

PHYLOGENETIC TREE OF LIFE

Michel Laurin

CR2P (UMR 7207), CNRS/MNHN/SU, Muséum National d'Histoire Naturelle, Paris, France

Keywords: Phylogeny, history of systematics, morphology, molecular biology, classification, evolution.

Contents

1. Oldest Written Records of Classification
 2. Great Explorers and the Rise of Rank-Based Nomenclature
 3. Evolution and Trees
 - 3.1. Taxonomy before Evolution
 - 3.2. The Advent of Evolution
 - 3.3. Evolution and the First Trees
 4. The Rise of Phylogenetics in the 20th Century and Beyond
 - 4.1. Cladistics (Parsimony)
 - 4.2. Phenetics (Numerical Taxonomy)
 - 4.3. Maximum Likelihood
 - 4.4. Bayesian Phylogenetics
 - 4.5. The New Frontier: Timetrees
 5. Phylogenetics and Its Impact on Taxonomy and Nomenclature
 6. Why are Trees Important?
 7. Conclusion
- Acknowledgments
Glossary
Bibliography
Biographical Sketch

Summary

Humans must have classified living beings very early in their history, and early traces of such classifications are found in book of *Genesis* and Homeric-era Greek texts. Aristotle laid the foundations upon which systematists would build for the next two millenia. As our knowledge of the living world expanded, there was an increasing need for an efficient, standardized nomenclature, and this led Linnaeus to publish two monographs (*Systema Naturae* and *Species Plantarum*) that laid the foundation of rank-based nomenclature in the mid-18th century. This new nomenclature, still used by most systematists, was based on a fixist, creationist world view, and its foundations were thus shaken by the rise of evolutionism in the 19th century. Phylogenetics progressed rapidly in the 20th century, with the rise of phylogenetic systematics (cladistics), model-based phylogenetics, phylogenetic analysis software, and molecular phylogenetics. These developments have reshaped our view of the living world, with detailed phylogenies of many major taxa now available. In parallel, they have prompted the development of a new nomenclatural system based on the tree metaphor to provide delimitation of taxa in a phylogenetic context, and hence, give a more precise meaning to taxon names. This

book presents the phylogeny of the main branches of the Tree of Life, as well as related phylogenetic topics.

1. Oldest Written Records of Classification

Our early ancestors must have felt a need to classify the living world, as shown by the fact that all peoples (including the numerous indigenes around the planet) have names for the animals and plants that occur in their environment. In fact, these folk taxonomies are fairly elaborate and share a hierarchical structure. According to a common view, they include ranks, namely folk-kingdom (e.g., plant, animal), life-form (e.g., tree, bush, bird, mammal), generic-specific (e.g., oak, shark, dog), folk-specific (e.g., white oak, poodle), and folk-varietal (i.e., swamp white oak, toy poodle). However, some anthropologists and systematists contend that these categories are merely constructs of Western scientists that have been applied to taxonomies of indigenous peoples who had no such concepts.

The earliest written records of classification systems we have are found in Homeric-era Greek literature (8th and 7th centuries BC), in which names of 71 animal taxa, mostly of domestic animals, occur. However, many of these taxa may have had much older now-lost names because several of them had been cultivated since the Neolithic, between 6000 and 8000 years ago. The book of *Genesis*, in the *Bible*, also attests to the early existence of biological nomenclature, as shown by this brief passage: “And out of the ground the LORD God formed every beast of the field, and every fowl of the air; and brought [them] unto Adam **to see what he would call them**: and whatsoever Adam called every living creature, that [was] the name thereof. [2:20] And Adam **gave names** to all cattle, and to the fowl of the air, and to every beast of the field; ...” (Genesis, quoted from the King James version).

