A Universal Phylogenetic Tree
Susan Offner

This paper presents a universal phylogenetic tree (Figure 1) suitable for use in high school and college biology classrooms. It is made possible by whole genome sequencing techniques that have been available since the mid 1990s. While it is likely that many of the details of the tree will change in the years ahead as more genome sequences become available, the basic ideas behind the tree are powerful and valid. This tree enables students to understand the entire sweep of the history of life. It vividly illustrates the antiquity of life, and the fact that all life is related, even if the relationship goes back about 3.5 billion years to a universal ancestor. It also illustrates the important principle that classification should reflect evolutionary relationships. The more recently two organisms had a common ancestor, the more closely they should be classified. And the best way to measure how recently two organisms had a common ancestor is to see how alike their DNA is. This turns the classification unit, usually a fairly dreary low point of the year, into an exciting high point since the classification students are learning is really the story of the history of life.

Small Subunit Ribosomal RNA Is the Basis of the Tree

Many lines of evidence have led to this phylogenetic tree. Most influential was the work of Carl Woese, a molecular biologist at the University of Illinois at Urbana-Champaign who has been comparing the small subunit ribosomal RNAs of a wide variety of organisms since the 1970s. Ribosomes are small organelles where the cell makes protein. Ribosomes themselves are physically made up of RNA and protein. Each ribosome has one large subunit and one small subunit. There is only one RNA molecule in the small subunit of the ribosome. This is called “small subunit ribosomal RNA,” abbreviated SSUrRNA. It is actually a fairly large molecule, containing more than 1500 bases in prokaryotes and about 2000 bases in eukaryotes. Notice that since it is RNA, it contains bases, and not base pairs. Since the SSUrRNA is read from genes in the DNA, the order of bases in the SSUrRNA is determined by the organism’s DNA.

After studying the SSUrRNA of hundreds of organisms, Woese found that they fell into three categories. In a landmark paper (Woese 1987), he named these bacteria, archaea and eukarya, and called them the three domains of life. There was much skepticism about this until 1996 when the whole genome of an archaean, Methanococcus jannaschii, was sequenced (Bult et al. 1996). An astonishing 56% of its genes were totally new to biology, found in neither bacteria nor eukarya! This gave important confirmation to Woese’s contention that archaeans belonged in a separate domain since they have not had a common ancestor with either bacteria or eukarya for a very long time.

While small subunit ribosomal RNA is the basis for the large outline of the tree, ribosomal RNA mutates too slowly to be useful for divergences that occurred relatively recently. Therefore, many of the details of the crown eukaryotes (i.e. in the 1.4 billion years after the appearance of meiosis), among others, are based on other genes in which mutations become fixed more rapidly. Note that mutations probably occur at random along the length of DNA. However, if the mutation is in a gene, such as ribosomal RNA that codes for a product in which virtually any change would be detrimental to the organism, then the individual will die and the mutation will not appear in future generations. We say that this mutation has not become fixed in a population. Since mutations in SSUrRNA become fixed very slowly, SSUrRNA is useful for studying ancient branchings of the tree, but not recent ones. By contrast, there is a great variation in hemoglobin genes among mammals. These mutations have become fixed in populations, because it is advantageous for mammals of different body size and living in different habitats to have hemoglobins with different affinities for oxygen. Thus, differences in hemoglobin genes would be useful in determining a phylogenetic tree of mammals, but not in determining a phylogenetic tree going back 3.5 billion years.

Three Domains Are More Accurate than Five Kingdoms

How, then, do three domains compare to five, or six, kingdoms? The five kingdom system was
Figure 1. A universal phylogenetic tree.

© 2000 Susan Offner
All Rights Reserved

Ga = Billion years ago
Ma = Million years ago
developed in the early 1960s by Robert H. Whittaker, an ecologist at Cornell. He used criteria that had been used to classify organisms for two centuries, since Linnaeus: anatomy, morphology, embryology and cellular structure. At the time, DNA sequencing was only a distant dream. The fact that, in nearly all cases, classification by DNA sequencing confirms the centuries of work of thousands of taxonomists, is a tribute to the intelligence, integrity and care these people brought to their work using the best techniques available to them in their time.

However, there are two serious problems with the five kingdom system. First, five kingdoms do not show the evolutionary relationships among the kingdoms. The system describes each one, but is mute on the subject of who evolved from whom. Second, two of the five kingdoms, the protists and monerans, are not proper monophyletic categories.

