Understanding phylogenies: Constructing and interpreting phylogenetic trees

Joseph L. Staton[1]

[1]*University of South Carolina Beaufort*

Interpretation of phylogenetic trees is fundamental in understanding the relationships between organisms, their traits or characteristics, their ecology and even their genomic and developmental biology. As trees appear more often in basic texts, many students, and even their teachers, clearly understand little of how they are constructed and even less about what can be inferred from them about the history of the representatives analyzed. Not only are these trees a source of confusion on what they do tell us, often non-specialists infer things wrongly or, worse, others misuse them in an attempt to negate the validity of evolutionary theory. In this brief introduction, I attempt to give a synopsis of basic tree-building methods, and more importantly demonstrate interpretation and dispel some common misconceptions about them.

Understanding a phylogeny, its construction and its interpretation, is at the core of the modern comparative method in biology. Life on Earth is diverse and seemingly impossible to comprehend. Even though objective methods to develop working hypotheses are central to understanding the evolutionary history for a group, these methods, or their resulting interpretation, are not immediately transparent to the majority of students or even researchers in the broader field of biology. Indeed, many outside of— or distrustful of— science are dubious that we can actually derive and study these histories. The use of modern phylogenetic methods in the life sciences has informed and revolutionized our understanding of the history of life on the planet and impacted diverse areas of research in forensic biology, biogeography, adaptation, and evolutionary biology.

There are multiple ways to construct a branching tree (Fig. 1) of organisms based on characteristics. Many of the primers available for learning these methods suffer, if not from philosophical complexity (not a trivial aspect of different methods employed), then at least from length. Even the more extensive presentations are geared toward the senior undergraduate or graduate levels (Baum & Offner 2008, Brooks & McLennan 1991, Hall 2011, Page & Holmes 1998). I will focus on a brief description of the different methods and a brief discussion of the relative strengths and weaknesses, but realize that entire books and even journals have been published dealing with the complex philosophies and intense computer-computational methods to produce these results. The three main areas of methodology for creating trees rely on **characteristic-based** (called **Cladistics**), **distance-based** (a mathematical index of relative similarity between taxa), or some complex combination of the two (**Maximum Likelihood**). For simplicity's sake, I will forgo the last category and focus on the details of the first two, as examples to be used to relate the information that can be and cannot be derived from a phylogeny.

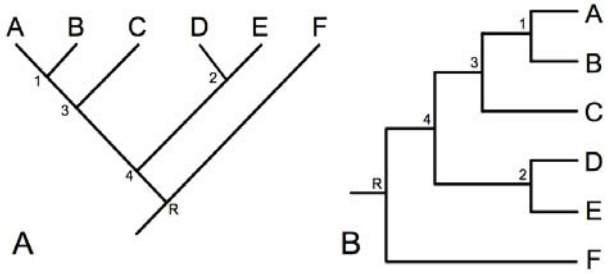

Figure 1. Two hypothetical trees. Trees progress from the tips (terminal "taxa") to the base of a tree when "rooted" (see section on rooting a tree). Each lineage is represented by a line that joins other lines at "nodes" (=a branch point representing a common ancestral condition; 1-5). This progresses further and further until all lines are joined to a single line which is the root of the tree (R). Note that the two trees are different only in their orientation and the branch representation; the branching pattern is the same for both.

Cladistics

Cladistics arose from the work of Hennig (1950, 1966). A dataset of characteristics (=alternative traits) is accumulated for a number of related organisms to be analyzed. For every organism or **taxon** (singular of **taxa**, an operational term for the tips of a tree, either a species, a genus, or higher), a set of characteristics (or characters) is compiled into a matrix, consisting of *T* rows x *C* columns, where *T* = the number of taxa and *C* = the number of characters. Each character has two or more possible types of condition called **states**. Trees are then constructed to minimize the number of changes between all **states** for all characters among all **taxa**. This is a deceptively simple statement. There is no *a priori* method for drawing one specific tree based on the variation among states in **character state matrix** (see Fig. 2, for example), but for simple datasets the solution can be readily apparent, even to non-scientists. An excellent hands-on exercise for demonstrating this method was published previously (Goldsmith 2003). For large datasets the patterns are often not obvious, so we construct all possible branching diagrams, count the number of state changes required for each, and retain those network(s) with the fewest changes. These minimum change solutions are considered to hold sway over more complex ones (*i.e.*, are more **parsimonious,** Sober 1981). An analogy that can illustrate this point to students is, as follows: a suspect that is

