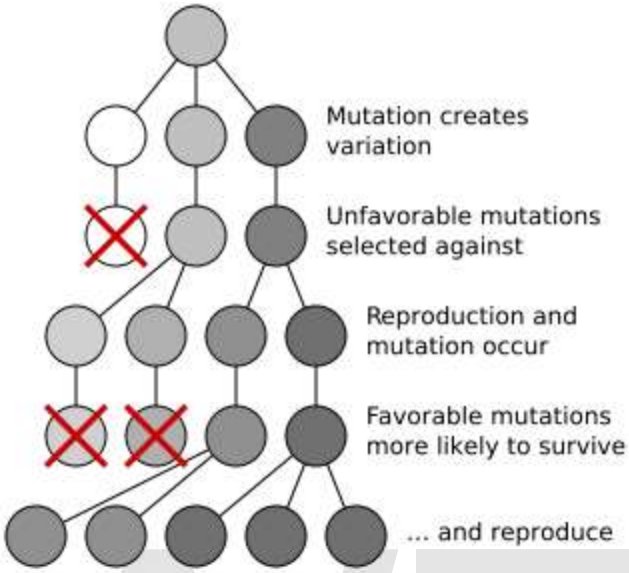
Hybridisation is, however, an important means of speciation in plants, since polyploidy (having more than two copies of each chromosome) is tolerated in plants more readily than in animals. Polyploidy is important in hybrids as it allows reproduction, with the two different sets of chromosomes each being able to pair with an identical partner during meiosis. Polyploids also have more genetic diversity, which allows them to avoid inbreeding depression in small populations.

Horizontal gene transfer is the transfer of genetic material from one organism to another organism that is not its offspring; this is most common among bacteria. In medicine, this contributes to the spread of antibiotic resistance, as when one bacteria acquires resistance genes it can rapidly transfer them to other species. Horizontal transfer of genes from bacteria to eukaryotes such as the yeast *Saccharomyces cerevisiae* and the adzuki bean beetle *Callosobruchus chinensis* may also have occurred. An example of larger-scale transfers are the eukaryotic bdelloid rotifers, which appear to have received a range of genes from bacteria, fungi, and plants. Viruses can also carry DNA between organisms, allowing transfer of genes even across biological domains. Large-scale gene transfer has also occurred between the ancestors of eukaryotic cells and prokaryotes, during the acquisition of chloroplasts and mitochondria.

Mechanisms

The two main mechanisms that produce evolution are natural selection and genetic drift. Natural selection is the process which favors genes that aid survival and reproduction. Genetic drift is the random change in the frequency of alleles, caused by the random sampling of a generation's genes during reproduction. The relative importance of natural selection and genetic drift in a population varies depending on the strength of the selection and the effective population size, which is the number of individuals capable of breeding. Natural selection usually predominates in large populations, whereas genetic drift dominates in small populations. The dominance of genetic drift in small populations can even lead to the fixation of slightly deleterious mutations. As a result, changing population size can dramatically influence the course of evolution. Population bottlenecks, where the population shrinks temporarily and therefore loses genetic variation, result in a more uniform population.

Natural selection



Natural selection of a population for dark colouration

Natural selection is the process by which genetic mutations that enhance reproduction become, and remain, more common in successive generations of a population. It has often been called a "self-evident" mechanism because it necessarily follows from three simple facts:

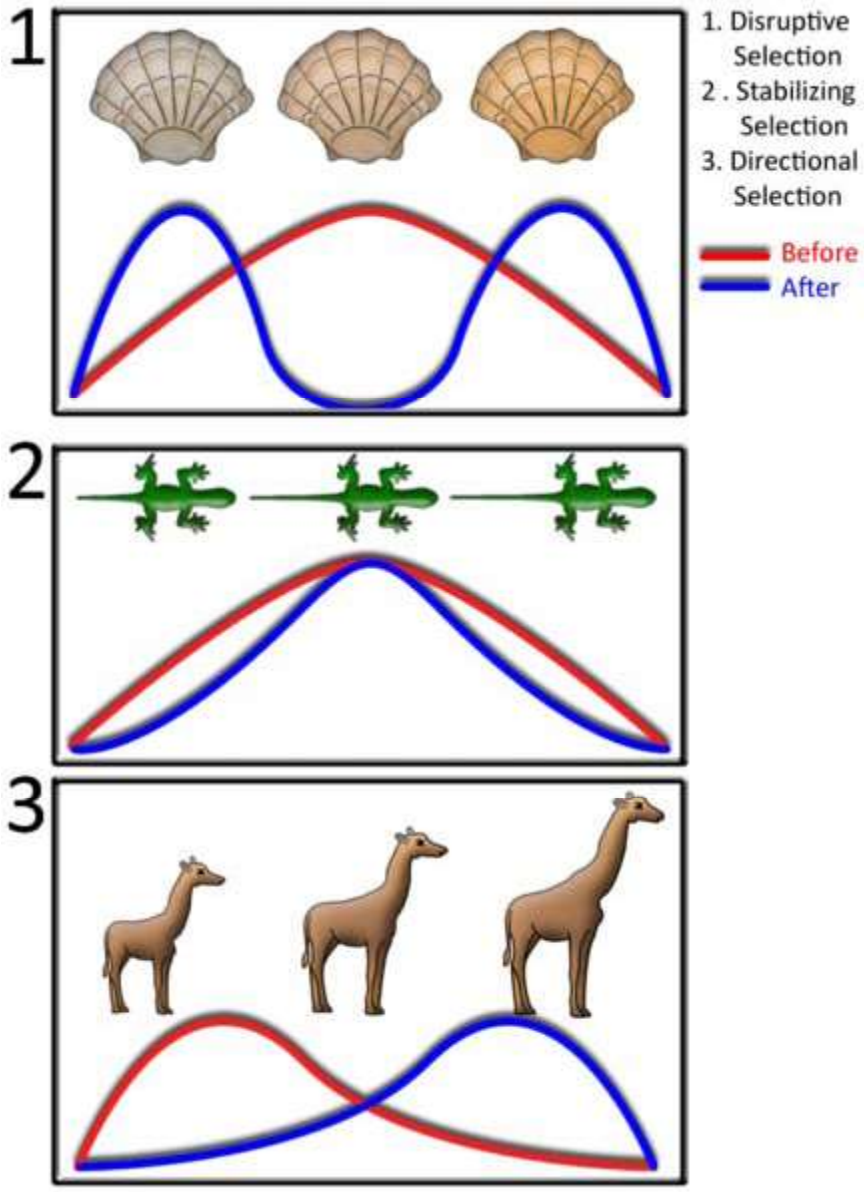
- Heritable variation exists within populations of organisms.
- Organisms produce more offspring than can survive.
- These offspring vary in their ability to survive and reproduce.

These conditions produce competition between organisms for survival and reproduction. Consequently, organisms with traits that give them an advantage over their competitors pass these advantageous traits on, while traits that do not confer an advantage are not passed on to the next generation.

The central concept of natural selection is the evolutionary fitness of an organism. Fitness is measured by an organism's ability to survive and reproduce, which determines the size of its genetic contribution to the next generation. However, fitness is not the same as the total number of offspring: instead fitness is indicated by the proportion of subsequent generations that carry an organism's genes. For example, if an organism could survive well and reproduce rapidly, but its offspring were all too small and weak to survive, this organism would make little genetic contribution to future generations and would thus have low fitness.

If an allele increases fitness more than the other alleles of that gene, then with each generation this allele will become more common within the population. These traits are

said to be "selected *for*". Examples of traits that can increase fitness are enhanced survival, and increased fecundity. Conversely, the lower fitness caused by having a less beneficial or deleterious allele results in this allele becoming rarer — they are "selected *against*". Importantly, the fitness of an allele is not a fixed characteristic; if the environment changes, previously neutral or harmful traits may become beneficial and previously beneficial traits become harmful. However, even if the direction of selection does reverse in this way, traits that were lost in the past may not re-evolve in an identical form.



A chart showing three types of selection. 1.Disruptive selection 2.Stabilizing selection 3.Directional selection

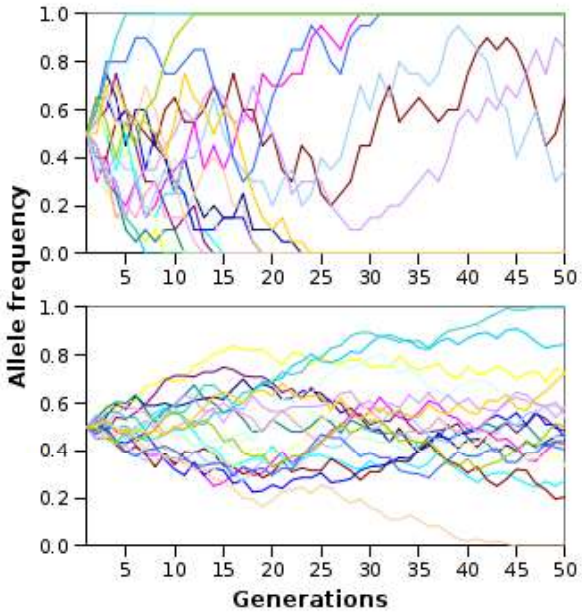
Natural selection within a population for a trait that can vary across a range of values, such as height, can be categorised into three different types. The first is directional selection, which is a shift in the average value of a trait over time — for example, organisms slowly getting taller. Secondly, disruptive selection is selection for extreme trait values and often results in two different values becoming most common, with selection against the average value. This would be when either short or tall organisms had an advantage, but not those of medium height. Finally, in stabilizing selection there is selection against extreme trait values on both ends, which causes a decrease in variance around the average value and less diversity. This would, for example, cause organisms to slowly become all the same height.

A special case of natural selection is sexual selection, which is selection for any trait that increases mating success by increasing the attractiveness of an organism to potential mates. Traits that evolved through sexual selection are particularly prominent in males of some animal species, despite traits such as cumbersome antlers, mating calls or bright colours that attract predators, decreasing the survival of individual males. This survival disadvantage is balanced by higher reproductive success in males that show these hard to fake, sexually selected traits.

Natural selection most generally makes nature the measure against which individuals, and individual traits, are more or less likely to survive. "Nature" in this sense refers to an ecosystem, that is, a system in which organisms interact with every other element, physical as well as biological, in their local environment. Eugene Odum, a founder of ecology, defined an ecosystem as: "Any unit that includes all of the organisms...in a given area interacting with the physical environment so that a flow of energy leads to clearly defined trophic structure, biotic diversity, and material cycles (ie: exchange of materials between living and nonliving parts) within the system." Each population within an ecosystem occupies a distinct niche, or position, with distinct relationships to other parts of the system. These relationships involve the life history of the organism, its position in the food chain, and its geographic range. This broad understanding of nature enables scientists to delineate specific forces which, together, comprise natural selection.

An active area of research is the unit of selection, with natural selection being proposed to work at the level of genes, cells, individual organisms, groups of organisms and species. None of these are mutually exclusive and selection can act on multiple levels simultaneously. An example of selection occurring below the level of the individual organism are genes called transposons, which can replicate and spread throughout a genome. Selection at a level above the individual, such as group selection, may allow the evolution of co-operation, as discussed below.

Genetic drift



Simulation of genetic drift of 20 unlinked alleles in populations of 10 (top) and 100 (bottom). Drift to fixation is more rapid in the smaller population.

Genetic drift is the change in allele frequency from one generation to the next that occurs because alleles in offspring are a random sample of those in the parents, as well as from the role that chance plays in determining whether a given individual will survive and reproduce. In mathematical terms, alleles are subject to sampling error. As a result, when selective forces are absent or relatively weak, allele frequencies tend to "drift" upward or downward randomly (in a random walk). This drift halts when an allele eventually becomes fixed, either by disappearing from the population, or replacing the other alleles entirely. Genetic drift may therefore eliminate some alleles from a population due to chance alone. Even in the absence of selective forces, genetic drift can cause two separate populations that began with the same genetic structure to drift apart into two divergent populations with different sets of alleles.

The time for an allele to become fixed by genetic drift depends on population size, with fixation occurring more rapidly in smaller populations. The precise measure of population that is important is called the effective population size. The effective population is always smaller than the total population since it takes into account factors such as the level of inbreeding, the number of animals that are too old or young to breed, and the lower probability of animals that live far apart managing to mate with each other.

An example when genetic drift is probably of central importance in determining a trait is the loss of pigments from animals that live in caves, a change that produces no obvious advantage or disadvantage in complete darkness. However, it is usually difficult to measure the relative importance of selection and drift, so the comparative importance of

these two forces in driving evolutionary change is an area of current research. These investigations were prompted by the neutral theory of molecular evolution, which proposed that most evolutionary changes are the result of the fixation of neutral mutations that do not have any immediate effects on the fitness of an organism. Hence, in this model, most genetic changes in a population are the result of constant mutation pressure and genetic drift. This form of the neutral theory is now largely abandoned, since it does not seem to fit the genetic variation seen in nature. However, a more recent and better-supported version of this model is the nearly neutral theory, where most mutations only have small effects on fitness.

Outcomes

Evolution influences every aspect of the form and behaviour of organisms. Most prominent are the specific behavioral and physical adaptations that are the outcome of natural selection. These adaptations increase fitness by aiding activities such as finding food, avoiding predators or attracting mates. Organisms can also respond to selection by co-operating with each other, usually by aiding their relatives or engaging in mutually beneficial symbiosis. In the longer term, evolution produces new species through splitting ancestral populations of organisms into new groups that cannot or will not interbreed.

These outcomes of evolution are sometimes divided into macroevolution, which is evolution that occurs at or above the level of species, such as extinction and speciation, and microevolution, which is smaller evolutionary changes, such as adaptations, within a species or population. In general, macroevolution is regarded as the outcome of long periods of microevolution. Thus, the distinction between micro- and macroevolution is not a fundamental one – the difference is simply the time involved. However, in macroevolution, the traits of the entire species may be important. For instance, a large amount of variation among individuals allows a species to rapidly adapt to new habitats, lessening the chance of it going extinct, while a wide geographic range increases the chance of speciation, by making it more likely that part of the population will become isolated. In this sense, microevolution and macroevolution might involve selection at different levels – with microevolution acting on genes and organisms, versus macroevolutionary processes such as species selection acting on entire species and affecting their rates of speciation and extinction.

A common misconception is that evolution has goals or long-term plans; realistically however, evolution has no long-term goal and does not necessarily produce greater complexity. Although complex species have evolved, they occur as a side effect of the overall number of organisms increasing, and simple forms of life still remain more common in the biosphere. For example, the overwhelming majority of species are microscopic prokaryotes, which form about half the world's biomass despite their small size, and constitute the vast majority of Earth's biodiversity. Simple organisms have therefore been the dominant form of life on Earth throughout its history and continue to be the main form of life up to the present day, with complex life only appearing more diverse because it is more noticeable. Indeed, the evolution of microorganisms is particularly important to modern evolutionary research, since their rapid reproduction

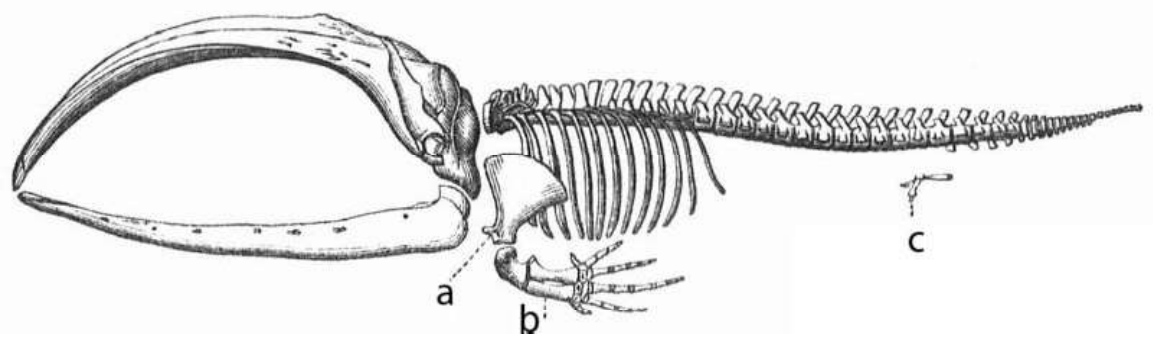
allows the study of experimental evolution and the observation of evolution and adaptation in real time.

Adaptation

Adaptation is one of the basic phenomena of biology, and is the *process* whereby an organism becomes better suited to its habitat. Also, the term adaptation may refer to a trait that is important for an organism's survival. For example, the adaptation of horses' teeth to the grinding of grass, or the ability of horses to run fast and escape predators. By using the term *adaptation* for the evolutionary process, and *adaptive trait* for the product (the bodily part or function), the two senses of the word may be distinguished. Adaptations are produced by natural selection. The following definitions are due to Theodosius Dobzhansky.

> 1. *Adaptation* is the evolutionary process whereby an organism becomes better able to live in its habitat or habitats.
> 2. *Adaptedness* is the state of being adapted: the degree to which an organism is able to live and reproduce in a given set of habitats.
> 3. An *adaptive trait* is an aspect of the developmental pattern of the organism which enables or enhances the probability of that organism surviving and reproducing.

Adaptation may cause either the gain of a new feature, or the loss of an ancestral feature. An example that shows both types of change is bacterial adaptation to antibiotic selection, with genetic changes causing antibiotic resistance by both modifying the target of the drug, or increasing the activity of transporters that pump the drug out of the cell. Other striking examples are the bacteria *Escherichia coli* evolving the ability to use citric acid as a nutrient in a long-term laboratory experiment, *Flavobacterium* evolving a novel enzyme that allows these bacteria to grow on the by-products of nylon manufacturing, and the soil bacterium *Sphingobium* evolving an entirely new metabolic pathway that degrades the synthetic pesticide pentachlorophenol. An interesting but still controversial idea is that some adaptations might increase the ability of organisms to generate genetic diversity and adapt by natural selection (increasing organisms' evolvability).



A baleen whale skeleton, *a* and *b* label flipper bones, which were adapted from front leg bones: while *c* indicates vestigial leg bones, suggesting an adaptation from land to sea.

Adaptation occurs through the gradual modification of existing structures. Consequently, structures with similar internal organisation may have different functions in related organisms. This is the result of a single ancestral structure being adapted to function in different ways. The bones within bat wings, for example, are very similar to those in mice feet and primate hands, due to the descent of all these structures from a common mammalian ancestor. However, since all living organisms are related to some extent, even organs that appear to have little or no structural similarity, such as arthropod, squid and vertebrate eyes, or the limbs and wings of arthropods and vertebrates, can depend on a common set of homologous genes that control their assembly and function; this is called deep homology.

During adaptation, some structures may lose their original function and become vestigial structures. Such structures may have little or no function in a current species, yet have a clear function in ancestral species, or other closely related species. Examples include pseudogenes, the non-functional remains of eyes in blind cave-dwelling fish, wings in flightless birds, and the presence of hip bones in whales and snakes. Examples of vestigial structures in humans include wisdom teeth, the coccyx, the vermiform appendix, and other behavioral vestiges such as goose bumps, and primitive reflexes.

However, many traits that appear to be simple adaptations are in fact exaptations: structures originally adapted for one function, but which coincidentally became somewhat useful for some other function in the process. One example is the African lizard *Holaspis guentheri*, which developed an extremely flat head for hiding in crevices, as can be seen by looking at its near relatives. However, in this species, the head has become so flattened that it assists in gliding from tree to tree—an exaptation. Within cells, molecular machines such as the bacterial flagella and protein sorting machinery evolved by the recruitment of several pre-existing proteins that previously had different functions. Another example is the recruitment of enzymes from glycolysis and xenobiotic metabolism to serve as structural proteins called crystallins within the lenses of organisms' eyes.

A critical principle of ecology is that of competitive exclusion: no two species can occupy the same niche in the same environment for a long time. Consequently, natural selection will tend to force species to adapt to different ecological niches. This may mean that, for example, two species of cichlid fish adapt to live in different habitats, which will minimise the competition between them for food.

An area of current investigation in evolutionary developmental biology is the developmental basis of adaptations and exaptations. This research addresses the origin and evolution of embryonic development and how modifications of development and developmental processes produce novel features. These studies have shown that evolution can alter development to create new structures, such as embryonic bone structures that develop into the jaw in other animals instead forming part of the middle ear in mammals. It is also possible for structures that have been lost in evolution to reappear due to changes in developmental genes, such as a mutation in chickens causing

embryos to grow teeth similar to those of crocodiles. It is now becoming clear that most alterations in the form of organisms are due to changes in a small set of conserved genes.

Co-evolution

Common garter snake (*Thamnophis sirtalis sirtalis*) which has evolved resistance to tetrodotoxin in its amphibian prey

Interactions between organisms can produce both conflict and co-operation. When the interaction is between pairs of species, such as a pathogen and a host, or a predator and its prey, these species can develop matched sets of adaptations. Here, the evolution of one species causes adaptations in a second species. These changes in the second species then, in turn, cause new adaptations in the first species. This cycle of selection and response is called co-evolution. An example is the production of tetrodotoxin in the rough-skinned newt and the evolution of tetrodotoxin resistance in its predator, the common garter snake. In this predator-prey pair, an evolutionary arms race has produced high levels of toxin in the newt and correspondingly high levels of toxin resistance in the snake.

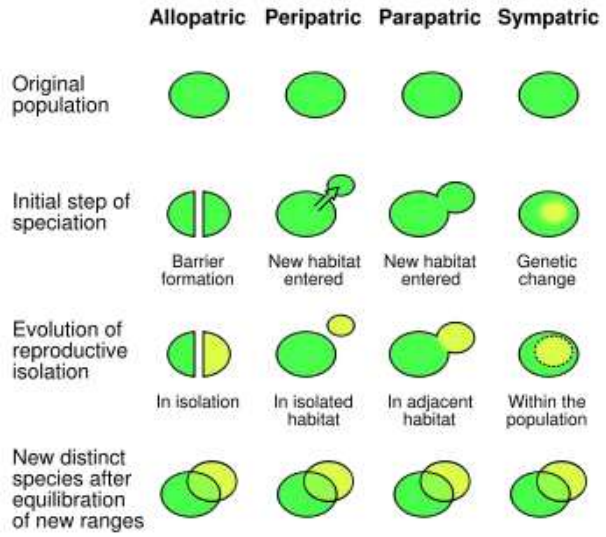
Co-operation

However, not all interactions between species involve conflict. Many cases of mutually beneficial interactions have evolved. For instance, an extreme cooperation exists between plants and the mycorrhizal fungi that grow on their roots and aid the plant in absorbing nutrients from the soil. This is a reciprocal relationship as the plants provide the fungi with sugars from photosynthesis. Here, the fungi actually grow inside plant cells, allowing them to exchange nutrients with their hosts, while sending signals that suppress the plant immune system.

Coalitions between organisms of the same species have also evolved. An extreme case is the eusociality found in social insects, such as bees, termites and ants, where sterile insects feed and guard the small number of organisms in a colony that are able to reproduce. On an even smaller scale, the somatic cells that make up the body of an animal limit their reproduction so they can maintain a stable organism, which then supports a small number of the animal's germ cells to produce offspring. Here, somatic cells respond to specific signals that instruct them whether to grow, remain as they are, or die. If cells ignore these signals and multiply inappropriately, their uncontrolled growth causes cancer.

Such cooperation within species may have evolved through the process of kin selection, which is where one organism acts to help raise a relative's offspring. This activity is selected for because if the *helping* individual contains alleles which promote the helping activity, it is likely that its kin will *also* contain these alleles and thus those alleles will be passed on. Other processes that may promote cooperation include group selection, where cooperation provides benefits to a group of organisms.

Speciation



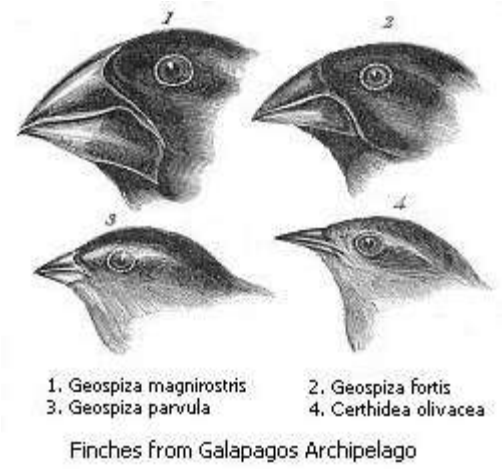
The four mechanisms of speciation

Speciation is the process where a species diverges into two or more descendant species. Darwinian theory involves a shift away from viewing species as static ideal types, defined by the presence of a particular trait, to viewing them as populations with a history that requires more of a statistical approach for the analysis of multiple variable traits. This view is counterintuitive since the classical idea of species is still widely held, with a species seen as a class of organisms exemplified by a "type specimen" that bears all the traits common to this species. Instead, a species is now defined as a separately evolving lineage that forms a single gene pool. Although properties such as genetics and morphology are used to help separate closely related lineages, this definition has fuzzy boundaries. Indeed, the exact definition of the term "species" is still controversial, particularly in prokaryotes, and this is called the species problem. Biologists have proposed a range of more precise definitions, but the definition used is a pragmatic choice that depends on the particularities of the species concerned. Typically the actual focus on biological study is the population, an observable *interacting* group of organisms, rather than a species, an observable *similar* group of individuals.

Speciation has been observed multiple times under both controlled laboratory conditions and in nature. In sexually reproducing organisms, speciation results from reproductive isolation followed by genealogical divergence. There are four mechanisms for speciation. The most common in animals is allopatric speciation, which occurs in populations initially isolated geographically, such as by habitat fragmentation or migration. Selection under these conditions can produce very rapid changes in the appearance and behaviour of organisms. As selection and drift act independently on populations isolated from the rest of their species, separation may eventually produce organisms that cannot interbreed.

The second mechanism of speciation is peripatric speciation, which occurs when small populations of organisms become isolated in a new environment. This differs from allopatric speciation in that the isolated populations are numerically much smaller than the parental population. Here, the founder effect causes rapid speciation through both rapid genetic drift and selection on a small gene pool.

The third mechanism of speciation is parapatric speciation. This is similar to peripatric speciation in that a small population enters a new habitat, but differs in that there is no physical separation between these two populations. Instead, speciation results from the evolution of mechanisms that reduce gene flow between the two populations. Generally this occurs when there has been a drastic change in the environment within the parental species' habitat. One example is the grass *Anthoxanthum odoratum*, which can undergo parapatric speciation in response to localised metal pollution from mines. Here, plants evolve that have resistance to high levels of metals in the soil. Selection against interbreeding with the metal-sensitive parental population produced a gradual change in the flowering time of the metal-resistant plants, which eventually produced complete reproductive isolation. Selection against hybrids between the two populations may cause *reinforcement*, which is the evolution of traits that promote mating within a species, as well as character displacement, which is when two species become more distinct in appearance.



Finches from Galapagos Archipelago

Geographical isolation of finches on the Galápagos Islands produced over a dozen new species.

Finally, in sympatric speciation species diverge without geographic isolation or changes in habitat. This form is rare since even a small amount of gene flow may remove genetic differences between parts of a population. Generally, sympatric speciation in animals requires the evolution of both genetic differences and non-random mating, to allow reproductive isolation to evolve.

One type of sympatric speciation involves cross-breeding of two related species to produce a new hybrid species. This is not common in animals as animal hybrids are usually sterile. This is because during meiosis the homologous chromosomes from each parent are from different species and cannot successfully pair. However, it is more common in plants because plants often double their number of chromosomes, to form polyploids. This allows the chromosomes from each parental species to form matching pairs during meiosis, since each parent's chromosomes are represented by a pair already. An example of such a speciation event is when the plant species *Arabidopsis thaliana* and *Arabidopsis arenosa* cross-bred to give the new species *Arabidopsis suecica*. This happened about 20,000 years ago, and the speciation process has been repeated in the laboratory, which allows the study of the genetic mechanisms involved in this process. Indeed, chromosome doubling within a species may be a common cause of reproductive isolation, as half the doubled chromosomes will be unmatched when breeding with undoubled organisms.

Speciation events are important in the theory of punctuated equilibrium, which accounts for the pattern in the fossil record of short "bursts" of evolution interspersed with relatively long periods of stasis, where species remain relatively unchanged. In this theory, speciation and rapid evolution are linked, with natural selection and genetic drift acting most strongly on organisms undergoing speciation in novel habitats or small populations. As a result, the periods of stasis in the fossil record correspond to the parental population, and the organisms undergoing speciation and rapid evolution are found in small populations or geographically restricted habitats, and therefore rarely being preserved as fossils.

Extinction

Tyrannosaurus rex. Non-avian dinosaurs died out in the Cretaceous–Tertiary extinction event at the end of the Cretaceous period.

Extinction is the disappearance of an entire species. Extinction is not an unusual event, as species regularly appear through speciation, and disappear through extinction. Nearly all animal and plant species that have lived on Earth are now extinct, and extinction appears to be the ultimate fate of all species. These extinctions have happened continuously throughout the history of life, although the rate of extinction spikes in occasional mass extinction events. The Cretaceous–Tertiary extinction event, during which the non-avian dinosaurs went extinct, is the most well-known, but the earlier Permian–Triassic extinction event was even more severe, with approximately 96 percent of species driven to extinction. The Holocene extinction event is an ongoing mass extinction associated with humanity's expansion across the globe over the past few thousand years. Present-day extinction rates are 100–1000 times greater than the background rate, and up to 30 percent of species may be extinct by the mid 21st century. Human activities are now the primary cause of the ongoing extinction event; global warming may further accelerate it in the future.

The role of extinction in evolution is not very well understood and may depend on which type of extinction is considered. The causes of the continuous "low-level" extinction events, which form the majority of extinctions, may be the result of competition between species for limited resources (competitive exclusion). If one species can out-compete another, this could produce species selection, with the fitter species surviving and the

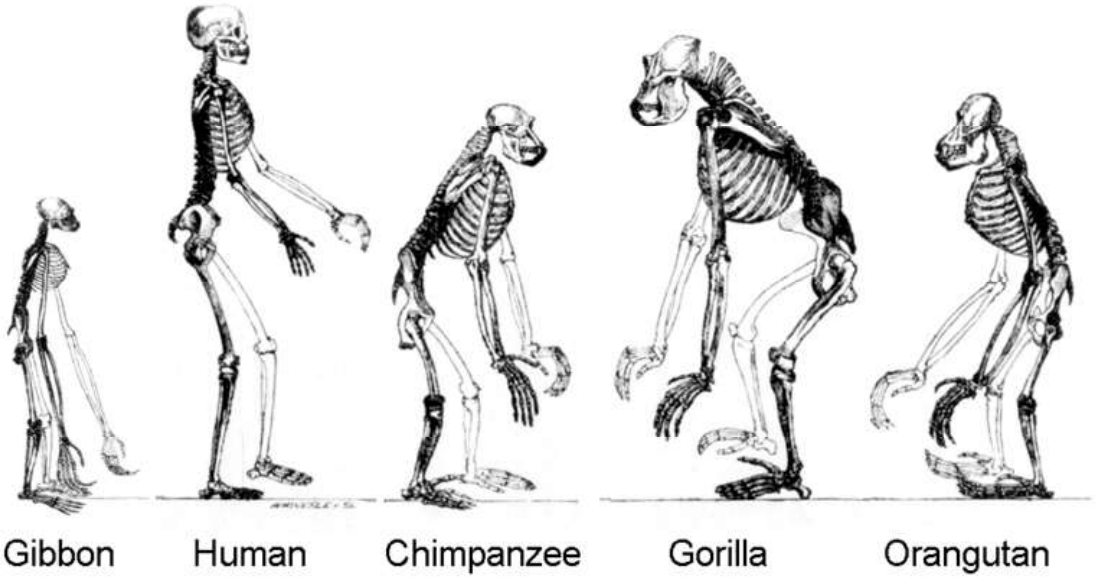
other species being driven to extinction. The intermittent mass extinctions are also important, but instead of acting as a selective force, they drastically reduce diversity in a nonspecific manner and promote bursts of rapid evolution and speciation in survivors.

Evolutionary history of life

Origin of life

The origin of life is a necessary precursor for biological evolution, but understanding that evolution occurred once organisms appeared and investigating how this happens does not depend on understanding exactly how life began. The current scientific consensus is that the complex biochemistry that makes up life came from simpler chemical reactions, but it is unclear how this occurred. Not much is certain about the earliest developments in life, the structure of the first living things, or the identity and nature of any last universal common ancestor or ancestral gene pool. Consequently, there is no scientific consensus on how life began, but proposals include self-replicating molecules such as RNA, and the assembly of simple cells.

Common descent



The hominoids are descendants of a common ancestor

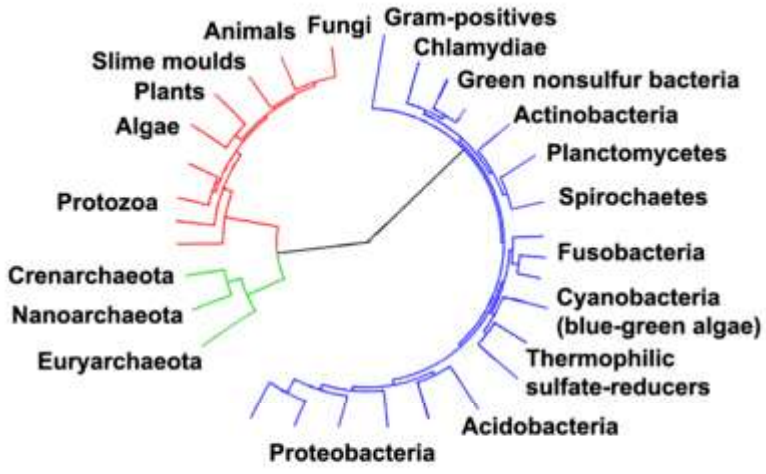
All organisms on Earth are descended from a common ancestor or ancestral gene pool. Current species are a stage in the process of evolution, with their diversity the product of a long series of speciation and extinction events. The common descent of organisms was first deduced from four simple facts about organisms: First, they have geographic distributions that cannot be explained by local adaptation. Second, the diversity of life is not a set of completely unique organisms, but organisms that share morphological

similarities. Third, vestigial traits with no clear purpose resemble functional ancestral traits, and finally, that organisms can be classified using these similarities into a hierarchy of nested groups – similar to a family tree. However, modern research has suggested that, due to horizontal gene transfer, this "tree of life" may be more complicated than a simple branching tree since some genes have spread independently between distantly related species.

Past species have also left records of their evolutionary history. Fossils, along with the comparative anatomy of present-day organisms, constitute the morphological, or anatomical, record. By comparing the anatomies of both modern and extinct species, paleontologists can infer the lineages of those species. However, this approach is most successful for organisms that had hard body parts, such as shells, bones or teeth. Further, as prokaryotes such as bacteria and archaea share a limited set of common morphologies, their fossils do not provide information on their ancestry.

More recently, evidence for common descent has come from the study of biochemical similarities between organisms. For example, all living cells use the same basic set of nucleotides and amino acids. The development of molecular genetics has revealed the record of evolution left in organisms' genomes: dating when species diverged through the molecular clock produced by mutations. For example, these DNA sequence comparisons have revealed that humans and chimpanzees share 96% of their genomes and analyzing the few areas where they differ helps shed light on when the common ancestor of these species existed.