The first known scientific attempt at classifying animals is preserved in Aristotle’s *History of Animals* (also known by its Latin name *Historia animalium*). The great Greek philosopher Aristotle (384–322 BCE, Figure 1) produced a hierarchical classification of animals that lacked fixed ranks (akin to the Linnaean categories that were developed about 2000 years later). Indeed, Aristotle’s work on animal classification uses the words *genos* and *eidos* to designate “kind” and “form”. These terms did not designate genera and species, as these terms are now employed in biology; on the contrary, they were relative terms because *eidos* clearly designated a subdivision (form) of a given *genos*, and a taxon that is considered an *eidos* in a given context can become a *genos* in another, and itself be divided into a number of *eide* (plural of *eidos*). This lack of fixed levels led some authors to doubt that Aristotle really produced a taxonomy. However, the latest works by zoologists support the existence of such a taxonomy in Aristotle’s writings, even though Aristotle apparently never summarized it into a simple tabular form, unlike his successors about two millennia later. Aristotle even apparently coined new names for taxa that he was the first to recognize, such as *Selachii* (elasmobranchs, a taxon that includes sharks, skates and rays) and *dithyra* (bivalve mollusks). Aristotle’s *History of Animals* was highly influential throughout the Middle Ages and the Renaissance, and many of the taxa that he erected or recognized are still used today, even though we now know that some, like his *ichthyas* (which became *Pisces*, “fishes”) are paraphyletic. In parallel with this, he arranged high-ranking taxa into his *Scala*

naturae (chain of beings), from simplest to most complex, and this exerted a lasting influence throughout the Middle Ages and later; in fact, many qualifiers used in taxonomic expressions still used by today's systematists, like "lower", "higher" or "advanced", hark back to Aristotle's *Scala naturae*.

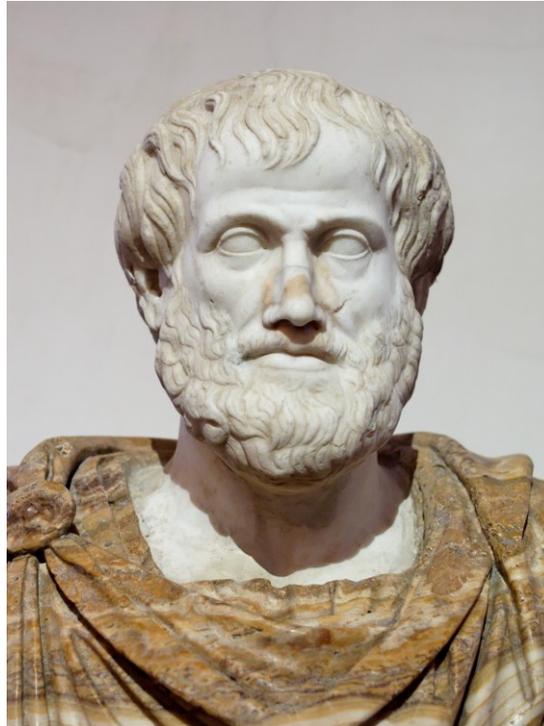


Figure 1. Bust of the Greek philosopher Aristotle. Roman marble copy of a Greek bronze bust of Aristotle by Lysippos, c. 330 BC; the alabaster mantle is modern. Picture taken by Jastrow for the Ludovisi Collection in 2006 of the bust in the Museo nazionale romano di palazzo Altemps (released into the public domain). Downloaded from https://commons.wikimedia.org/wiki/File:Aristotle_Altemps_Inv8575.jpg on 1-2-2021.

2. Great Explorers and the Rise of Rank-Based Nomenclature

Our knowledge of the living world expanded quickly in the Renaissance, partly because of the great explorers, who often brought naturalists along to collect and describe new taxa. It quickly became necessary to develop a better nomenclatural system because of the confusion that prevailed (the same taxon could be described several times under different names by different authors) and because the names had become unwieldy. Indeed, pre-Linnaean names were largely descriptive. This worked well when we knew relatively few taxa, but this nomenclature became difficult to use as more taxa were discovered. This can be illustrated by a hypothetical example. Suppose that only two kinds of a flower (let's say a rose) are known: a red and a white one. It is easy enough to call them "white rose" and "red rose" or their Latin equivalents (Latin was then the international scientific language). But as more diversity was recognized, this required longer names, to better describe the taxon (in the case of the rose, we might have added a descriptive term for the shape of the leaves or stems, or the shape of petals, etc.). This led some systematists to propose very long names, a problem that was discussed by the