The cornerstone of a phylogenetic tree is the concept of a clade. A clade is a lineage or a monophyletic group. It is a group of organisms consisting of a common ancestor and all the descendants of the common ancestor. If you look at the five kingdoms, only plants, animals and fungi qualify as clades. Protists are not a clade because the common ancestor of protists is also the common ancestor of plants, animals and fungi. Protists consist of all eukarya except for plants. All eukarya and fungi. Similarly, monerans consist of the domains bacteria and archaea and are not a clade since the common ancestor of monerans is the universal ancestor, which is also the common ancestor of the eukarya. Two of the three domains, bacteria and archaea, are organisms with prokaryotic cells. Therefore, under the new three domain system of classification, the mere presence of prokaryotic cells is not enough to place an organism in one of the domains. One advantage of using the three domain system is that it gives some perspective to the relative importance of different organisms. Prokaryotes, the bacteria and archaea, got here first and still dominate the planet. Animals, our favorite group, are just a little twig off in a recently evolved corner of eukaryotes, as are plants and fungi. Thus, three of the five kingdoms are really recently evolved groups of the domain eukarya.

**History of Life**

What, then, were some of the major events in the history of life? First, we have the origin of life, between 3.7 and 4 billion years ago. There are no cellular fossils available from this time, and it is impossible to infer properties of organisms that lived before the universal ancestor by studying extant organisms. However, there are a few molecular clues. Most of us think of fossils in terms of physical parts or imprints of animals. However, many organisms leave stable telltale molecules after their demise (Knoll 1999). These molecules, usually lipids, that are known to be made only by organisms or only by particular kinds of organisms, are known as **molecular fossils**. They are important in studying very ancient life since the single-celled prokaryotes of this time did not often leave cellular fossils. The earliest molecular fossils go back between 3.7 and 4 billion years. It is therefore inferred that life originated during this time. It is also likely that the first organisms were chemolithothrophic, using inorganic materials without light or photosynthesis. They may have obtained their energy in reactions that split H2S rather than H2O, thus producing sulfur rather than molecular oxygen. It takes considerably less energy to split H2S than to split H2O, so even though splitting H2O (as is done in oxygen-producing photosynthesis) yields more energy, it was not possible to split water until proteins evolved that permitted the cell to perform this energetically more difficult reaction. As we will see later, oxygen-producing photosynthesis evolved only once in the history of life, with the appearance of the cyanobacteria. Exactly when this happened is an important area of research.

**Characteristics of the Universal Ancestor**

The characteristics of the universal ancestor are a fascinating area of speculation. Woese (1998) sees the universal ancestor not as a single organism, or even a single species, but as a community of organisms with inaccurate DNA duplicating mechanisms (there were no repair enzymes yet) and inaccurate protein translation. Therefore, he says, genes and proteins had to be short since a long protein would have too many errors in it to be functional. He also hypothesizes that cells carrying many copies of each gene (to compensate for these inaccuracies) share their genes freely with each other. Other properties of the universal ancestor can be inferred by looking at properties shared by all three domains: DNA as genetic material, proteins made of L-amino acids, ribosomes for making proteins, the genetic code, and transfer RNAs, among others. Gunter Wachtershauser (1994) suggests that the universal ancestor was chemosynthetic, fixing CO2 and running the Krebs Cycle backwards while splitting H2S for energy. Some bacteria today have this metabolism. The earliest branches of the phylogenetic tree, the Aquificales, Thermotogales, and many early (and archaeans) are all hyperthermophiles, living at very high temperatures. This strongly suggests that the universal ancestor lived at high temperatures as well.

**Rooting the Tree**

The universal ancestor is inferred to have lived between 3.3 and 3.5 billion years ago. There was
then a major division between the bacteria on one hand and the archaea and eukarya on the other. This division is supported by analysis of a protein called elongation factor. This protein is about 300 amino acids long that interacts with ribosomes to add amino acids to the growing protein. It is the protein that pushes the incoming amino acid against the growing polypeptide chain and also makes sure that the codon on the messenger RNA and anti-codon of the transfer RNA match. All organisms have two genes for the elongation factor protein. These are called EF-Tu/1 and EF-G/2. Each gene codes for one protein. The two proteins are similar to each other in their amino acid sequence and function as a dimer in the cell. Since all organisms have these two genes, it is inferred that the genes came from a single gene that duplicated before the time of the universal ancestor. Analysis of both of these genes gives the universal tree shown in Figure 1. This is the evidence that the root of the tree is between the bacteria and archaea.

After the bacteria split from the archaea and eukarya, the archaea and eukarya split from each other. Before this second split, it is likely that some kind of histone proteins evolved to support DNA since these are found in archaea and eukarya, but not in bacteria.