Evolution of life



Evolutionary tree showing the divergence of modern species from their common ancestor in the center. The three domains are coloured, with bacteria blue, archaea green, and eukaryotes red.

Despite the uncertainty on how life began, it is generally accepted that prokaryotes inhabited the Earth from approximately 3–4 billion years ago. No obvious changes in morphology or cellular organisation occurred in these organisms over the next few billion years.

The eukaryotes were the next major change in cell structure. These came from ancient bacteria being engulfed by the ancestors of eukaryotic cells, in a cooperative association called endosymbiosis. The engulfed bacteria and the host cell then underwent co-evolution, with the bacteria evolving into either mitochondria or hydrogenosomes. An independent second engulfment of cyanobacterial-like organisms led to the formation of chloroplasts in algae and plants. It is unknown when the first eukaryotic cells appeared though they first emerged between 1.6 – 2.7 billion years ago.

The history of life was that of the unicellular eukaryotes, prokaryotes, and archaea until about 610 million years ago when multicellular organisms began to appear in the oceans in the Ediacaran period. The evolution of multicellularity occurred in multiple independent events, in organisms as diverse as sponges, brown algae, cyanobacteria, slime moulds and myxobacteria.

Soon after the emergence of these first multicellular organisms, a remarkable amount of biological diversity appeared over approximately 10 million years, in an event called the Cambrian explosion. Here, the majority of types of modern animals appeared in the fossil record, as well as unique lineages that subsequently became extinct. Various triggers for the Cambrian explosion have been proposed, including the accumulation of oxygen in the atmosphere from photosynthesis.

About 500 million years ago, plants and fungi colonised the land, and were soon followed by arthropods and other animals. Insects were particularly successful and even today make up the majority of animal species. Amphibians first appeared around 300 million years ago, followed by early amniotes, then mammals around 200 million years ago and birds around 100 million years ago (both from "reptile"-like lineages). However, despite the evolution of these large animals, smaller organisms similar to the types that evolved early in this process continue to be highly successful and dominate the Earth, with the majority of both biomass and species being prokaryotes.

Applications

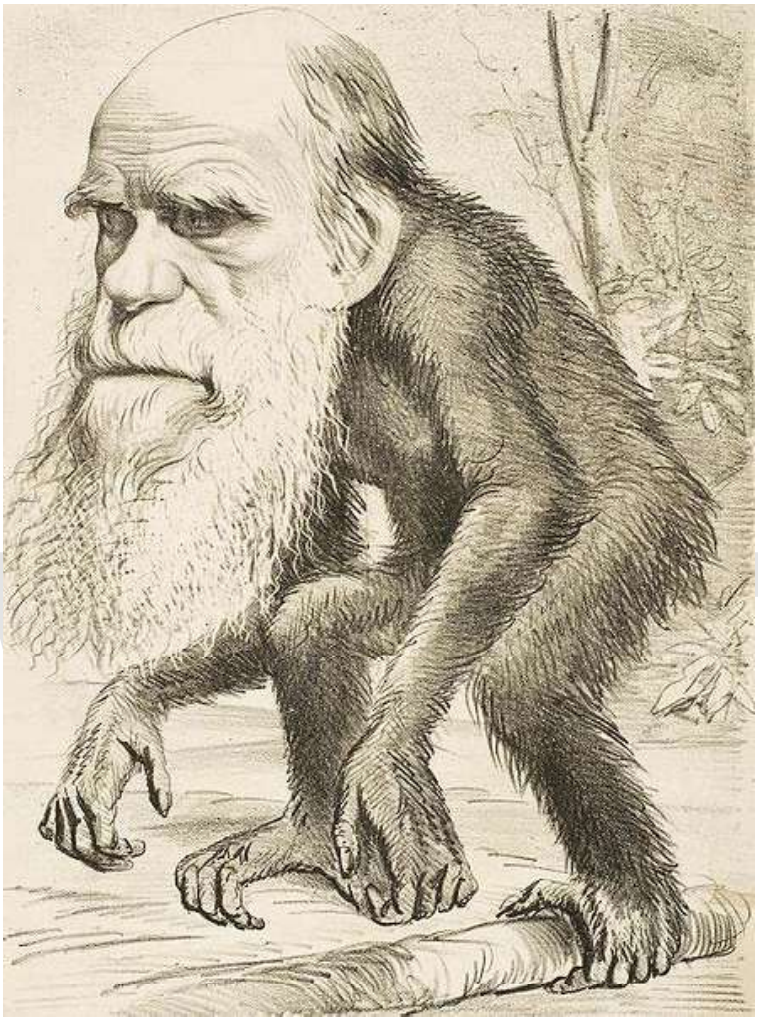
Evolutionary biology, and in particular the understanding of how organisms evolve through natural selection, is an area of science with many practical applications. A major technological application of evolution is artificial selection, which is the intentional selection of certain traits in a population of organisms. Humans have used artificial selection for thousands of years in the domestication of plants and animals. More recently, such selection has become a vital part of genetic engineering, with selectable markers such as antibiotic resistance genes being used to manipulate DNA in molecular biology. It is also possible to use repeated rounds of mutation and selection to evolve

proteins with particular properties, such as modified enzymes or new antibodies, in a process called directed evolution.

Understanding the changes that have occurred during organism's evolution can reveal the genes needed to construct parts of the body, genes which may be involved in human genetic disorders. For example, the Mexican tetra is an albino cavefish that lost its eyesight during evolution. Breeding together different populations of this blind fish produced some offspring with functional eyes, since different mutations had occurred in the isolated populations that had evolved in different caves. This helped identify genes required for vision and pigmentation, such as crystallins and the melanocortin 1 receptor. Similarly, comparing the genome of the Antarctic icefish, which lacks red blood cells, to close relatives such as the Antarctic rockcod revealed genes needed to make these blood cells.

As evolution can produce highly optimised processes and networks, it has many applications in computer science. Here, simulations of evolution using evolutionary algorithms and artificial life started with the work of Nils Aall Barricelli in the 1960s, and was extended by Alex Fraser, who published a series of papers on simulation of artificial selection. Artificial evolution became a widely recognised optimisation method as a result of the work of Ingo Rechenberg in the 1960s and early 1970s, who used evolution strategies to solve complex engineering problems. Genetic algorithms in particular became popular through the writing of John Holland. As academic interest grew, dramatic increases in the power of computers allowed practical applications, including the automatic evolution of computer programs. Evolutionary algorithms are now used to solve multi-dimensional problems more efficiently than software produced by human designers, and also to optimise the design of systems.

Social and cultural responses



As evolution became widely accepted in the 1870s, caricatures of Charles Darwin with an ape or monkey body symbolised evolution.

In the 19th century, particularly after the publication of *On the Origin of Species* in 1859, the idea that life had evolved was an active source of academic debate centered on the philosophical, social and religious implications of evolution. Nowadays, the fact that organisms evolve is uncontested in the scientific literature and the modern evolutionary synthesis is widely accepted by scientists. However, evolution remains a contentious concept for some theists.

While various religions and denominations have reconciled their beliefs with evolution through concepts such as theistic evolution, there are creationists who believe that evolution is contradicted by the creation myths found in their respective religions and who raise various objections to evolution. As had been demonstrated by responses to the publication of *Vestiges of the Natural History of Creation* in 1844, the most controversial

aspect of evolutionary biology is the implication of human evolution that human mental and moral faculties, which had been thought purely spiritual, are not distinctly separated from those of other animals. In some countries—notably the United States—these tensions between science and religion have fueled the current creation-evolution controversy, a religious conflict focusing on politics and public education. While other scientific fields such as cosmology and Earth science also conflict with literal interpretations of many religious texts, evolutionary biology experiences significantly more opposition from religious literalists.

The teaching of evolution in American secondary school biology classes was uncommon in most of the first half of the 20th century. The Scopes Trial decision of 1925 caused the subject to become very rare in American secondary biology textbooks for a generation, but it was gradually re-introduced about a generation later and legally protected with the 1968 *Epperson v. Arkansas* decision. Since then, the competing religious belief of creationism was legally disallowed in secondary school curricula in various decisions in the 1970s and 1980s, but it returned in the form of intelligent design, to be excluded once again in the 2005 *Kitzmiller v. Dover Area School District* case.

Another example somewhat associated with evolutionary theory that is now widely regarded as unwarranted is "Social Darwinism", a derogatory term associated with the 19th century Malthusian theory developed by Whig philosopher Herbert Spencer. It was later expanded by others into ideas about "survival of the fittest" in commerce and human societies as a whole, and led to claims that social inequality, sexism, racism, and imperialism were justified. However, these ideas contradict Darwin's own views, and contemporary scientists and philosophers consider these ideas to be neither mandated by evolutionary theory nor supported by data.

Chapter- 3

Evolutionary History of Life

The **evolutionary history of life** on Earth traces the processes by which living and fossil organisms evolved. It stretches from the origin of life on Earth, thought to be over 3,500 million years ago, to the present day. The similarities between all present day organisms indicate the presence of a common ancestor from which all known species have diverged through the process of evolution.

Microbial mats of coexisting bacteria and archaea were the dominant form of life in the early Archean and many of the major steps in early evolution are thought to have taken place within them. The evolution of oxygenic photosynthesis, around 3,500 million years ago, eventually led to the oxygenation of the atmosphere, beginning around 2,400 million years ago. The earliest evidence of eukaryotes (complex cells with organelles), dates from 1,850 million years ago, and while they may have been present earlier, their diversification accelerated when they started using oxygen in their metabolism. Later, around 1,700 million years ago, multicellular organisms began to appear, with differentiated cells performing specialised functions.

The earliest land plants date back to around 450 million years ago, though evidence suggests that algal scum formed on the land as early as 1,200 million years ago. Land plants were so successful that they are thought to have contributed to the late Devonian extinction event. Invertebrate animals appear during the Vendian period, while vertebrates originated about 525 million years ago during the Cambrian explosion.

During the Permian period, synapsids, including the ancestors of mammals, dominated the land, but the Permian–Triassic extinction event 251 million years ago came close to wiping out all complex life. During the recovery from this catastrophe, archosaurs became the most abundant land vertebrates, displacing therapsids in the mid-Triassic. One archosaur group, the dinosaurs, dominated the Jurassic and Cretaceous periods, with the ancestors of mammals surviving only as small insectivores. After the Cretaceous–Tertiary extinction event 65 million years ago killed off the non-avian dinosaurs mammals increased rapidly in size and diversity. Such mass extinctions may have accelerated evolution by providing opportunities for new groups of organisms to diversify.

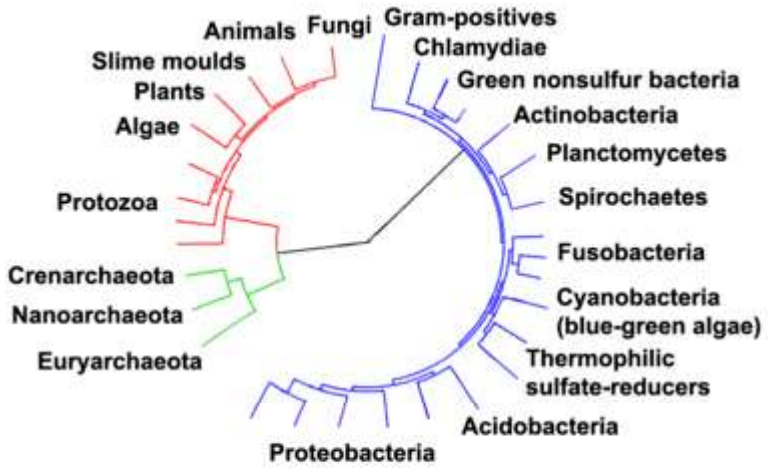
Fossil evidence indicates that flowering plants appeared and rapidly diversified in the Early Cretaceous, between 130 million years ago and 90 million years ago, probably

helped by coevolution with pollinating insects. Flowering plants and marine phytoplankton are still the dominant producers of organic matter. Social insects appeared around the same time as flowering plants. Although they occupy only small parts of the insect "family tree", they now form over half the total mass of insects. Humans evolved from a lineage of upright-walking apes whose earliest fossils date from over 6 million years ago. Although early members of this lineage had chimpanzee-sized brains, there are signs of a steady increase in brain size after about 3 million years ago.

Earliest evidence for life on Earth

The earliest identified organisms were minute and relatively featureless, their fossils look like small rods, which are very difficult to tell apart from structures that arise through abiotic physical processes. The oldest undisputed evidence of life on Earth, interpreted as fossilized bacteria, dates to 3,000 million years ago. Other finds in rocks dated to about 3,500 million years ago have been interpreted as bacteria, with geochemical evidence also seeming to show the presence of life 3,800 million years ago. However these analyses were closely scrutinized, and non-biological processes were found which could produce all of the "signatures of life" that had been reported. While this does not prove that the structures found had a non-biological origin, they cannot be taken as clear evidence for the presence of life. Currently, the oldest unchallenged evidence for life is geochemical signatures from rocks deposited 3,400 million years ago, although these statements have not been thoroughly examined by critics.

Origins of life on Earth



Evolutionary tree showing the divergence of modern species from their common ancestor in the center. The three domains are colored, with bacteria blue, archaea green, and eukaryotes red.

Biologists reason that all living organisms on Earth must share a single last universal ancestor, because it would be virtually impossible that two or more separate lineages could have independently developed the many complex biochemical mechanisms common to all living organisms. As previously mentioned the earliest organisms for which fossil evidence is available are bacteria, cells far too complex to have arisen directly from non-living materials. The lack of fossil or geochemical evidence for earlier organisms has left plenty of scope for hypotheses, which fall into two main groups: 1) that life arose spontaneously on Earth or 2) that it was "seeded" from elsewhere in the universe.

Life "seeded" from elsewhere

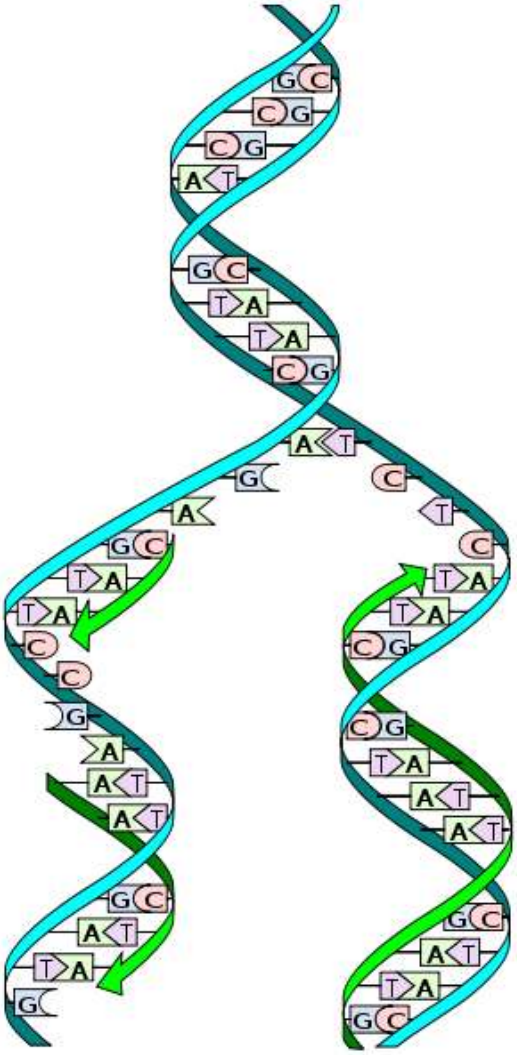
The idea that life on Earth was "seeded" from elsewhere in the universe dates back at least to the fifth century BCE. In the twentieth century it was proposed by the physical chemist Svante Arrhenius, by the astronomers Fred Hoyle and Chandra Wickramasinghe, and by molecular biologist Francis Crick and chemist Leslie Orgel. There are three main versions of the "seeded from elsewhere" hypothesis: from elsewhere in our Solar system via fragments knocked into space by a large meteor impact, in which case the only credible source is Mars; by alien visitors, possibly as a result of accidental contamination by micro-organisms that they brought with them; and from outside the Solar system but by natural means. Experiments suggest that some micro-organisms can survive the shock of being catapulted into space and some can survive exposure to radiation for several days, but there is no proof that they can survive in space for much longer periods. Scientists are divided over the likelihood of life arising independently on Mars, or on other planets in our galaxy.

Independent emergence on Earth

Life on Earth is based on carbon and water. Carbon provides stable frameworks for complex chemicals and can be easily extracted from the environment, especially from carbon dioxide. The only other element with similar chemical properties, silicon, forms much less stable structures and, because most of its compounds are solids, would be more difficult for organisms to extract. Water is an excellent solvent and has two other useful properties: the fact that ice floats enables aquatic organisms to survive beneath it in winter; and its molecules have electrically negative and positive ends, which enables it to form a wider range of compounds than other solvents can. Other good solvents, such as ammonia, are liquid only at such low temperatures that chemical reactions may be too slow to sustain life, and lack water's other advantages. Organisms based on alternative biochemistry may however be possible on other planets.

Research on how life might have emerged unaided from non-living chemicals focuses on three possible starting points: self-replication, an organism's ability to produce offspring that are very similar to itself; metabolism, its ability to feed and repair itself; and external cell membranes, which allow food to enter and waste products to leave, but exclude unwanted substances. Research on abiogenesis still has a long way to go, since theoretical and empirical approaches are only beginning to make contact with each other.

Replication first: RNA world



The replicator in virtually all known life is deoxyribonucleic acid. DNA's structure and replication systems are far more complex than those of the original replicator.

Even the simplest members of the three modern domains of life use DNA to record their "recipes" and a complex array of RNA and protein molecules to "read" these instructions and use them for growth, maintenance and self-replication. This system is far too complex to have emerged directly from non-living materials. The discovery that some RNA molecules can catalyze both their own replication and the construction of proteins led to the hypothesis of earlier life-forms based entirely on RNA. These ribozymes could have formed an RNA world in which there were individuals but no species, as mutations and horizontal gene transfers would have meant that the offspring in each generation were quite likely to have different genomes from those that their parents started with. RNA would later have been replaced by DNA, which is more stable and therefore can build longer genomes, expanding the range of capabilities a single organism can have.

Ribozymes remain as the main components of ribosomes, modern cells' "protein factories".

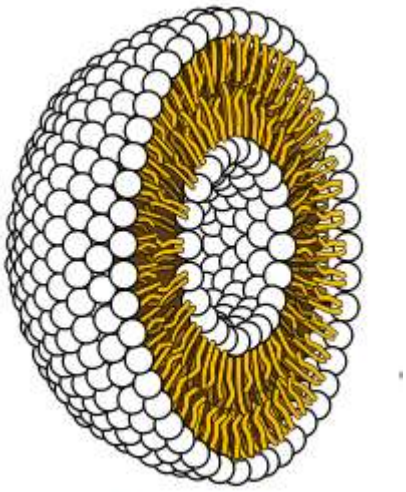
Although short self-replicating RNA molecules have been artificially produced in laboratories, doubts have been raised about where natural non-biological synthesis of RNA is possible. The earliest "ribozymes" may have been formed of simpler nucleic acids such as PNA, TNA or GNA, which would have been replaced later by RNA.

In 2003 it was proposed that porous metal sulfide precipitates would assist RNA synthesis at about 100 °C (212 °F) and ocean-bottom pressures near hydrothermal vents. In this hypothesis lipid membranes would be the last major cell components to appear and until then the proto-cells would be confined to the pores.

Metabolism first: Iron-sulfur world

A series of experiments starting in 1997 showed that early stages in the formation of proteins from inorganic materials including carbon monoxide and hydrogen sulfide could be achieved by using iron sulfide and nickel sulfide as catalysts. Most of the steps required temperatures of about 100 °C (212 °F) and moderate pressures, although one stage required 250 °C (482 °F) and a pressure equivalent to that found under 7 kilometres (4.3 mi) of rock. Hence it was suggested that self-sustaining synthesis of proteins could have occurred near hydrothermal vents.

Membranes first: Lipid world



= water-attracting heads of lipid molecules

= water-repellent tails

Cross-section through a liposome

It has been suggested that double-walled "bubbles" of lipids like those that form the external membranes of cells may have been an essential first step. Experiments that simulated the conditions of the early Earth have reported the formation of lipids, and these can spontaneously form liposomes, double-walled "bubbles", and then reproduce themselves. Although they are not intrinsically information-carriers as nucleic acids are, they would be subject to natural selection for longevity and reproduction. Nucleic acids such as RNA might then have formed more easily within the liposomes than they would have outside.

The clay theory

RNA is complex and there are doubts about whether it can be produced non-biologically in the wild. Some clays, notably montmorillonite, have properties that make them plausible accelerators for the emergence of an RNA world: they grow by self-replication of their crystalline pattern; they are subject to an analog of natural selection, as the clay "species" that grows fastest in a particular environment rapidly becomes dominant; and they can catalyze the formation of RNA molecules. Although this idea has not become the scientific consensus, it still has active supporters.

Research in 2003 reported that montmorillonite could also accelerate the conversion of fatty acids into "bubbles", and that the "bubbles" could encapsulate RNA attached to the clay. These "bubbles" can then grow by absorbing additional lipids and then divide. The formation of the earliest cells may have been aided by similar processes.

A similar hypothesis presents self-replicating iron-rich clays as the progenitors of nucleotides, lipids and amino acids.

Environmental and evolutionary impact of microbial mats

Modern stromatolites in Shark Bay, Western Australia

Microbial mats are multi-layered, multi-species colonies of bacteria and other organisms that are generally only a few millimeters thick, but still contain a wide range of chemical environments, each of which favors a different set of micro-organisms. To some extent each mat forms its own food chain, as the by-products of each group of micro-organisms generally serve as "food" for adjacent groups.

Stromatolites are stubby pillars built as microbes in mats slowly migrate upwards to avoid being smothered by sediment deposited on them by water. There has been vigorous debate about the validity of alleged fossils from before 3,000 million years ago, with critics arguing that so-called stromatolites could have been formed by non-biological processes. In 2006 another find of stromatolites was reported from the same part of Australia as previous ones, in rocks dated to 3,500 million years ago.

In modern underwater mats the top layer often consists of photosynthesizing cyanobacteria which create an oxygen-rich environment, while the bottom layer is oxygen-free and often dominated by hydrogen sulfide emitted by the organisms living there. It is estimated that the appearance of oxygenic photosynthesis by bacteria in mats increased biological productivity by a factor of between 100 and 1,000. The reducing agent used by oxygenic photosynthesis is water, which is much more plentiful than the geologically-produced reducing agents required by the earlier non-oxygenic

photosynthesis. From this point onwards life itself produced significantly more of the resources it needed than did geochemical processes. Oxygen is toxic to organisms that are not adapted to it, but greatly increases the metabolic efficiency of oxygen-adapted organisms. Oxygen became a significant component of Earth's atmosphere about 2,400 million years ago. Although eukaryotes may have been present much earlier, the oxygenation of the atmosphere was a prerequisite for the evolution of the most complex eukaryotic cells, from which all multicellular organisms are built. The boundary between oxygen-rich and oxygen-free layers in microbial mats would have moved upwards when photosynthesis shut down overnight, and then downwards as it resumed on the next day. This would have created selection pressure for organisms in this intermediate zone to acquire the ability to tolerate and then to use oxygen, possibly via endosymbiosis, where one organism lives inside another and both of them benefit from their association.

Cyanobacteria have the most complete biochemical "toolkits" of all the mat-forming organisms. Hence they are the most self-sufficient of the mat organisms and were well-adapted to strike out on their own both as floating mats and as the first of the phytoplankton, providing the basis of most marine food chains.

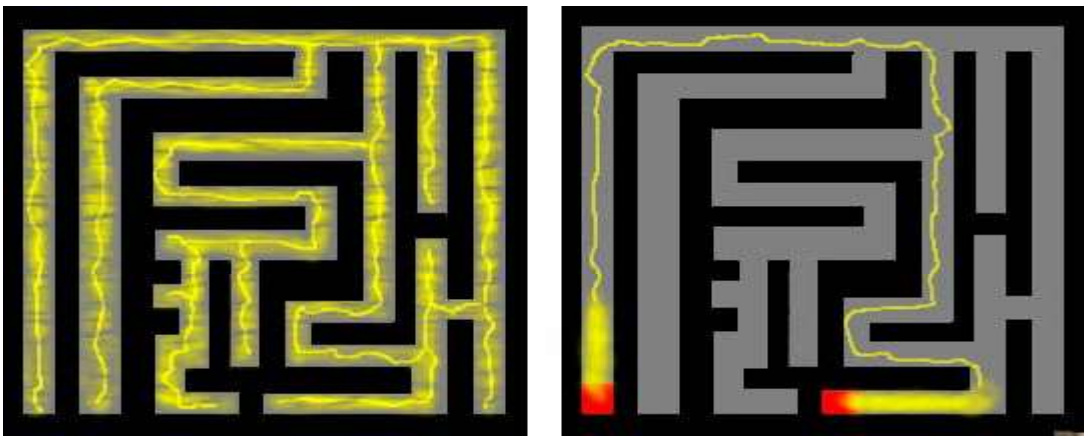
Diversification of eukaryotes

Eukaryotes may have been present long before the oxygenation of the atmosphere, but most modern eukaryotes require oxygen, which their mitochondria use to fuel the production of ATP, the internal energy supply of all known cells. In the 1970s it was proposed and, after much debate, widely accepted that eukaryotes emerged as a result of a sequence of endosymbioses between "procaryotes". For example: a predatory micro-organism invaded a large procaryote, probably an archaean, but the attack was neutralized, and the attacker took up residence and evolved into the first of the mitochondria; one of these chimeras later tried to swallow a photosynthesizing cyanobacterium, but the victim survived inside the attacker and the new combination became the ancestor of plants; and so on. After each endosymbiosis began, the partners would have eliminated unproductive duplication of genetic functions by re-arranging their genomes, a process which sometimes involved transfer of genes between them. Another hypothesis proposes that mitochondria were originally sulfur- or hydrogen-metabolising endosymbionts, and became oxygen-consumers later. On the other hand mitochondria might have been part of eukaryotes' original equipment.

There is a debate about when eukaryotes first appeared: the presence of steranes in Australian shales may indicate that eukaryotes were present 2,700 million years ago; however an analysis in 2008 concluded that these chemicals infiltrated the rocks less than 2,200 million years ago and prove nothing about the origins of eukaryotes. Fossils of the alga *Grypania* have been reported in 1,850 million-year-old rocks (originally dated to 2,100 million years ago but later revised), and indicates that eukaryotes with organelles had already evolved. A diverse collection of fossil algae were found in rocks dated between 1,500 million years ago and 1,400 million years ago. The earliest known fossils of fungi date from 1,430 million years ago.

Multicellular organisms and sexual reproduction

Multicellularity



A slime mold solves a maze. The mold (yellow) explored and filled the maze (left). When the researchers placed sugar (red) at two separate points, the mold concentrated most of its mass there and left only the most efficient connection between the two points (right).

The simplest definitions of "multicellular", for example "having multiple cells", could include colonial cyanobacteria like Nostoc. Even a professional biologist's definition such as "having the same genome but different types of cell" would still include some genera of the green alga Volvox, which have cells that specialize in reproduction. Multicellularity evolved independently in organisms as diverse as sponges and other animals, fungi, plants, brown algae, cyanobacteria, slime moulds and myxobacteria. For the sake of brevity, here focuses on the organisms that show the greatest specialization of cells and variety of cell types, although this approach to the evolution of complexity could be regarded as "rather anthropocentric".

The initial advantages of multicellularity may have included: increased resistance to predators, many of which attacked by engulfing; the ability to resist currents by attaching to a firm surface; the ability to reach upwards to filter-feed or to obtain sunlight for photosynthesis; the ability to create an internal environment that gives protection against the external one; and even the opportunity for a group of cells to behave "intelligently" by sharing information. These features would also have provided opportunities for other organisms to diversify, by creating more varied environments than flat microbial mats could.

Multicellularity with differentiated cells is beneficial to the organism as a whole but disadvantageous from the point of view of individual cells, most of which lose the opportunity to reproduce themselves. In an asexual multicellular organism, rogue cells which retain the ability to reproduce may take over and reduce the organism to a mass of

undifferentiated cells. Sexual reproduction eliminates such rogue cells from the next generation and therefore appears to be a prerequisite for complex multicellularity.

The available evidence indicates that eukaryotes evolved much earlier but remained inconspicuous until a rapid diversification around 1,000 million years ago. The only respect in which eukaryotes clearly surpass bacteria and archaea is their capacity for variety of forms, and sexual reproduction enabled eukaryotes to exploit that advantage by producing organisms with multiple cells that differed in form and function.

Evolution of sexual reproduction

The **evolution of sexual reproduction** is currently described by several competing scientific hypotheses. Many groups of organisms, notably the majority of animals and plants, reproduce sexually. The evolution of sex contains two related, yet distinct, themes: its *origin* and its *maintenance*. However, since the hypotheses for the origins of sex are difficult to test experimentally, most current work has been focused on the maintenance of sexual reproduction.

It seems that a sexual cycle is maintained because it improves the quality of progeny (fitness), despite reducing the overall number of offspring (the twofold cost of sex). In order for sex to be evolutionarily advantageous, it must be associated with a significant increase in the fitness of offspring. One of the most widely accepted explanations for the advantage of sex lies in the *creation of genetic variation*. Another explanation is based on two molecular advantages. First is the advantage of *recombinational DNA repair* (promoted during meiosis because homologous chromosomes pair at that time), and second is the advantage of *complementation* (also known as hybrid vigor, heterosis or masking of mutations).

For the advantage due to creation of genetic variation, there are three possible reasons this might happen. First, sexual reproduction can bring together mutations that are beneficial into the same individual (*sex aids in the spread of advantageous traits*). Second, sex acts to bring together currently deleterious mutations to create severely unfit individuals that are then eliminated from the population (*sex aids in the removal of deleterious genes*). Last, sex creates new gene *combinations* that may be more fit than previously existing ones, or may simply lead to reduced competition among relatives.

For the advantage due to *DNA repair*, there is an immediate large benefit to removal of DNA damage by recombinational DNA repair during meiosis, since this removal allows greater survival of progeny with undamaged DNA. The advantage of *complementation* to each sexual partner is avoidance of the bad effects of their deleterious recessive genes in progeny by the masking effect of normal dominant genes contributed by the other partner.

The classes of hypotheses based on the *creation of variation* are further broken down below. It is important to realise that any number of these hypotheses may be true in any given species (they are not mutually exclusive), and that different hypotheses may apply

in different species. However, a research framework based on *creation of variation* has yet to be found that allows one to determine whether the reason for sex is universal for all sexual species, and, if not, which mechanism is acting in each species.

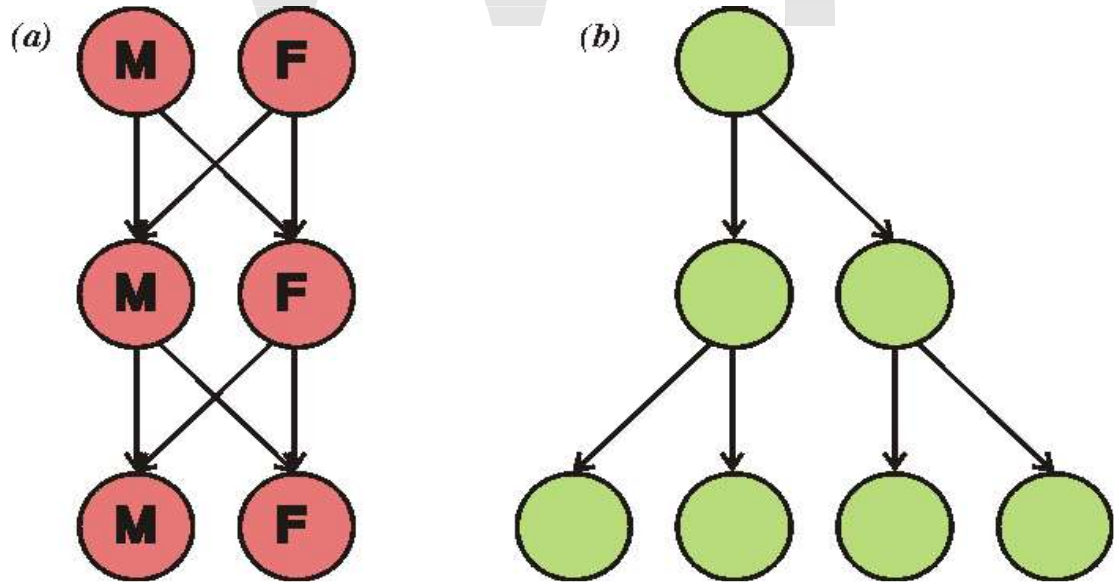
On the other hand, the maintenance of sex based on DNA repair and complementation applies widely to all sexual species. This explanation for the maintenance of sex is explored further in Section 6.2.

Historical perspective

Modern philosophical-scientific thinking on the problem can be traced back to Erasmus Darwin in the 18th century; it also features in Aristotle's writings. The thread was later picked up by August Weismann in 1889, who understood that the purpose of sex was to generate genetic variation, as is detailed in the majority of the explanations below. On the other hand, Charles Darwin, also in 1889, concluded that the effects of hybrid vigor (complementation) "is amply sufficient to account for the ... genesis of the two sexes." This is consistent with the repair and complementation hypothesis, given below under "Other explanations."

Several explanations have been suggested by biologists including W. D. Hamilton, Alexey Kondrashov, George C. Williams, Harris Bernstein, Carol Bernstein, Michael M. Cox, Frederic A. Hopf and Richard E. Michod to explain how sexual reproduction is maintained in a vast array of different living organisms.

Two-fold cost of sex



This diagram illustrates the *twofold cost of sex*. If each individual were to contribute to the same number of offspring (two), *(a)* the sexual population remains the same size each generation, where the *(b)* asexual population doubles in size each generation.

In all sexual species, the population consists of two sexes, only one of which is capable of bearing young (with the exception of simultaneous hermaphrodites). In an asexual species, each member of the population is capable of bearing young. This implies that an asexual population has an intrinsic capacity to grow more rapidly each generation. The cost was first described in mathematical terms by John Maynard Smith. He imagined an asexual mutant arising in a sexual population, half of which comprises males that cannot themselves produce offspring. With female-only offspring, the asexual lineage doubles its representation in the population each generation, all else being equal. Often all else is not equal, however, in which case the realised fitness cost to sex may be much less than this intrinsic twofold cost of producing males. For example, an asexual mutant arising in a sexual population occupies a niche frozen to that of its parental genotype because the asexual descendants are genetically self-identical. Analysis of competitive Lotka-Volterra equations suggests that the asexual lineage may never realise its full twofold advantage in population growth capacity, if the broader niche of the sexual population confers even a small competitive edge.

An additional cost is that males and females must search for each other in order to mate, and sexual selection often favours traits that reduce the survival of individuals.

Evidence that the cost is not insurmountable comes from George C. Williams, who noted the existence of species which are capable of both asexual and sexual reproduction. These species time their sexual reproduction with periods of environmental uncertainty, and reproduce asexually when conditions are more favourable. The important point is that these species are observed to reproduce sexually when they could choose not to, implying that there is a selective advantage to sexual reproduction.

Another disadvantage of sexual reproduction is that a sexually reproducing organism will only be able to pass on 50% of its genes to each offspring. This is a consequence of the fact that gametes from sexually reproducing species are haploid.