French botanist Joseph Pitton de Tournefort (1656-1708; Figure 2) in his *Elemens de botanique* (published in 1694), in which he stated (translation mine):

“Plant names must be as short and clear as possible; but they must contain, in their brevity, what is most singular and sensitive in each species. ... Morison has given such long names that you lose your breath reciting them. ... A modern author has given the following name to an African plant:

Mesembrianthemum Africanum frutescens minus, erectum, triagularis foliis viridibus cornuum taurinorum in modum inflexis, fructu turbinato, parvo, pentagono, lignescente, flore albo.

If we needed this herb to treat a patient, would we dare to fill a prescription with that name?”



Figure 2. Portrait of the French botanist Joseph Pitton de Tournefort (1656-1708). File uploaded by Wikipedia user Materialschemist on April 9, 2012 into Wikipedia Commons (public domain). Downloaded from https://commons.wikimedia.org/wiki/File:Tournefort_Joseph_Pitton_de_1656-1708.jpg

It is in this context that the Swedish botanist Linnaeus (1707-1778) proposed his binominal nomenclature, in which each species name consisted of two words: a *genus name* and the *specific epithet*. This simple system allowed trimming down the names to a manageable size and separated the description from the name. In fact, such a system had been proposed and used earlier, though in a more flexible format, and may have been influenced by Aristotle's concepts of *genos* and *eidos*, which gave rise to *genus* and *species*, even though in Aristotle, these terms did not correspond to these Linnaean categories, or even to fixed taxonomic levels. But as we discovered more biodiversity, the names had expanded to remain unique and descriptive. Linnaeus advocated using a strictly binomial nomenclature associated with brief descriptions, which later became

diagnoses. This explicit dissociation between names and descriptions, as well as the shorter names, contributed to the success of Linnaean nomenclature.

Linnaeus' nomenclatural system also featured fixed ranks, which came to be known as "Linnaean categories", even though Linnaeus proposed only some of them. In addition to *species* and *genera*, these included *varieties* (for plants) or *races* (for animals) below the species rank, as well as *order*, *class*, and *regnum* (kingdom). Additional ranks were later inserted, such as *family* (between genus and order) and *phylum* (between *class* and *regnum*), and prefixes (such as sub- and super-) were used to subdivide these ranks.

This system brought some order into what had been nomenclatural chaos, but soon, new problems arose because authors disagreed about the correct name for a given taxon and its delimitation. Thus, codes of biological nomenclature were developed, starting with the Strickland code, which was published in the 1840s. This was an important development because these codes regulated names through two new principles: chronological priority and the use of types. Linnaeus had not tackled these issues because he thought that he should decide alone the name and delimitation of taxa. But as an increasing number of systematists published competing taxonomies and nomenclatures, this problem became acute and was tackled (incompletely) by the rise of rank-based codes of nomenclature. Under these codes, a name is defined by a *type* and a *rank*. For species, types are specimens, but for higher taxa, types are a lower-ranking taxon. For instance, the taxon *Hominidae* is defined, under the Zoological code, as the family-level taxon that has the genus *Homo* as its type, and the taxon *Homo* is defined as the genus-level taxon that has *Homo sapiens* as its type. Note that taxa thus defined are inherently undelimited, but some proponents of rank-based nomenclature view this as an advantage, as explained in the preface of the Zoological code: "Nomenclature does not determine the inclusiveness or exclusiveness of any taxon, nor the rank to be accorded to any assemblage of animals, but, rather, provides the name that is to be used for a taxon whatever taxonomic limits and rank are given to it." As we will see below, we would have to wait for the rise of phylogenetic nomenclature in the late 1980s to see the development of an alternative nomenclatural system that seeks to delimit taxa precisely.

3. Evolution and Trees

3.1. Taxonomy before Evolution

Early classifications of living beings did not rely on a phylogenetic tree. However, folk taxonomies (many of which were developed by societies that did not know writing until they established contacts with technically more advanced civilizations) are mostly hierarchical, and can thus easily be represented by a tree, even though this idea is foreign to indigenous thought.