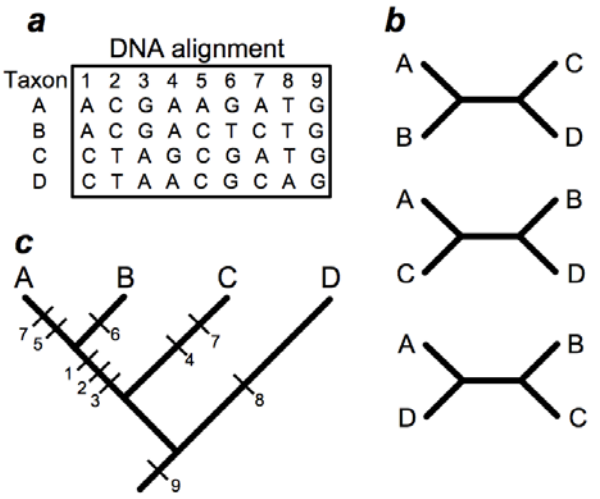

Figure 2. A brief example of a simple matrix with the resulting networks. A) An abbreviated data matrix for 9 DNA positions of a gene. B) Three possible unrooted networks for four taxa. The top network is the preferred as it has the least number of changes required when mapping the state changes on the network. C) That network is redrawn and "rooted" using taxon D as the outgroup. The root is then a stem introduced to the network that isolates the outgroup from the remaining ingroup and represents the basal ancestral condition for the phylogeny.

arrested near a murder scene fits the general description of a person who was witnessed committing a murder and has similar fingerprints to those found at the murder scene. Barring further evidence of another such person existing (i.e., with similar appearance and fingerprint pattern) and being near the vicinity of the murder at the same time, the suspect in custody is the most likely culprit. Computers usually work with networks of taxa to complete these computations (see Fig. 2), and subsequently a network can be "rooted" (that is, shown with an estimated point of origin) when one considers the ancestral condition of the group under consideration (see "*Rooting a Tree*").

With relatively few taxa, this is trivial. For only three taxa, there is only one unrooted network (but three possible "**rooted**" trees, where the root isolates two taxa from the other one, what would that look like?). With four taxa, there are only three possible arrangements (see Fig. 2). After the number of changes (=evolutionary events) needed are noted for each, the network that requires the fewest evolutionary events (= state changes) is chosen as the preferred network, given the notion of parsimony (i.e., "Ockham's razor"). It is the shared character states that support the branching patterns of a tree, whereas those states unique to a single branch or shared among all taxa (while evolutionarily of interest) do not shed light on relationships in the network or tree.

Rooting a Tree and Analyzing Characters

Once the best network is chosen, it is usually **rooted** at a "trunk," which represents the ancestral condition for all members within the tree. Several ways have been suggested over the years of phylogenetic analysis, but by far the commonly accepted practice for creating a root is by using an **outgroup** (Wheeler 1990). The outgroup is considered a taxon (or sometimes a group of taxa) that is closely related to the group of interest (i.e., the **ingroup**), but not a part of that group. As a result, a network can be "rooted" by dividing the outgroup(s) from the remaining ingroup taxa (see Fig. 1, 2c), with a basal branch or **root** (**R**). One can envision this as "grabbing" the branch between the outgroup (singular or a cluster of taxa) and remaining network (the ingroup) and pulling that branch "down" to create a new basal branch (trunk) of the tree. For example, the network in Figure 2c was rooted by using "D" as the outgroup for one unrooted network in Figure 2b (which one was used?).

Rooting the tree provides perspective to the history of character states and order that new states came into existence. In a network, one can distinguish how character states transition along branches of the network, but with no root—there is no historical perspective of the order in which the traits arose. A root identifies those traits common to all descendants in the study and those that are shared among only subsets of the groups. With a root, we can ascribe new characteristics to states to identify their status, either an **autapomorphy,** a **symplesiomorphy,** or a **synapomorphy.** In Cladistics, synapomorphies are considered to be the "phylogenetically informative" character states in a study, because they support the pattern of branches between taxa. One last group of characteristics is called **homoplasy**. When the most parsimonious (= shortest) tree is found, a small subset of the characteristics analyzed may not be consistent with the overall tree, which is driven by the largest congruent subset of synapomorphies. These non-congruent changes are called homoplasies. They are not "bad" data, but represent either characters that are thought to be misrecorded (and need to be re-evaluated), or that the hypothesis of their homology may be flawed, as traits can evolve more than once (parallel evolution). This case is especially well known for individual DNA bases over long time periods. Indeed, classic homoplasies are often trivially obvious in many taxa (*e.g.*, the wing of a bird and the wing of a bat), but many not so obvious in terms of subtle characteristics in a cladistic analysis.