Promotion of genetic variation

August Weismann proposed in 1889 an explanation for the evolution of sex, where the advantage of sex is the creation of variation among siblings. It was then subsequently explained in genetics terms by Fisher and Muller and has been recently summarised by Burt in 2000.

George C. Williams gave an example based around the elm tree. In the forest of this example, empty patches between trees can support one individual each. When a patch becomes available because of the death of a tree, other trees' seeds will compete to fill the patch. Since the chance of a seed's success in occupying the patch depends upon its genotype, and a parent cannot anticipate which genotype is most successful, each parent will send many seeds, creating competition between siblings. Natural selection therefore favours parents which can produce a variety of offspring.

A similar hypothesis is named the *tangled bank hypothesis* after a passage in Charles Darwin's *The Origin of Species*:

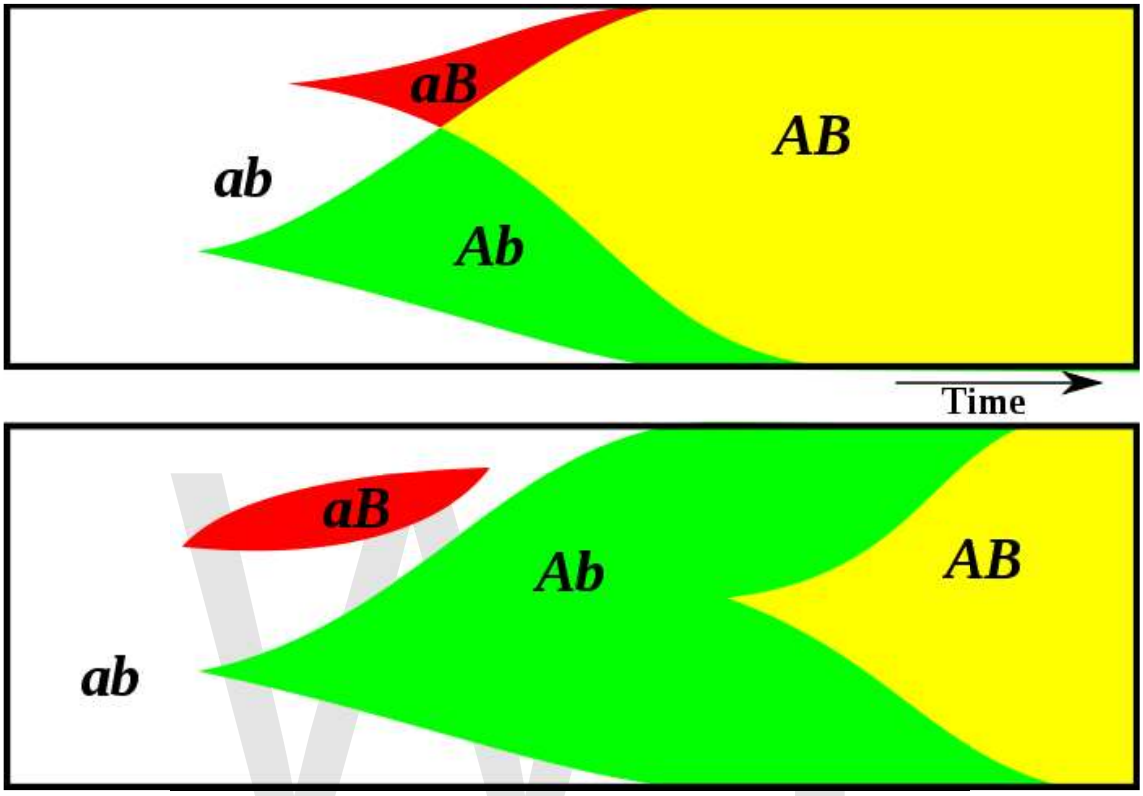
"It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us."

The hypothesis, proposed by Michael Ghiselin in his 1974 book, *The Economy of Nature and the Evolution of Sex*, suggests that a diverse set of siblings may be able to extract more food from its environment than a clone, because each sibling uses a slightly different niche. One of the main proponents of this hypothesis is Graham Bell of McGill University. The hypothesis has been criticised for failing to explain how asexual species developed sexes. In his book, *Evolution and Human Behavior* (MIT Press, 2000), John Cartwright comments:

"Although once popular, the tangled bank hypothesis now seems to face many problems, and former adherents are falling away. The theory would predict a greater interest in sex among animals that produce lots of small offspring that compete with each other. In fact, sex is invariably associated with organisms that produce a few large offspring, whereas organisms producing small offspring frequently engage in parthenogenesis [asexual reproduction]. In addition, the evidence from fossils suggests that species go for vast periods of [geologic] time without changing much."

Spread of advantageous traits

Novel genotypes



This diagram illustrates how sex might create novel genotypes more rapidly. Two advantageous alleles *A* and *B* occur at random. The two alleles are recombined rapidly in a sexual population (top), but in an asexual population (bottom) the two alleles must independently arise because of clonal interference.

Sex could be a method by which novel genotypes are created. Since sex combines genes from two individuals, sexually reproducing populations can more easily combine advantageous genes than can asexual populations. If, in a sexual population, two different advantageous alleles arise at different loci on a chromosome in different members of the population, a chromosome containing the two advantageous alleles can be produced within a few generations by recombination. However, should the same two alleles arise in different members of an asexual population, the only way that one chromosome can develop the other allele is to independently gain the same mutation, which would take much longer.

Ronald Fisher also suggested that sex might facilitate the spread of advantageous genes by allowing them to escape their genetic surroundings, if they should arise on a chromosome with deleterious genes.

But these explanations depend upon the rate of mutation. If favourable mutations are so rare that each will become fixed in the population before the next arises (bearing in mind that mutation is a Poisson process), then sexual and asexual populations would evolve at the same rate.

Additionally, these explanations depend upon group selection, which most theorists maintain is a weak selective force relative to individual selection – sex is still disadvantageous to the individual due to the twofold cost of sex. Therefore, these explanations do not explain why heterogonic species choose to adopt sexual reproduction, as George C. Williams pointed out in his *balance argument*, and hence are insufficient to explain the evolution of sex.

Supporters of these theories respond to the balance argument that the individuals produced by sexual and asexual reproduction may differ in other respects too – which may influence the persistence of sexuality. For example, in water fleas of the genus *Cladocera*, sexual offspring form eggs which are better able to survive the winter.

Increased resistance to parasites

One of the most widely accepted theories to explain the persistence of sex is that it is maintained to assist sexual individuals in resisting parasites, also known as the Red Queen's Hypothesis.

When an environment changes, previously neutral or deleterious alleles can become favourable. If the environment changed sufficiently rapidly (i.e. between generations), these changes in the environment can make sex advantageous for the individual. Such rapid changes in environment are caused by the co-evolution between hosts and parasites.

Imagine, for example that there is one gene in parasites with two alleles *p* and *P* conferring two types of parasitic ability, and one gene in hosts with two alleles *h* and *H*, conferring two types of parasite resistance, such that parasites with allele *p* can attach themselves to hosts with the allele *h*, and *P* to *H*. Such a situation will lead to cyclic changes in allele frequency - as *p* increases in frequency, *h* will be disfavoured.

In reality, there will be several genes involved in the relationship between hosts and parasites. In an asexual population of hosts, offspring will only have the different parasitic resistance if a mutation arises. In a sexual population of hosts, however, offspring will have a new combination of parasitic resistance alleles.

In other words, like Lewis Carroll's Red Queen, sexual hosts are continually adapting in order to stay ahead of their parasites.

Evidence for this explanation for the evolution of sex is provided by comparison of the rate of molecular evolution of genes for kinases and immunoglobulins in the immune system with genes coding other proteins. The genes coding for immune system proteins evolve considerably faster.

Critics of the Red Queen hypothesis question whether the constantly-changing environment of hosts and parasites is sufficiently common to explain the evolution of sex.

Deleterious mutation clearance

Mutations can have many different effects upon an organism. It is generally believed that the majority of non-neutral mutations are deleterious, which means that they will cause a decrease in the organism's overall fitness. If a mutation has a deleterious effect, it will then usually be removed from the population by the process of natural selection. Sexual reproduction is believed to be more efficient than asexual reproduction in removing those mutations from the genome.

There are two main hypotheses which explain how sex may act to remove deleterious genes from the genome.

Maintenance of mutation-free individuals

In a finite asexual population under the pressure of deleterious mutations, the random loss of individuals without such mutations is inevitable. This is known as Muller's ratchet. In a sexual population, however, mutation-free genotypes can be restored by recombination of genotypes containing deleterious mutations.

This comparison will only work for a small population - in a large population, random loss of the most fit genotype becomes unlikely even in an asexual population.

Removal of deleterious genes

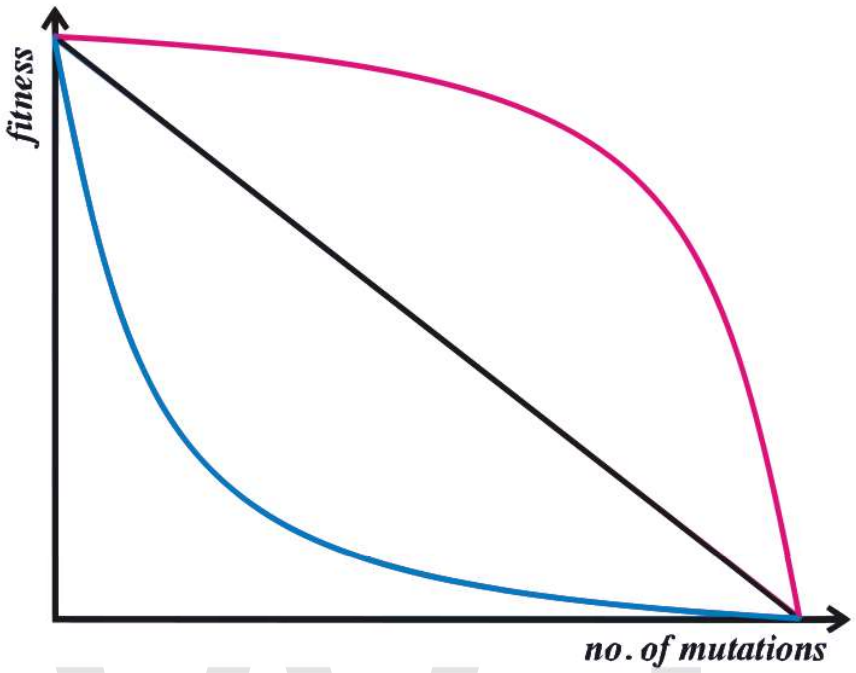


Diagram illustrating different relationships between numbers of mutations and fitness. Kondrashov's model requires *synergistic epistasis*, which is represented by the blue line - each mutation has a disproproportionately large effect on the organism's fitness.

This hypothesis was proposed by Alexey Kondrashov, and is sometimes known as the *deterministic mutation hypothesis*. It assumes that the majority of deleterious mutations are only slightly deleterious, and affect the individual such that the introduction of each additional mutation has an increasingly large effect on the fitness of the organism. This relationship between number of mutations and fitness is known as *synergistic epistasis*.

By way of analogy, think of a car with several minor faults. Each is not sufficient alone to prevent the car from running, but in combination, the faults combine to prevent the car from functioning.

Similarly, an organism may be able to cope with a few defects, but the presence of many mutations could overwhelm its backup mechanisms.

Kondrashov argues that the slightly deleterious nature of mutations means that the population will tend to be composed of individuals with a small number of mutations. Sex will act to recombine these genotypes, creating some individuals with fewer deleterious mutations, and some with more. Because there is a major selective disadvantage to individuals with more mutations, these individuals die out. In essence, sex compartmentalises the deleterious mutations.

There has been much criticism of Kondrashov's theory, since it relies on two key restrictive conditions. The first requires that the rate of deleterious mutation should exceed one per genome per generation in order to provide a substantial advantage for sex. While there is some empirical evidence for it (for example in Drosophila and E. coli), there is also strong evidence against it. Secondly, there should be strong interactions among loci (synergistic epistasis), a mutation-fitness relation for which there is only limited evidence. Conversely, there is also the same amount of evidence that mutations show no epistasis (purely additive model) or antagonistic interactions (each additional mutation has a disproportionally *small* effect).

Other explanations

Speed of evolution

Ilan Eshel suggested that sex prevents rapid evolution. He suggests that recombination breaks up favourable gene combinations more often than it creates them, and sex is maintained because it ensures selection is longer-term than in asexual populations - so the population is less affected by short-term changes. This explanation is not widely accepted, as its assumptions are very restrictive.

It has recently been shown in experiments with *Chlamydomonas* algae that sex can remove the speed limit on evolution.

DNA repair and complementation

As discussed in the earlier, sexual reproduction is conventionally explained as an adaptation for producing genetic variation through allelic recombination. As acknowledged above, however, serious problems with this explanation have led many biologists to conclude that the benefit of sex is a major unsolved problem in evolutionary biology.

An alternative “informational” approach to this problem has led to the view that the two fundamental aspects of sex, genetic recombination and outcrossing, are adaptive responses to the two major sources of “noise” in transmitting genetic information. Genetic noise can occur as either physical damage to the genome (e.g. chemically altered bases of DNA or breaks in the chromosome) or replication errors (mutations) This alternative view is referred to as the repair and complementation hypothesis, to distinguish it from the traditional variation hypothesis.

The repair and complementation hypothesis assumes that genetic recombination is fundamentally a DNA repair process, and that when it occurs during meiosis it is an adaptation for repairing the genomic DNA which is passed on to progeny. Recombinational repair is the only repair process known which can accurately remove double-strand damages in DNA, and such damages are both common in nature and ordinarily lethal if not repaired. Recombinational repair is prevalent from the simplest viruses to the most complex multicellular eukaryotes. It is effective against many

different types of genomic damage, and in particular is highly efficient at overcoming double-strand damages. Studies of the mechanism of meiotic recombination indicate that meiosis is an adaptation for repairing DNA. These considerations form the basis for the first part of the repair and complementation hypothesis.

In some lines of descent from the earliest organisms, the diploid stage of the sexual cycle, which was at first transient, became the predominant stage, because it allowed complementation — the masking of deleterious recessive mutations (i.e. hybrid vigor or heterosis). Outcrossing, the second fundamental aspect of sex, is maintained by the advantage of masking mutations and the disadvantage of inbreeding (mating with a close relative) which allows expression of recessive mutations (commonly observed as inbreeding depression). This is in accord with Charles Darwin, who concluded that the adaptive advantage of sex is hybrid vigor; or as he put it, "the offspring of two individuals, especially if their progenitors have been subjected to very different conditions, have a great advantage in height, weight, constitutional vigor and fertility over the self fertilised offspring from either one of the same parents."

However, outcrossing may be abandoned in favor of parthogenesis or selfing (which retain the advantage of meiotic recombinational repair) under conditions in which the costs of mating are very high. For instance, costs of mating are high when individuals are rare in a geographic area, such as when there has been a forest fire and the individuals entering the burned area are the initial ones to arrive. At such times mates are hard to find, and this favors parthenogenic species.

In the view of the repair and complementation hypothesis, the removal of DNA damage by recombinational repair produces a new, less deleterious form of informational noise, allelic recombination, as a by-product.

Origin of sexual reproduction

All sexually reproducing organisms derive from a common ancestor which was a single celled eukaryotic species. Many protists reproduce sexually, as do the multicellular plants, animals, and fungi. There are a few species which have secondarily lost this feature, such as Bdelloidea and some parthenocarpic plants.

Organisms need to replicate their genetic material in an efficient and reliable manner. The necessity to repair genetic damage is one of the leading theories explaining the origin of sexual reproduction. Diploid individuals can repair a mutated section of its DNA via homologous recombination, since there are two copies of the gene in the cell and one copy is presumed to be undamaged. A mutation in an haploid individual, on the other hand, is more likely to become resident, as the DNA repair machinery has no way of knowing what the original undamaged sequence was. The most primitive form of sex may have been one organism with damaged DNA replicating an undamaged strand from a similar organism in order to repair itself.

Another theory is that sexual reproduction originated from selfish parasitic genetic elements that exchange genetic material (that is: copies of their own genome) for their transmission and propagation. In some organisms, sexual reproduction has been shown to enhance the spread of parasitic genetic elements (e.g.: yeast, filamentous fungi). Bacterial conjugation, a form of genetic exchange that some sources describe as sex, is not a form of reproduction, but rather an example of horizontal gene transfer. However, it does support the selfish genetic element theory, as it is propagated through such a "selfish gene", the F-plasmid. Similarly, it has been proposed that sexual reproduction evolved from ancient haloarchaea through a combination of jumping genes, and swapping plasmids.

A third theory is that sex evolved as a form of cannibalism. One primitive organism ate another one, but rather than completely digesting it, some of the 'eaten' organism's DNA was incorporated into the 'eater' organism.

Sex may also be derived from prokaryotic processes. A comprehensive 'origin of sex as vaccination' theory proposes that eukaryan sex-as-syngamy (fusion sex) arose from prokaryan unilateral sex-as-infection when infected hosts began swapping nuclearised genomes containing coevolved, vertically transmitted symbionts that provided protection against horizontal superinfection by more virulent symbionts. Sex-as-meiosis (fission sex) then evolved as a host strategy to uncouple (and thereby emasculate) the acquired symbiont genomes.

Mechanistic origin of sexual reproduction

Though theories positing benefits that lead to the origin of sex are often problematic, several credible theories on the evolution of the mechanisms of sexual reproduction have been proposed.

Viral eukaryogenesis

The viral eukaryogenesis (VE) theory proposes that eukaryotic cells arose from a combination of a lysogenic virus, an archaeon and a bacterium. This model suggests that the nucleus originated when the lysogenic virus incorporated genetic material from the archaeon and the bacterium and took over the role of information storage for the amalgam. The archaeal host transferred much of its functional genome to the virus during the evolution of cytoplasm but retained the function of gene translation and general metabolism. The bacterium transferred most of its functional genome to the virus as it transitioned into a mitochondrion.

For these transformations to lead to the eukaryotic cell cycle, the VE hypothesis specifies a pox-like virus as the lysogenic virus. A pox-like virus is a likely ancestor because of its fundamental similarities with eukaryotic nuclei. These include a double stranded DNA genome, a linear chromosome with short telomeric repeats, a complex membrane bound capsid, the ability to produce capped mRNA, and the ability to export the capped mRNA across the viral membrane into the cytoplasm. The presence of a lysogenic pox-like virus

ancestor explains the development of meiotic division, an essential component of sexual reproduction.

Meiotic division in the VE hypothesis arose because of the evolutionary pressures placed on the lysogenic virus as a result of its inability to enter into the lytic cycle. This selective pressure resulted in the development of processes allowing the viruses to spread horizontally throughout the population. The outcome of this selection was cell-to-cell fusion. (This is distinct from the conjugation methods used by bacterial plasmids under evolutionary pressure, with important consequences.) The possibility of this kind of fusion is supported by the presence of fusion proteins in the envelopes of the pox viruses that allow them to fuse with host membranes. These proteins could have been transferred to the cell membrane during viral reproduction, enabling cell-to-cell fusion between the virus host and an uninfected cell. The theory proposes meiosis originated from the fusion between two cells infected with related but different viruses which recognised each other as uninfected. After the fusion of the two cells, incompatibilities between the two viruses result in a meiotic-like cell division.

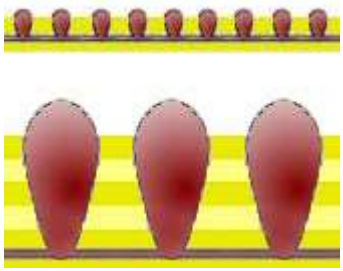
The two viruses established in the cell would initiate replication in response to signals from the host cell. A mitosis-like cell cycle would proceed until the viral membranes dissolved, at which point linear chromosomes would be bound together with centromeres. The homologous nature of the two viral centromeres would incite the grouping of both sets into tetrads. It is speculated that this grouping may be the origin of crossing over, characteristic of the first division in modern meiosis. The partitioning apparatus of the mitotic-like cell cycle the cells used to replicate independently would then pull each set of chromosomes to one side of the cell, still bound by centromeres. These centromeres would prevent their replication in subsequent division, resulting in four daughter cells with one copy of one of the two original pox-like viruses. The process resulting from combination of two similar pox viruses within the same host closely mimics meiosis.

Neomuran revolution

An alternative theory, proposed by Thomas Cavalier-Smith, was labeled the Neomuran revolution. The designation "Neomuran revolution" refers to the appearances of the common ancestors of eukaryotes and archaea. Cavalier-Smith proposes that the first neomurans emerged 850 million years ago. Other molecular biologists assume that this group appeared much earlier, but Cavalier-Smith dismisses these claims because they are based on the "theoretically and empirically" unsound model of molecular clocks. Cavalier-Smith's theory of the Neomuran revolution has implications for the evolutionary history of the cellular machinery for recombination and sex. It suggests that this machinery evolved in two distinct bouts separated by a long period of stasis; first the appearance of recombination machinery in a bacterial ancestor which was maintained for 3 Gy, until the neomuran revolution when the mechanics were adapted to the presence of nucleosomes. The archaeal products of the revolution maintained recombination machinery that was essentially bacterial, whereas the eukaryotic products broke with this bacterial continuity. They introduced cell fusion and ploidy cycles into cell life histories.

Cavalier-Smith argues that both bouts of mechanical evolution were motivated by similar selective forces: the need for accurate DNA replication without loss of viability.

Fossil evidence for multicellularity and sexual reproduction



Horodyskia may have been an early metazoan, or a colonial foraminiferan

The Francevillian Group Fossil, dated to 2,100 million years ago, is the earliest known fossil organism that is clearly multicellular. They may have had differentiated cells. Another early multicellular fossil, *Qingshania*, dated to 1,700 million years ago, appears to consist of virtually identical cells. The red alga called *Bangiomorpha*, dated at 1,200 million years ago, is the earliest known organism which certainly has differentiated, specialized cells, and is also the oldest known sexually-reproducing organism. The 1,430 million-year-old fossils interpreted as fungi appear to have been multicellular with differentiated cells. The "string of beads" organism *Horodyskia*, found in rocks dated from 1,500 million years ago to 900 million years ago, may have been an early metazoan; however it has also been interpreted as a colonial foraminiferan.

Emergence of animals

Animals are multicellular eukaryotes and are distinguished from plants, algae, and fungi by lacking cell walls. All animals are motile, if only at certain life stages. All animals except sponges have bodies differentiated into separate tissues, including muscles, which move parts of the animal by contracting, and nerve tissue, which transmits and processes signals.

The earliest widely-accepted animal fossils are rather modern-looking cnidarians (the group that includes jellyfish, sea anemones and hydras), possibly from around 580 million years ago, although fossils from the Doushantuo Formation can only be dated approximately. Their presence implies that the cnidarian and bilaterian lineages had already diverged.

The Ediacara biota, which flourished for the last 40 million years before the start of the Cambrian, were the first animals more than a very few centimeters long. Many were flat and had a "quilted" appearance, and seemed so strange that there was a proposal to classify them as a separate kingdom, Vendozoa. Others, however, been interpreted as

early molluscs (*Kimberella*), echinoderms (*Arkarua*), and arthropods (*Spriggina*, *Parvancorina*). There is still debate about the classification of these specimens, mainly because the diagnostic features which allow taxonomists to classify more recent organisms, such as similarities to living organisms, are generally absent in the Ediacarans. However there seems little doubt that *Kimberella* was at least a triploblastic bilaterian animal, in other words significantly more complex than cnidarians.

The small shelly fauna are a very mixed collection of fossils found between the Late Ediacaran and Mid Cambrian periods. The earliest, *Cloudina*, shows signs of successful defense against predation and may indicate the start of an evolutionary arms race. Some tiny Early Cambrian shells almost certainly belonged to molluscs, while the owners of some "armor plates", *Halkieria* and *Microdictyon*, were eventually identified when more complete specimens were found in Cambrian lagerstätten that preserved soft-bodied animals.

Opabinia made the largest single contribution to modern interest in the Cambrian explosion.

In the 1970s there was already a debate about whether the emergence of the modern phyla was "explosive" or gradual but hidden by the shortage of Pre-Cambrian animal fossils. A re-analysis of fossils from the Burgess Shale lagerstätte increased interest in the issue when it revealed animals, such as *Opabinia*, which did not fit into any known

phylum. At the time these were interpreted as evidence that the modern phyla had evolved very rapidly in the "Cambrian explosion" and that the Burgess Shale's "weird wonders" showed that the Early Cambrian was a uniquely experimental period of animal evolution. Later discoveries of similar animals and the development of new theoretical approaches led to the conclusion that many of the "weird wonders" were evolutionary "aunts" or "cousins" of modern groups – for example that *Opabinia* was a member of the lobopods, a group which includes the ancestors of the arthropods, and that it may have been closely related to the modern tardigrades. Nevertheless there is still much debate about whether the Cambrian explosion was really explosive and, if so, how and why it happened and why it appears unique in the history of animals.

Acanthodians were among the earliest vertebrates with jaws

Most of the animals at the heart of the Cambrian explosion debate are protostomes, one of the two main groups of complex animals. One deuterostome group, the echinoderms, many of which have hard calcite "shells", are fairly common from the Early Cambrian small shelly fauna onwards. Other deuterostome groups are soft-bodied, and most of the significant Cambrian deuterostome fossils come from the Chengjiang fauna, a lagerstätte in China. The Chengjiang fossils *Haikouichthys* and *Myllokunmingia* appear to be true vertebrates, and *Haikouichthys* had distinct vertebrae, which may have been slightly mineralized. Vertebrates with jaws, such as the Acanthodians, first appeared in the Late Ordovician.

Colonization of land

Adaptation to life on land is a major challenge: all land organisms need to avoid drying-out and all those above microscopic size have to resist gravity; respiration and gas exchange systems have to change; reproductive systems cannot depend on water to carry eggs and sperm towards each other. Although the earliest good evidence of land plants and animals dates back to the Ordovician period (488 to 444 million years ago), modern land ecosystems only appeared in the late Devonian, about 385 to 359 million years ago.

Evolution of soil

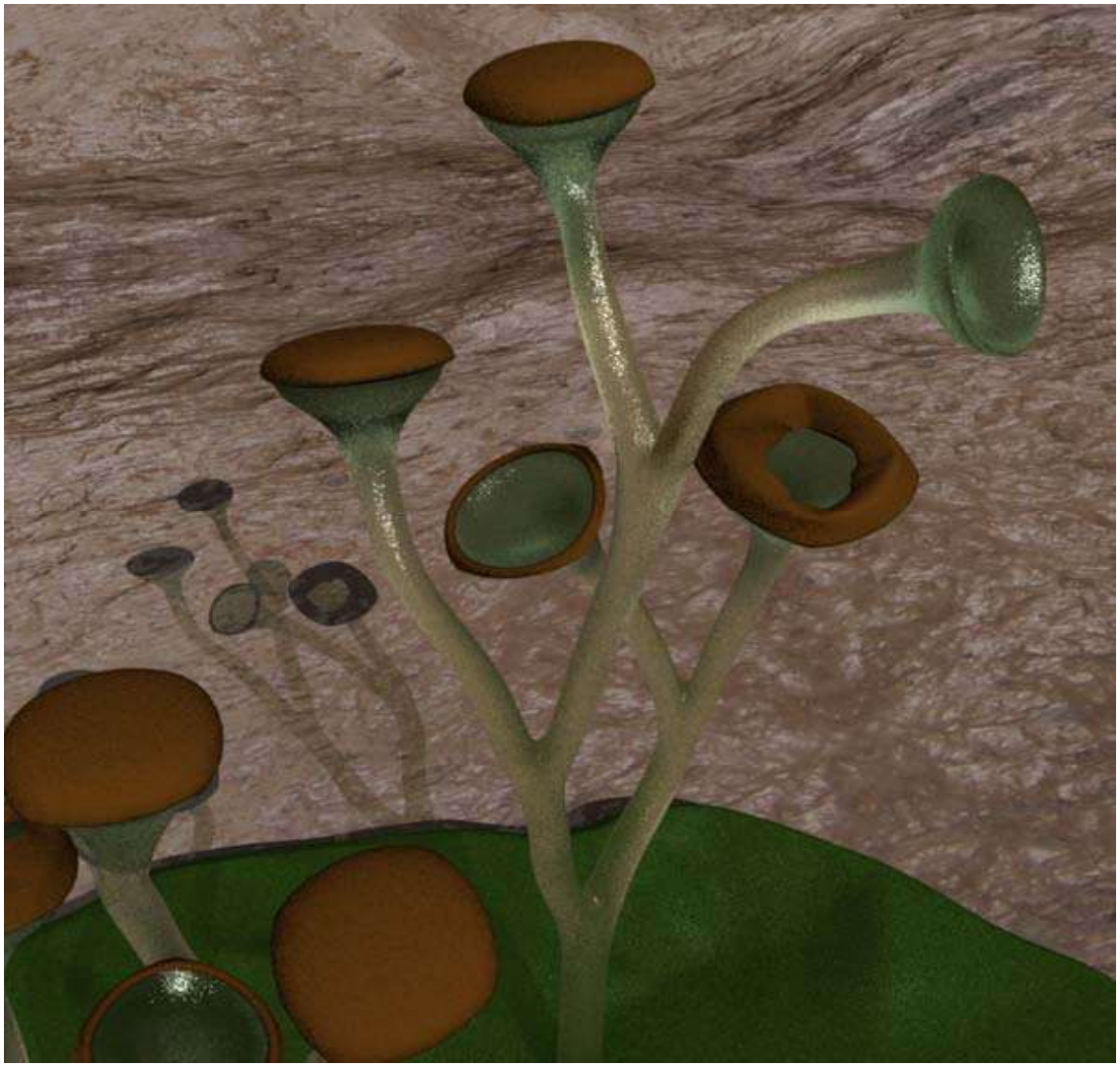
Before the colonization of land, soil, a combination of mineral particles and decomposed organic matter, did not exist. Land surfaces would have been either bare rock or unstable sand produced by weathering. Water and any nutrients in it would have drained away very quickly.

Lichens growing on concrete

Films of cyanobacteria, which are not plants but use the same photosynthesis mechanisms, have been found in modern deserts, and only in areas that are unsuitable for vascular plants. This suggests that microbial mats may have been the first organisms to colonize dry land, possibly in the Precambrian. Mat-forming cyanobacteria could have gradually evolved resistance to desiccation as they spread from the seas to tidal zones and then to land. Lichens, which are symbiotic combinations of a fungus (almost always an ascomycete) and one or more photosynthesizers (green algae or cyanobacteria), are also important colonizers of lifeless environments, and their ability to break down rocks contributes to soil formation in situations where plants cannot survive. The earliest known ascomycete fossils date from 423 to 419 million years ago in the Silurian.

Soil formation would have been very slow until the appearance of burrowing animals, which mix the mineral and organic components of soil and whose feces are a major source of the organic components. Burrows have been found in Ordovician sediments, and are attributed to annelids ("worms") or arthropods.

Plants and the Late Devonian wood crisis



Reconstruction of *Cooksonia*, a vascular plant from the Silurian

Fossilized trees from the Mid-Devonian Gilboa fossil forest

In aquatic algae, almost all cells are capable of photosynthesies and are nearly independent. Life on land required plants to become internally more complex and specialized: photosynthesis was most efficient at the top; roots were required in order to extract water from the ground; the parts in between became supports and transport systems for water and nutrients.

Spores of land plants, possibly rather like liverworts, have been found in Mid Ordovician rocks dated to about 476 million years ago. In Mid Silurian rocks 430 million years ago there are fossils of actual plants including clubmosses such as *Baragwanathia*; most were under 10 centimetres (3.9 in) high, and some appear closely related to vascular plants, the group that includes trees.

By the Late Devonian 370 million years ago, trees such as *Archaeopteris* were so abundant that they changed river systems from mostly braided to mostly meandering, because their roots bound the soil firmly. In fact they caused a "Late Devonian wood crisis", because:

- They removed more carbon dioxide from the atmosphere, reducing the greenhouse effect and thus causing an ice age in the Carboniferous period. In later ecosystems the carbon dioxide "locked up" in wood is returned to the atmosphere by decomposition of dead wood. However the earliest fossil evidence of fungi that can decompose wood also comes from the Late Devonian.
- The increasing depth of plants' roots led to more washing of nutrients into rivers and seas by rain. This caused algal blooms whose high consumption of oxygen caused anoxic events in deeper waters, increasing the extinction rate among deep-water animals.

Land invertebrates

Animals had to change their feeding and excretory systems, and most land animals developed internal fertilization of their eggs. The difference in refractive index between water and air required changes in their eyes. On the other hand in some ways movement and breathing became easier, and the better transmission of high-frequency sounds in air encouraged the development of hearing.

Some trace fossils from the Cambrian-Ordovician boundary about 490 million years ago are interpreted as the tracks of large amphibious arthropods on coastal sand dunes, and may have been made by euthycarcinoids, which are thought to be evolutionary "aunts" of myriapods. Other trace fossils from the Late Ordovician a little over 445 million years ago probably represent land invertebrates, and there is clear evidence of numerous arthropods on coasts and alluvial plains shortly before the Silurian-Devonian boundary, about 415 million years ago, including signs that some arthropods ate plants. Arthropods were well pre-adapted to colonise land, because their existing jointed exoskeletons provided protection against desiccation, support against gravity and a means of locomotion that was not dependent on water.

The fossil record of other major invertebrate groups on land is poor: none at all for non-parasitic flatworms, nematodes or nemerteans; some parasitic nematodes have been fossilized in amber; annelid worm fossils are known from the Carboniferous, but they may still have been aquatic animals; the earliest fossils of gastropods on land date from the Late Carboniferous, and this group may have had to wait until leaf litter became abundant enough to provide the moist conditions they need.

The earliest confirmed fossils of flying insects date from the Late Carboniferous, but it is thought that insects developed the ability to fly in the Early Carboniferous or even Late Devonian. This gave them a wider range of ecological niches for feeding and breeding, and a means of escape from predators and from unfavorable changes in the environment. About 99% of modern insect species fly or are descendants of flying species.

Land vertebrates

Acanthostega changed views about the early evolution of tetrapods

Tetrapods, vertebrates with four limbs, evolved from other rhipidistians over a relatively short timespan during the Late Devonian, between 370 million years ago and 360 million years ago. From the 1950s to the early 1980s it was thought that tetrapods evolved from fish that had already acquired the ability to crawl on land, possibly in order to go from a pool that was drying out to one that was deeper. However in 1987 nearly-complete fossils of *Acanthostega* from about 363 million years ago showed that this Late Devonian transitional animal had legs and both lungs and gills, but could never have survived on land: its limbs and its wrist and ankle joints were too weak to bear its weight; its ribs were too short to prevent its lungs from being squeezed flat by its weight; its fish-like tail fin would have been damaged by dragging on the ground. The current hypothesis is that *Acanthostega*, which was about 1 metre (3.3 ft) long, was a wholly aquatic predator that hunted in shallow water. Its skeleton differed from that of most fish, in ways that enabled it to raise its head to breathe air while its body remained submerged, including: its jaws show modifications that would have enabled it to gulp air; the bones at the back of its skull are locked together, providing strong attachment points for muscles that raised its head; the head is not joined to the shoulder girdle and it has a distinct neck.