In the Judeo-Christian tradition, life forms had been created by God, as written in book of *Genesis*. Thus, biological classification was thought to reflect the plans of the Creator.

The oldest known classification of living organisms, produced by Aristotle, was hierarchical (see section 1). There was no tree in Aristotle, but his *Scala Naturae*, came in time to be interpreted in terms of evolutionary progress, a connotation that it did not have in Aristotle's work, given that the concept of organic evolution was formulated two millennia later.

3.2. The Advent of Evolution

The idea of organic evolution was developed slowly in the 17th and 18th centuries with limited transformism (evolution within a genus), which we see, among others, in the late work by Linnaeus (who had been a creationist early in his career). Limited transformism explained diversity within genera: one species had been created in each genus and subsequently diversified to give several closely-related species. This idea imposed itself to explain the obvious similarities between closely-related species.

The works of Lamarck and (later) Darwin established general transformism, which explained the evolution of life from the simplest to the most complex forms. Lamarck is generally considered to have been the first true evolutionist. He proposed an explicit (and often misrepresented) theory of evolution. Lamarck suggested that over geological time, changes in the environment of animals led to changes in their needs, and that this triggered changes in their activity. This then led to changes in their morphology (or other changes) through use and disuse, and to the inheritance of these changes to the offspring. This last step has proven most problematic, as it subsequently became clear that acquired characters are very rarely transmitted to the offspring. Lamarck's acceptance of spontaneous generation and inheritance of acquired characters were shared by most of his contemporaries; in fact, the idea of inheritance of acquired characters harks back to Antiquity, and so Lamarck never considered it his idea.

3.3. Evolution and the First Trees

In Lamarck's work, evolution is rather directional and proceeds to an extent in a linear fashion. However, Lamarck realized that not all of biological diversity could be organized into a simple linear series, contrary to Aristotle's *Scala naturae*. This was a first important step in the development of the *Tree of Life* metaphor, which would be elaborated later by Darwin. Thus, Lamarck's schematic evolutionary tree shows the inferred relationships between high-ranking taxa, from worms to various mammal taxa. In his *Philosophie zoologique* (published in 1809), he even boldly proposed (correctly) that humans fit in this scheme as a derivative from great apes.

However, Lamarck's trees lacked a clear temporal division, unlike subsequent evolutionary trees. This would come later, with Darwin's famous tree (published in his *Origin of species* in 1859), first as a schematic representation of evolutionary time in a hypothetical tree.

With Darwin, the tree (with a time dimension) became the dominant paradigm to explain biodiversity, and the idea of massively parallel progress disappeared from evolutionary thought (even though other workers, notably some paleontologists, who did not accept Darwinism or neo-Darwinism would continue to assume orthogenesis for

several decades). However, the sole evolutionary tree published by Darwin was a purely theoretical one.

The first true empirical evolutionary trees (with real taxa and a geological timescale) were published in 1866 by two zoologists, French paleontologist Albert Gaudry (1827–1908) and German evolutionist, marine biologist, and developmental zoologist Ernst Haeckel (1834–1919). In his monograph on the Cenozoic (mostly Neogene) placental mammals of Pikermi (Attika, Greece) published in 1866, Gaudry produced several phylogenetic trees of Cenozoic (mostly Neogene) mammals placed in a geological timescale. Haeckel also published several trees, which were shown in a naturalistic way (like real trees, whereas Gaudry had used only straight lines to link putative ancestors to their descendants). In most of Haeckel's trees, the time axis is only implied, but his tree of vertebrates (like Gaudry's) differs in having a geological timescale on which we can read the divergence times between various taxa, some of which were represented by many extinct terminal taxa, like Labyrinthodontia (a paraphyletic taxon that contained, at the time, mostly Permian and Triassic temnospondyls, which are often called “amphibians” but may actually be stem-tetrapods). These were thus genuine evolutionary trees that were conceptually very close to those that we still infer today (but with more sophisticated methods). The paleontological timetree was born!