The Use of Cladistic Methods

The use of morphology in cladistics is appealing since it has been at the core of evolutionary study for over a hundred years. The further idea that cladistics methods are objective and can remove researcher bias adds to that appeal. By using a large number of character states, cladistics methods can provide a robust understanding of traits across an entire group with some level of completeness.

Additionally, a phylogenetic tree can aid in understanding processes that yield the patterns in it. One can "map" the character state changes onto the tree to see when specific traits evolved and in which taxa they occur, or which descendants may have secondarily lost them. For all those descendants possessing a trait, it must have been present in the common ancestor—therefore it evolved at some point prior to the speciation of that ancestor (i.e., the node) and from the previous ancestor that gave rise to it (i.e., the next deeper node). In a humorous treatment elsewhere (Staton 1998), I argued that the characteristic of "tastes

like chicken" confuses a deep symplesiomorphic trait (here "chicken flavor") with a derived condition (the evolution of adapations that are present in birds, alone). Clearly, a "chicken-like" flavor is present in many other related organisms (the Tetrapods, *e.g.*, reptiles, amphibians, some mammals, etc.)— therefore the phrase should not be "tastes like chicken" but more aptly "tastes like tetrapod." Similarly, if we think about biology in a phylogenetic sense of first appearance of a specific characteristic, we can answer the age-old pseudo-philosophical question, "Which came first, the chicken or the egg?" since several species predating the origin of birds laid eggs. Such humorous approaches can get even the most jaded student engaged in discussions about phylogenetic trees.

Sometimes, organisms that share a close biological relationship, like that of parasites and their hosts, can be analyzed separately and compared for similar branching patterns among the host/parasite trees. When there is close agreement of branching pattern, it is taken as evidence of a shared, co-evolutionary history (Brooks & McLennan 1991). There are also more complex analyses that relate biogeography to phylogeny (termed **phylogeography**; see Avise 1998, 2000) so that researchers can understand how species evolve within a geographical landscape and do or do not spread over the different biotic zones of the planet.

Up to this point, I have not discussed the use of phylogenetic methods with molecular data, but the real growth in these analyses has been in their use with DNA data to produce phylogenies, since the development of new sequencing technologies in late 1980s. Such methods allow for hypotheses of deeper phylogenies, as well as comparison of morphologic to molecular trees to assess agreement in these results. There is no succinct way of reviewing the diverse ways that phylogenetic trees can be employed in different research programs, here— but it is without doubt that these methods are in widespread use across most biological fields.

Issues that Impact Cladistic Methods

Several difficulties are inherent in application of a cladistic analysis. I will try to give a very brief listing of them, not as an indictment against their use, but more as an explanation as to why there is no single accepted method for producing phylogenies.

Since cladistic methodology works backwards to hypothesize an evolutionary process, we approach the problem by constructing all possible networks and saving the shortest network as the best estimate of the correct one. The problem with this method is that the number of networks possible increases exponentially as the number of taxa increases linearly. The formula for the number of possible unrooted networks for n taxa is $(2n\text{-}5)! / [2^{n-3} \cdot (n\text{-}3)!]$ (Eq. 5.1; Li 1997). For example, there are $\sim 5 \cdot 10^{94}$ possible networks for 60 taxa, which is more than the estimated number of atoms in the universe. It is, in fact, a computational impossibility to search all of these networks merely due to time limitations. Researchers have developed shortcuts of searching a reasonable subset many of the most probable networks. Still, the methods are computationally intensive and can provide misleading results or even miss the most parsimonious tree.