The Devonian proliferation of land plants may help to explain why air-breathing would have been an advantage: leaves falling into streams and rivers would have encouraged the growth of aquatic vegetation; this would have attracted grazing invertebrates and small fish that preyed on them; they would have been attractive prey but the environment was unsuitable for the big marine predatory fish; air-breathing would have been necessary because these waters would have been short of oxygen, since warm water holds less dissolved oxygen than cooler marine water and since the decomposition of vegetation would have used some of the oxygen.

Later discoveries revealed earlier transitional forms between *Acanthostega* and completely fish-like animals. Unfortunately there is then a gap of about 30 million years between the fossils of ancestral tetrapods and Mid Carboniferous fossils of vertebrates that look well-adapted for life on land. Some of these look like early relatives of modern amphibians, most of which need to keep their skins moist and to lay their eggs in water, while others are accepted as early relatives of the amniotes, whose water-proof skins and eggs enable them to live and breed far from water.

Dinosaurs, birds and mammals

Amniotes, whose eggs can survive in dry environments, probably evolved in the Late Carboniferous period, between 330 million years ago and 314 million years ago. The earliest fossils of the two surviving amniote groups, synapsids and sauropsids, date from around 313 million years ago. The synapsid pelycosaurs and their descendants the therapsids are the most common land vertebrates in the best-known Permian fossil beds, between 229 million years ago and 251 million years ago. However at the time these were all in temperate zones at middle latitudes, and there is evidence that hotter, drier environments nearer the Equator were dominated by sauropsids and amphibians.

The Permian-Triassic extinction wiped out almost all land vertebrates, as well as the great majority of other life. During the slow recovery from this catastrophe, estimated to be 30M years, a previously obscure sauropsid group became the most abundant and diverse terrestrial vertebrates: a few fossils of archosauriformes ("shaped like archosaurs") have been found in Late Permian rocks, but by the Mid Triassic archosaurs were the dominant land vertebrates. Dinosaurs distinguished themselves from other archosaurs in the Late Triassic, and became the dominant land vertebrates of the Jurassic and Cretaceous periods, between 199 million years ago and 65 million years ago.

During the Late Jurassic, birds evolved from small, predatory theropod dinosaurs. The first birds inherited teeth and long, bony tails from their dinosaur ancestors, but some developed horny, toothless beaks by the very Late Jurassic and short pygostyle tails by the Early Cretaceous.

While the archosaurs and dinosaurs were becoming more dominant in the Triassic, the mammaliform successors of the therapsids could only survive as small, mainly nocturnal insectivores. This apparent set-back may actually have promoted the evolution of mammals, for example nocturnal life may have accelerated the development of

endothermy ("warm-bloodedness") and hair or fur. By 195 million years ago in the Early Jurassic there were animals that were very nearly mammals. Unfortunately there is a gap in the fossil record throughout the Mid Jurassic. However fossil teeth discovered in Madagascar indicate that true mammals existed at least 167 million years ago. After dominating land vertebrate niches for about 150 million years, the dinosaurs perished 65 million years ago in the Cretaceous–Tertiary extinction along with many other groups of organisms. Mammals throughout the time of the dinosaurs had been restricted to a narrow range of taxa, sizes and shapes, but increased rapidly in size and diversity after the extinction, with bats taking to the air within 13 million years, and cetaceans to the sea within 15 million years.

Flowering plants

The 250,000 to 400,000 species of flowering plants outnumber all other ground plants combined, and are the dominant vegetation in most terrestrial ecosystems. There is fossil evidence that flowering plants diversified rapidly in the Early Cretaceous, between 130 million years ago and 90 million years ago, and that their rise was associated with that of pollinating insects. Among modern flowering plants Magnolias are thought to be close to the common ancestor of the group. However paleontologists have not succeeded in identifying the earliest stages in the evolution of flowering plants.

Social insects

The social insects are remarkable because the great majority of individuals in each colony are sterile. This appears contrary to basic concepts of evolution such as natural selection and the selfish gene. In fact there are very few eusocial insect species: only 15 out of approximately 2,600 living families of insects contain eusocial species, and it seems that eusociality has evolved independently only 12 times among arthropods, although some eusocial lineages have diversified into several families. Nevertheless social insects have been spectacularly successful; for example although ants and termites account for only about 2% of known insect species, they form over 50% of the total mass of insects. Their ability to control a territory appears to be the foundation of their success.

These termite mounds have survived a bush fire

The sacrifice of breeding opportunities by most individuals has long been explained as a consequence of these species' unusual haplodiploid method of sex determination, which has the paradoxical consequence that two sterile worker daughters of the same queen share more genes with each other than they would with their offspring if they could breed. However Wilson and Hölldobler argue that this explanation is faulty: for example, it is based on kin selection, but there is no evidence of nepotism in colonies that have multiple queens. Instead, they write, eusociality evolves only in species that are under strong pressure from predators and competitors, but in environments where it is possible to build "fortresses"; after colonies have established this security, they gain other advantages though co-operative foraging. In support of this explanation they cite the appearance of eusociality in bathyergid mole rats, which are not haplodiploid.

The earliest fossils of insects have been found in Early Devonian rocks from about 400 million years ago, which preserve only a few varieties of flightless insect. The Mazon Creek lagerstätten from the Late Carboniferous, about 300 million years ago, include about 200 species, some gigantic by modern standards, and indicate that insects had occupied their main modern ecological niches as herbivores, detritivores and insectivores. Social termites and ants first appear in the Early Cretaceous, and advanced social bees have been found in Late Cretaceous rocks but did not become abundant until the Mid Cenozoic.

Humans

Modern humans evolved from a lineage of upright-walking apes that has been traced back over 6 million years ago to *Sahelanthropus*. The first known stone tools were made about 2.5 million years ago, apparently by *Australopithecus garhi*, and were found near animal bones that bear scratches made by these tools. The earliest hominines had chimp-sized brains, but there has been a fourfold increase in the last 3 million years; a statistical analysis suggests that hominine brain sizes depend almost completely on the date of the fossils, while the species to which they are assigned has only slight influence. There is a long-running debate about whether modern humans evolved all over the world simultaneously from existing advanced hominines or are descendants of a single small population in Africa, which then migrated all over the world less than 200,000 years ago and replaced previous hominine species. There is also debate about whether anatomically-modern humans had an intellectual, cultural and technological "Great Leap Forward" under 100,000 years ago and, if so, whether this was due to neurological changes that are not visible in fossils.

Mass extinctions

Life on Earth has suffered occasional mass extinctions at least since 542 million years ago. Although they are disasters at the time, mass extinctions have sometimes accelerated the evolution of life on Earth. When dominance of particular ecological niches passes from one group of organisms to another, it is rarely because the new dominant group is "superior" to the old and usually because an extinction event eliminates the old dominant group and makes way for the new one.

The fossil record appears to show that the gaps between mass extinctions are becoming longer and the average and background rates of extinction are decreasing. Both of these phenomena could be explained in one or more ways:

- The oceans may have become more hospitable to life over the last 500 million years and less vulnerable to mass extinctions: dissolved oxygen became more widespread and penetrated to greater depths; the development of life on land reduced the run-off of nutrients and hence the risk of eutrophication and anoxic events; and marine ecosystems became more diversified so that food chains were less likely to be disrupted.

- Reasonably complete fossils are very rare, most extinct organisms are represented only by partial fossils, and complete fossils are rarest in the oldest rocks. So paleontologists have mistakenly assigned parts of the same organism to different genera which were often defined solely to accommodate these finds – the story of *Anomalocaris* is an example of this. The risk of this mistake is higher for older fossils because these are often unlike parts of any living organism. Many of the "superfluous" genera are represented by fragments which are not found again and the "superfluous" genera appear to become extinct very quickly.

Biodiversity in the fossil record, which is

> "the number of distinct genera alive at any given time; that is, those whose first occurrence predates and whose last occurrence postdates that time"

shows a different trend: a fairly swift rise from 542 to 400 million years ago; a slight decline from 400 to 200 million years ago, in which the devastating Permian–Triassic extinction event is an important factor; and a swift rise from 200 million years ago to the present.

The present

Oxygenic photosynthesis accounts for virtually all of the production of organic matter from non-organic ingredients. Production is split about evenly between land and marine plants, and phytoplankton are the dominant marine producers.

The processes that drive evolution are still operating. Well-known examples include the changes in coloration of the peppered moth over the last 200 years and the more recent appearance of pathogens that are resistant to antibiotics. There is even evidence that humans are still evolving, and possibly at an accelerating rate over the last 40,000 years.

Chapter- 4

Cladistics

Cladistics (Ancient Greek: *κλάδος*, *klados*, "branch") is a method of classifying species of organisms into groups called **clades**, which consist only of firstly, all the descendants of an ancestral organism and secondly, the ancestor itself. For example, birds, dinosaurs, crocodiles, and all descendants (living or extinct) of their most recent common ancestor form a clade. In the terms of biological systematics, a clade is a single "branch" on the "tree of life", a monophyletic group.

Cladistics can be distinguished from other taxonomic systems, such as phenetics, by its focus on shared derived characters (synapomorphies). Systems developed earlier usually employed overall morphological similarity to group species into genera, families and other higher level groups (taxa); cladistic classifications (usually in the form of trees called cladograms) are intended to reflect the relative recency of common ancestry or the sharing of homologous features. Cladistics is also distinguished by an emphasis on parsimony and hypothesis testing (particularly falsificationism), leading to a claim that cladistics is more objective than systems which rely on subjective judgements of relationship based on similarity.

Cladistics originated in the work of the German entomologist Willi Hennig, who referred to it as "phylogenetic systematics" (also the name of his 1966 book); the use of the terms "cladistics" and "clade" was popularized by other researchers. The technique and sometimes the name have been successfully applied in other disciplines: for example, to determine the relationships between the surviving manuscripts of the *Canterbury Tales*.

Cladists use *cladograms*, diagrams which show ancestral relations between species, to represent the monophyletic relationships of species, termed sister-group relationships. This is interpreted as representing phylogeny, or evolutionary relationships. Although traditionally such cladograms were generated largely on the basis of morphological characters, genetic sequencing data and computational phylogenetics are now very commonly used in the generation of cladograms.

Cladistics, either generally or in specific applications, has been criticized from its beginnings. A decision as to whether a particular character is a synapomorphy or not may be challenged as involving subjective judgements, raising the issue of whether cladistics as actually practised is as objective as has been claimed. Formal classifications based on cladistic reasoning are said to emphasize ancestry at the expense of descriptive

characteristics, and thus ignore biologically sensible, clearly defined groups which do not fall into clades (e.g. reptiles as traditionally defined or prokaryotes).

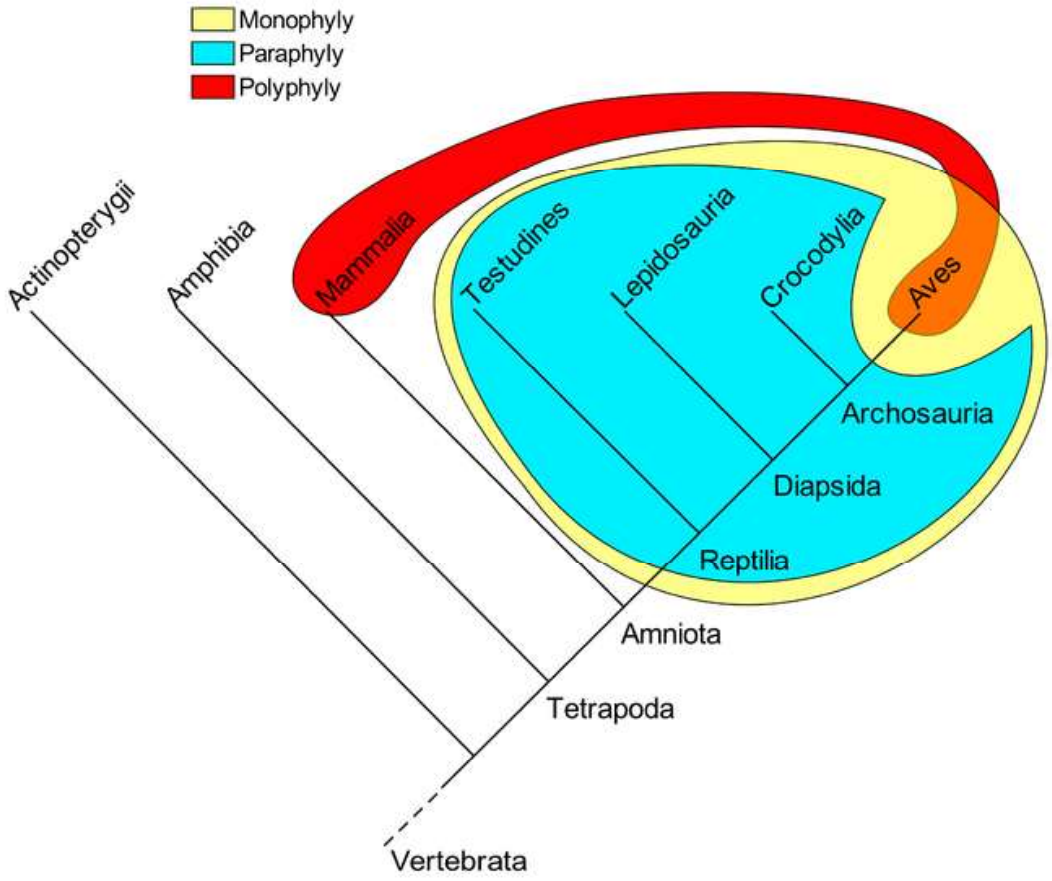
History of cladistics

The term *clade* was introduced in 1958 by Julian Huxley, *cladistic* by Cain and Harrison in 1960, and *cladist* (for an adherent of Hennig's school) by Mayr in 1965. Hennig referred to his own approach as *phylogenetic systematics*. From the time of his original formulation until the end of the 1980s cladistics remained a minority approach to classification. However in the 1990s it rapidly became the dominant method of classification in evolutionary biology. Computers made it possible to process large quantities of data about organisms and their characteristics. At about the same time the development of effective polymerase chain reaction techniques made it possible to apply cladistic methods of analysis to biochemical and molecular genetic features of organisms as well as to anatomical ones.

Cladistics as a successor to phenetics

For some decades in the mid to late twentieth century, a commonly used methodology was phenetics ("numerical taxonomy"). This can be seen as a predecessor to some methods of today's cladistics (namely distance matrix methods such as neighbor-joining), but made no attempt to resolve phylogeny, only similarities.

Clades



Partial Evolutionary Tree of the Vertebrates

The yellow group (sauropsids) is monophyletic, the blue group (traditional reptiles) is paraphyletic, and the red group (warm-blooded animals) is polyphyletic.

A clade is a group of taxa consisting only of an ancestor taxon and all of its descendant taxa. In the diagram provided (a **cladogram**), it is hypothesized that all vertebrates, including ray-finned fishes (Actinopterygii), had a common ancestor, and so form a clade. Within the vertebrates, all tetrapods, including amphibians, mammals, reptiles (as traditionally defined) and birds, are hypothesized to have had a common ancestor, and so also form a clade. The tetrapod ancestor was a descendant of the original vertebrate ancestor, but is not an ancestor of any ray-finned fish living today.

An important caution is that any cladogram is a provisional hypothesis. Although unlikely, future genetic or morphological evidence might suggest that ray-finned fish and amphibians share a common ancestor that was not an ancestor of the other tetrapods. The new information would cause us to define a ray-finned-fish-and-amphibian clade, altering the cladogram.

The relationship between clades can be described in several ways:

- A clade is *basal* to another clade if it contains that other clade as a subset within it. In the example, the vertebrate clade is basal to the tetrapod and ray-finned fish clades. (Some authors have used "basal" differently to mean a clade that is less species-rich than a sister clade, with such a deficit being taken as an indication of 'primitiveness'. Others consider this usage to be incorrect.)
- A clade located within a clade is said to be *nested* within that clade. In the diagram, the tetrapod clade is nested within the vertebrate clade.
- Two clades are *sisters* if they have an immediate common ancestor. In the diagram, crocodiles and birds are sister clades, as are amphibians and amniotes.

Terminology for characters

The following terms are used to identify shared or distinct characters among groups:

- *Plesiomorphy* ("close form") or *ancestral state*, also *symplesiomorphy* ("shared plesiomorphy", i.e. "shared close form"), is a characteristic that is present at the base of a tree (cladogram). Since a plesiomorphy that is inherited from the common ancestor may appear anywhere in a tree, its presence provides no evidence of relationships within the tree. The traditional definition of reptiles (the blue group in the diagram) includes being cold-blooded (i.e. not maintaining a constant high body temperature), whereas birds are warm-blooded. Since cold-bloodedness is a plesiomorphy, inherited from the common ancestor of traditional reptiles and birds, it should not be used to define a group in a system based on cladistics.
- *Apomorphy* ("separate form") or *derived state* is a characteristic believed to have evolved within the tree. It can thus be used to separate one group in the tree from the rest. Within the group which shares the apomorphy it is a *synapomorphy* ("shared apomorphy", i.e. "shared separate form"). For example, within the vertebrates, all tetrapods (and only tetrapods) have four limbs; thus, having four limbs is an synapomorphy for tetrapods. All the tetrapods can legitimately be grouped together because they have four limbs.
- *Homoplasy* is a characteristic shared by members of a tree but not present in their common ancestor. It arises by convergence or reversion. Both mammals and birds are able to maintain a high constant body temperature (i.e. they are 'warm-blooded'). However, the ancestors of each group did not share this character, so it must have evolved independently. Mammals and birds should not be grouped together on the basis that they are warm-blooded.

The terms (sym)plesiomorphy and (syn)apomorphy are relative and their application depends on the position of a group within a tree. An apomorphy of one clade is a plesiomorphy of another contained within it. For example, when trying to decide whether tetrapods should form a clade, an important question is whether having four limbs is a synapomorphy of all the taxa to be included within Tetrapoda: did all the possible members of the Tetrapoda inherit four limbs from a common ancestor, whereas all other

vertebrates did not? By contrast, for a group within the tetrapods, such as birds, having four limbs is a plesiomorphy. The fact that ostriches and rheas both have four limbs does not provide any support for putting them into a separate group of 'flightless birds'. Using these two terms allows a greater precision in the discussion of homology, in particular allowing clear expression of the hierarchical relationships among different homologies.

It can be difficult to decide whether a character is in fact the same, and thus can classified as a synapomorphy which may identify a group, or whether it only appears to be the same, and is thus a homoplasy which cannot identify a group. There is a danger of circular reasoning: assumptions about the shape of a phylogenetic tree are used to justify decisions about characters, which are then used as evidence for the shape of the tree. It has been argued that this kind of reasoning has been used by proponents of the view that birds are nested within the theropod dinosaur clade.

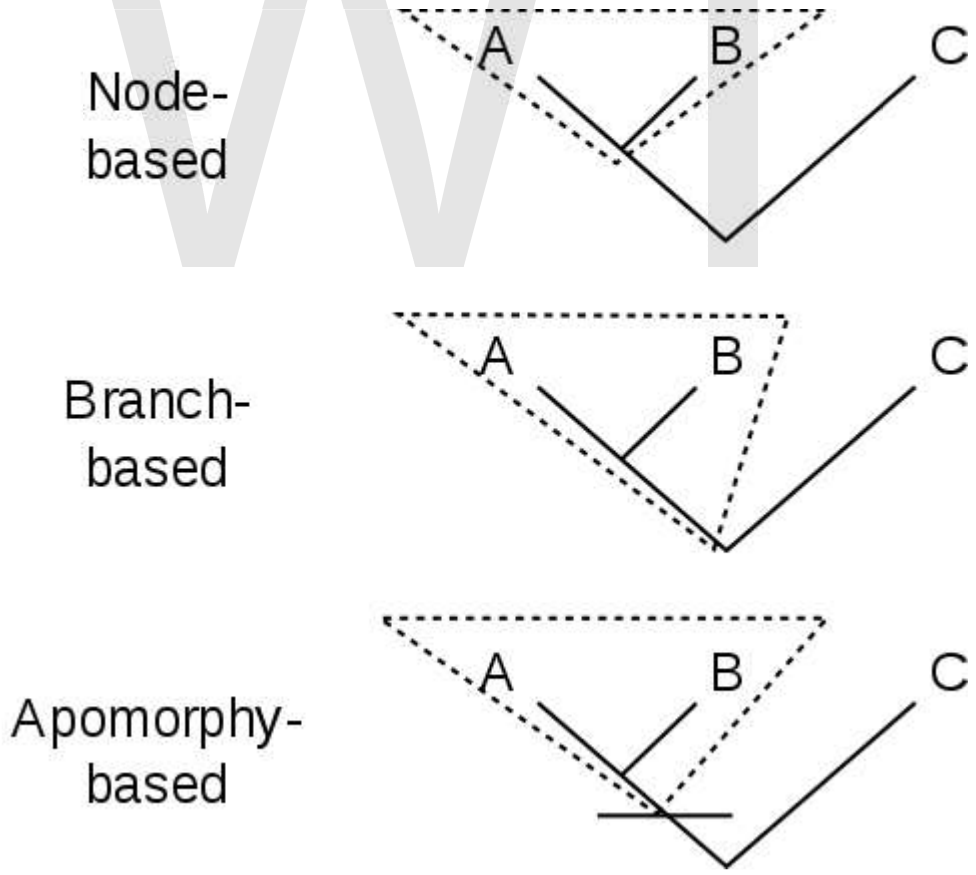
Terminology for groups

Three main types of group can be identified on the basis of their relationships in cladograms. The three can be defined in two different but related ways, as shown in the table below. The first is in terms of the shape of a set of nodes taken from a cladogram. In this approach, an 'ancestor node' is simply a branching point in the diagram; it may or may not correspond to an actual ancestor. The second is in terms of the characters of the taxa being classified and how these characters have been inherited. In this approach, an ancestor is an actual taxon, whether currently known or not.

Term	Node-based definition	Character-based definition
Monophyly	A monophyletic group of nodes in a tree is one which includes all the nodes descended from their most recent common ancestor node, plus the most recent common ancestor node, but no other nodes.	A monophyletic group of taxa is characterized by one or more **synapomorphies**: derived characters inherited by all members of the group from ancestors and not inherited by any other taxa. A monophyletic group is a 'clade'. A 'crown group' is an example of a monophyletic group.
Paraphyly	A paraphyletic group of nodes in a tree is one which is constructed by taking a monophyletic group and removing one or more smaller monophyletic groups. (Removing one group produces a singly paraphyletic group, removing two a doubly paraphylectic group, and so on.) A paraphyletic group is necessarily non-monophyletic.	A paraphyletic group of taxa is characterized by one or more **(sym)plesiomorphies**: characters inherited from ancestors but not present in all of their descendants. As a consequence, a paraphyletic group is truncated, in that it excludes one or more monophyletic taxa from an initially monophyletic group. An alternative name is an 'evolutionary grade', refering to the ancestral character state within the group. A

		'stem group' is an example of a paraphyletic group.
Polyphyly	A polyphyletic group of nodes in a tree is one which is neither monophyletic nor paraphyletic.	A polyphyletic group of taxa is characterized by by one or more **homoplasies**: characters which have converged or reverted so as to appear to be the same but which have not been inherited from common ancestors. As a consequence, polyphyletic groups of taxa are totally artificial.

Branch-based definitions of clade



Three alternative ways to define a clade

The node-based definition of a monophyletic group (i.e. a clade) given above regards the lines in the cladogram only as a way of showing connections between taxa. This is appropriate when considering only living (extant) taxa; however, when extinct taxa are to be included in a cladogram, lines correspond to sequences of ancestors. There are two alternative ways of defining a clade which explicitly take into account the line below the branching point at the base of a clade. These definitions are most notably set out in the PhyloCode.

Consider how a clade combining A and B in the diagram can be defined.

- *Node-based*: The node-based definition specifies A+B as the *last* common ancestor of A and B, and all descendants of that ancestor. It thus excludes from the clade the line below the junction of A and B. Crown groups are a type of node-based clade.
- *Branch-based*: A branch-based definition specifies A+B as the *first* ancestor of A which is not also an ancestor of C, and all descendants of that ancestor. It thus includes in the clade the line below the junction of A and B. (This type of definition was originally called "stem-based", but this was changed to avoid confusion with the term "stem group", which is parapyletic.) Total groups are a type of branch-based clade.
- *Apomorphy-based*: An apomorphy-based definition specifies A+B as the first ancestor of A to possess derived trait M homologously (that is, synapomorphically) with that trait in A, and all descendants of that ancestor. It thus includes in the clade only that part of the line below the junction of A and B which corresponds to ancestors possessing the apomorphy. The process of identifying and naming groups based on apomorphies is the method that most resembles classical systematics, with the proviso that cladistic taxa always denote a clade.

Note that these alternative definitions do not alter the classification of the tips of the tree, and so are equivalent if only living (extant) taxa are being considered.

Cladograms

Cladists use *cladograms*, diagrams which show ancestral relations between taxa, to represent the evolutionary tree of life. Although traditionally such cladograms were generated largely on the basis of morphological characters, molecular sequencing data and computational phylogenetics are now very commonly used in the generation of cladograms.

The starting point of cladistic analysis is a group of species and molecular, morphological, or other data characterizing those species. The end result is a tree-like relationship diagram called a cladogram, or sometimes a *dendrogram* (Greek for "tree drawing"). The cladogram graphically represents a hypothetical evolutionary process. Cladograms are subject to revision as additional data become available.

The terms "evolutionary tree", and sometimes "phylogenetic tree" are often used synonymously with cladogram but others treat phylogenetic tree as a broader term that includes trees generated with a nonevolutionary emphasis. In cladograms, all species lie at the leaves. The two taxa on either side of a split, with a common ancestor and no additional descendents, are called "sister taxa" or "sister groups". Each subtree, whether it contains only two or a hundred thousand items, is called a "clade". Many cladists require that all forks in a cladogram be 2-way forks. Some cladograms include 3-way or 4-way forks when there are insufficient data to resolve the forking to a higher level of detail.

For a given set of taxa, the number of distinct cladograms that can be drawn (ignoring which cladogram best matches the taxon characteristics) is:

Number of taxa	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**N**
Number of rooted cladograms	1	3	15	105	945	10,395	135,135	2,027,025	34,459,425	1*3*5*7*...*(2N-3)

This superexponential growth of the number of possible cladograms explains why manual creation of cladograms becomes very difficult when the number of taxa is large. If a cladogram represents N taxa, the number of levels (the "depth") in the cladogram is on the order of $\log_2(N)$. For example, if there are 32 species of deer, a cladogram representing deer could be around 5 levels deep (because $2^5 = 32$), although this is really just the lower limit. A cladogram representing the complete tree of life, with about 10 million species, could be about 23 levels deep. This formula gives a lower limit, with the actual depth generally a larger value, because the various branches of the cladogram will not be uniformly deep. Conversely, the depth may be shallower if forks larger than 2-way forks are permitted.

A cladogram tree has an implicit time axis, with time running forward from the base of the tree to the leaves of the tree. If the approximate date (for example, expressed as millions of years ago) of all the evolutionary forks were known, those dates could be captured in the cladogram. Thus, the time axis of the cladogram could be assigned a time scale (e.g. 1 cm = 1 million years), and the forks of the tree could be graphically located along the time axis. Such cladograms are called *scaled cladograms*. Many cladograms are not of this type, for a variety of reasons:

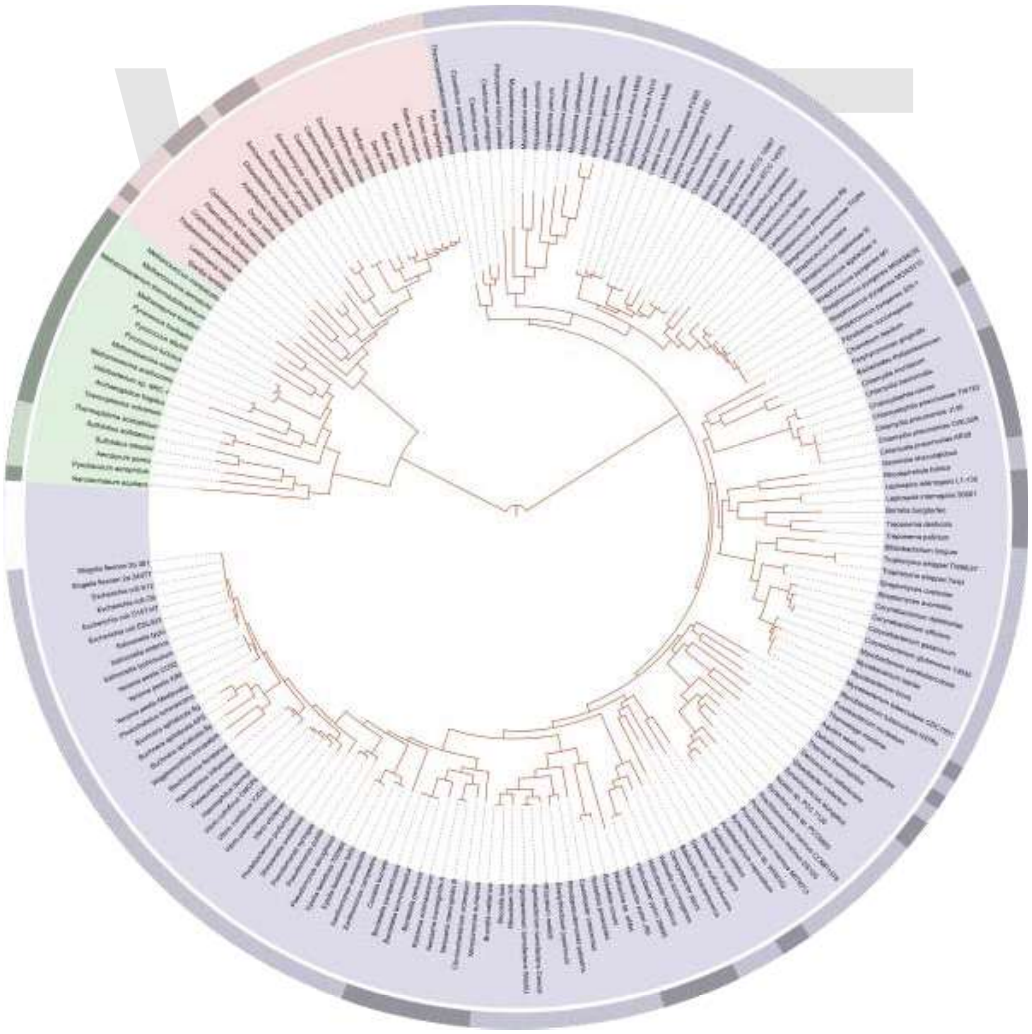
- They are built from species characteristics that cannot be readily dated (e.g. morphological data in the absence of fossils or other dating information)
- When the characteristic data are DNA/RNA sequences, it is feasible to use sequence differences to establish the relative ages of the forks, but converting those ages into actual years requires a significant approximation of the rate of change

- Even when the dating information is available, positioning the cladogram's forks along the time axis in proportion to their dates may cause the cladogram to become difficult to understand or hard to fit within a human-readable format

Cladistics makes no distinction between extinct and extant species, and it is appropriate to include extinct species in the group of organisms being analyzed. Cladograms that are based on DNA/RNA generally do not include extinct species because DNA/RNA samples from extinct species are rare. Cladograms based on morphology, especially morphological characteristics that are preserved in fossils, are more likely to include extinct species.

Cladistics in taxonomy

Phylogenetic nomenclature contrasted with traditional taxonomy



A highly resolved, automatically generated tree of life based on completely sequenced genomes

Most taxonomists have used the traditional approaches of Linnaean taxonomy and later Evolutionary taxonomy to organize life forms. These approaches use several fixed levels of a hierarchy, such as kingdom, phylum, class, order, and family. Phylogenetic nomenclature does not feature those terms, because the evolutionary tree is so deep and so complex that it is inadvisable to set a fixed number of levels.

Evolutionary taxonomy insists that groups reflect phylogenies. In contrast, Linnaean taxonomy allows both monophyletic and paraphyletic groups as taxa. Since the early 20th century, Linnaean taxonomists have generally attempted to make at least family- and lower-level taxa (i.e. those regulated by the codes of nomenclature) monophyletic. Ernst Mayr in 1985 drew a distinction between the terms cladistics and phylogeny: "It would seem to me to be quite evident that the two concepts of phylogeny (and their role in the construction of classifications) are sufficiently different to require terminological distinction. The term *phylogeny* should be retained for the broad concept of phylogeny, promoted by Darwin and adopted by most students of phylogeny in the ensuing 90 years. The concept of phylogeny as mere genealogy should be terminologically distinguished as *cladistics*. To lump the two concepts together terminologically could not help but produce harmful equivocation."

Willi Hennig's pioneering work provoked a spirited debate about the relative merits of phylogenetic nomenclature versus Linnaean or evolutionary taxonomy, which has continued down to the present; however Hennig did not advocate abandoning the Linnaean nomenclatural system. Some of the debates in which the cladists were engaged had been running since the 19th century, but they were renewed fervor, as can be seen from the *Foreword* to Hennig (1979) by Rosen, Nelson, and Patterson:

"Encumbered with vague and slippery ideas about adaptation, fitness, biological species and natural selection, neo-Darwinism (summed up in the "evolutionary" systematics of Mayr and Simpson) not only lacked a definable investigatory method, but came to depend, both for evolutionary interpretation and classification, on consensus or authority."

Phylogenetic nomenclature strictly and exclusively follows phylogeny and has arbitrarily deep trees with binary branching: each taxon corresponds to a clade. Linnaean taxonomy, while since the advent of evolutionary theory following phylogeny, also may subjectively consider similarity and has a fixed hierarchy of taxonomic ranks, and its taxa are not required to correspond to clades.

Paraphyletic groups discouraged

Many cladists discourage the use of paraphyletic groups in classification of organisms, because they detract from cladistics' emphasis on clades (monophyletic groups). In contrast, proponents of the use of paraphyletic groups argue that any dividing line in a cladogram creates both a monophyletic section above and a paraphyletic section below. They also contend that paraphyletic taxa are necessary for classifying earlier sections of the tree – for instance, the early vertebrates that would someday evolve into the family

Hominidae cannot be placed in any other monophyletic family. They also argue that paraphyletic taxa provide information about significant changes in organisms' morphology, ecology, or life history – in short, that both paraphyletic groups and clades are valuable notions with separate purposes.

Complexity of the Tree of Life

The cladistic tree of life is a fractal:

"The tree of life is inherently fractal-like in its complexity, ... Look closely at the 'lineage' of a phylogeny ... and it dissolves into many smaller lineages, and so on, down to a very fine scale."

The overall shape of a dichotomous (bifurcating) tree is recursive; as a viewpoint zooms into the tree of life, the same type of tree appears no matter what the scale. When extinct species are considered (both known and unknown), the complexity and depth of the tree can be very large. Moreover the tree continues to recreate itself by bifurcation, a series of events called fractal evolution. Every single speciation event, including all the species that are now extinct, represents an additional fork on the hypothetical, complete cladogram of the tree of life.

The tree of life is a quasi-self-similar fractal; that is, the deep reconstruction is not as regular as the shallow reconstruction. By shallow Mishler means the most recent branching toward and at the tips, and by deep the more ancient branches further back, which are harder to reconstruct and are missing unknown extinct lines. In the shallow part of the tree, branching events are relatively regular; it is often possible to estimate the times between them. In the deep part of the tree, "homology assessments" are "difficult" and the times vary widely. At this level Eldredge's and Gould's punctuated equilibrium applies, which hypothesizes long periods of stability followed by punctuations of rapid speciation, based on the fossil record.