**The Endosymbiont Hypothesis**

It has been suggested over the past years that some eukaryotic organelles evolved when a bacterium was engulfed by a eukaryotic cell. This was thought to be particularly true for mitochondria and chloroplasts, both of which have their own DNA and small, prokaryotic-like ribosomes. Whole genome sequencing has confirmed this, showing that mitochondria came from a group of alpha purple bacteria and chloroplasts came from cyanobacteria. This is based on the DNA sequences of the genes in mitochondria and chloroplasts that are similar to the DNA sequences of the bacteria from which they are derived.

**The Cytoskeleton May Have Been the Key Innovation in the Eukaryotic Cell**

The split between archaea and eukarya is marked by milestones that led to what we now know as the eukaryotic cell. While we think of the nucleus as the defining characteristic of the eukaryotic cell, I have listed the development of the cytoskeleton as the first event. This is because the presence of a cytoskeleton (tubulin, actin, etc.) makes endocytosis possible. The nucleus is probably the result of endocytosis. Mitchell Sogin (1994) offers a fascinating speculation that the original eukaryotic cell came about when a primitive cell that had evolved a cytoskeleton engulfed another primitive cell that had a well-developed chromosome, and used the genome of this engulfed organism as its nucleus. It is likely that the engulfed organism was an archaean since both archaea and eukarya have histone proteins associated with their chromosomes. Also characteristic of eukarya, as shown in Figure 1, are membrane-bound organelles, linear chromosomes with telomeres (as opposed to circular chromosomes in bacteria and archaea), and cilia and flagella made of microtubules in a “9 + 2” arrangement.

**Origin of Mitochondria**

The earliest eukarya, the diplomonads and trichomonads, do not contain mitochondria. Therefore, it is inferred that mitochondria entered a eukaryotic cell by endosymbiosis after these two lineages split off, and that all remaining eukarya evolved from descendants of that event. DNA analysis has shown that a branch of purple bacteria known as the alpha subdivision of the purple bacteria is the source of mitochondria.

**Crown Eukaryotes Follow the Appearance of Meiosis**

Looking at the eukarya domain, some startling things are apparent. First, about 1.4 billion years ago, there was a tremendous diversification of eukaryotes. Eukaryotic clades appearing after this time are known as crown eukaryotes. As can be seen from the phylogenetic tree, this diversification occurred after the appearance of meiosis. We can speculate on reasons, but it is likely that the ability to easily shuffle genes in offspring confers a significant survival advantage to the organisms possessing it. Meiosis makes it possible to easily do two things that seem to confer an enormous survival advantage on organisms:

1. A gene that makes an organism better adapted to its environment can be easily separated from a gene that makes the organism less adapted to its environment.
2. Two genes, each of which makes the organism better adapted to its environment, that arise in different individuals can be combined in one of their descendants.

Thus, it is not surprising that a large adaptive radiation should have followed the appearance of meiosis. Looking at the crown eukaryotes offers further surprises. Animals are most closely related to fungi, having shared a common ancestor between 650 and 900 million years ago. By contrast, we last had a common ancestor with plants between 1 and 1.2 billion years ago. Plants branched off from a clade of green algae about 450 million years ago, making them relative newcomers to the planet.
Further back in the crown eukaryotes are the alveolates. These consist of ciliates, including the familiar paramecium, stentor and blepharisma we use in class; the dinoflagellates, including the red tide organism, and the apicomplexans that include the malaria plasmodium. Closely related to the alveolates are the stramenopiles consisting of diatoms, brown algae (kelp), and golden brown algae.

Among the non-crown eukaryotes, the flagellates include the euglena. These are closely related to the trypanosomes, a group that includes the organism which causes sleeping sickness. The diplomonads are, unfortunately, well known to hikers who thoughtlessly drink the water in a crystal clear mountain spring and become sick with Giardia, a diplomonad.

At this writing, the details of the crown eukaryote clade are still very much in flux. The adaptive radiation that followed the origin of meiosis is real. What is in question are details such as whether flagellates should be included in the crown eukaryote clade. This is an active area of research. Also unknown, but interesting, are the origins of spindle formation, mitosis, paired chromosomes and recombination.