In the recent decades, more phylogenetic work has focused on DNA data. Although morphological homology may be difficult to assess between a lobster and a human, both species contain genes that are homologous at the molecular level. Phylogenetic analyses of DNA is a powerful tool, however its analysis with cladistic methods can result in a unique problem called "long branch attraction" (described originally by Felsenstein 1978). Some DNA datasets contain broadly related taxa with highly divergent DNA sequences. Random patterns of species formation and extinction may mean that one or a few taxa might be quite different from those in the rest of the study. After DNA is aligned at the nucleotide level— unique taxa will differ greatly from all others (> 30-50% of variable bases at all nucleotide positions). When this happens, these taxa tend to branch off at a node near the **least-related** taxon— not from common ancestry but due to a few similar (convergent) mutations that randomly accumulate between sequences over time (i.e., they share chance similar homoplasies). This is most common for DNA because an adenine (A) at position 132 looks like another adenine at that position (A), whether it was inherited via common ancestry or converged to the same base by a separate mutational event (in this case A). Since the algorithms are developed to find the shortest tree by making branching patterns, they will make all taxa branch with one another, even if badly. The most maddening fact about this miscalculation is that the addition of more sequence data will only make the algorithm find— with even greater support—that a "bad" tree is the most parsimonious (see Felsenstein 1978). In such cases, other methods (such as distance-based methods) provide a different means of developing a phylogeny that do not suffer from these issues.

Building Distance-based Trees

Oddly enough, the first methods for making trees were phenetic methods (Sokal & Sneath 1963) based on some metric of similarity (or its converse—distance) between datasets to construct a branching network. This often worked poorly for morphological data (true phenetics), but worked arguably better for molecular data, where there was an implied mode of evolution between pairs of species DNA. These methods are computationally faster and usually yield a single tree (a feat not always accomplished by cladistic methods). Many argue that since the data are transformed into a comparative metric (e.g., percentage similarity or weighed similarity), the results are not as reliable as cladistic methods. However, with the aforementioned "long branch" problem and with large datasets involving hundreds of taxa (often the case in many studies), cladistic methods are not accurate or practical. As previously noted, to review all possibilities of methods here would not be possible. There are several competing tree/network building algorithms and many distance metrics that are reviewed elsewhere (Li 1997, Nei 1996, for example). For the sake of teaching the essence of these methods, I will focus on the classic analysis of variation in the protein cytochrome *c* (data from Fitch & Margoliash 1967) using a simple metric—minimum mutational distance— and branching algorithm—the unweighted pair-group method based on arithmetic means (or **UPGMA**, Sokal & Michener 1958). In this study, amino-acid sequences were aligned for each species

from previously published data, and a minimum mutational distance was calculated based on the fewest base substitutions required in the redundant genetic code to change each amino acid between each pair (Table 1A). The largest difference in the subset of their data we analyze (Table 1A) is Tuna/Moth pairwise distance of 41, meaning the minimum number of nucleotide changes needed to account for the paired Tuna/Moth cytochrome *c* differences at the amino-acid level locus is 41 (i.e., each changed codon could differ by either 1, 2, or all 3 bases in the translation code). Methods that use pairwise distances seek to produce a single branching "tree" so that length of branches between all taxon pairs is proportional to each branch distance for taxa in the "tree." In the case that all distances are not strictly additive, as can happen with distance datasets, UPGMA averages the non-equal distances to reach a balanced compromise among values across all pairs. For UPGMA, the smallest difference between a pair of taxa is used as the starting point—here between Man (*Homo sapiens*) and Monkey (a Rhesus monkey [*Macaca mullata*])—and is taken to be the complete distance between the two taxa. In this case, the minimum distance of one nucleotide between them is considered as one-half of one nucleotide difference (on average) from their common ancestor (see Fig. 3). In the next (and all subsequent steps), the original matrix of pairwise distances is collapsed one level so that the new group (in step 2 "Man/Monkey") becomes a single column (or a combined taxon group), and all remaining individual distances are collapsed so that each collection of distances (in step 2, those paired with "Man" and "Monkey") now becomes an average for the collapsed group to each remaining taxon (e.g., in this case the "Man-Turtle" distance of 19 differences gets averaged with the "Monkey-Turtle" distance of 18 to make an average difference of "Man/Monkey" and "Turtle" now 18.5). This is continued for all pairwise distances associated with "Man" or "Monkey" until a second matrix is complete, which has one less column and row (see Table 1B). This procedure is continued with the next smallest distance in the recalculated table ("Turtle/Chicken" in Table 1B). All subsequent averages (means) involve defining the mean for all possible pairwise distances for each new cluster from those in the **original** distance matrix. While calculating this by hand for the first time seems complicated, it is a simple repetitive algorithm and can be completed in milliseconds by computer even for large datasets. The end result of this process is a tree where pairs or groups of taxa will have branches where each pair of taxa is approximately 0.5 the total distance to the node as they have between one another (i.e., "Man" to "Monkey" = 1, Table 1A; "Man" or "Monkey" to node 1 = 0.5, Fig. 3). With the averaging process, the deeper (i.e., more distant nodes) are less reflective of the original data ("Moth" to "Tuna" = 41, Table 1A; "Moth" or "Tuna" to node 6 = 17 [not 20.5]). Other tree methods have been developed to circumvent these types of averaging errors (e.g., neighbor-joining, Saitou & Nei 1987), and results from these methods of tree construction can be demonstrated with a computer program in the classroom—but are not as amenable to direct calculation by teacher or student at the introductory level.