This account has left out the fixists (those who do not believe in biological evolution) and creationists (those who think that the universe was created more or less as we now see it by a god) who clung to their beliefs throughout this period. Indeed, at any given time, a variety of points of view are present in the scientific community. Thus, the famous paleontologists Georges Cuvier (1769–1832) and Louis Agassiz (1807–1873) remained fixist until the ends of their careers. However, by the late 19th century, most if not all paleontologists had accepted the reality of biological evolution.

Note that with the acceptance of the idea of biological evolution and of the Tree of Life as a metaphor, biological classification now had to reflect evolutionary patterns (as depicted in the Tree of Life), rather than a plan of the Creator or simple similarities. The main problem to solve was: how to infer these patterns? Progress on this front was slow, but it accelerated greatly in the second half of the 20th century, and this remains an active area of research.

-
-

TO ACCESS ALL THE 25 PAGES OF THIS CHAPTER,
Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>

Bibliography

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars, *Behavioral and Brain Sciences* 21: 547-569. [A good review of studies of folk biology that argues that folk taxonomies include five main ranks.]

Berlin, B. (2014). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*, 354 pp. Princeton, NJ, Princeton University Press. [This book summarizes decades of research in the field of folk taxonomy, which gives us an indirect glimpse at the first classifications of the living beings that were developed.]

Cantino, P.D., de Queiroz, K. (2020). *International Code of Phylogenetic Nomenclature (PhyloCode): A Phylogenetic Code of Biological Nomenclature*, xl + 149 pp. Boca Raton, Florida, CRC Press. [The only code of phylogenetic nomenclature, this proposes rules to define clade names in a phylogenetic context; such definitions provide unambiguous taxon delimitation even when the phylogeny changes.]

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection Or the Preservation of Favoured Races in the Struggle for Life*, 502 pp. London, John Murray. [This is the first edition of Darwin's famous book in which he explained in detail the theory of evolution through natural selection, which he had presented briefly in a paper co-authored with Alfred Russell Wallace.]

de Queiroz, K., Gauthier, J. (1990). Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names, *Systematic Zoology* 39: 307-322. [One of the founding papers of phylogenetic nomenclature, which explains why this system was developed.]

Didier, G., Laurin, M. (2020). Exact distribution of divergence times from fossil ages and tree topologies, *Systematic Biology* 69: 1068–1087. [This paper develops the FBD to date trees and applies it to show that the Amniota, defined as a crown group, is older than assumed by most previous studies.]

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity, *Biological Conservation* 61: 1–10. [This is the initial presentation of phylogenetic diversity, which is a great alternative to taxon counts to assess biodiversity. The paper contains worked examples that show how taking the tree into consideration could lead to different conservation priorities.]

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap, *Evolution* 39: 783–791. [An important paper that propose a widely-used method to assess the relative support of clades, using character sampling with replacement.]

Felsenstein, J. (1985). Phylogenies and the comparative method, *The American Naturalist* 125: 1-15. [This is the original presentation of phylogenetic independent contrasts, the first statistical method that took into consideration the fact that comparative data are not statistically independent of each other.]

Felsenstein, J. (2001). The troubled growth of statistical phylogenetics, *Systematic Biology* 50: 465-467. [A humorous account of the difficult beginnings of statistical phylogenetics, including phenetics and maximum likelihood, especially as used by molecular systematists, from the perspective of a major participants of this colorful episode in the history of systematics.]

Gaudry, A. (1866). *Considérations générales sur les animaux fossiles de Pikermi*, F. Savy. [This paper was the first one to include evolutionary trees of actual taxa inserted into a geological timescale.]

Geraads, D., Didier, G., Barr, A., Reed, D., Laurin, M. (2020). The fossil record of camelids demonstrates a late divergence between Bactrian camel and dromedary, *Acta Palaeontologica Polonica* 65: 251–260. [This study uses the FBD to show that the divergence between Bactrian camel and dromedary probably occurred about 1 Ma and at the most 2 Ma, rather than 4-8 as previously suggested by molecular dating studies.]