PhyloCode approach to naming species

A formal code of phylogenetic nomenclature, the PhyloCode, is currently under development. It is intended for use by both those who would like to abandon Linnaean taxonomy and those who would like to use taxa and clades side by side. In several instances it has been employed to clarify uncertainties in Linnaean systematics so that in combination they yield a taxonomy that unambiguously places problematic groups in the evolutionary tree in a way that is consistent with current knowledge.

Example

For example, Linnaean taxonomy contains the taxon Tetrapoda, defined morphologically as vertebrates with four limbs (as well as animals with four-limbed ancestors, such as snakes), which is often given the rank of superclass, and divides into the classes Amphibia, Reptilia, Aves, Mammalia.

Phylogenetic nomenclature also contains the taxon Tetrapoda, whose living members can be classified phylogenically as "the clade defined by the common ancestor of amphibians and mammals", or more precisely the clade defined by the common ancestor of a specific amphibian and mammal (or bird or snake). This definition gives us the Crown group tetrapods (or Crown-Tetrapoda). A few primitive four legged ancestors (the Ichthyostegalia) fall outside Crown-Tetrapoda. An alternative is to define tetrapoda as all animals more closely related to mammals than to lungfish (our nearest living non-tetrapod relatives). In this definition, the ichthyostegalians are included, together with a host of fossil animals usually classed as crossopterygian fish. This wider definition is termed Pan-Tetrapoda. A third option is to define Tetrapoda according to their apomorphy (their unique trait, i.e. having legs rather than fins), a definition that yield the same group as the Linnaean taxon.

Non of the phylogenetic taxa as described above have a rank, and neither do its subtaxa. All the subclades are contained within one another. The clades are not divided into several non-overlapping taxa (as in traditional taxonomy), rather the clade is split into two clades at the first branching, a process repeated throughout. With regards to the traditional classes, Aves and Mammalia are subclades, contained in the subclade Amniota, while Reptilia and Amphibia are paraphyletic taxa, not clades. Instead of classifying non-mammalian, non-avian amniotes as reptiles, Amniota is divided into the two clades Sauropsida (which contains birds and all living amniotes other than mammals, including all living traditional reptiles) and Theropsida (mammals and the extinct mammal-like reptiles). Similarly, Amphibia can be split into the Batrachomorpha (fossil amphibians more closely related to modern amphibians) and Reptiliomorpha, the latter of which the amiotes is a sub-clade. Ichthyostegalians and other Stem-tetrapods represent sister groups from splits predating the Batrachomorpha/Reptilopmorpha split.

Summary of advantages of phylogenetic nomenclature

Proponents of phylogenetic nomenclature enumerate key distinctions between phylogenetic nomenclature and Linnaean taxonomy as follows:

Phylogenetic Nomenclature	Linnaean Taxonomy
Handles arbitrarily deep trees.	Often must invent new level names (such as superorder, suborder, infraorder, parvorder, magnorder) to accommodate new discoveries. Biased towards trees about 4 to 12 levels deep.
Discourages naming or use of groups that are not monophyletic	Acceptable to name and use paraphyletic groups
Primary goal is to reflect actual process of evolution	Primary goal is to group species based on morphological similarities

Assumes that the shape of the tree will change frequently with new discoveries	New discoveries often require renaming or releveling of Classes, Orders, and Kingdoms

Summary of criticisms of phylogenetic nomenclature

Critics of phylogenetic nomenclature include Ashlock, Mayr, and Williams. Some of their criticisms include:

Phylogenetic Nomenclature	**Linnaean Taxonomy**
Limited to entities related by evolution or ancestry	Supports groupings without reference to evolution or ancestry
Does not include a process for naming species	Includes a process for giving unique names to species
Clade definitions emphasize ancestry at the expense of descriptive characteristics	Taxa definitions based on tangible characteristics
Ignores sensible, clearly defined paraphyletic groups such as reptiles	Permits clearly defined groups such as reptiles
Difficult to determine if a given species is in a clade or not (e.g. if clade X is defined as "most recent common ancestor of A and B along with its descendants", then the only way to determine if species Y is in the clade is to perform a complex evolutionary analysis)	Straightforward process to determine if a given species is in a taxon or not
Limited to organisms that evolved by inherited traits; not applicable to organisms that evolved via complex gene sharing or lateral transfer	Applicable to all organisms, regardless of evolutionary mechanism

Application to other disciplines

The comparisons used to acquire data on which cladograms can be based are not limited to the field of biology. Any group of individuals or classes, hypothesized to have a common ancestor, and to which a set of common characteristics may or may not apply, can be compared pairwise. Cladograms can be used to depict the hypothetical descent relationships within groups of items in many different academic realms. The only requirement is that the items have characteristics that can be identified and measured.

Recent attempts to use cladistic methods outside of biology address the reconstruction of lineages in:

- Anthropology and archeology. Compares cultures or artifacts using groups of cultural traits or artifact features.
- Linguistics. Compares languages using groups of linguistic features.
- Textual criticism or Stemmatics. Compares manuscripts of the same work (original lost) using groups of distinctive copying errors.
- Ethology. Compares animal species using behavioral traits presumed hereditary.

Chapter- 5

Evolutionary Developmental Biology

Evolutionary developmental biology (**evolution of development** or informally, **evo-devo**) is a field of biology that compares the developmental processes of different organisms to determine the ancestral relationship between them, and to discover how developmental processes evolved. It addresses the origin and evolution of embryonic development; how modifications of development and developmental processes lead to the production of novel features, such as the evolution of feathers; the role of developmental plasticity in evolution; how ecology impacts in development and evolutionary change; and the developmental basis of homoplasy and homology.

Although interest in the relationship between ontogeny and phylogeny extends back to the nineteenth century, the contemporary field of evo-devo has gained impetus from the discovery of genes regulating embryonic development in model organisms. General hypotheses remain hard to test because organisms differ so much in shape and form.

Nevertheless, it now appears that just as evolution tends to create new genes from parts of old genes (molecular economy), evo-devo demonstrates that evolution alters developmental processes (genes and gene networks) to create new and novel structures from the old gene networks (such as bone structures of the jaw deviating to the ossicles of the middle ear) or will conserve (molecular economy) a similar program in a host of organisms such as eye development genes in molluscs, insects, and vertebrates. Initially the major interest has been in the evidence of homology in the cellular and molecular mechanisms that regulate body plan and organ development. However more modern approaches include developmental changes associated with speciation.

Basic principles

Charles Darwin's theory of evolution is based on three principles: natural selection, heredity, and variation. At the time that Darwin wrote, the principles underlying heredity and variation were poorly understood. In the 1940s, however, biologists incorporated Gregor Mendel's principles of genetics to explain both, resulting in the modern synthesis. It was not until the 1980s and 1990s, however, when more comparative molecular

sequence data between different kinds of organisms was amassed and detailed, that an understanding of the molecular basis of the developmental mechanisms has arisen.

Currently, it is well understood how genetic mutation occurs. However, developmental mechanisms are not understood sufficiently to explain which kinds of phenotypic variation can arise in each generation from variation at the genetic level. Evolutionary developmental biology studies how the dynamics of development determine the phenotypic variation arising from genetic variation and how that affects phenotypic evolution (especially its direction). At the same time evolutionary developmental biology also studies how development itself evolves.

Thus, the origins of evolutionary developmental biology come from both an improvement in molecular biology techniques as applied to development, and the full appreciation of the limitations of classic neo-Darwinism as applied to phenotypic evolution. Some evo-devo researchers see themselves as extending and enhancing the modern synthesis by incorporating into it findings of molecular genetics and developmental biology. Others, drawing on findings of discordances between genotype and phenotype and epigenetic mechanisms of development, are mounting an explicit challenge to neo-Darwinism.

Evolutionary developmental biology is not yet a unified discipline, but can be distinguished from earlier approaches to evolutionary theory by its focus on a few crucial ideas. One of these is modularity: as has been long recognized, plants and animal bodies are modular: they are organized into developmentally and anatomically distinct parts. Often these parts are repeated, such as fingers, ribs, and body segments. Evo-devo seeks the genetic and evolutionary basis for the division of the embryo into distinct modules, and for the partly independent development of such modules.

Another central idea is that some gene products function as switches whereas others act as diffusible signals. Genes specify proteins, some of which act as structural components of cells and others as enzymes that regulate various biochemical pathways within an organism. Most biologists working within the modern synthesis assumed that an organism is a straightforward reflection of its component genes. The modification of existing, or evolution of new, biochemical pathways (and, ultimately, the evolution of new species of organisms) depended on specific genetic mutations. In 1961, however, Jacques Monod, Jean-Pierre Changeux and François Jacob discovered within the bacterium Escherichia coli a gene that functioned only when "switched on" by an environmental stimulus. Later, scientists discovered specific genes in animals, including a subgroup of the genes which contain the homeobox DNA motif, called Hox genes, that act as switches for other genes, and could be induced by other gene products, morphogens, that act analogously to the external stimuli in bacteria. These discoveries drew biologists' attention to the fact that genes can be selectively turned on and off, rather than being always active, and that highly disparate organisms (for example, fruit flies and human beings) may use the same genes for embryogenesis (e.g., the genes of the "developmental-genetic toolkit", see below), just regulating them differently.

Similarly, organismal form can be influenced by mutations in promoter regions of genes, those DNA sequences at which the products of some genes bind to and control the activity of the same or other genes, not only protein-specifying sequences. In addition to providing new support for Darwin's assertion that all organisms are descended from a common ancestor, this finding suggested that the crucial distinction between different species (even different orders or phyla) may be due less to differences in their content of gene products than to differences in spatial and temporal *expression* of conserved genes. The implication that large evolutionary changes in body morphology are associated with changes in gene regulation, rather than the evolution of new genes, suggested that the action of natural selection on promoters responsive to Hox and other "switch" genes may play a major role in evolution.

Another focus of evo-devo is developmental plasticity, the basis of the recognition that organismal phenotypes are not uniquely determined by their genotypes. If generation of phenotypes is conditional, and dependent on external or environmental inputs, evolution can proceed by a "phenotype-first" route, with genetic change following, rather than initiating, the formation of morphological and other phenotypic novelties. The case for this was argued for by Mary Jane West-Eberhard in her 2003 book *Developmental plasticity and evolution*.

History

An early version of recapitulation theory, also called the *biogenetic law* or *embryological parallelism*, was put forward by Étienne Serres in 1824–26 as what became known as the "Meckel-Serres Law" which attempted to provide a link between comparative embryology and a "pattern of unification" in the organic world. It was supported by Étienne Geoffroy Saint-Hilaire as part of his ideas of idealism, and became a prominent part of his version of Lamarckism leading to disagreements with Georges Cuvier. It was widely supported in the Edinburgh and London schools of higher anatomy around 1830, notably by Robert Edmond Grant, but was opposed by Karl Ernst von Baer's embryology of divergence in which embryonic parallels only applied to early stages where the embryo took a general form, after which more specialised forms diverged from this shared unity in a branching pattern. The anatomist Richard Owen used this to support his idealist concept of species as showing the unrolling of a divine plan from an archetype, and in the 1830s attacked the transmutation of species proposed by Lamarck, Geoffroy and Grant. In the 1850s Owen began to support an evolutionary view that the history of life was the gradual unfolding of a teleological divine plan, in a continuous "ordained becoming", with new species appearing by natural birth.

In *On the Origin of Species* (1859), Charles Darwin proposed evolution through natural selection, a theory central to modern biology. Darwin recognised the importance of embryonic development in the understanding of evolution, and the way in which von Baer's branching pattern matched his own idea of descent with modification:

> We can see why characters derived from the embryo should be of equal importance with those derived from the adult, for a natural classification of course includes all ages.

Ernst Haeckel (1866), in his endeavour to produce a synthesis of Darwin's theory with Lamarckism and Naturphilosophie, proposed that "ontogeny recapitulates phylogeny," that is, the development of the embryo of every species (ontogeny) fully repeats the evolutionary development of that species (phylogeny), in Geoffroy's linear model rather than Darwin's idea of branching evolution. Haeckel's concept explained, for example, why humans, and indeed all vertebrates, have gill slits and tails early in embryonic development. His theory has since been discredited. However, it served as a backdrop for a renewed interest in the evolution of development after the modern evolutionary synthesis was established (roughly 1936 to 1947).

Stephen Jay Gould called this approach to explaining evolution as *terminal addition*; as if every evolutionary advance was added as new stage by reducing the duration of the older stages. The idea was based on observations of neoteny. This was extended by the more general idea of heterochrony (changes in timing of development) as a mechanism for evolutionary change.

D'Arcy Thompson postulated that differential growth rates could produce variations in form in his 1917 book *On Growth and Form*. He showed the underlying similarities in *body plans* and how geometric *transformations* could be used to explain the variations.

Edward B. Lewis discovered homeotic genes, rooting the emerging discipline of evo-devo in molecular genetics. In 2000, a special section of the Proceedings of the National Academy of Sciences (PNAS) was devoted to "evo-devo", and an entire 2005 issue of the Journal of Experimental Zoology Part B: Molecular and Developmental Evolution was devoted to the key evo-devo topics of evolutionary innovation and morphological novelty.

The developmental-genetic toolkit

The developmental-genetic toolkit consists of a small fraction of the genes in an organism's genome whose products control its development. These genes are highly conserved among Phyla. Differences in deployment of toolkit genes affect the body plan and the number, identity, and pattern of body parts. The majority of toolkit genes are components of signaling pathways, and encode for the production of transcription factors, cell adhesion proteins, cell surface receptor proteins, and secreted morphogens, all of these participate in defining the fate of undifferentiated cells, generating spatial and

temporal patterns, which in turn form the body plan of the organism. Among the most important of the toolkit genes are those of the **Hox** gene cluster, or complex. Hox genes, transcription factors containing the more broadly distributed homeobox protein-binding DNA motif, function in patterning the body axis. Thus, by combinatorial specifying the identity of particular body regions, Hox genes determine where limbs and other body segments will grow in a developing embryo or larva. A paragon of a toolbox gene is *Pax6/eyeless*, which controls eye formation in all animals. It has been found to produce eyes in mice and *Drosophila*, even if mouse *Pax6/eyeless* was expressed in *Drosophila*.

This means that a big part of the morphological evolution undergone by organisms is a product of variation in the genetic toolkit, either by the genes changing their expression pattern or acquiring new functions. A good example of the first is the enlargement of the beak in Darwin's Large Ground-finch (*Geospiza magnirostris*), in which the gene *BMP* is responsible for the larger beak of this bird, relative to the other finches.

The loss of legs in snakes and other squamates is another good example of genes changing their expression pattern. In this case the gene *Distal-less* is very under-expressed, or not expressed at all, in the regions where limbs would form in other tetrapods. This same gene determines the spot pattern in butterfly wings, which shows that the toolbox genes can change their function.

Toolbox genes, as well as being highly conserved, also tend to evolve the same function convergently or in parallel. Classic examples of this are the already mentioned *Distal-less* gene, which is responsible for appendage formation in both tetrapods and insects, or, at a finer scale, the generation of wing patterns in the butterflies *Heliconius erato* and *Heliconius melpomene*. These butterflies are Müllerian mimics whose coloration pattern arose in different evolutionary events, but is controlled by the same genes. The previous supports Kirschner and Gerhardt's theory of Facilitated Variation, which states that morphological evolutionary novelty is generated by regulatory changes in various members of a large set of conserved mechanisms of development and physiology.

Development and the origin of novelty

Among the more surprising and, perhaps, counterintuitive (from a neo-Darwinian viewpoint) results of recent research in evolutionary developmental biology is that the diversity of body plans and morphology in organisms across many phyla are not necessarily reflected in diversity at the level of the sequences of genes, including those of the developmental genetic toolkit and other genes involved in development. Indeed, as Gerhart and Kirschner have noted, there is an apparent paradox: "where we most expect to find variation, we find conservation, a lack of change".

Even within a species, the occurrence of novel forms within a population does not generally correlate with levels of genetic variation sufficient to account for all morphological diversity. For example, there is significant variation in limb morphologies amongst salamanders and in differences in segment number in centipedes, even when the respective genetic variation is low.

A major question then, for evo-devo studies, is: If the morphological novelty we observe at the level of different clades is not always reflected in the genome, where does it come from? Apart from neo-Darwinian mechanisms such as mutation, translocation and duplication of genes, novelty may also arise by mutation-driven changes in gene regulation. The finding that much biodiversity is not due to differences in genes, but rather to alterations in gene regulation, has introduced an important new element into evolutionary theory. Diverse organisms may have highly conserved developmental genes, but highly divergent regulatory mechanisms for these genes. Changes in gene regulation are "second-order" effects of genes, resulting from the interaction and timing of activity of gene networks, as distinct from the functioning of the individual genes in the network.

The discovery of the homeotic Hox gene family in vertebrates in the 1980s allowed researchers in developmental biology to empirically assess the relative roles of gene duplication and gene regulation with respect to their importance in the evolution of morphological diversity. Several biologists, including Sean B. Carroll of the University of Wisconsin–Madison suggest that "changes in the cis-regulatory systems of genes" are more significant than "changes in gene number or protein function". These researchers argue that the combinatorial nature of transcriptional regulation allows a rich substrate for morphological diversity, since variations in the level, pattern, or timing of gene expression may provide more variation for natural selection to act upon than changes in the gene product alone.

Epigenetic alterations of gene regulation or phenotype generation that are subsequently consolidated by changes at the gene level constitute another class of mechanisms for evolutionary innovation. Epigenetic changes include modification of the genetic material due to methylation and other reversible chemical alteration, as well as nonprogrammed remolding of the organism by physical and other environmental effects due to the inherent plasticity of developmental mechanisms. The biologists Stuart A. Newman and Gerd B. Müller have suggested that organisms early in the history of multicellular life were more susceptible to this second category of epigenetic determination than are modern organisms, providing a basis for early macroevolutionary changes.

Chapter- 6

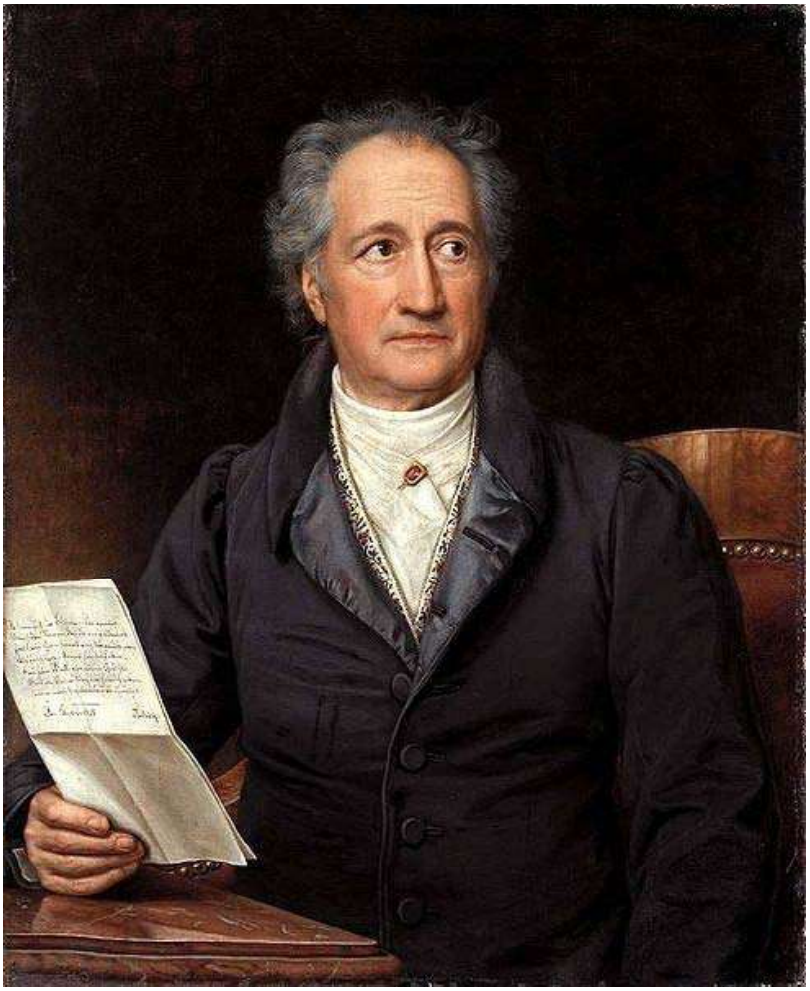
Plant Evolutionary Developmental Biology

Evolutionary developmental biology (evo-devo) refers to the study of developmental programs and patterns from an evolutionary perspective. It seeks to understand the various influences shaping the form and nature of life on the planet. Evo-devo arose as a separate branch of science only in the last decade. Most of the synthesis in evo-devo has

been in the field of animal evolution, one reason being the presence of elegant model systems like *Drosophila melanogaster*, *C. elegans*, zebrafish and *Xenopus laevis*. However, in the past couple of decades, a wealth of information on plant morphology, coupled with modern molecular techniques has helped shed light on the conserved and unique developmental patterns in the plant kingdom also.

Historical perspective

Before 1900



Johann Wolfgang von Goethe

The origin of the term "morphology" is generally attributed to Johann Wolfgang von Goethe (1749–1832). He was of the opinion that there is an underlying fundamental organisation *(Bauplan)* in the diversity of flowering plants. In his book titled The Metamorphosis of Plants, he proposed that the *Bauplan* enabled us to predict the forms of plants that had not yet been discovered. Goethe also was the first to make the perceptive suggestion that flowers consist of modified leaves.

In the middle centuries, several basic foundations of our current understanding of plant morphology were laid down. Nehemiah Grew, Marcello Malpighi, Robert Hooke, Antonie van Leeuwenhoek, Wilhelm von Nageli were just some of the people who helped build knowledge on plant morphology at various levels of organisation. It was the taxonomical classification of Carolus Linnaeus in the eighteenth century though, that generated a firm base for the knowledge to stand on and expand. The introduction of the concept of Darwinism in contemporary scientific discourse also had had an effect on the thinking on plant forms and their evolution.

Wilhelm Hofmeister, one of the most brilliant botanists of his times, was the one to diverge away from the idealist way of pursuing botany. Over the course of his life, he brought an interdisciplinary outlook into botanical thinking. He came up with biophysical explanations on phenomena like phototaxis and geotaxis, and also discovered the alternation of generations in the plant life cycle.

1900 to the present

Arabidopsis thaliana. This flowering plant has been a model system for most of plant molecular studies

The past century witnessed a rapid progress in the study of plant anatomy. The focus shifted from the population level to more reductionist levels. While the first half of the century saw expansion in developmental knowledge at the tissue and the organ level, in the latter half, especially since the 1990s, there has also been a strong impetus on gaining molecular information.

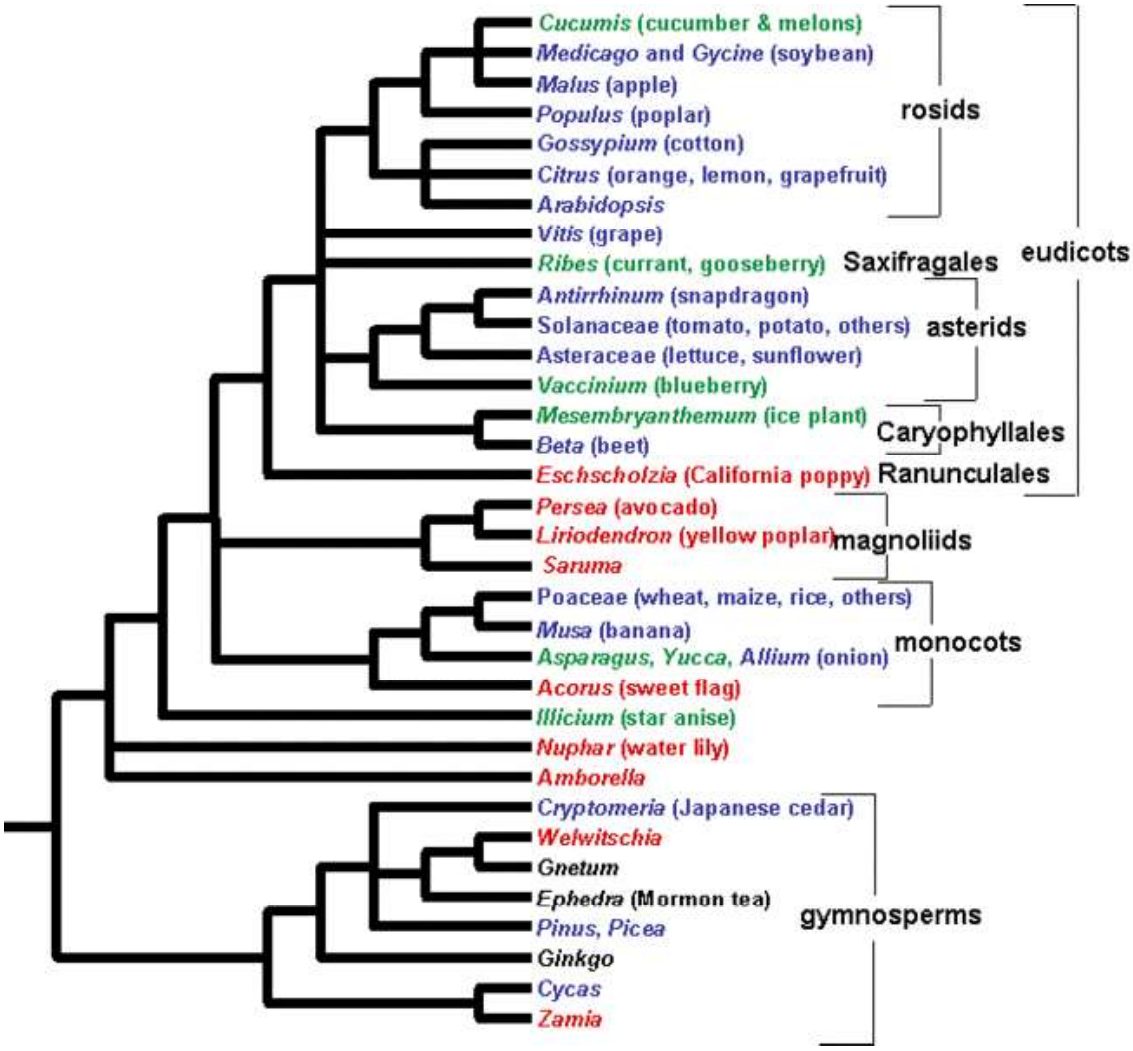
Edward Charles Jeffrey was one of the early evo devo researchers of the 20th century. He performed a comparative analyses of the vasculatures of living and fossil Gymnosperms and came to the conclusion that the storage parenchyma has been derived from tracheids. His research focussed primarily on plant anatomy in the context of phylogeny. This tradition of evolutionary analyses of plant architectures was further advanced by Katherine Esau, best known for her book *The Plant Anatomy*. Her work focussed on the origin and development of various tissues in different plants. Working with Vernon Cheadle, she also explained the evolutionary specialization of the phloem tissue with respect to its function.

In the meantime, by the beginning of the latter half of 1900s, *Arabidopsis thaliana* had begun to be used in some developmental studies. The first collection of *Arabidopsis thaliana* mutants were made around 1945. However it formally became established as a model organism only in 1998.

The recent spurt in information on various plant-related processes has largely been a result of the revolution in molecular biology. Powerful techniques like mutagenesis and complementation were made possible in *Arabidopsis thaliana* via generation of T-DNA containing mutant lines, recombinant plasmids, techniques like Transposon Tagging etc. Availability of complete physical and genetic maps, RNAi vectors, and rapid transformation protocols are some of the technologies that have significantly altered the scope of the field. Recently, there has also been a massive increase in the genome and EST sequences of various non-model species, which, coupled with the Bioinformatics tools existing today, generate interesting opportunities in the field of plant evo devo research.

Cusset provided a detailed in-depth analysis of the history of plant morphology, including plant development and evolution, from its beginnings to the end of the 20th century.

Organisms, databases and tools



The sampling of the Floral Genome Project

The most important model systems in plant development have been arabidopsis and maize. Maize has traditionally been the favorite of plant geneticists, while extensive resources in almost every area of plant physiology and development are available for *Arabidopsis thaliana*. Apart from these, rice, *Antirrhinum majus*, *Brassica*, tomato are also being used in a variety of studies. The genomes of *Arabidopsis thaliana* and rice have been completely sequenced, while the others are in process.. It must be emphasized here that the information from these "model" organisms form the basis of our developmental knowledge. While *Brassica* has been used primarily because of its convenient location in the phylogenetic tree in the mustard family, *Antirrhinum majus* is a convenient system for studying leaf architecture. Rice has been traditionally used for studying responses to hormones like abscissic acid and gibberelin as well as responses to stress. However, recently, not just the domesticated rice strain, but also the wild strains have been studied for their underlying genetic architectures.

Some people have objected against extending the results of model organisms to the plant world. One argument is that the effect of gene knockouts in lab conditions wouldn't truly reflect even the same plant's response in the natural world. Also, these supposedly *crucial* genes might not be responsible for the evolutionary origin of that character. For these reasons, a comparative study of plant traits has been proposed as the way to go now.

Since the past few years, researchers have indeed begun looking at non-model, "non-conventional" organisms using modern genetic tools. One example of this is the Floral Genome Project, which envisages to study the evolution of the current patterns in the genetic architecture of the flower through comparative genetic analyses, with a focus on EST sequences. Like the FGP, there are several such ongoing projects that aim to find out conserved and diverse patterns in evolution of the plant shape. Expressed sequence tag (EST) sequences of quite a few non-model plants like Sugarcane, Apple, Lotus, Barley, Cycas, Coffee, to name a few, are available freely online. The Cycad Genomics Project, for example, aims to understand the differences in structure and function of genes between gymnosperms and angiosperms through sampling in the order Cycadales. In the process, it intends to make available information for the study of evolution of structures like seeds, cones and evolution of life cycle patterns. Presently the most important sequenced genomes from an evo-devo point of view include those of *A.thaliana (a flowering plant)*, Poplar *(a woody plant)*, *Physcomitrella patens (a bryophyte)*, Maize *(extensive genetic information)*, and *Chlamydomonas reinhardtii (a green alga)*. The impact of such a vast amount of information on understanding common underlying developmental mechanisms can easily be realised.

Apart from EST and genome sequences, several other tools like PCR, Yeast two hybrid system, microarrays, RNA Interference, SAGE, QTL mapping etc. permit the rapid study of plant developmental patterns. Recently, cross-species hybridization has begun to be employed on microarray chips, to study the conservation and divergence in mRNA expression patterns between closely related species. Techniques for analyzing this kind of data have also progressed over the past decade. We now have better models for molecular evolution, more refined analysis algorithms and better computing power as a result of advances in computer sciences.

Evolution of plant morphology

Overview of plant evolution

Evidence suggests that an algal scum formed on the land 1,200 million years ago, but it was not until the Ordovician period, around 500 million years ago, that land plants appeared. These begun to diversify in the late Silurian period, around 420 million years ago, and the fruits of their diversification are displayed in remarkable detail in an early Devonian fossil assemblage known as the Rhynie chert. This chert preserved early plants in cellular detail, petrified in volcanic springs. By the middle of the Devonian period most of the features recognised in plants today are present, including roots, leaves and seeds. By the late Devonian, plants had reached a degree of sophistication that allowed them to form forests of tall trees. Evolutionary innovation continued after the Devonian

period. Most plant groups were relatively unscathed by the Permo-Triassic extinction event, although the structures of communities changed. This may have set the scene for the evolution of flowering plants in the Triassic (~200 million years ago), which exploded the Cretaceous and Tertiary. The latest major group of plants to evolve were the grasses, which became important in the mid Tertiary, from around 40 million years ago. The grasses, as well as many other groups, evolved new mechanisms of metabolism to survive the low CO_2 and warm, dry conditions of the tropics over the last 10 million years.

Evolution of meristems

The meristematic cells give rise to various organs of the plant, and keep the plant growing. The Shoot Apical Meristem (SAM) gives rise to organs like the leaves and flowers. The cells of the apical meristems - SAM and RAM (Root Apical Meristem)- divide rapidly and are considered to be indeterminate, in that they do not possess any defined end fate. In that sense, the meristematic cells are frequently compared to the stem cells in animals, that have an analogous behavior and function.

Diversity in meristem architectures

Is the mechanism of being *indeterminate* conserved in the SAM's of the plant world? The SAM contains a population of stem cells that also produce the lateral meristems while the stem elongates. It turns out that the mechanism of regulation of the stem cell number might indeed be evolutionarily conserved. The *CLAVATA* gene *CLV2* responsible for maintaining the stem cell population in *Arabidopsis thaliana* is very closely related to the Maize gene *FASCIATED EAR 2*(*FEA2*) also involved in the same function. Similarly, in Rice, the *FON1-FON2* system seems to bear a close relationship with the CLV signaling system in *Arabidopsis thaliana*. These studies suggest that the regulation of stem cell number, identity and differentiation might be an evolutionarily conserved mechanism in monocots, if not in angiosperms. Rice also contains another genetic system distinct from *FON1-FON2*, that is involved in regulating stem cell number. This example underlines the innovation that goes about in the living world all the time.

Role of the KNOX-family genes

Note the long spur of the above flower. Spurs attract pollinators and confer pollinator specificity. *(Flower:Linaria dalmatica)*

Complex leaves of *C. hirsuta* are a result of KNOX gene expression

Genetic screens have identified genes belonging to the KNOX family in this function. These genes essentially maintain the stem cells in an undifferentiated state. The KNOX family has undergone quite a bit of evolutionary diversification, while keeping the overall mechanism more or less similar. Members of the KNOX family have been found in plants as diverse as Arabidopsis thaliana, rice, barley and tomato. KNOX-like genes are also present in some algae, mosses, ferns and gymnosperms. Misexpression of these genes leads to formation of interesting morphological features. For example, among members of *Antirrhinae*, only the species of genus *Antirrhinum* lack a structure called spur in the floral region. A spur is considered an evolutionary innovation because it defines pollinator specificity and attraction. Researchers carried out transposon mutagenesis in *Antirrhinum majus*, and saw that some insertions led to formation of spurs that were very similar to the other members of *Antirrhinae*, indicating that the loss of spur in wild *Antirrhinum majus* populations could probably be an evolutionary innovation.

The KNOX family has also been implicated in leaf shape evolution. One study looked at the pattern of KNOX gene expression in *A. thaliana*, that has simple leaves and *Cardamine hirsuta*, a plant having complex leaves. In *A. thaliana*, the KNOX genes are completely turned off in leaves, but in *C.hirsuta*, the expression continued, generating complex leaves. Also, it has been proposed that the mechanism of KNOX gene action is conserved across all vascular plants, because there is a tight correlation between KNOX expression and a complex leaf morphology.

Evolution of the meristem architecture

The meristem architectures do differ between angiosperms, gymnosperms and pteridophytes. The gymnosperm vegetative meristem lacks organization into distinct tunica and corpus layers. They possess large cells called Central Mother Cells in the meristem. In angiosperms, the outermost layer of cells divides anticlinally to generate the new cells, while in gymnosperms, the plane of division in the meristem differs for different cells. However, the apical cells do contain organelles like large vacuoles and starch grains, like the angiosperm meristematic cells.