The resulting tree is close to the evolutionary tree that many would predict based on morphological, physiological or biochemical similarities (Fig. 3). The two primates are shown closest together (node 1), mammals form a clade (node 3), birds and reptiles form a group with a common ancestor exclusive of other vertebrates (node 2), and all organisms believed to have descended from a common amniote-egg layer cluster at a deeper level (node 4) than do the nodes of more recent divergence (nodes 1, 2 & 3). The fact that molecular evolution parallels that of other evolutionary theories based on different data (e.g., morphology) is not surprising, but it is impressive that even a small protein sequence can accurately capture these hierarchical patterns, in this case.

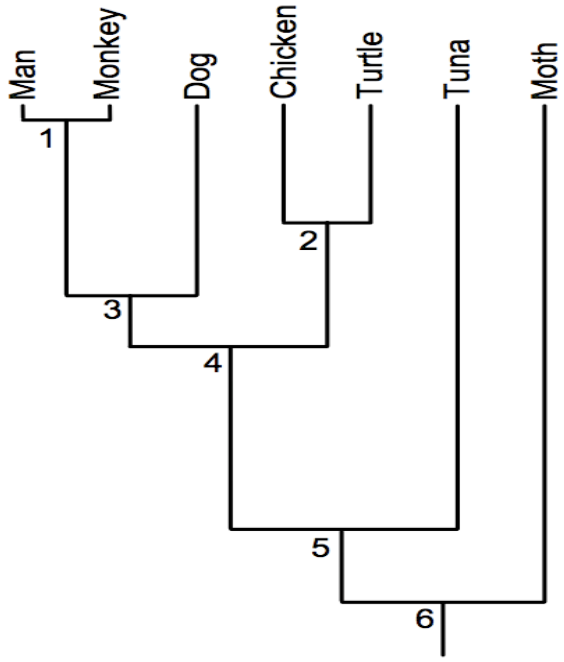

Figure 3. A UPGMA tree based on the minimal mutational distance for each pair of taxa listed in Table 1 (after Fitch & Margoliash 1967). Note that the distance between Man or Monkey to node 1 is half the total distance between Man and Monkey and represents an average distance from their most recent common ancestor. Each node is labeled in ascending order of the calculation from the original matrix. The next calculation in the iterative process is node 2 joining "Turtle" and "Chicken".

Interpretation of trees

Certainly, the interpretation of a phylogenetic tree is the basis for most of comparative evolutionary biology. The inferred pattern of branching is a road map to the understanding of any other hierarchy of traits that is possessed by those groups. Such traits can then be assessed as to relatively when they arose in the spectrum of ancestor-descendant relationships within the groups and their subgroups.

Misconceptions in tree interpretation

There are several misconceptions that can arise in the interpretation of phylogenetic trees. Some of these will already seem foolish to you, the reader, if you have followed the narrative to this point, but others may not be as straightforward.

Table 1. A) The minimal mutational distance between cytochrome *c* amino acid sequences of different taxa from Fitch & Margoliash (1967). The top half of the matrix is omitted, as it would be a duplicate of the data shown. The diagonal is the identity value for each species. B) This matrix shows the recalculation and reduction of the data in A to a six by six matrix, with Man-Monkey as a single row/column and each pairwise difference now an average of each distance (*e.g.*, [AB + AF]/n or [19 + 18]/2 = 18.5) from the original columns containing "Man" and "Monkey", respectively (see shaded cells). The next lowest pairwise distance (here the "Turtle/Chicken" distance of 8) will be the next pair of taxa to be collapsed into a single matrix row/column. Note that in many cases the next smaller group can be between a single taxon and a previously collapsed row/column.