Klimis, G. (2008). *Medicinal Herbs and Plants in Homer. History of Mechanism and Machine Science 6 Science and Technology in Homeric Epics*: (Ed. Paipetis, S. A.), pp. 283–291. Berlin: Springer. [This paper lists and tries to identify the species of animals and plants mentioned in Homer's Iliad and Odyssey, which are among the earliest written records we have of early biological nomenclature.]

Laurin, M., Humar, M. (2022). Phylogenetic signal in characters from Aristotle's History of Animals, *Comptes rendus Palevol* 21: 1-16. [This paper analyzes a data matrix compiled from Homer's History of Animals and shows that this produces a tree that includes many taxa recognized by Aristotle, some of which are still valid today.]

Pyron, R.A. (2011). Divergence-time estimation using fossils as terminal taxa and the origins of Lissamphibia, *Systematic Biology* 60: 466–481. [This is the first tip dating study; it uses molecular, morphological and paleontological data to date the evolutionary radiation of amphibians, with emphasis on extant amphibians (lissamphibians).]

Rineau, V., Zaragüeta I Bagils, R., Laurin, M. (2018). Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis, *Contributions to Zoology* **87**: 25-40. [This simulation-based study shows that character states that represent the discretization of continuous traits should be ordered in phylogenetic analysis. This yields better results, even if a few minor ordering errors are made.]

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., Rasnitsyn, A. (2012). A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera, *Systematic Biology* **61**: 973–999. [The second tip dating study, but the first one to use the new version of MrBayes, which greatly facilitates implementing such studies. Performed by a multi-disciplinary team combining molecular systematists and a paleontologist.]

Sanderson, M.J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach, *Molecular Biology and Evolution* **19**: 101-109. [This paper introduces penalized likelihood, a relaxed molecular clock method that penalizes rate changes between adjacent branches and uses a cross-validation procedure to obtain the optimal penalization setting.]

Tassy, P. (2011). Trees before and after Darwin, *Journal of Zoological Systematics and Evolutionary Research* **49**: 89–101. [This is a good account of the progressive development of evolutionary trees through the 19th century and of their predecessors, the biological classifications, in the late 17th and 18th centuries.]

Zuckerkandl, E., Pauling, L. (1965). *Evolutionary divergence and convergence in proteins. Evolving Genes and Proteins*: (Ed. Bryson, V. and Vogel, H. J.), pp. 97–166. New York: Academic Press. [This is one of the landmark studies that initiated molecular dating, in which the hypothesis of a molecular clock is explicitly formulated and justified based on theoretical principles.]

Biographical Sketch

Michel Laurin received his PhD at the University of Toronto in 1994, and did a first postdoc at Berkeley (1994-1996). He then moved to Paris (1996-1997), where he worked as associate invited professor in the University Paris 7, to lecture on paleontology, the evolution of animals, biological nomenclature, and other related topics. He then did a last postdoc in Berlin (1997-1998), in the Natural History museum. He was hired by the CNRS as a research scientist 1998. He has published several scientific papers, in addition to a book on paleontology and systematics in French and English. He has led the team “Squelette des vertébrés” for two years (2007–2008) and has been leading the team “Phylogénie et Diversification des Métazoaires” of the CR2P (Paris) since 2014. He is a member of several scientific societies, and has served the *ISPN (International Society for Phylogenetic Nomenclature)* as both Secretary (several terms, including a current one) and President. He currently serves on seven editorial boards, including for the *Journal of Evolutionary Biology*. In January 2011, he has become Chief Editor of the *Comptes Rendus Palevol*. He has studied the evolution of vertebrates from the Devonian to the present, and has covered most of their evolution, from early jawless vertebrates to extant mammals. His current interests include comparative methods and dating the Tree of Life, and he has tackled problems as diverse as the invasion of land by vertebrates, the appearance of the tympanum (ear drum), and origin of extant amphibians. He has introduced innovations in paleontology and paleobiology, notably in the form of computer programs, some of which can be used to perform paleontological dating of the Tree of Life or to analyze bone microanatomical data.