Pteridophytes, like fern, on the other hand, do not possess a multicellular apical meristem. They possess a tetrahedral apical cell, which goes on to form the plant body. Any somatic mutation in this cell can lead to hereditary transmission of that mutation. The earliest meristem-like organization is seen in an algal organism from group *Charales* that has a single dividing cell at the tip, much like the pteridophytes, yet simpler. One can thus see a clear pattern in evolution of the meristematic tissue, from pteridophytes to angiosperms. Pteridophytes, with a single meristematic cell; gymnosperms with a multicellular, but less defined organization and finally, angiosperms, with the highest degree of organization. The genetic innovations that contributed to this evolution are yet not clearly known.

Evolution of leaves

Origins of the leaf

Leaf lamina. The leaf architecture probably arose multiple times in the plant lineage

Leaves are the primary photosynthetic organs of a plant. Based on their structure, they are classified into two types - microphylls, that lack complex venation patterns and megaphylls, that are large and with a complex venation. It has been proposed that these structures arose independently. Megaphylls, according to the Telome hypothesis, have evolved from plants that showed a three dimensional branching architecture, through three transformations—**plantation**, which involved formation of a planar architecture, **webbing**, or formation of the outgrowths between the planar branches and **fusion**, where these webbed outgrowths fused to form a proper leaf lamina. Studies have revealed that these three steps happened multiple times in the evolution of today's leaves.

It has been proposed that the before the evolution of leaves, plants had the photosynthetic apparatus on the stems. Today's megaphyll leaves probably became commonplace some 360mya, about 40my after the simple leafless plants had colonized the land in the early Devonian period. This spread has been linked to the fall in the atmospheric carbon dioxide concentrations in the Late Paleozoic era associated with a rise in density of stomata on leaf surface. This must have allowed for better transpiration rates and gas

exchange. Large leaves with less stomata would have gotten heated up in the sun's heat, but an increased stomatal density allowed for a better-cooled leaf, thus making its spread feasible.

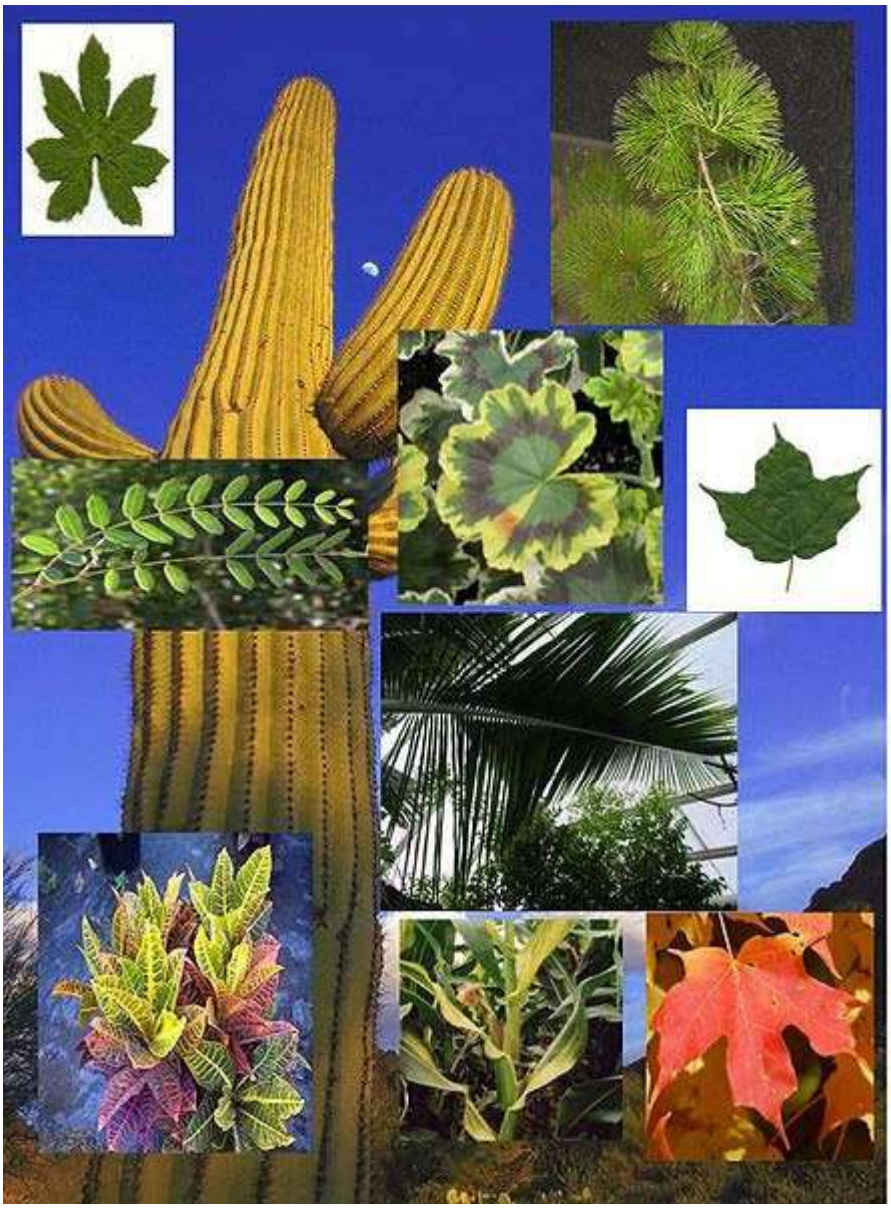
Factors influencing leaf architectures

Spiny leaves of *Aciphylla squarrosa*. It is thought that these leaves evolved as an adaptation against the now extinct Moas

Various physical and physiological forces like light intensity, humidity, temperature, wind speeds etc. are thought to have influenced evolution of leaf shape and size. It is observed that high trees rarely have large leaves, owing to the obstruction they generate for winds. This obstruction can eventually lead to the tearing of leaves, if they are large. Similarly, trees that grow in temperate or taiga regions have pointed leaves, presumably to prevent nucleation of ice onto the leaf surface and reduce water loss due to transpiration. Herbivory, not only by large mammals, but also small insects has been implicated as a driving force in leaf evolution, an example being plants of the genus *Aciphylla*, that are commonly found in New Zealand. The now extinct Moas fed upon these plants, and its seen that the leaves have spines on their bodies, which probably functioned to discourage the moas from feeding on them. Other members of *Aciphylla* that did not co-exist with the moas, do not have these spines.

Genetic evidences for leaf evolution

At the genetic level, developmental studies have shown that repression of the KNOX genes is required for initiation of the leaf primordium. This is brought about by *ARP* genes, which encode transcription factors. Genes of this type have been found in many plants studied till now, and the mechanism i.e. repression of KNOX genes in leaf primordia, seems to be quite conserved. Interestingly, expression of KNOX genes in leaves produces complex leaves. It is speculated that the *ARP* function arose quite early in vascular plant evolution, because members of the primitive group Lycophytes also have a functionally similar gene Other players that have a conserved role in defining leaf primordia are the phytohormone auxin, gibberelin and cytokinin.



The diversity of leaves

One interesting feature of a plant is its phyllotaxy. The arrangement of leaves on the plant body is such that the plant can maximally harvest light under the given constraints, and hence, one might expect the trait to be genetically robust. However, it may not be so. In maize, a mutation in only one gene called *abphyl (ABNORMAL PHYLLOTAXY)* was enough to change the phyllotaxy of the leaves. It implies that sometimes, mutational tweaking of a single locus on the genome is enough to generate diversity. The *abphyl* gene was later on shown to encode a cytokinin response regulator protein.

Once the leaf primordial cells are established from the SAM cells, the new axes for leaf growth are defined, one important (and more studied) among them being the abaxial-adaxial (lower-upper surface) axes. The genes involved in defining this, and the other axes seem to be more or less conserved among higher plants. Proteins of the *HD-ZIPIII* family have been implicated in defining the adaxial identity. These proteins deviate some cells in the leaf primordium from the default abaxial state, and make them adaxial. It is believed that in early plants with leaves, the leaves just had one type of surface - the abaxial one. This is the underside of today's leaves. The definition of the adaxial identity occurred some 200 million years after the abaxial identity was established. One can thus imagine the early leaves as an intermediate stage in evolution of today's leaves, having just arisen from spiny stem-like outgrowths of their leafless ancestors, covered with stomata all over, and not optimized as much for light harvesting.

How the infinite variety of plant leaves is generated is a subject of intense research. Some common themes have emerged. One of the most significant is the involvement of KNOX genes in generating compound leaves, as in tomato. But this again is not universal. For example, pea uses a different mechanism for doing the same thing. Mutations in genes affecting leaf curvature can also change leaf form, by changing the leaf from flat, to a crinky shape, like the shape of cabbage leaves. There also exist different morphogen gradients in a developing leaf which define the leaf's axis. Changes in these morphogen gradients may also affect the leaf form. Another very important class of regulators of leaf development are the microRNAs, whose role in this process has just begun to be documented. The coming years should see a rapid development in comparative studies on leaf development, with many EST sequences involved in the process coming online.

Evolution of flowers

The pollen bearing organs of the early flower *Crossotheca*

A flower is, arguably, one of the most beautiful products of evolution. Flower-like structures first appear in the fossil records some ~130 mya, in the Cretaceous era.

The flowering plants have long been assumed to have evolved from within the gymnosperms; according to the traditional morphological view, they are closely allied to the gnetales. However, recent molecular evidence is at odds to this hypothesis, and further suggests that gnetales are more closely related to some gymnosperm groups than angiosperms, and that gymnosperms form a distinct clade to the angiosperms,. Molecular clock analysis predicts the divergence of flowering plants (anthophytes) and gymnosperms to ~300 mya

The main function of a flower is reproduction, which, before the evolution of the flower and angiosperms, was the job of microsporophylls and megasporophylls. A flower can be considered a powerful evolutionary innovation, because its presence allowed the plant world to access new means and mechanisms for reproduction.

Origins of the flower

Amborella trichopoda : Amborellaceae is considered the sister family of all flowering plants *(magnified image)*

The family Amborellaceae is regarded as the sister family of all living flowering plants. That means members of this family were most likely the first flowering plants.

It seems that on the level of the organ, the leaf may be the ancestor of the flower, or at least some floral organs. When we mutate some crucial genes involved in flower development, we end up with a cluster of leaf-like structures. Thus, sometime in history, the developmental program leading to formation of a leaf must have been altered to generate a flower. There probably also exists an overall robust framework within which the floral diversity has been generated. A example of that is a gene called *LEAFY (LFY)*, which is involved in flower development in *Arabidopsis thaliana*. The homologs of this gene are found in angiosperms as diverse as tomato, snapdragon, pea, maize and even gymnosperms. Interestingly, expression of *Arabidopsis thaliana* LFY in distant plants like poplar and citrus also results in flower-production in these plants. The *LFY* gene regulates the expression of some gene belonging to the MADS-box family. These genes, in turn, act as direct controllers of flower development.

Evolution of the MADS-box family

The members of the MADS-box family of transcription factors play a very important and evolutionarily conserved role in flower development. According to the ABC Model of flower development, three zones - A,B and C - are generated within the developing flower primordium, by the action of some transcription factors, that are members of the MADS-box family. Among these, the functions of the B and C domain genes have been evolutionarily more conserved than the A domain gene. Many of these genes have arisen through gene duplications of ancestral members of this family. Quite a few of them show redundant functions.

The evolution of the MADS-box family has been extensively studied. These genes are present even in pteridophytes, but the spread and diversity is many times higher in angiosperms. There appears to be quite a bit of pattern into how this family has evolved. Consider the evolution of the C-region gene *AGAMOUS (AG)*. It is expressed in today's flowers in the stamens, and the carpel, which are reproductive organs. It's ancestor in gymnosperms also has the same expression pattern. Here, it is expressed in the strobili, an organ that produces pollens or ovules. Similarly, the B-genes' *(AP3 and PI)* ancestors are expressed only in the male organs in gymnosperms. Their descendants in the modern angiosperms also are expressed only in the stamens, the male reproductive organ. Thus, the same, then-existing components were used by the plants in a novel manner to generate the first flower. This is a recurring pattern in evolution.

Factors influencing floral diversity

The various shapes and colors of flowers

How is the enormous diversity in the shape, color and sizes of flowers established? There is enormous variation in the developmental program in different plants. For example, monocots possess structures like lodicules and palea, that were believed to be analogous to the dicot petals and carpels respectively.It turns out that this is true, and the variation is due to slight changes in the MADS-box genes and their expression pattern in the monocots. Another example is that of a plant called *Linaria vulgaris*, which has two kinds of flower symmetries-radial and bilateral. These symmetries are due to epigenetic changes in just one gene called *CYCLOIDEA*.

Large number of petals in roses has probably been a result of human selection

Arabidopsis thaliana has a gene called *AGAMOUS* that plays an important role in defining how many petals and sepals and other organs are generated. Mutations in this gene give rise to the floral meristem obtaining an indeterminate fate, and many floral organs keep on getting produced. We have flowers like roses, carnations and morning glory, for example, that have very dense floral organs. These flowers have been selected by horticulturists since long for increased number of petals. Researchers have found that the morphology of these flowers is because of strong mutations in the *AGAMOUS* homolog in these plants, which leads to them making a large number of petals and sepals. Several studies on diverse plants like petunia, tomato, Impatiens, maize etc. have suggested that the enormous diversity of flowers is a result of small changes in genes controlling their development.

Some of these changes also cause changes in expression patterns of the developmental genes, resulting in different phenotypes. The Floral Genome Project looked at the EST data from various tissues of many flowering plants. The researchers confirmed that the ABC Model of flower development is not conserved across all angiosperms. Sometimes expression domains change, as in the case of many monocots, and also in some basal angiosperms like *Amborella*. Different models of flower development like the *The fading boundaries model*, or the *Overlapping-boundaries model* which propose non-rigid domains of expression, may explain these architectures. There is a possibility that from the basal to the modern angiosperms, the domains of floral architecture have gotten more and more fixed through evolution.

Flowering time

Another floral feature that has been a subject of natural selection is flowering time. Some plants flower early in their life cycle, others require a period of vernalization before flowering. This decision is based on factors like temperature, light intensity, presence of pollinators and other environmental signals. We know that genes like *CONSTANS (CO)*, *Flowering Locus C* (*FLC*) and *FRIGIDA* regulate integration of environmental signals into the pathway for flower development. Variations in these loci have been associated with flowering time variations between plants. For example, *Arabidopsis thaliana* ecotypes that grow in the cold, temperate regions require prolonged vernalization before they flower, while the tropical varieties, and the most common lab strains, don't. We now know that this variation is due to mutations in the *FLC* and *FRIGIDA* genes, rendering them non-functional.

Quite a few players in this process are conserved across all the plants studied. Sometimes though, despite genetic conservation, the mechanism of action turns out to be different. For example, rice is a short-day plant, while *Arabidopsis thaliana* is a long-day plant. Now, in both plants, the proteins *CO* and *FLOWERING LOCUS T (FT)* are present. But in *Arabidopsis thaliana*, *CO* enhances *FT* production, while in rice, the *CO* homolog represses *FT* production, resulting in completely opposite downstream effects.

Theories of flower evolution

There are many theories that propose how flowers evolved. Some of them are described below.

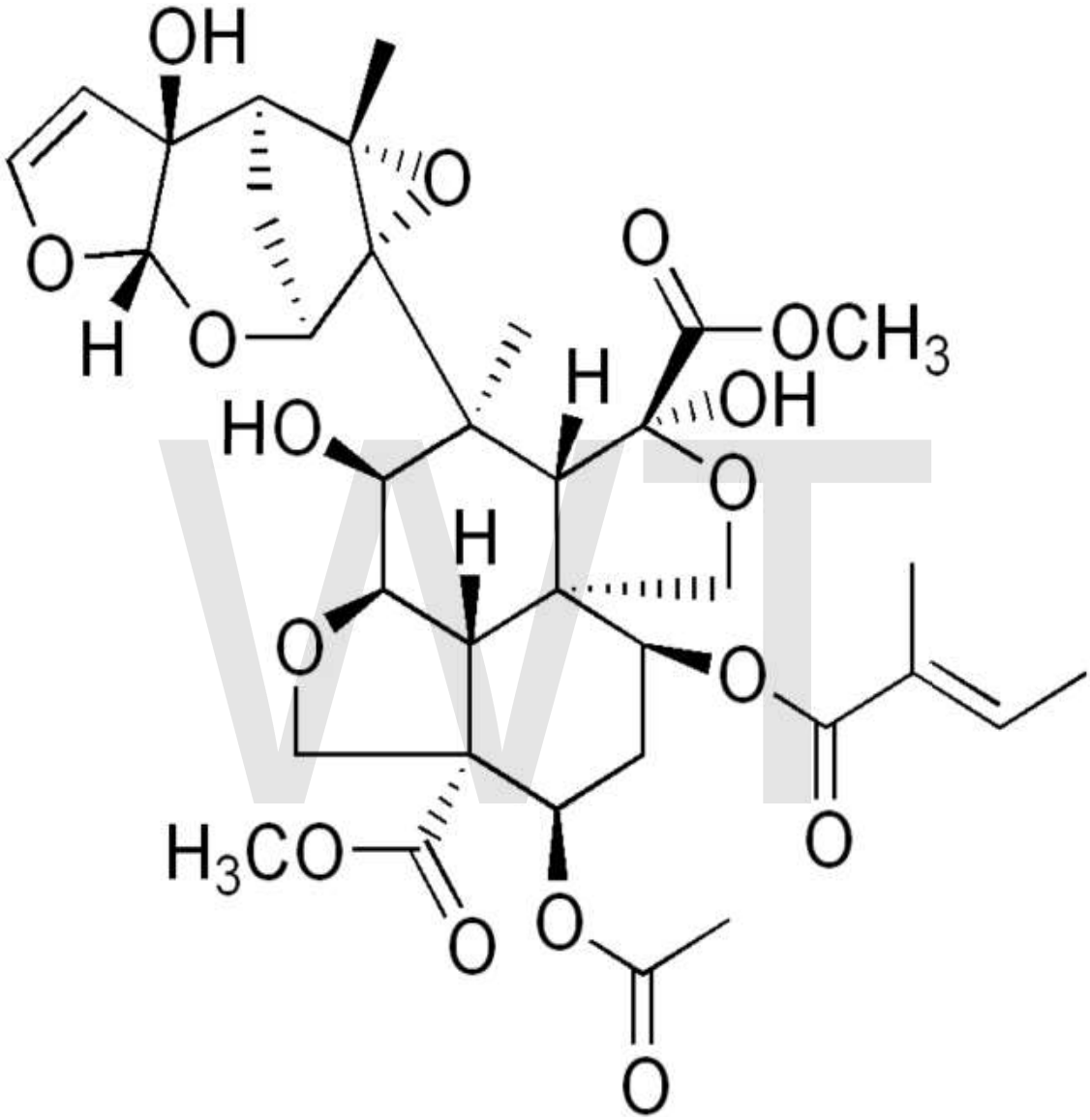
The *Anthophyte Theory* was based upon the observation that a gymnospermic group Gnetales has a flower-like ovule. It has partially developed vessels as found in the angiosperms, and the megasporangium is covered by three envelopes, like the ovary structure of angiosperm flowers. However, many other lines of evidence show that Gnetales is not related to angiosperms.

The *Mostly Male Theory* has a more genetic basis. Proponents of this theory point out that the gymnosperms have two very similar copies of the gene *LFY* while angiosperms just one. Molecular clock analysis has shown that the other *LFY* paralog was lost in angiosperms around the same time as flower fossils become abundant, suggesting that this event might have led to floral evolution. According to this theory, loss of one of the *LFY* paralog led to flowers that were more male, with the ovules being expressed ectopically. These ovules initially performed the function of attracting pollinators, but sometime later, may have been integrated into the core flower.

One theory also suggests that humans have been one of the reasons for the diversity of flowers. This theory suggests that since the early settlers found flowers beautiful, they may have started selecting for them artificially. The flowers may have evolved to exploit the ecological niche being opened because of humans finding them attractive. The

validity of this theory, however, is debatable, not least because flowers started diversifying long before they came into contact with humans.

Evolution of secondary metabolism



Structure of Azadirachtin, a terpenoid produced by the Neem plant, which helps ward off microbes and insects. Many secondary metabolites have complex structures

Although we know many secondary metabolites produced by plants, the extent of the same is still unfathomable. Secondary metabolites are essentially low molecular weight compounds, sometimes having complex structures. They function in processes as diverse as immunity, anti-herbivory, pollinator attraction, communication between plants, maintaining symbiotic associations with soil flora, enhancing the rate of fertilization etc., and hence are significant from the evo-devo perspective. The structural and functional diversity of these secondary metabolites across the plant kingdom is so huge that it is

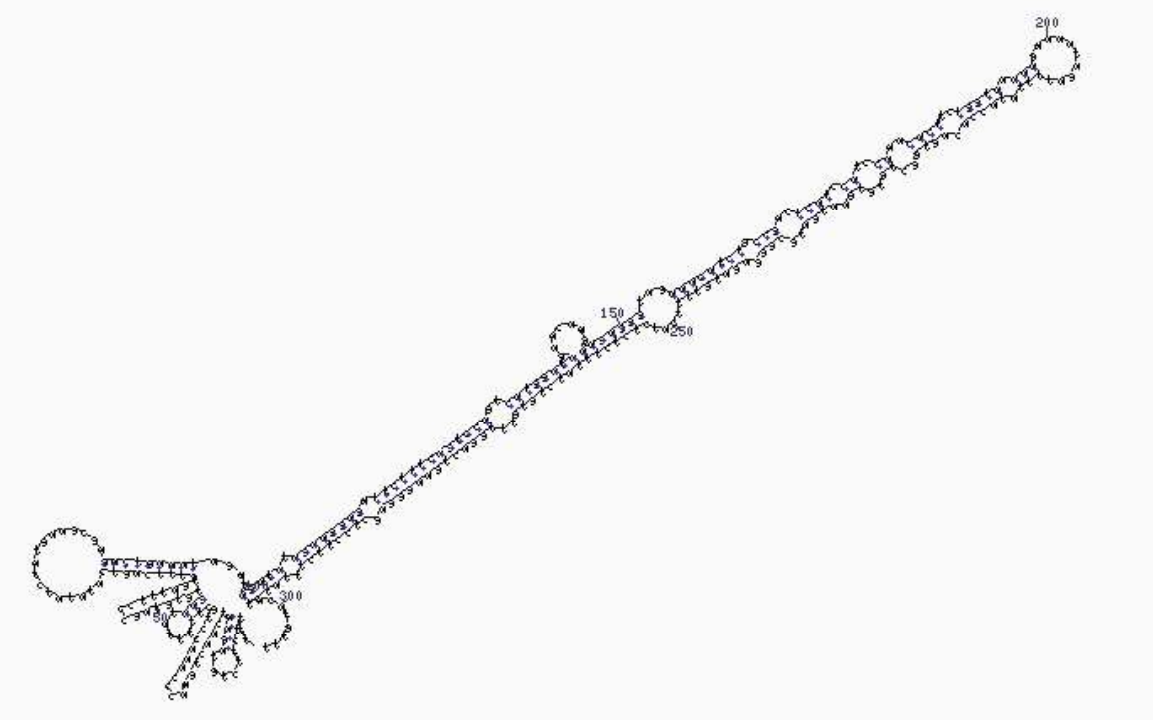
estimated that hundreds of thousands of enzymes might be involved in this process in the entire of the plant kingdom, with about 15–25% of the coding genome coding for these enzymes. Despite this, every species has its unique arsenal of secondary metabolites. Many of these metabolites are of enormous medical significance to humans.

What is the purpose of having so many secondary metabolites being produced, with a significant chunk of the metabolome devoted to this activity? It is hypothesized that most of these chemicals help in generating immunity, and in consequence, the diversity of these metabolites is a result of a constant war between plants and their parasites. There is evidence that this may be true in many cases. The big question here is the reproductive cost involved in maintaining such an impressive inventory. Various models have been suggested that probe into this aspect of the question, but a consensus on the extent of the cost is lacking. We still cannot predict whether a plant with more secondary metabolites would be better off than other plants in its vicinity.

Secondary metabolite production seems to have arisen quite early during evolution. Even bacteria possess the ability to make these compounds. But they assume more significant roles in life from fungi onwards to plants. In plants they seem to have spread out using different mechanisms like gene duplications, evolution of novel genes etc. Furthermore, studies have shown that diversity in some of these compounds may be positively selected for.

Although the role of novel gene evolution in the evolution of secondary metabolism cannot be denied, there are several examples where new metabolites have been formed by small changes in the reaction. For example, cyanogen glycosides have been proposed to have evolved multiple times in different plant lineages. There are several such instances of convergent evolution. For example, we now know that enzymes for synthesis of limonene – a terpene – are more similar between angiosperms and gymnosperms than to their own terpene synthesis enzymes. This suggests independent evolution of the limonene biosynthetic pathway in these two lineages.

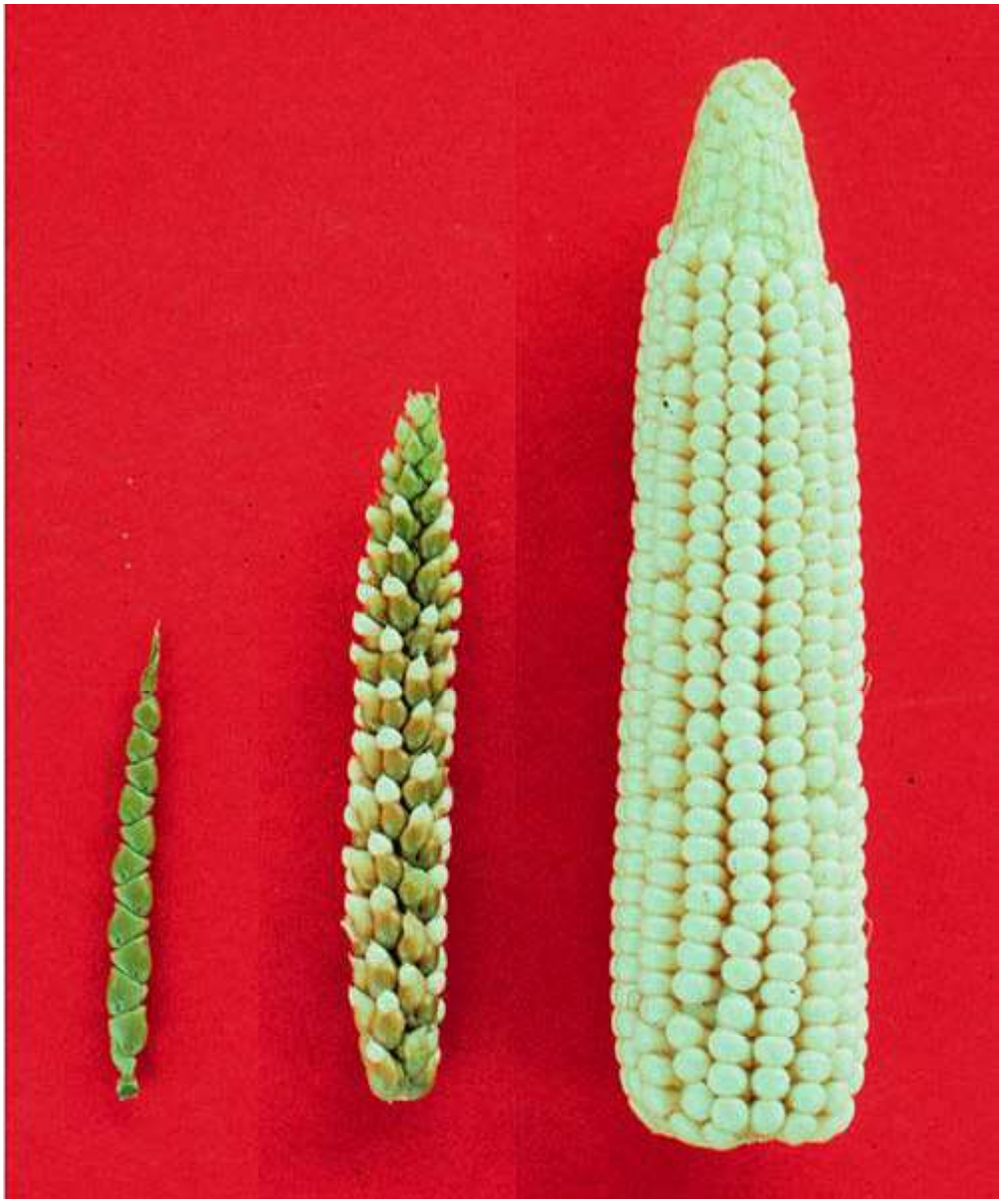
Mechanisms and players in evolution



The stem-loop secondary structure of a pre-microRNA from Brassica oleracea

While environmental factors are significantly responsible for evolutionary change, they act merely as agents for natural selection. Change is inherently brought about via phenomena at the genetic level - mutations, chromosomal rearrangements and epigenetic changes. While the general types of mutations hold true across the living world, in plants, some other mechanisms have been implicated as highly significant.

Polyploidy is a very common feature in plants. It is believed that at least half *(and probably all)* plants are or have been polyploids. Polyploidy leads to genome doubling, thus generating functional redundancy in most genes. The duplicated genes may attain new function, either by changes in expression pattern or changes in activity. Polyploidy and gene duplication are believed to be among the most powerful forces in evolution of plant form. It is not know though, why genome doubling is such a frequent process in plants. One probable reason is the production of large amounts of secondary metabolites in plant cells. Some of them might interfere in the normal process of chromosomal segregation, leading to polypoidy.



Extreme left: teosinte, Extreme right: maize, middle: maize-teosinte hybrid

In recent times, plants have been shown to possess significant microRNA families, which are conserved across many plant lineages. In comparison to animals, while the number of plant miRNA families are lesser than animals, the size of each family is much larger. The miRNA genes are also much more spread out in the genome than those in animals, where we find them clustered. It has been proposed that these miRNA families have expanded by duplications of chromosomal regions. Many miRNA genes involved in regulation of plant development have been found to be quite conserved between plants studied.

Domestication of plants like maize, rice, barley, wheat etc. has also been a significant driving force in their evolution. Some studies have tried to look at the origins of the maize plant and it turns out that maize is a domesticated derivative of a wild plant from

Mexico called teosinte. Teosinte belongs to the genus *Zea*, just as maize, but bears very small inflorescence, 5-10 hard cobs and a highly branched and spread out stem.

Cauliflower : *Brassica oleracea var botrytis*

Interestingly, crosses between a particular teosinte variety and maize yields fertile offsprings that are intermediate in phenotype between maize and teosinte. QTL analysis has also revealed some loci that when mutated in maize yield a teosinte-like stem or teosinte-like cobs. Molecular clock analysis of these genes estimates their origins to some 9000 years ago, well in accordance with other records of maize domestication. It is believed that a small group of farmers must have selected some maize-like natural mutant of teosinte some 9000 years ago in Mexico, and subjected it to continuous selection to yield the maize plant as we know today.

Another interesting case is that of cauliflower. The edible cauliflower is a domesticated version of the wild plant *Brassica oleracea*, which does not possess the dense undifferentiated inflorescence called the curd, that cauliflower possesses.

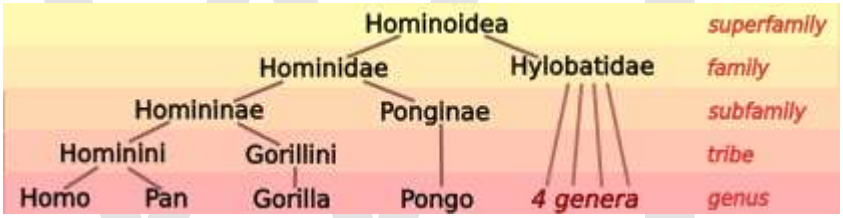
Cauliflower possesses a single mutation in a gene called *CAL*, controlling meristem differentiation into inflorescence. This causes the cells at the floral meristem to gain an undifferentiated identity, and instead of growing into a flower, they grow into a lump of undifferentiated cells. This mutation has been selected through domestication at least since the Greek empire.

Chapter- 7

Human Evolutionary Genetics

Human evolutionary genetics studies how one human genome differs from the other, the evolutionary past that gave rise to it, and its current effects. Differences between genomes have anthropological, medical and forensic implications and applications. Genetic data can provide important insight into human evolution.

Origin of apes

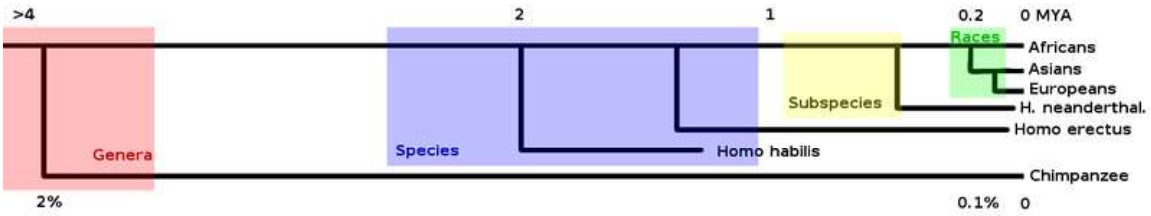


Taxonomic relationships of hominoids

Biologists classify humans, along with only a few other species, as great apes (species in the family Hominidae). The Hominidae include two distinct species of chimpanzee (the bonobo, *Pan paniscus*, and the common chimpanzee, *Pan troglodytes*), two species of gorilla (the western gorilla, *Gorilla gorilla*, and the eastern gorilla, *Gorilla graueri*), and two species of orangutan (the Bornean orangutan, *Pongo pygmaeus*, and the Sumatran orangutan, *Pongo abelii*).

Apes, in turn, belong to the primates order (>400 species). Data from both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) indicates that primates belong to the group of Euarchontoglires, together with Rodentia, Lagomorpha, Dermoptera, and Scandentia. This is further supported by Alu-like short interspersed nuclear elements (SINEs) which have been found only in members of the Euarchontoglires.

Cladistics



A phylogenetic tree like the one shown above is usually derived from DNA or protein sequences from populations. Often mitochondrial DNA or Y chromosome sequences are used to study ancient human demographics. These single-locus sources of DNA do not recombine and are almost always inherited from a single parent, with only one known exception in mtDNA (Schwartz and Vissing 2002). Individuals from the various continental groups tend to be more similar to one another than to people from other continents. The tree is rooted in the common ancestor of chimpanzees and humans, which is believed to have originated in Africa. Horizontal distance in the diagram corresponds to two things:

1. **Genetic distance**. Given below the diagram, the genetic difference between humans and chimps is less than 2%, or 20 times larger than the variation among modern humans.
2. **Temporal remoteness** of the most recent common ancestor. Rough estimates are given above the diagram, in millions of years. The mitochondrial most recent common ancestor of modern humans lived roughly 200,000 years ago, latest common ancestors of humans and chimps between four and seven million years ago.

Chimpanzees and humans belong to different genera, indicated in red. Formation of species and subspecies is also indicated, and the formation of "races" is indicated in the green rectangle to the right (note that only a very rough representation of human phylogeny is given). Note that vertical distances are not meaningful in this representation.

Speciation of humans and the African apes

The separation of humans from their closest relatives, the African apes (chimpanzees and gorillas), has been studied extensively for more than a century. Five major questions have been addressed:

- Which apes are our closest ancestors?
- When did the separations occur?
- What was the effective population size of the common ancestor before the split?
- Are there traces of population structure (subpopulations) preceding the speciation or partial admixture succeeding it?