A:		Turtle	Man	Tuna	Chicken	Moth	Monkey	Dog
		A	B	C	D	E	F	G
Turtle	A	--						
Man	B	19	--					
Tuna	C	27	31	--				
Chicken	D	8	18	26	--			
Moth	E	33	36	41	31	--		
Monkey	F	18	1	32	17	35	--	
Dog	G	13	13	29	14	28	12	--

Man to node 1 = 0.5 Monkey to node 1 = 0.5

B:		Turtle	Man-Monkey	Tuna	Chicken	Moth	Dog
		A	BF	C	D	E	G
Turtle	A	--					
Man-Monkey	BF	18.5	--				
Tuna	C	27	31.5	--			
Chicken	D	**8**	17.5	26	--		
Moth	E	33	35.5	41	31	--	
Dog	G	13	12.5	29	14	28	--

Chicken to node 2 = 4.0; Turtle to node 2 = 4.0 (8 changes total)

The intuition of some students can lead to erroneous thoughts that the gene in question is the source or driver of speciation, since they are loosely taught that "mutation the source of genetic change." However, even though every gene has the potential to capture unique mutational/molecular events that parallel a speciation event within individuals, the temptation is for the individual gene under scrutiny to be the sole focus of the speciation process for the novice. Molecular evolution is only a measure of the speciation pattern which has been captured in a molecular context of any given stretch of DNA. Students might aver that "a Dog would be a Monkey, if not for those 12 changes!" Or, wrongly assert that 12 mutations directly mutated a dog ancestor into a monkey, like some perverse science fiction movie. It is a difficult concept to grasp that DNA mutations in the cytochrome *c* record, in parallel, other changes in these lineages, and that only 12 of these changes were accumulated between dog and monkey cytochrome *c*, since their last common ancestor. In this sense, different genes are changing faster or slower in the evolutionary process, and sometimes they give us insight into the process as a whole.

There is no real expectation that each node represents a specific known species, or worse is some combined (chimeric) "monster" (i.e., a "Dog-Man-Key" for node 3, Fig. 3). Each node merely represents that some lineage of varying individuals that had a specific cytochrome *c* sequence that was ancestral prior of the formation of each descendant species. The idea that the study of evolution focuses on finding direct "missing links" is a strange hold over from earlier times when people believed in almost "alchemistic" transformation of one species into another (see Saint-Hilaire 1822). At fine scales of micro-evolution we often find transitions, but not finding specific transitions between larger taxonomic groups is not an indictment of the absence of the evolutionary process, as has been suggested by some factions critical of evolution.

Lastly, the casual observer often attributes importance to the order of the tips of the tree. While these are grouped in some order by the author, their specific pattern has no inherent meaning, necessarily. The fact that our example tree has "Dog" positioned next to "Monkey" has no special relevance (Fig. 3), nor does "Turtle" being next to "Tuna"— although novices will be confused by this pattern. Dogs are equally distant or close to the Man/Monkey common ancestor, but that is all. Likewise, "Chicken" is no closer or more distant than "Turtle" to the mammalian ancestor (node 3, Fig. 3). In fact, the final tree could be drawn with "Moth" next to "Man," however the branch of "Moth" would still connect at the base of the tree, at node 6 (Fig. 3). The order of the tips should be considered fluid, like membrane-bound proteins floating in an unconstrained lipid-bilayer, even though they might be connected within the "cytoplasm." Each node is like a frictionless turnbuckle/pivot;

free to rotate like arms of a mobile viewed upside down. Only the relative (internal) branching pattern is important, which is why trees can easily be represented by Venn diagrams or as a nesting code (called Newick format)— (((((Man, Monkey) Dog)(Chicken, Turtle) Tuna) Moth)—where the parentheses represent the clustering of taxa. The ordering of the clusters is unimportant: (((((Man, Monkey) Dog)(Chicken, Turtle) Tuna) Moth) is equal to (Moth (Tuna (Dog (Monkey, Man)) (Turtle, Chicken))). The relative grouping of taxa is the same.