Origin of Animals

The origin of animals is shown to have occurred between 900 and 650 million years ago. There is a disagreement in the scientific community about the exact date (Conway-Morris 2000). The oldest animal fossils are 650 million years old. Paleontologists (i.e. fossil collectors) contend that the fossil record is reasonably complete, and therefore that the animal and fungus lineages diverged from each other at this time. On the other hand, the molecular chronologists say that the differences between animal and fungus DNA point to a divergence around 900 million years ago. They say that the paleontologists need to go back and look for animal fossils in the 650 to 900 million year range. The paleontologists disagree with this. They say that the molecular dates are too old. The molecular dates assume that mutations in DNA become fixed in the population at a constant, measurable rate. (Everybody agrees that they occur at a constant rate.) However, the paleontologists say that during an adaptive radiation, when new species are rapidly evolving, it is likely that many mutations will become fixed quickly in the newly emerging species. Thus, a calculation that assumes they become fixed at a constant rate would give too old a divergence date. I personally think that the paleontologists are right, but have listed both dates in the phylogenetic tree pending the resolution of the question in the research community.

Domain Archaea

The archaeans are the least familiar, and in some ways the most fascinating, of the domains. Many, but not all, archaeans live under extreme conditions, making them exciting subjects for the classroom and excellent examples to use to illustrate the tremendous diversity of life. These organisms were not discovered until 1977 and are divided into two major clades, the euryarchaeota and the crenarchaeota. The euryarchaeota are methane producers. Methanococcus jan纳斯ii, the first archaean to be sequenced, is an euryarchaeotan. It was obtained by the submarine Alvin from a hydrothermal vent deep in the Pacific Ocean (Morell 1996). It lives at 200 atmospheres of pressure, at a depth of 2.6 km; it is an obligate anaerobe with an optimal temperature of 85 degrees Celsius. It dies when the temperature falls below 45 degrees Celsius. And it has an unusual chemolithotrophic metabolism. It uses carbon dioxide, nitrogen and hydrogen and produces methane. This is autotrophy without light or photosynthesis. This concept will become more important in our teaching since it is likely that the first organisms were chemolithoautotrophs. The euryarchaeota also include many species of halophiles, organisms that live in concentrated salt solutions.

The crenarchaeota oxidize H₂ with sulfur being the acceptor. Thus, they produce H₂S. The most thermophilic organism known is a crenarchaeote. This is Pyrolobus fumarius, an organism that lives at an astonishing 113 degrees Celsius. There are also many crenarchaeota that live at cold temperatures, but none of these has been cultivated yet.

Origin of Oxygen-Producing Photosynthesis

A significant event in the history of the domain bacteria was the development of oxygen-producing photosynthesis, the familiar kind that occurs in green plants and produces all the oxygen in the atmosphere. This tremendous innovation evolved only once, in the ancestor of today’s cyanobacteria. The date of the first oxygen-producing photosynthesis is not known with certainty. Molecular fossils (molecules produced only by cyanobacteria) indicating the presence of cyanobacteria have been recently found in rocks 2.7 billion years old (Brocks et al. 1999; Knoll 1999), meaning that oxygen-producing photosynthesis is at least that old. At the other end, Schopf (1994) thinks it is possible that some cellular fossils 3.5 billion years old are cyanobacteria. When oxygen-producing photosynthesis evolved is an important area of research. Interestingly, the oxygen level in the atmosphere did not increase substantially until about 2 billion years ago (Kerr 1999) when it reached approximately 10% of the current level. Before then, the oxygen level was about one ten-thousandth of the present level. What happened to all the oxygen from photosynthesis before this? The most likely speculation was that it oxidized the enormous iron
deposits dissolved in the ocean, forming banded iron formations (described as the “Earth rusting” by Schopf in a 1997 lecture sponsored by the Wright Center at Boston’s Museum of Science).

The evolution of oxygen-producing photosynthesis was one of the most important things to happen in the history of life. First, it unleashed a highly efficient source of energy for use by organisms—the ability to split water. Second, by eventually leading to an accumulation of oxygen in the atmosphere, it made aerobic life as we know it today possible. Third, the eventual accumulation of oxygen in the atmosphere was a highly toxic event that killed off much of the early life on Earth, organisms that were obligate anaerobes. Some descendants of these organisms survive today, such as the organisms that cause syphilis, tetanus and botulism, but they are confined to marginal, anaerobic habitats.