- What were the specific events (including fusion of chromosomes 2a and 2b) prior to and subsequent to the separation?

General observations

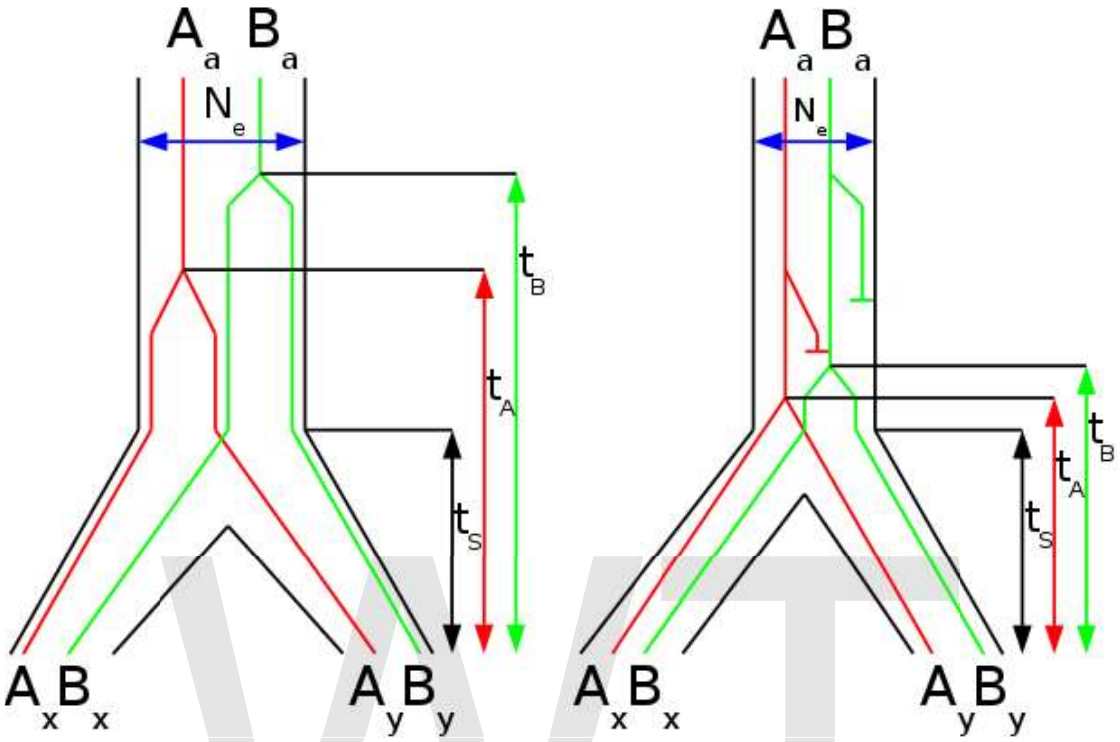
As discussed before, different parts of the genome show different sequence divergence between different hominoids. It has also been shown that the sequence divergence between DNA from humans and chimpanzees varies greatly. For example the sequence divergence varies between 0% to 2.66% between non-coding, non-repetitive genomic regions of humans and chimpanzees. Additionally gene trees, generated by comparative analysis of DNA segments, do not always fit the species tree. Summing up:

- The sequence divergence varies significantly between humans, chimpanzees and gorillas.
- For most DNA sequences, humans and chimpanzees appear to be most closely related, but some point to a human-gorilla or chimpanzee-gorilla clade.
- The human genome has been sequenced, as well as the chimpanzee genome. Humans have 23 pairs of chromosomes, while chimpanzees, gorillas, and orangutans have 24. Human chromosome 2 is a fusion between two chromosomes that remained separate in the other primates.

Divergence times

The divergence time of humans from other apes is of great interest. One of the first molecular studies, published in 1967 measured immunological distances (IDs) between different primates. Basically the study measured the strength of immunological response that an antigen from one species (human albumin) induces in the immune system of another species (human, chimpanzee, gorilla and Old World monkeys). Closely related species should have similar antigens and therefore weaker immunological response to each other's antigens. The immunological response of a species to its own antigens (e.g. human to human) was set to be 1. The ID between humans and gorillas was determined to be 1.09, that between humans and chimpanzees was determined as 1.14. However the distance to six different Old World monkeys was on average 2.46 indicating that the African apes are far closer related to humans than to monkeys. The authors consider the divergence time between Old World monkeys and hominoids to be 30 million years ago (MYA), based on fossil data, and the immunological distance was considered to grow at a constant rate. They concluded that divergence time of humans and the African apes to be roughly ~5 MYA. That was a surprising result. Most scientists at that time thought that humans and great apes diverged much earlier (>15 MYA). The gorilla was, in ID terms, closer to human than to chimpanzees, however the difference was so slight that the trichotomy could not be resolved with certainty. Later studies based on molecular genetics were able to resolve the trichotomy: chimpanzees are phylogenetically closer to humans than to gorillas. However, the divergence times estimated later (using much more sophisticated methods in molecular genetics) do not substantially differ from the very first estimate in 1967.

Divergence times and ancestral effective population size

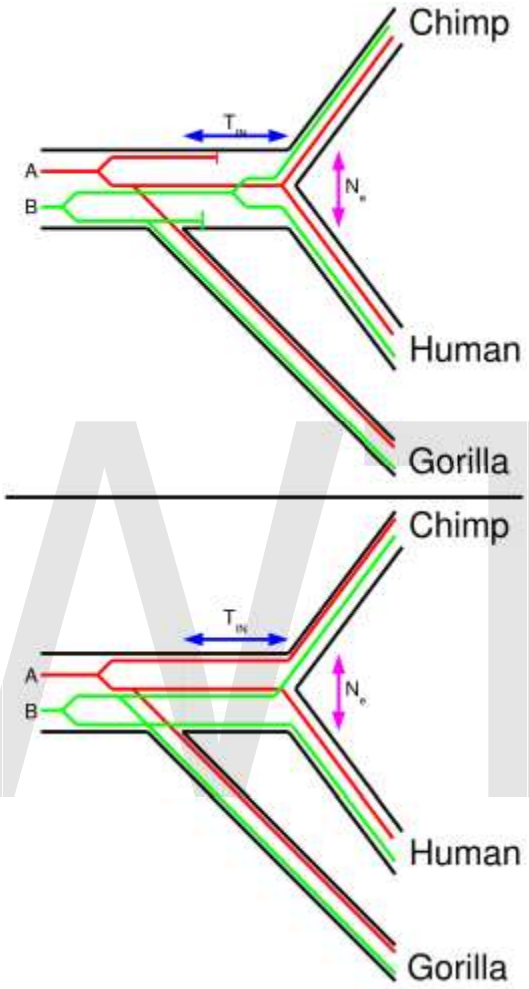


The sequences of the DNA segments diverge earlier than the species. A large effective population size in the ancestral population (left) preserves different variants of the DNA segments (=alleles) for a longer period of time. Therefore, on average, the gene divergence times (t_A for DNA segment A; t_B for DNA segment B) will deviate more from the time the species diverge (t_S) compared to a small ancestral effective population size (right).

Current methods to determine divergence times use DNA sequence alignments and molecular clocks. Usually the molecular clock is calibrated assuming that the orangutan split from the African apes (including humans) 12-16 MYA. Some studies also include some old world monkeys and set the divergence time of them from hominoids to 25-30 MYA. Both calibration points are based on very little fossil data and have been criticized. If these dates are revised, the divergence times estimated from molecular data will change as well. However, the relative divergence times are unlikely to change. Even if we can't tell absolute divergence times exactly, we can be pretty sure that the divergence time between chimpanzees and humans is about sixfold shorter than between chimpanzees (or humans) and monkeys.

One study (Takahata *et al.*, 1995) used 15 DNA sequence from different regions of the genome from human and chimpanzee and 7 DNA sequences from human, chimpanzee and gorilla. They determined that chimpanzees are more closely related to humans than gorillas. Using various statistical methods, they estimated the divergence time human-chimp to be 4.7 MYA and the divergence time between gorillas and humans (and

chimps) to be 7.2 MYA. Additionally they estimated the effective population size of the common ancestor of humans and chimpanzees to be ~100,000. This was somewhat surprising since the present day effective population size of humans is estimate to be only ~10,000. If true that means that the human lineage would have experienced an immense decrease of its effective population size (and thus genetic diversity) in its evolution.



A and B are two different loci. In the upper figure they fit to the species tree. The DNA that is present in today's gorillas diverged earlier from the DNA that is present in today's humans and chimps. Thus both loci should be more similar between human and chimp than between gorilla and chimp or gorilla and human. In the lower graph, locus A has a more recent common ancestor in human and gorilla compared to the chimp sequence. Whereas chimp and gorilla have a more recent common ancestor for locus B. Here the gene trees are incongruent to the species tree.

Another study (Chen & Li, 2001) sequenced 53 non-repetitive, intergenic DNA segments from a human, a chimpanzee, a gorilla, and orangutan. When the DNA sequences were concatenated to a single long sequence, the generated neighbor-joining tree supported the *Homo-Pan* clade with 100% bootstrap (that is that humans and chimpanzees are the

closest related species of the four). When three species are fairly closely related to each other (like human, chimpanzee and gorilla), the trees obtained from DNA sequence data may not be congruent with the tree that represents the speciation (species tree). The shorter internodal time span (T_{IN}) the more common are incongruent gene trees. The effective population size (N_e) of the internodal population determines how long genetic lineages are preserved in the population. A higher effective population size causes more incongruent gene trees. Therefore, if the internodal time span is known, the ancestral effective population size of the common ancestor of humans and chimpanzees can be calculated.

When each segment was analyzed individually, 31 supported the *Homo-Pan* clade, 10 supported the *Homo-Gorilla* clade, and 12 supported the *Pan-Gorilla* clade. Using the molecular clock the authors estimated that gorillas split up first 6.2-8.4 MYA and chimpanzees and humans split up 1.6-2.2 million years later (internodal time span) 4.6-6.2 MYA. The internodal time span is useful to estimate the ancestral effective population size of the common ancestor of humans and chimpanzees.

A parsimonious analysis revealed that 24 loci supported the *Homo-Pan* clade, 7 supported the *Homo-Gorilla* clade, 2 supported the *Pan-Gorilla* clade and 20 gave no resolution. Additionally they took 35 protein coding loci from databases. Of these 12 supported the *Homo-Pan* clade, 3 the *Homo-Gorilla* clade, 4 the *Pan-Gorilla* clade and 16 gave no resolution. Therefore only ~70% of the 52 loci that gave a resolution (33 intergenic, 19 protein coding) support the 'correct' species tree. From the fraction of loci which did not support the species tree and the internodal time span they estimated previously, the effective population of the common ancestor of humans and chimpanzees was estimated to be ~52 000 to 96 000. This value is not as high as that from the first study (Takahata), but still much higher than present day effective population size of humans.

A third study (Yang, 2002) used the same dataset that Chen and Li used but estimated the ancestral effective population of 'only' ~12,000 to 21,000, using a different statistical method.

Genetic differences between humans and other great apes

The alignable sequences within genomes of humans and chimpanzees differ by about 35 million single nucleotide substitutions. Additionally about 3% of the complete genomes differ by deletions, insertions and duplications.

Since mutation rate is relatively constant, roughly one half of these changes occurred in the human lineage. Only a very tiny fraction of those fixed differences gave rise to the different phenotypes of humans and chimpanzees and finding those is a great challenge. The vast majority of the differences are neutral and do not affect the phenotype.

Molecular evolution may act in different ways, through protein evolution, gene loss, differential gene regulation and RNA evolution. All are thought to have played some part in human evolution.

Gene loss

Many different mutations can inactivate a gene, but few will change its function in a specific way. Inactivation mutations will therefore be readily available for selection to act on. Gene loss could thus be a common mechanism of evolutionary adaptation (the "less-is-more" hypothesis).

80 genes were lost in the human lineage after separation from the last common ancestor with the chimpanzee. 36 of those were for olfactory receptors. Genes involved in chemoreception and immune response are overrepresented. Another study estimated that 86 genes had been lost.

Hair keratin gene KRTHAP1

A gene for type I hair keratin was lost in the human lineage. Keratins are a major component of hairs. Humans still have nine functional type I hair keratin genes but the loss of that particular gene may have caused the thinning of human body hair. The gene loss occurred relatively recently in human evolution—less than 240,000 years ago.

Myosin gene MYH16

Stedman *et al.* (2004) stated that the loss of the sarcomeric myosin gene MYH16 in the human lineage led to smaller masticatory muscles. They estimated that the mutation that led to the inactivation (a two base pair deletion) occurred 2.4 million years ago, predating the appearance of *Homo ergaster/erectus* in Africa. The period that followed was marked by a strong increase in cranial capacity, promoting speculation that the loss of the gene may have removed an evolutionary constraint on brain size in the genus *Homo*.

Another estimate for the loss of the MYH16 gene is 5.3 million years ago, long before *Homo* appeared.

Other

- CASPASE12, a cysteinyl aspartate proteinase

Gene addition

Segmental duplications (SDs or LCRs) have had roles in creating new primate genes and shaping human genetic variation.

Selection pressures

Human accelerated regions are areas of the genome that differ between humans and chimpanzees to a greater extent than can be explained by genetic drift over the time since the two species shared a common ancestor. These regions show signs of being subject to natural selection, leading to the evolution of distinctly human traits. Two examples are HAR1F, which is believed to be related to brain development and HAR2 (a.k.a HACNS1) that may have played a role in the development of the opposable thumb.

Genetic differences between humans and Neanderthals

An international group of scientists has completed a draft of 63 percent of the Neanderthal genome, consisting of 3.7 billion base pairs. Some parts of the Neanderthal genome have more in common with chimps than humans , while other parts are shared by both Neanderthals and humans. Neanderthals and most humans share a lactose-intolerant variant of the lactase gene that encodes an enzyme that is unable to break down lactose in milk after weaning. Humans and Neanderthals also share the FOXP2 gene variant associated with brain development and with speech in humans, indicating that Neanderthals may have been able to speak. Chimps have two amino acid differences in FOXP2 compared with human FOXP2. In May 2010 the entire Neanderthal genome was completed and the results were published in the journal Nature. The results indicate some breeding between humans and Neanderthals as the genomes of non-African humans are 1-4% Neanderthal.

Sequence divergence between humans and apes

The draft sequence of the common chimpanzee genome published in the summer 2005 showed the regions that are similar enough to be aligned with one another account for 2400 million of the human genome's 3164.7 million bases – that is, 75.8% of the genome. This 75.8% of the human genome is 1.23% different from the chimpanzee genome in single nucleotide polymorphisms (changes of single DNA "letters" in the genome). Another type of difference, called indels (insertions/deletions) account for another ~3 % difference between the alignable sequences. In addition, variation in copy number of large segments (> 20 kb) of similar DNA sequence provides a further 2.7% difference between the two species. Hence the total similarity of the genomes could be as low as about 70%.

The figures above do not take into account differences in the organization of the alignable sequences within the genomes of humans and chimps. Short stretches of alignable sequence may be in very different orders and locations within the two genomes. At present we cannot fully assess the difference in structure of the two genomes, because the human genome was used as a scaffold when the chimpanzee draft genome was assembled. When genomes are sequenced, relatively short sequences of DNA are produced, and these sequences have to be fitted together like a jigsaw puzzle. This requires multiple overlapping reads to accurately assemble the overall sequence. The

human genome sequence is relatively accurate, with 8 to 9-fold coverage, but the chimpanzee draft genome only has 3.6-fold coverage. The human genome was sequenced using a hierarchical shotgun method which can deal with duplications and difficult-to-assemble sequences better than the whole genome shotgun method that was used for the chimpanzee draft genome. The human genome was used as a template for the assembly of the draft chimpanzee genome, on the assumption that the two genomes would be similar.

Almost half of that 1.23% SNP change belongs to the human at 0.53%, whose genetic variance is lower than a chimp, and just over half to the chimp at 0.7%. If we also take into account that random "genetic drift" takes up the bulk of the 0.54% difference, then that percentage difference where SNPs have a potential positive impact on human abilities, is between 0.01% and 0.02%. The bonobo is a sibling species of common chimpanzee and is genetically about as different from humans as are common chimps.

Percentage sequence divergence between humans and other hominids

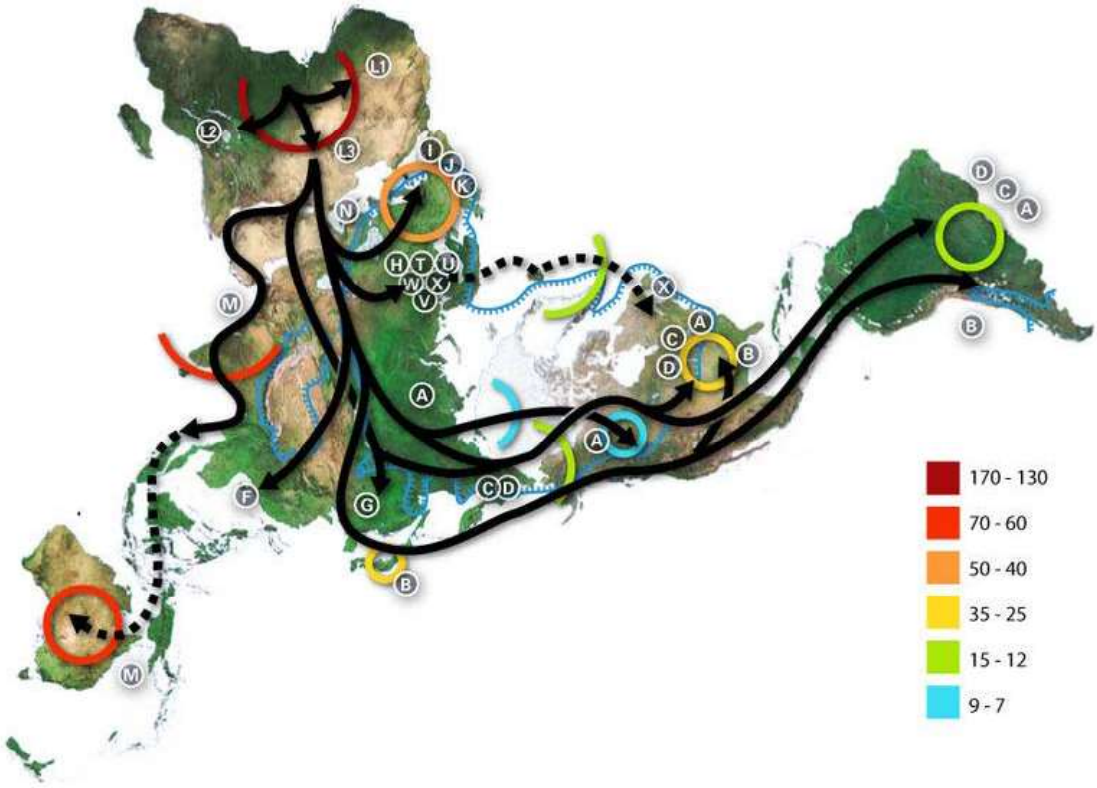
Locus	Human-Chimp	Human-Gorilla	Human-Orangutan
Alu elements	2	-	-
Non-coding (Chr. Y)	1.68 ± 0.19	2.33 ± 0.2	5.63 ± 0.35
Pseudogenes (autosomal)	1.64 ± 0.10	1.87 ± 0.11	-
Pseudogenes (Chr. X)	1.47 ± 0.17	-	-
Noncoding (autosomal)	1.24 ± 0.07	1.62 ± 0.08	3.08 ± 0.11
Genes (K_s)	1.11	1.48	2.98
Introns	0.93 ± 0.08	1.23 ± 0.09	-
Xq13.3	0.92 ± 0.10	1.42 ± 0.12	3.00 ± 0.18
Subtotal for X chromosome	1.16 ± 0.07	1.47 ± 0.08	-
Genes (K_a)	0.8	0.93	1.96

The sequence divergence has generally the following pattern: Human-Chimp < Human-Gorilla << Human-Orangutan, highlighting the close kinship between humans and the African apes. Alu elements diverge quickly due to their high frequency of CpG dinucleotides which mutate roughly 10 times more often than the average nucleotide in the genome. The mutation rate is higher in the male germ line, therefore the divergence in the Y chromosome—which is inherited solely from the father—is higher than in autosomes. The X chromosome is inherited twice as often through the female germ line as through the male germ line and therefore shows slightly lower sequence divergence. The sequence divergence of the Xq13.3 region is surprisingly low between humans and chimpanzees.

Mutations altering the amino acid sequence of proteins (K_a) are the least common. In fact ~29% of all orthologous proteins are identical between human and chimpanzee. The typical protein differs by only two amino acids.

The measures of sequence divergence shown in the table only take the substitutional differences, for example from an A (adenine) to a G (guanine), into account. DNA sequences may however also differ by insertions and deletions (indels) of bases. These are usually stripped from the alignments before the calculation of sequence divergence is performed. The overall sequence divergence between humans and chimpanzees for example is close to 5% if indels would be included.

Modern humans



Map of the migration of modern humans out of Africa, based on mitochondrial DNA. Coloured rings indicate years before present, in thousands.

Molecular biologists starting with Wesley Brown on mtDNA and Allan Wilson on mtDNA have produced observations relevant to human evolution.

Age of the common ancestor

By estimating the rate at which mutations occur in mtDNA, the age of the common ancestral mtDNA type can be estimated: "the common ancestral mtDNA (type a) links mtDNA types that have diverged by an average of nearly 0.57%. Assuming a rate of 2%-4% per million years, this implies that the common ancestor of all surviving mtDNA types existed 140,000-290,000 years ago." This observation is robust, and this common

direct female line ancestor (or mitochondrial most recent common ancestor (mtMRCA)) of all extant humans has become known as Mitochondrial Eve. The observation that the mtMRCA is the direct matrilineal ancestor of all living humans does not mean either that she was the first anatomically modern human, nor that no other female humans lived concurrently with her. Other women would have lived at the same time and passed nuclear genes down to living humans, but their mitochondrial lineages were lost over time. This could be due to random events such as producing only male children.

African origin for modern humans

There is evidence that modern human mtDNA has an African origin: "We infer from the tree of minimum length... that Africa is a likely source of the human mitochondrial gene pool. This inference comes from the observation that one of the two primary branches leads exclusively to African mtDNAs... while the second primary branch also leads to African mtDNAs... By postulating that the common ancestral mtDNA... was African, we minimize the number of intercontinental migrations needed to account for the geographic distribution of mtDNA types."

The broad study of African genetic diversity headed by Sarah Tishkoff found the San people to express the greatest genetic diversity among the 113 distinct populations sampled, making them one of 14 "ancestral population clusters". The research also located the origin of modern human migration in south-western Africa, near the coastal border of Namibia and Angola.

Y chromosome findings

The Y chromosome is much larger than mtDNA, and is relatively homogeneous; therefore it has taken much longer to find distinct lineages and to analyse them. Conversely, because the Y chromosome is so large by comparison, it holds more genetic information. Y chromosome studies show similar findings to those made with mtDNA. The estimate for the age of the ancestral Y chromosome for all extant Y chromosomes is given at about 70,000 years ago and is also placed in Africa; the individual who contributed this Y chromosomal heritage is sometimes referred to as Y chromosome Adam. The difference in dates between Y chromosome Adam and mitochondrial Eve is usually attributed to a higher extinction rate for Y chromosomes due to greater differential reproductive success between individual men, which means that a small number of very successful men may produce many children, while a larger number of less successful men will produce far fewer children.

Chapter- 8

Molecular Evolution

Molecular evolution is the process of evolution at the scale of DNA, RNA, and proteins. Molecular evolution emerged as a scientific field in the 1960s as researchers from molecular biology, evolutionary biology and population genetics sought to understand recent discoveries on the structure and function of nucleic acids and protein. Some of the key topics that spurred development of the field have been the evolution of enzyme function, the use of nucleic acid divergence as a "molecular clock" to study species divergence, and the origin of noncoding DNA.

Recent advances in genomics, including whole-genome sequencing, high-throughput protein characterization, and bioinformatics have led to a dramatic increase in studies on the topic. In the 2000s, some of the active topics have been the role of gene duplication in the emergence of novel gene function, the extent of adaptive molecular evolution versus neutral drift, and the identification of molecular changes responsible for various human characteristics especially those pertaining to infection, disease, and cognition.

Principles of molecular evolution

Mutations

In molecular biology and genetics, **mutations** are changes in a genomic sequence: the DNA sequence of a cell's genome or the DNA or RNA sequence of a virus. Mutations are caused by radiation, viruses, transposons and mutagenic chemicals, as well as errors that occur during meiosis or DNA replication. They can also be induced by the organism itself, by cellular processes such as hypermutation.

Mutation can result in several different types of change in DNA sequences; these can either have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely. Studies in the fly *Drosophila melanogaster* suggest that if a mutation changes a protein produced by a gene, this will probably be harmful, with about 70 percent of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial. Due to the damaging effects that mutations can have on genes, organisms have evolved mechanisms such as DNA repair to remove mutations.

Therefore, the optimal mutation rate for a species is a trade-off between costs of a high mutation rate, such as deleterious mutations, and the metabolic costs of maintaining systems to reduce the mutation rate, such as DNA repair enzymes. Viruses that use RNA as their genetic material have rapid mutation rates, which can be an advantage since these viruses will evolve constantly and rapidly, and thus evade the defensive responses of e.g. the human immune system.

Description

Mutations can involve large sections of DNA becoming duplicated, usually through genetic recombination. These duplications are a major source of raw material for evolving new genes, with tens to hundreds of genes duplicated in animal genomes every million years. Most genes belong to larger families of genes of shared ancestry. Novel genes are produced by several methods, commonly through the duplication and mutation of an ancestral gene, or by recombining parts of different genes to form new combinations with new functions.

Here, domains act as modules, each with a particular and independent function, that can be mixed together to produce genes encoding new proteins with novel properties. For example, the human eye uses four genes to make structures that sense light: three for color vision and one for night vision; all four arose from a single ancestral gene. Another advantage of duplicating a gene (or even an entire genome) is that this increases redundancy; this allows one gene in the pair to acquire a new function while the other copy performs the original function. Other types of mutation occasionally create new genes from previously noncoding DNA.

Changes in chromosome number may involve even larger mutations, where segments of the DNA within chromosomes break and then rearrange. For example, two chromosomes in the *Homo* genus fused to produce human chromosome 2; this fusion did not occur in the lineage of the other apes, and they retain these separate chromosomes. In evolution, the most important role of such chromosomal rearrangements may be to accelerate the divergence of a population into new species by making populations less likely to interbreed, and thereby preserving genetic differences between these populations.

Sequences of DNA that can move about the genome, such as transposons, make up a major fraction of the genetic material of plants and animals, and may have been important in the evolution of genomes. For example, more than a million copies of the Alu sequence are present in the human genome, and these sequences have now been recruited to perform functions such as regulating gene expression. Another effect of these mobile DNA sequences is that when they move within a genome, they can mutate or delete existing genes and thereby produce genetic diversity.

A mutation has caused this garden moss rose to produce flowers of different colors. This is a somatic mutation that may also be passed on in the germ line.

In multicellular organisms with dedicated reproductive cells, mutations can be subdivided into germ line mutations, which can be passed on to descendants through their reproductive cells, and somatic mutations (also called acquired mutations), which involve cells outside the dedicated reproductive group and which are not usually transmitted to descendants. If the organism can reproduce asexually through mechanisms such as cuttings or budding the distinction can become blurred.

For example, plants can sometimes transmit somatic mutations to their descendants asexually or sexually where flower buds develop in somatically mutated parts of plants. A new mutation that was not inherited from either parent is called a *de novo* mutation.

The source of the mutation is unrelated to the consequence, although the consequences are related to which cells were mutated.

Nonlethal mutations accumulate within the gene pool and increase the amount of genetic variation. The abundance of some genetic changes within the gene pool can be reduced by natural selection, while other "more favorable" mutations may accumulate and result in adaptive evolutionary changes.

For example, a butterfly may produce offspring with new mutations. The majority of these mutations will have no effect; but one might change the color of one of the butterfly's offspring, making it harder (or easier) for predators to see. If this color change is advantageous, the chance of this butterfly surviving and producing its own offspring are a little better, and over time the number of butterflies with this mutation may form a larger percentage of the population.

Neutral mutations are defined as mutations whose effects do not influence the fitness of an individual. These can accumulate over time due to genetic drift. It is believed that the overwhelming majority of mutations have no significant effect on an organism's fitness. Also, DNA repair mechanisms are able to mend most changes before they become permanent mutations, and many organisms have mechanisms for eliminating otherwise permanently mutated somatic cells.

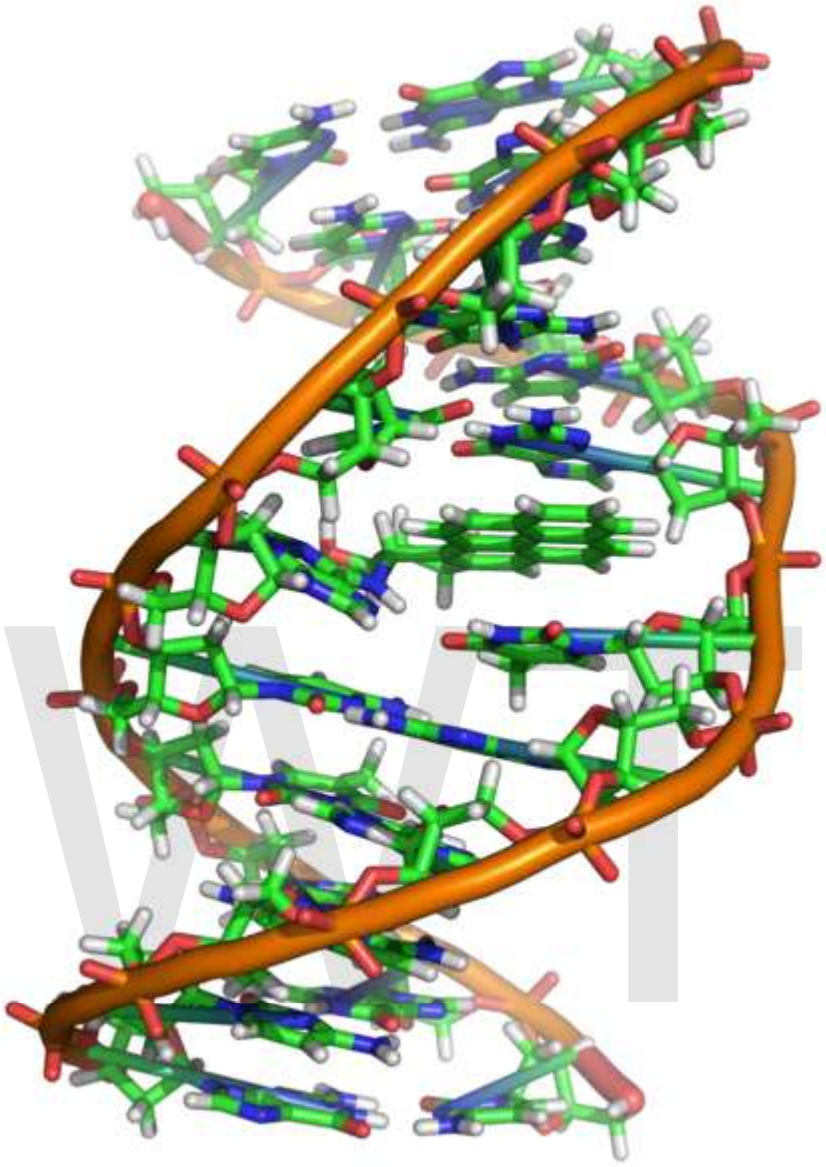
Mutation is generally accepted by biologists as the mechanism by which natural selection acts, generating advantageous new traits that survive and multiply in offspring as well as disadvantageous traits, in less fit offspring, that tend to die out.

Causes

Two classes of mutations are spontaneous mutations (molecular decay) and induced mutations caused by mutagens.

Spontaneous mutations on the molecular level can be caused by:

- Tautomerism – A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base resulting in incorrect base pairing during replication.
- Depurination – Loss of a purine base (A or G) to form an apurinic site (AP site).
- Deamination – Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.
- Slipped strand mispairing - Denaturation of the new strand from the template during replication, followed by renaturation in a different spot ("slipping"). This can lead to insertions or deletions.



A covalent adduct between benzo[*a*]pyrene, the major mutagen in tobacco smoke, and DNA

Induced mutations on the molecular level can be caused by:

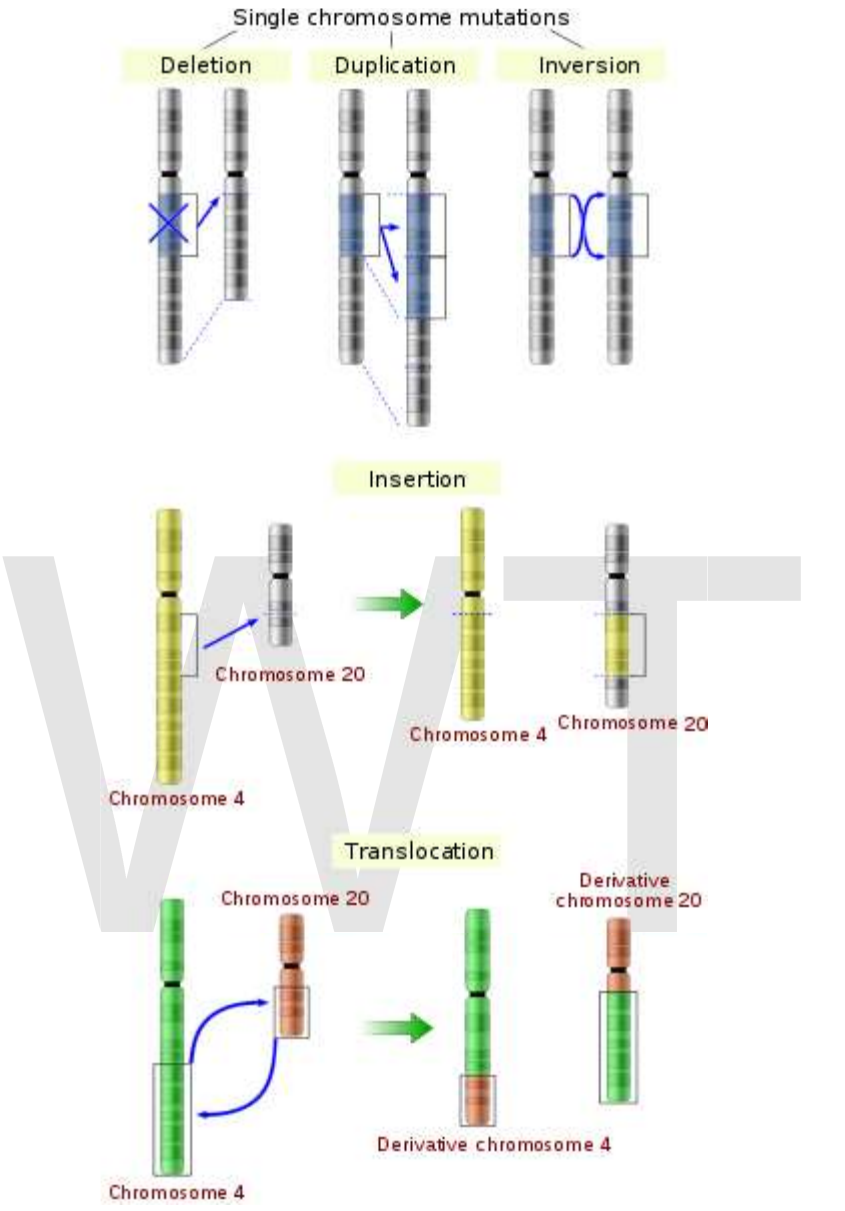
- Chemicals
  - Hydroxylamine NH_2OH
  - Base analogs (e.g. BrdU)
  - Alkylating agents (e.g. *N*-ethyl-*N*-nitrosourea) These agents can mutate both replicating and non-replicating DNA. In contrast, a base analog can only mutate the DNA when the analog is incorporated in replicating the DNA. Each of these classes of chemical mutagens has certain effects that then lead to transitions, transversions, or deletions.