Misrepresentations of phylogenetic trees

Lastly, it is necessary to point out that our phylogenetic knowledge is not complete and hypotheses about relationships change over time. Some groups are well-studied and this is less likely, but the more research data is gathered on different groups—the clearer the tree of life becomes. As we progress toward a better understanding of evolutionary patterns, we often admit that parts of our understanding are incomplete. Texts will publish trees with representative question marks or dotted lines to represent poorly understood or unstudied/understudied regions of a phylogeny. This is not a representation of the ineffectiveness of the method, but an honest assessment of what can being substantiated by the method at present. That is the strength of phylogenetics, we can list objective criteria for the tree-branching patterns that we uncover. The relative strengths and weaknesses can be objectively presented. Unfortunately, non-scientists that object to the principle of evolutionary biology point to this as a weakness of the process of evolution or that somehow lack of complete resolution, in fact, renders all of evolutionary theory questionable. Nothing could be farther from the truth. Evolution could be invalidated if an objective team of researchers found a human fossil in the Cambrian, or that land plants predated ocean algal species. Mere incompleteness of understanding has never been a challenge to science or evolution as a process, only a challenge pointing out the need for better studies. Absence of evidence is not the same as evidence of absence. In fact, modern molecular methods and classical understanding of morphology are showing us in ways never before thought possible the unity of life on this planet.

Conclusions

Understanding of phylogenetic trees and the underlying process of creation has become central to the understanding of comparative biology at multiple levels. Phylogenies put vast amounts of biological data into an integrated whole (Baum & Offner 2008) not unlike GIS databases integrate geography with demographics and economics of a region. It adds depth and connectivity to data that to a non-specialist might seem completely unrelated. As Theodor Dobzhansky (1973) said, "Nothing in biology makes sense, except in the light of evolution."

Acknowledgements: I would like to thank Dr. T.M. (Mac) James, Dr. Mark Maloney, and others, for comments that greatly improved this manuscript. All remaining errors are mine. I would also like to thank Dr. Alan Harvey, who first taught me to *understand* cladistics.

Notes and references

**Corresponding author E-mail: jstaton@uscb.edu*

Avise JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology* **7**: 371-379.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA. 484 pp.

Baum DA and Offner S. 2008. Phylogenics & tree-thinking. *The American Biology Teacher* 70:222-229.

Brooks DR & McLennan DA (1991) *Phylogeny, Ecology, and Behavior: a Research Program in Comparative Biology*. Chicago: University of Chicago Press. 441 pp.

Dobzhansky T (1973) Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher* **35**:125-129.

Felsenstein J (1978). Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401–410.

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* **155**:279-84.

Goldsmith DW (2003) The great clade race: presenting cladistic thinking to biology majors & general science students. *The American Biology Teacher* **65**: 679-682

Hall BG (2011) *Phylogenetic Trees Made Easy: A How To Manual*. Sinauer Associates, Inc.; Fourth edition. 255 pp.

Hennig W (1950) *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin. 370 pp.

Hennig W (1966) *Phylogenetic Systematics*. Urbana, IL: University of Illinois Press.

Li W-H (1997) *Molecular Evolution*. Sinauer Associates, Inc. 487 pp.

Nei M (1996) Phylogenetic analysis in molecular evolutionary genetics. *Annual Review of Genetics* **30**: 371-403

Page DM & Holmes EC (1998) *Molecular Evolution: A Phylogenetic Approach*. Wiley-Blackwell. 352 pp.

Saint-Hilaire GE (1822) Considérations générales sur la vertebre. *Mémoires du Muséum d'Histoire Naturelle* **9**: 89-119

Saitou N & Nei M (1987) The neighbor-joining method: a new method for constructing phylogenetic trees. *Molecular Biology & Evolution* **4**: 406-425.

Sober E (1981) The principle of parsimony. *British Journal for the Philosophy of Science* **32**: 145–156.

Sokal RR & Michener CD (1958**)** A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**: 1409-1438.

Sokal RR & Sneath PHA (1963) *Principles of Numerical Taxonomy*. Freeman & Co., San Francisco. 359 pp.

Staton J (1998) Tastes like chicken? *Annals of Improbable Research* **4**(4): 5-9.

Wheeler WC (1990) Nucleic acid sequence phylogeny and random outgroups. *Cladistics* **6**:363 – 368.