**Origin of Chloroplasts**

The appearance of chloroplasts in eukaryotic cells was more complicated than the appearance of mitochondria. Chloroplasts are found in red algae, green algae, plants, alveolates, stramenopiles and flagellates, but not in animals, fungi, or cellular slime molds. In addition, different eukaryotic chloroplasts have different numbers of membranes around them. Red algae, green algae and plants have chloroplasts surrounded by two membranes. Diatoms and brown algae (stramenopiles) have chloroplasts surrounded by four membranes and flagellates, and dinoflagellates have chloroplasts surrounded by three membranes. Hence, the presence of chloroplasts cannot be explained by a single endosymbiotic event, as can the presence of mitochondria. The actual events are not totally clear, but the following is a likely scenario, based both on the number of membranes around the chloroplasts, and the sequencing of the DNA in the chloroplasts of the different organisms. A common ancestor of red algae, green algae and plants engulfed a cyanobacterium, giving rise to the chloroplasts with two membranes found in red algae, green algae and plants. This was the primary endosymbiotic event. There followed subsequent secondary endosymbiotic events in which eukaryotic descendants of this primary event were engulfed by the ancestors of alveolates, stramenopiles and flagellates, giving rise to chloroplasts surrounded by additional membranes. Thus, the current distribution of chloroplasts is best explained by a series of independent endosymbiotic events, many of which involve one eukaryote engulfing another, although the original chloroplasts came from a cyanobacterium that was engulfed by a eukaryote.

**Domain Bacteria**

A look at the bacteria domain reveals many familiar organisms, as well as some that are not so familiar. Some highlights follow. Both the thermotogales and aquificales are thermophilic, living at high temperatures in hot springs in Iceland and Yellowstone National Park. They are also chemolithoautotrophs. The spirochetes are one group of bacteria that can be distinguished by their corkscrew-like morphology. The syphilis bacteria belong to this group, as do the lyme disease bacteria. Chlamydiae is a group containing the bacteria that cause chlamydia. The cyanobacteria have already been discussed. Formerly known as “blue-green algae,” they are the organisms in which oxygen-producing photosynthesis first developed. Most, but not all, of them continue to be photosynthetic. The gram positives are organisms whose cell wall retains a gram stain. This group includes some familiar lab organisms such as *Bacillus subtilis*. However, it also contains some of the most terrible pathogens in human history. This is likely because their thick cell wall makes them less susceptible to attack by the immune system. Bacteria causing staphylococcus and streptococcus infections are gram positive, along with those causing diphtheria, tuberculosis, botulism (*Clostridium*), scarlet fever, toxic shock...
syndrome, and pneumonia. Mycoplasmas, the smallest known independently living organisms, are in the gram positive group, closely related to *Clostridium*. In addition, some gram positives (actinomycetes and streptomycoses) produce important antibiotics. The purple bacteria, sometimes known as proteobacteria, are most of the familiar gram negative organisms. This large group contains *Escherichia coli*, *Rhizobium* (the nitrogen fixers in legumes), *Helio bacter pylori* (the cause of ulcers), *Legionella* (Legionnaire’s disease), and *Neisseria* (gonorrhea). This group is also the source of mitochondria.

**A Final Caution: Lateral Gene Transfer**

This picture of three domains is so pretty, I am sorry to have to ruin it with an important caveat. A tree-like phylogeny like this one assumes that once species branch off from each other, there is no further exchange of genetic material between them. In the endosymbiotic acquisition of mitochondria and chloroplasts, we have already seen that this is not entirely true. We are also intuitively familiar with the concept of lateral transfer of genes. This is the exchange of genetic material between organisms of different species. The most obvious case is the troubling spread of plasmids containing genes for antibiotic resistance (Offner 1998). This freely occurs across species boundaries, as bacteria exchange these life saving (for them) genes with each other. What is virtually certain is that this kind of lateral gene transfer is not confined to genes for bacterial resistance. As more genome sequences become available, it appears that there is large-scale lateral gene transfer between members of the bacteria and archaea domains. One species of bacteria, for example, was found to have 24% archaea genes. An important area of research in the next decade will be to look at genome sequences and try to determine just how widespread this lateral gene transfer is. If a universal ancestor lived 3.5 billion years ago, there have been billions, or even trillions of generations between that universal ancestor and present day organisms. It is possible that during this time there has been so much lateral gene transfer that the tree we are looking at represents the history of ribosomal RNA genes and no more. I, personally, am more optimistic. To me, the fact that many recognizable organisms (spirochetes, cyanobacteria, gram positives) fall together on the rRNA tree means that, even if there has been extensive lateral gene transfer, the tree is still genetically meaningful. However, the final answer will come from the hundreds of whole genome sequences that will be done in the coming years.

**Acknowledgments**

I would like to thank Dr. David Baum for many helpful conversations and insights during the preparation of this paper.

**References**


---

**2001 National Convention**

**November 7-10, 2001**

Palais des Congrès de Montréal
Centre Hotel Le Sheraton
Montréal, Québec, Canada