    - Agents that form DNA adducts (e.g. ochratoxin A metabolites)
    - DNA intercalating agents (e.g. ethidium bromide)
    - DNA crosslinkers
    - Oxidative damage
    - Nitrous acid converts amine groups on A and C to diazo groups, altering their hydrogen bonding patterns which leads to incorrect base pairing during replication.
- Radiation
    - Ultraviolet radiation (nonionizing radiation). Two nucleotide bases in DNA – cytosine and thymine – are most vulnerable to radiation that can change their properties. UV light can induce adjacent pyrimidine bases in a DNA strand to become covalently joined as a pyrimidine dimer. UV radiation, particularly longer-wave UVA, can also cause oxidative damage to DNA.
    - Ionizing radiation
- Viral infections

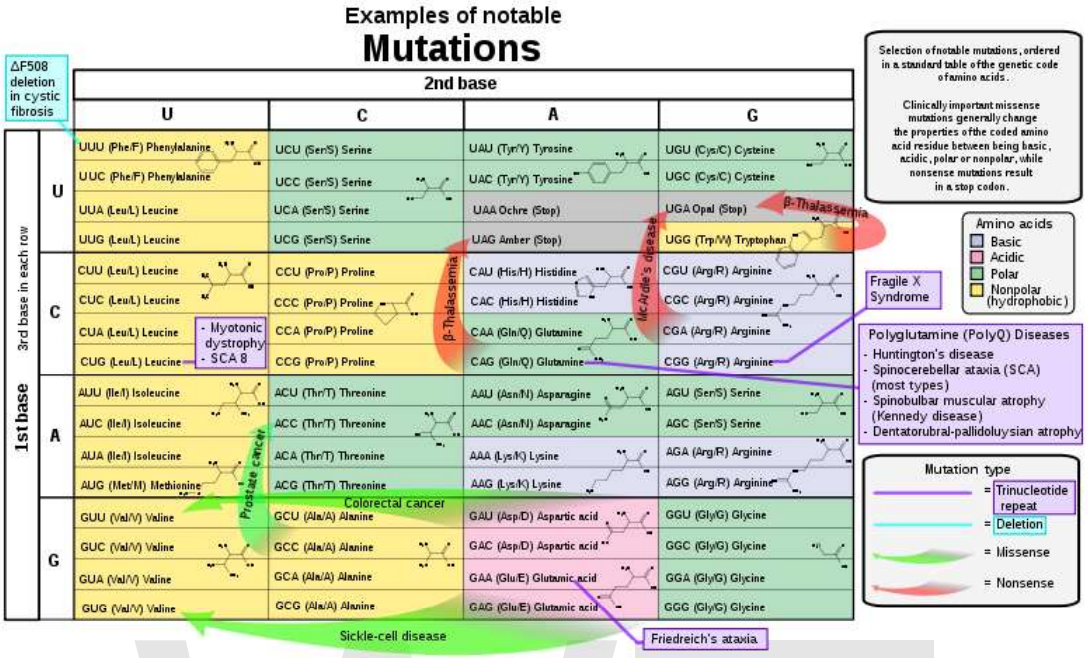
DNA has so-called hotspots, where mutations occur up to 100 times more frequently than the normal mutation rate. A hotspot can be at an unusual base, e.g., 5-methylcytosine.

Mutation rates also vary across species. Evolutionary biologists have theorized that higher mutation rates are beneficial in some situations, because they allow organisms to evolve and therefore adapt more quickly to their environments. For example, repeated exposure of bacteria to antibiotics, and selection of resistant mutants, can result in the selection of bacteria that have a much higher mutation rate than the original population (mutator strains).

Classification of mutation types



Illustrations of five types of chromosomal mutations.



Selection of disease-causing mutations, in a standard table of the genetic code of amino acids.

By effect on structure

The sequence of a gene can be altered in a number of ways. Gene mutations have varying effects on health depending on where they occur and whether they alter the function of essential proteins. Mutations in the structure of genes can be classified as:

- Small-scale mutations, such as those affecting a small gene in one or a few nucleotides, including:
    - **Point mutations**, often caused by chemicals or malfunction of DNA replication, exchange a single nucleotide for another. These changes are classified as transitions or transversions. Most common is the transition that exchanges a purine for a purine (A ↔ G) or a pyrimidine for a pyrimidine, (C ↔ T). A transition can be caused by nitrous acid, base mis-pairing, or mutagenic base analogs such as 5-bromo-2-deoxyuridine (BrdU). Less common is a transversion, which exchanges a purine for a pyrimidine or a pyrimidine for a purine (C/T ↔ A/G). An example of a transversion is adenine (A) being converted into a cytosine (C). A point mutation can be reversed by another point mutation, in which the nucleotide is changed back to its original state (true reversion) or by second-site reversion (a complementary mutation elsewhere that results in regained gene functionality). Point mutations that occur within the protein coding region of a gene may be classified into three kinds, depending upon what the erroneous codon codes for:
        - Silent mutations: which code for the same amino acid.

        - Missense mutations: which code for a different amino acid.
        - Nonsense mutations: which code for a stop and can truncate the protein.
    - **Insertions** add one or more extra nucleotides into the DNA. They are usually caused by transposable elements, or errors during replication of repeating elements (e.g. AT repeats). Insertions in the coding region of a gene may alter splicing of the mRNA (splice site mutation), or cause a shift in the reading frame (frameshift), both of which can significantly alter the gene product. Insertions can be reverted by excision of the transposable element.
    - **Deletions** remove one or more nucleotides from the DNA. Like insertions, these mutations can alter the reading frame of the gene. They are generally irreversible: though exactly the same sequence might theoretically be restored by an insertion, transposable elements able to revert a very short deletion (say 1–2 bases) in *any* location are either highly unlikely to exist or do not exist at all. Note that a deletion is not the exact opposite of an insertion: the former is quite random while the latter consists of a specific sequence inserting at locations that are not entirely random or even quite narrowly defined.
- Large-scale mutations in chromosomal structure, including:
    - **Amplifications** (or gene duplications) leading to multiple copies of all chromosomal regions, increasing the dosage of the genes located within them.
    - **Deletions** of large chromosomal regions, leading to loss of the genes within those regions.
    - Mutations whose effect is to juxtapose previously separate pieces of DNA, potentially bringing together separate genes to form functionally distinct fusion genes (e.g. bcr-abl). These include:
        - **Chromosomal translocations**: interchange of genetic parts from nonhomologous chromosomes.
        - **Interstitial deletions**: an intra-chromosomal deletion that removes a segment of DNA from a single chromosome, thereby apposing previously distant genes. For example, cells isolated from a human astrocytoma, a type of brain tumor, were found to have a chromosomal deletion removing sequences between the "fused in glioblastoma" (fig) gene and the receptor tyrosine kinase "ros", producing a fusion protein (FIG-ROS). The abnormal FIG-ROS fusion protein has constitutively active kinase activity that causes oncogenic transformation (a transformation from normal cells to cancer cells).
        - **Chromosomal inversions**: reversing the orientation of a chromosomal segment.
    - **Loss of heterozygosity**: loss of one allele, either by a deletion or recombination event, in an organism that previously had two different alleles.

By effect on function

- **Loss-of-function mutations** are the result of gene product having less or no function. When the allele has a complete loss of function (null allele) it is often called an **amorphic mutation**. Phenotypes associated with such mutations are most often recessive. Exceptions are when the organism is haploid, or when the reduced dosage of a normal gene product is not enough for a normal phenotype (this is called haploinsufficiency).
- **Gain-of-function mutations** change the gene product such that it gains a new and abnormal function. These mutations usually have dominant phenotypes. Often called a neomorphic mutation.
- **Dominant negative mutations** (also called **antimorphic mutations**) have an altered gene product that acts antagonistically to the wild-type allele. These mutations usually result in an altered molecular function (often inactive) and are characterised by a dominant or semi-dominant phenotype. In humans, Marfan syndrome is an example of a dominant negative mutation occurring in an autosomal dominant disease. In this condition, the defective glycoprotein product of the fibrillin gene (FBN1) antagonizes the product of the normal allele.
- **Lethal mutations** are mutations that lead to the death of the organisms which carry the mutations.
- A **back mutation** or **reversion** is a point mutation that restores the original sequence and hence the original phenotype.

By effect on fitness

In applied genetics it is usual to speak of mutations as either harmful or beneficial.

- A **harmful mutation** is a mutation that decreases the fitness of the organism.
- A **beneficial mutation** is a mutation that increases fitness of the organism, or which promotes traits that are desirable.

In theoretical population genetics, it is more usual to speak of such mutations as deleterious or advantageous. In the neutral theory of molecular evolution, genetic drift is the basis for most variation at the molecular level.

- A **neutral mutation** has no harmful or beneficial effect on the organism. Such mutations occur at a steady rate, forming the basis for the molecular clock.
- A **deleterious mutation** has a negative effect on the phenotype, and thus decreases the fitness of the organism.
- An **advantageous mutation** has a positive effect on the phenotype, and thus increases the fitness of the organism.
- A **nearly neutral mutation** is a mutation that may be slightly deleterious or advantageous, although most nearly neutral mutations are slightly deleterious.

In reality, viewing the fitness effects of mutations in these discrete categories is an oversimplification. Attempts have been made to infer the distribution of fitness effects

using mutagenesis experiments or theoretical models applied to molecular sequence data. However, the current distribution is still uncertain, and some aspects of the distribution likely vary between species.

By inheritance

- inheritable generic in pro-generic tissue or cells on path to be changed to gametes.
- non inheritable **somatic** (e.g., carcinogenic mutation)
- non inheritable post mortem aDNA mutation in decaying remains.

By pattern of inheritance The human genome contains two copies of each gene – a paternal and a maternal allele.

- A **heterozygous mutation** is a mutation of only one allele.
- A **homozygous mutation** is an identical mutation of both the paternal and maternal alleles.
- **Compound heterozygous** mutations or a **genetic compound** comprises two different mutations in the paternal and maternal alleles.
- A **wildtype** or **homozygous non-mutated** organism is one in which neither allele is mutated. (Just not a mutation)

By impact on protein sequence

- A **frameshift mutation** is a mutation caused by insertion or deletion of a number of nucleotides that is not evenly divisible by three from a DNA sequence. Due to the triplet nature of gene expression by codons, the insertion or deletion can disrupt the reading frame, or the grouping of the codons, resulting in a completely different translation from the original. The earlier in the sequence the deletion or insertion occurs, the more altered the protein produced is.

- A **nonsense mutation** is a point mutation in a sequence of DNA that results in a premature stop codon, or a *nonsense codon* in the transcribed mRNA, and possibly a truncated, and often nonfunctional protein product.

- **Missense mutations** or *nonsynonymous mutations* are types of point mutations where a single nucleotide is changed to cause substitution of a different amino acid. This in turn can render the resulting protein nonfunctional. Such mutations are responsible for diseases such as Epidermolysis bullosa, sickle-cell disease, and SOD1 mediated ALS (Boillée 2006, p. 39).

- A **neutral mutation** is a mutation that occurs in an amino acid codon which results in the use of a different, but chemically similar, amino acid. The similarity between the two is enough that little or no change is often rendered in the protein. For example, a change from AAA to AGA will encode arginine, a chemically similar molecule to the intended lysine.

- **Silent mutations** are mutations that do not result in a change to the amino acid sequence of a protein. They may occur in a region that does not code for a protein, or they may occur within a codon in a manner that does not alter the final amino acid sequence. The phrase *silent mutation* is often used interchangeably with the phrase *synonymous mutation*; however, synonymous mutations are a subcategory of the former, occurring only within exons. The name silent could be a misnomer. For example, a silent mutation in the exon/intron border may lead to alternative splicing by changing the splice site, thereby leading to a changed protein.

Special classes

- **Conditional mutation** is a mutation that has wild-type (or less severe) phenotype under certain "permissive" environmental conditions and a mutant phenotype under certain "restrictive" conditions. For example, a temperature-sensitive mutation can cause cell death at high temperature (restrictive condition), but might have no deleterious consequences at a lower temperature (permissive condition).

Nomenclature

Nomenclature of mutations specify the type of mutation and base or amino acid changes.

- Nucleotide substitution (e.g. 76A>T) - The number is the position of the nucleotide from the 5' end, the first letter represents the wild type nucleotide, and the second letter represents the nucleotide which replaced the wild type. In the given example, the adenine at the 76th position was replaced by a thymine.
    - If it becomes necessary to differentiate between mutations in genomic DNA, mitochondrial DNA, and RNA, a simple convention is used. For example, if the 100th base of a nucleotide sequence mutated from G to C, then it would be written as g.100G>C if the mutation occurred in genomic DNA, m.100G>C if the mutation occurred in mitochondrial DNA, or r.100g>c if the mutation occurred in RNA. Note that for mutations in RNA, the nucleotide code is written in lower case.
- Amino acid substitution (e.g. D111E) – The first letter is the one letter code of the wild type amino acid, the number is the position of the amino acid from the N terminus, and the second letter is the one letter code of the amino acid present in the mutation. Nonsense mutations are represented with an X for the second amino acid (e.g. D111X).
- Amino acid deletion (e.g. ΔF508) – The Greek letter Δ (delta) indicates a deletion. The letter refers to the amino acid present in the wild type and the number is the position from the N terminus of the amino acid were it to be present as in the wild type.

Harmful mutations

Changes in DNA caused by mutation can cause errors in protein sequence, creating partially or completely non-functional proteins. To function correctly, each cell depends on thousands of proteins to function in the right places at the right times. When a mutation alters a protein that plays a critical role in the body, a medical condition can result. A condition caused by mutations in one or more genes is called a genetic disorder. Some mutations alter a gene's DNA base sequence but do not change the function of the protein made by the gene. Studies of the fly *Drosophila melanogaster* suggest that if a mutation does change a protein, this will probably be harmful, with about 70 percent of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial. However, studies in yeast have shown that only 7% of mutations that are not in genes are harmful.

If a mutation is present in a germ cell, it can give rise to offspring that carries the mutation in all of its cells. This is the case in hereditary diseases. On the other hand, a mutation may occur in a somatic cell of an organism. Such mutations will be present in all descendants of this cell within the same organism, and certain mutations can cause the cell to become malignant, and thus cause cancer.

Often, gene mutations that could cause a genetic disorder are repaired by the DNA repair system of the cell. Each cell has a number of pathways through which enzymes recognize and repair mistakes in DNA. Because DNA can be damaged or mutated in many ways, the process of DNA repair is an important way in which the body protects itself from disease.

Beneficial mutations

Although most mutations that change protein sequences are neutral or harmful, some mutations have a positive effect on an organism. In this case, the mutation may enable the mutant organism to withstand particular environmental stresses better than wild-type organisms, or reproduce more quickly. In these cases a mutation will tend to become more common in a population through natural selection.

For example, a specific 32 base pair deletion in human CCR5 (CCR5-Δ32) confers HIV resistance to homozygotes and delays AIDS onset in heterozygotes. The CCR5 mutation is more common in those of European descent. One possible explanation of the etiology of the relatively high frequency of CCR5-Δ32 in the European population is that it conferred resistance to the bubonic plague in mid-14th century Europe. People with this mutation were more likely to survive infection; thus its frequency in the population increased. This theory could explain why this mutation is not found in southern Africa, where the bubonic plague never reached. A newer theory suggests that the selective pressure on the CCR5 Delta 32 mutation was caused by smallpox instead of the bubonic plague.

Another example, is Sickle cell disease which is a blood disorder in which the body produces an abnormal type of the oxygen-carrying substance hemoglobin in the red blood cells. One-third of all indigenous inhabitants of Sub-Saharan Africa carry the gene, because in areas where malaria is common, there is a survival value in carrying only a single sickle-cell gene (sickle cell trait). Those with only one of the two alleles of the sickle-cell disease are more resistant to malaria, since the infestation of the malaria plasmodium is halted by the sickling of the cells which it infests.

Prion mutation

Prions are proteins and do not contain genetic material. However, prion replication has been shown to be subject to mutation and natural selection just like other forms of replication.

Mutations are permanent, transmissible changes to the genetic material (usually DNA or RNA) of a cell. Mutations can be caused by copying errors in the genetic material during cell division and by exposure to radiation, chemicals, or viruses, or can occur deliberately under cellular control during the processes such as meiosis or hypermutation. Mutations are considered the driving force of evolution, where less favorable (or *deleterious*) mutations are removed from the gene pool by natural selection, while more favorable (or *beneficial*) ones tend to accumulate. Neutral mutations do not affect the organism's chances of survival in its natural environment and can accumulate over time, which might result in what is known as punctuated equilibrium; the modern interpretation of classic evolutionary theory.

Causes of change in allele frequency

There are three known processes that affect the survival of a characteristic; or, more specifically, the frequency of an allele (variant of a gene):

- Genetic drift describes changes in gene frequency that cannot be ascribed to selective pressures, but are due instead to events that are unrelated to inherited traits. This is especially important in small mating populations, which simply cannot have enough offspring to maintain the same gene distribution as the parental generation.
- Gene flow or Migration: or gene admixture is the only one of the agents that makes populations closer genetically while building larger gene pools.
- Selection, in particular natural selection produced by differential mortality and fertility. Differential mortality is the survival rate of individuals before their reproductive age. If they survive, they are then selected further by differential fertility – that is, their total genetic contribution to the next generation. In this way, the alleles that these surviving individuals contribute to the gene pool will increase the frequency of those alleles. Sexual selection, the attraction between mates that results from two genes, one for a feature and the other determining a preference for that feature, is also very important.

Molecular study of phylogeny

Molecular systematics is a product of the traditional field of systematics and molecular genetics. It is the process of using data on the molecular constitution of biological organisms' DNA, RNA, or both, in order to resolve questions in systematics, i.e. about their correct scientific classification or taxonomy from the point of view of evolutionary biology.

Molecular systematics has been made possible by the availability of techniques for DNA sequencing, which allow the determination of the exact sequence of nucleotides or *bases* in either DNA or RNA. At present it is still a long and expensive process to sequence the entire genome of an organism, and this has been done for only a few species. However, it is quite feasible to determine the sequence of a defined area of a particular chromosome. Typical molecular systematic analyses require the sequencing of around 1000 base pairs.

The driving forces of evolution

Depending on the relative importance assigned to the various forces of evolution, three perspectives provide evolutionary explanations for molecular evolution.

While recognizing the importance of random drift for silent mutations, **selectionists hypotheses** argue that balancing and positive selection are the driving forces of molecular evolution. Those hypotheses are often based on the broader view called panselectionism, the idea that selection is the only force strong enough to explain evolution, relaying random drift and mutations to minor roles.

Neutralists hypotheses emphasize the importance of mutation, purifying selection and random genetic drift. The introduction of the neutral theory by Kimura, quickly followed by King and Jukes' own findings, lead to a fierce debate about the relevance of neodarwinism at the molecular level. The Neutral theory of molecular evolution states that most mutations are deleterious and quickly removed by natural selection, but of the remaining ones, the vast majority are neutral with respect to fitness while the amount of advantageous mutations is vanishingly small. The fate of neutral mutations are governed by genetic drift, and contribute to both nucleotide polymorphism and fixed differences between species.

Mutationists hypotheses emphasize random drift and biases in mutation patterns. Sueoka was the first to propose a modern mutationist view. He proposed that the variation in GC content was not the result of positive selection, but a consequence of the GC mutational pressure.

Related fields

An important area within the study of molecular evolution is the use of molecular data to determine the correct biological classification of organisms. This is called molecular systematics or molecular phylogenetics.

Tools and concepts developed in the study of molecular evolution are now commonly used for comparative genomics and molecular genetics, while the influx of new data from these fields has been spurring advancement in molecular evolution.

Key researchers in molecular evolution

Some researchers who have made key contributions to the development of the field:

- Motoo Kimura — Neutral theory
- Masatoshi Nei — Adaptive evolution
- Walter M. Fitch — Phylogenetic reconstruction
- Walter Gilbert — RNA world
- Joe Felsenstein — Phylogenetic methods
- Susumu Ohno — Gene duplication
- John H. Gillespie — Mathematics of adaptation

Journals and societies

Journals dedicated to molecular evolution include *Molecular Biology and Evolution*, *Journal of Molecular Evolution*, and *Molecular Phylogenetics and Evolution*. Research in molecular evolution is also published in journals of genetics, molecular biology, genomics, systematics, or evolutionary biology. The Society for Molecular Biology and Evolution publishes the journal "Molecular Biology and Evolution" and holds an annual international meeting.

Chapter- 9

Phylogenetics

In biology, **phylogenetics** is the study of evolutionary relatedness among various groups of organisms (for example, species or populations), which is discovered through molecular sequencing data and morphological data matrices. The term *phylogenetics* is of Greek origin from the terms *phyle/phylon* (φυλή/φῦλον), meaning "tribe, race", and *genetikos* (γενετικός), meaning "relative to birth" from *genesis* (γένεσις, "birth"). Taxonomy, the classification, identification, and naming of organisms, has been richly informed by phylogenetics but remains methodologically and logically distinct.

The fields overlap however in the science of phylogenetic systematics – often called "cladism" or "cladistics" – where only phylogenetic trees are used to delimit taxa, which represent groups of lineage-connected individuals. In biological systematics as a whole, phylogenetic analyses have become essential in researching the evolutionary tree of life.

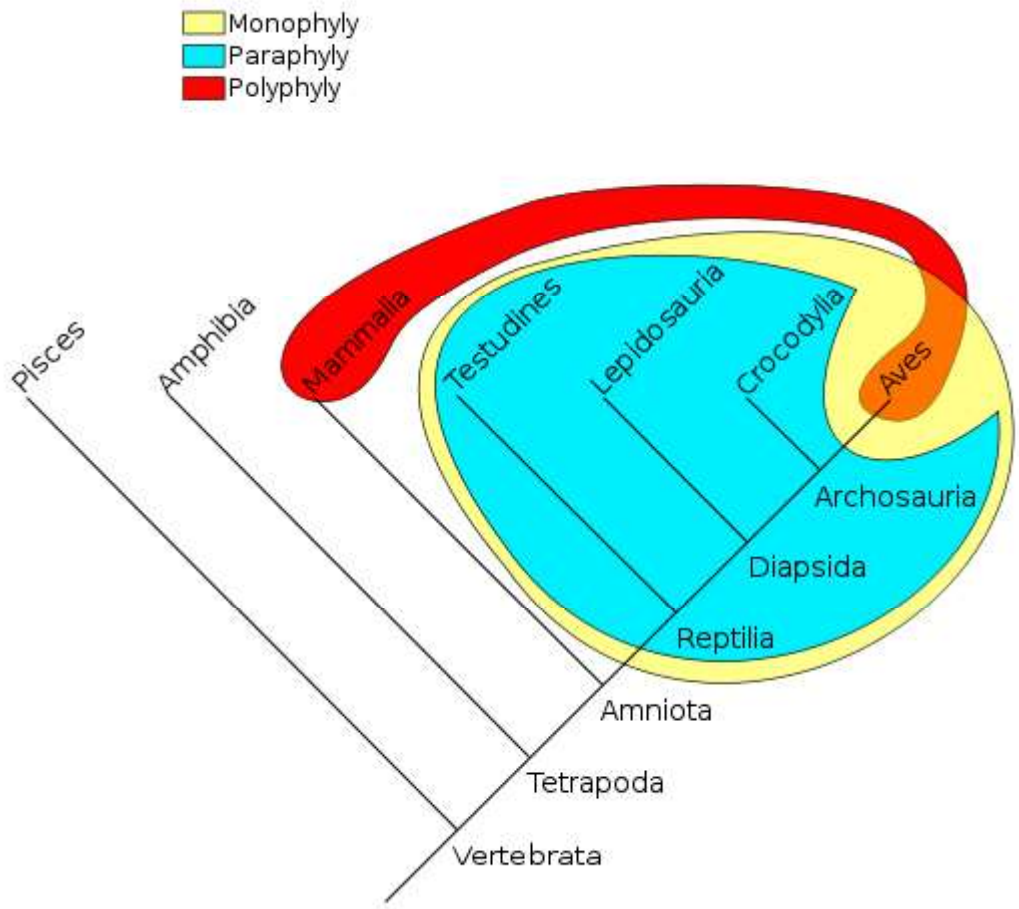
Construction of a phylogenetic tree

Evolution is regarded as a branching process, whereby populations are altered over time and may speciate into separate branches, hybridize together, or terminate by extinction. This may be visualized in a phylogenetic tree.

The problem posed by phylogenetics is that genetic data are only available for living taxa, and the fossil records (osteometric data) contains less data and more-ambiguous morphological characters. A phylogenetic tree represents a hypothesis of the order in which evolutionary events are assumed to have occurred.

Cladistics is the current method of choice to infer phylogenetic trees. The most commonly-used methods to infer phylogenies include parsimony, maximum likelihood, and MCMC-based Bayesian inference. Phenetics, popular in the mid-20th century but now largely obsolete, uses distance matrix-based methods to construct trees based on overall similarity, which is often assumed to approximate phylogenetic relationships. All methods depend upon an implicit or explicit mathematical model describing the evolution of characters observed in the species included, and are usually used for molecular phylogeny, wherein the characters are aligned nucleotide or amino acid sequences.

Grouping of organisms



Phylogenetic groups, or *taxa*, can be monophyletic, paraphyletic, or polyphyletic.

There are some terms that describe the nature of a grouping in such trees. For instance, all birds and reptiles are believed to have descended from a single common ancestor, so this taxonomic grouping (yellow in the diagram below) is called monophyletic. "Modern reptile" (cyan in the diagram) is a grouping that contains a common ancestor, but does not contain all descendants of that ancestor (birds are excluded).

This is an example of a paraphyletic group. A grouping such as warm-blooded animals would include only mammals and birds (red/orange in the diagram) and is called polyphyletic because the members of this grouping do not include the most recent common ancestor.

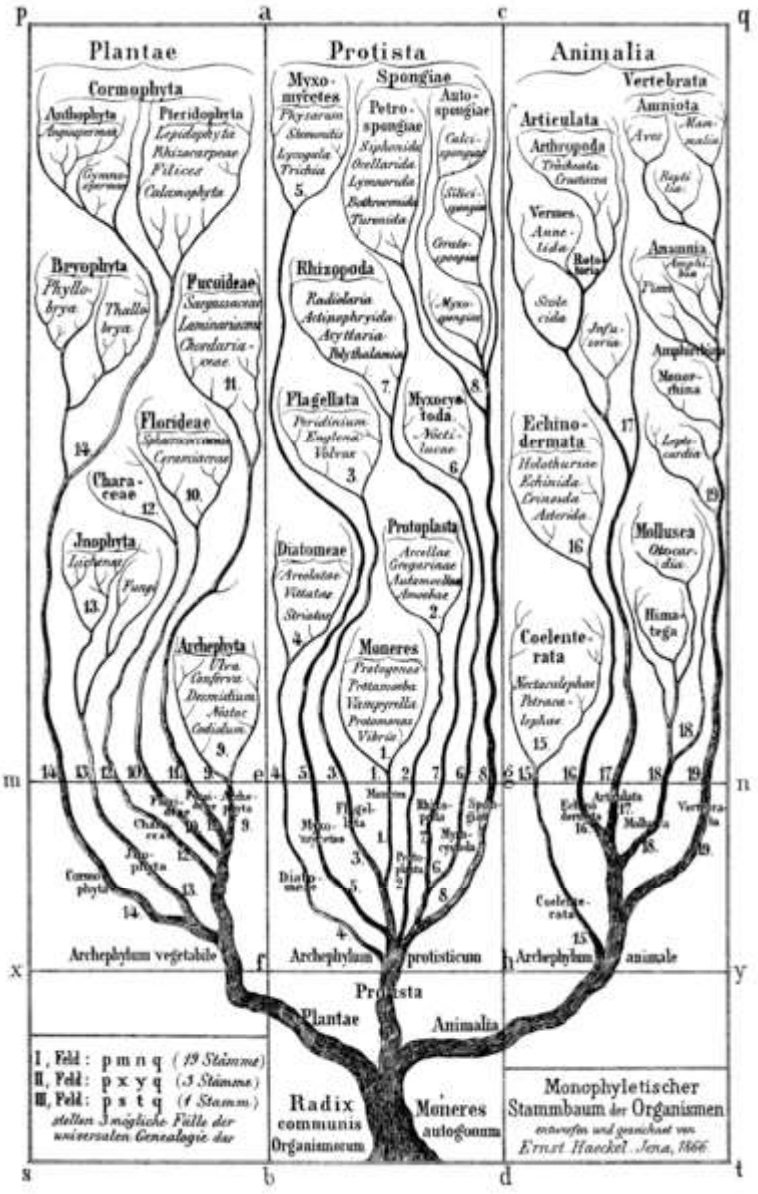
Molecular phylogenetics

The evolutionary connections between organisms are represented graphically through phylogenetic trees. Due to the fact that evolution takes place over long periods of time

that cannot be observed directly, biologists must reconstruct phylogenies by inferring the evolutionary relationships among present-day organisms. Fossils can aid with the reconstruction of phylogenies; however, fossil records are often too poor to be of good help. Therefore, biologists tend to be restricted with analysing present-day organisms to identify their evolutionary relationships. Phylogenetic relationships in the past were reconstructed by looking at phenotypes, often anatomical characteristics. Today, molecular data, which includes protein and DNA sequences, are used to construct phylogenetic trees.

The overall goal of National Science Foundation's Assembling the Tree of Life activity (AToL) is to resolve evolutionary relationships for large groups of organisms throughout the history of life, with the research often involving large teams working across institutions and disciplines. Investigators are typically supported for projects in data acquisition, analysis, algorithm development and dissemination in computational phylogenetics and phyloinformatics. For example, RedToL aims at reconstructing the Red Algal Tree of Life.

Ernst Haeckel's recapitulation theory



Genealogical tree suggested by Haeckel (1866)

During the late 19th century, Ernst Haeckel's recapitulation theory, or biogenetic law, was widely accepted. This theory was often expressed as "ontogeny recapitulates phylogeny", i.e. the development of an organism exactly mirrors the evolutionary development of the species. Haeckel's early version of this hypothesis [that the embryo mirrors *adult* evolutionary ancestors] has since been rejected, and the hypothesis amended as the embryo's development mirroring *embryos* of its evolutionary ancestors. He was accused by five professors of falsifying his images of embryos. Most modern biologists recognize numerous connections between ontogeny and phylogeny, explain

them using evolutionary theory, or view them as supporting evidence for that theory. Donald I. Williamson suggested that larvae and embryos represented adults in other taxa that have been transferred by hybridization (the larval transfer theory). However, Williamson's views do not represent mainstream thought in molecular biology, and there is a significant body of evidence against the larval transfer theory.

Gene transfer

In general, organisms can inherit genes in two ways: vertical gene transfer and horizontal gene transfer. Vertical gene transfer is the passage of genes from parent to offspring, and horizontal gene transfer or lateral gene transfer occurs when genes jump between unrelated organisms, a common phenomenon in prokaryotes.

Horizontal gene transfer has complicated the determination of phylogenies of organisms, and inconsistencies in phylogeny have been reported among specific groups of organisms depending on the genes used to construct evolutionary trees.

Carl Woese came up with the three-domain theory of life (eubacteria, archaea and eukaryotes) based on his discovery that the genes encoding ribosomal RNA are ancient and distributed over all lineages of life with little or no horizontal gene transfer. Therefore, rRNAs are commonly recommended as molecular clocks for reconstructing phylogenies.

This has been particularly useful for the phylogeny of microorganisms, to which the species concept does not apply and which are too morphologically simple to be classified based on phenotypic traits.

Taxon sampling and phylogenetic signal

Owing to the development of advanced sequencing techniques in molecular biology, it has become feasible to gather large amounts of data (DNA or amino acid sequences) to infer phylogenetic hypotheses. For example, it is not rare to find studies with character matrices based on whole mitochondrial genomes (~16,000 nucleotides, in many animals). However, it has been proposed that it is more important to increase the number of taxa in the matrix than to increase the number of characters, because the more taxa the more robust is the resulting phylogenetic tree.

This may be partly due to the breaking up of long branches. It has been argued that this is an important reason to incorporate data from fossils into phylogenies where possible. Of course, phylogenetic data that include fossil taxa are generally based on morphology, rather than DNA data. Using simulations, Derrick Zwickl and David Hillis found that increasing taxon sampling in phylogenetic inference has a positive effect on the accuracy of phylogenetic analyses.

Another important factor that affects the accuracy of tree reconstruction is whether the data analyzed actually contain a useful phylogenetic signal, a term that is used generally to denote whether related organisms tend to resemble each other with respect to their genetic material or phenotypic traits. Ultimately, however, there is no way to measure whether a particular phylogenetic hypothesis is accurate or not, unless the "true" relationships among the taxa being examined are already known. The best result an empirical systematist can hope to attain is a tree with branches well-supported by the available evidence.

Importance of missing data

In general, the more data that is available when constructing a tree, the more accurate and reliable the resulting tree will be. Missing data is no less detrimental than simply having less data, although its impact is greatest when most of the missing data is in a small number of taxa. The fewer characters that have missing data, the better; concentrating the missing data across a small number of character states produces a more robust tree.

Role of fossils

Because many morphological characters involve embryological or soft-tissue characters that cannot be fossilized, and the interpretation of fossils is more ambiguous than living taxa, it is sometimes difficult to incorporate fossil data into phylogenies. However, despite these limitations, the inclusion of fossils is invaluable, as they can provide information in sparse areas of trees, breaking up long branches and constraining intermediate character states; thus, fossil taxa contribute as much to tree resolution as modern taxa.

Molecular phylogenies can reveal rates of diversification, but in order to track rates of origination, extinction and patterns in diversification, fossil data must be incorporated. Molecular techniques assume a constant rate of diversification, which is rarely likely to be true; in some (but by no means all) cases, the assumptions inherent in interpreting the fossil record (e.g. a complete and unbiased record) are closer to being true than the assumption of a constant rate, making fossil insights more accurate than molecular reconstructions.

Homoplasy weighting

Certain characters are more likely to be evolved convergently than others; logically, such characters should be given less weight in the reconstruction of a tree. Unfortunately the only objective way to determine convergence is by the construction of a tree – a somewhat circular method. Even so, weighting homoplasious characters does indeed lead to better-supported trees. Further refinement can be brought by weighting changes in one direction higher than changes in another; for instance, the presence of thoracic wings almost guarantees placement among the pterygote insects, although because wings are often lost secondarily, their absence does not exclude a taxon from the group.