8-31-2015

Model of the Use of Evolutionary Trees (MUET)

Yi Kong
Purdue University, ykong@purdue.edu

Trevor Anderson
Purdue University, ander333@purdue.edu

Nancy Pelaez
Purdue University, npelaez@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/pibergiim

 Part of the Curriculum and Instruction Commons, and the Science and Mathematics Education Commons

A Glossary of Terms for the Model of the Use of Evolutionary Trees (MUET)

by Yi Kong, Trevor Anderson, and Nancy J. Pelaez
Purdue University
[*Note to instructors: Edit bracketed sections*]
Modified by [*Your name, your institute, and the date*]

This Glossary and the interactive Model of the Use of Evolutionary Trees (MUET) can be found online at http://bilbo.bio.purdue.edu/~npelaez/TreeThinking

Absolute time scale: Estimation of the time in history when the entities on the tree appeared and/or disappeared.

Agglomerative: A "bottom up" approach where "each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy" (Rai & Singh, 2010, p. 3).

Amino acid: An amino acid, which contains an amino group (-NH2) and an acidic carboxyl group (-COOH), is the smallest unit of a protein. The order of amino acids determines the structure and function of proteins (Raven, Johnson, Losos, Mason, & Singer, 2014).

Average linkage clustering: This commonly used linkage criteria is for merging clusters based on the mean distance between elements of each cluster (Krebs, 1999). UPGMA is a method that uses this criterion.

Bayesian inference: In statistical analysis, Bayesian inference is a method of inference that used Bayes' rule based on prior knowledge of probabilities to update the probability estimate

for a hypothesis (Box & Tiao, 2011). In phylogeny, Bayesian inference is based on the

posterior probabilities of phylogenetic trees (Huelsenbeck & Ronquist, 2001). This

method differs from maximum likelihood methods only in the use of a prior distribution

of the inferred quantity (Felsenstein, 2004).

Bifurcating trees: In bifurcating trees, every interior node is connected to three other branches

and every tip node is connected to only one branch (Figure 1 and Figure 2). From each

interior node of these trees, there are two branches leading toward a tip.
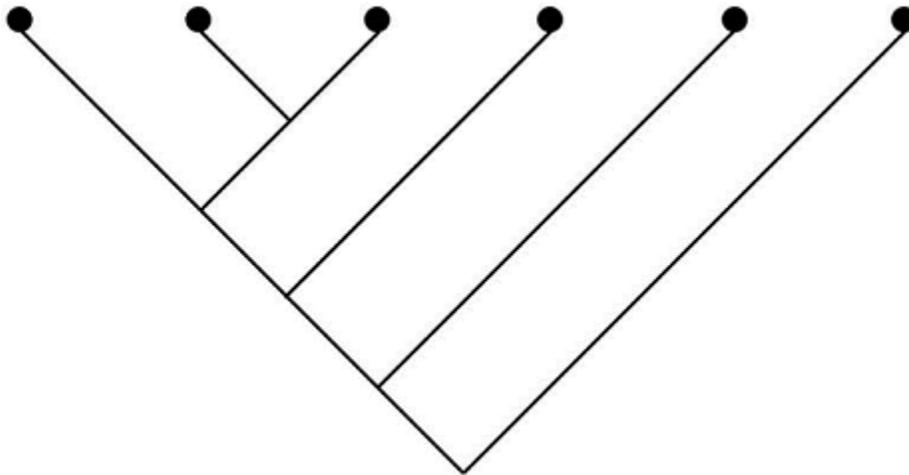
Figure 1 Rooted bifurcating tree
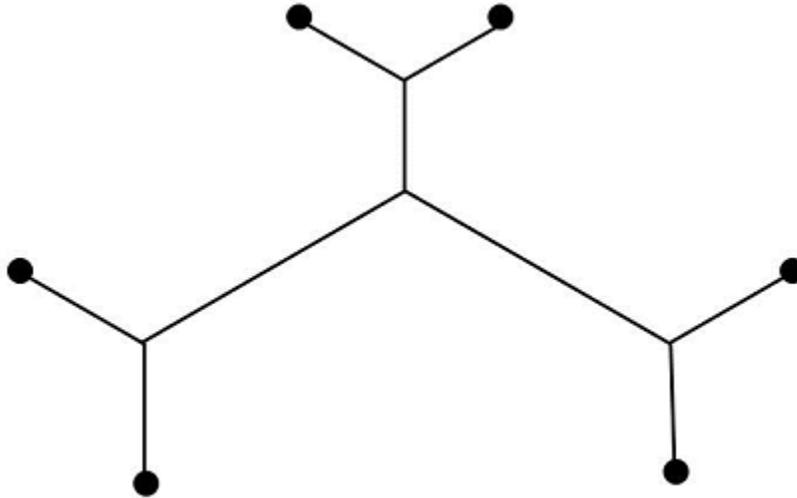
Figure 2 Unrooted bifurcating tree

Branches: The lines that connect tips with root and/or nodes are called branches.

Cell: The basic unit of life (Raven, Johnson, Losos, Mason, & Singer, 2014). Every known living

organism is either a single cell or consists of an ensemble of many cells (Morrs, Hartl,

Knoll, & Lue, 2013).

Cell data: The data are identified as cell as well as cell lines of organisms.

Cell level entities: These entities are identified as cell as well as cell lines of organisms.

Cell line: A cell culture with an indefinite life span is called a cell line (Lodish et al., 2008).

Chromosome: In eukaryotes, the vehicle to physically transmit hereditary information from one

generation to the next; each chromosome consists of a large DNA molecule together with

a small number of proteins (Freeman, 2008; Raven, Johnson, Losos, Mason, & Singer,

2014).

Chronology: The science of ordering events by occurrence in time (O'Hara, 1988) is chronology.

Tree-thinking attempts to show a temporal sequence in the appearance of features

identified in the targeted data sources.

Clade: A part of a rooted tree that can be separated from the rest of the tree by cutting a single

branch (Baum & Smith, 2012). This group of organisms will share a common ancestor as

well as derived characteristics that are typical for the clade (Raven, Johnson, Losos,

Mason, & Singer, 2014).

Cladistics: Cladistics is an approach to classify organisms based on whether the organisms share

common characteristics inherited from their latest common ancestor (Janvier, 1984).

Tree-thinking attempts to show how new traits relate to the divergence in the targeted

data sources on the tree.

Cladogram: Cladogram is a kind of evolutionary tree that communicates the pattern of branching

but not the lengths of branches. It provides all the information relevevant to determining

the degree of evolutionary relatedness between taxa (Baum & Smith, 2012).

Cluster analysis: This technique is for classifying information into manageable meaningful

clusters without having prior knowledge about which elements belong to which cluster

(Burns & Burns, 2008). The same technique broadly used in mathematics, computer

sciences, statistics, biology and economics is used as an exploratory tool to group genes

based upon the similarity of gene expression responses across several conditions

(Gordon, 1999). Trees illustrate or represent cluster analysis.

Cluster: A group of things or organisms that share similar traits.

Compare with published trees: Scientists conduct research to support or oppose a hypothesis by

testing expectations or predictions based on the hypothesis or on trees previously

published in the literature.

Common ancestry: "The existence of ancestral organisms or taxa that are ancestral to more than

one taxon (living or extinct)" (Baum & Smith, 2012, p.442).

Complete-linkage clustering: This commonly used linkage criteria is for merging clusters based

on the maximum distance between elements of each cluster (Krebs, 1999).

Construction of trees: According to our analysis of papers published in *Science* journals from

2012 to 2013, this refers to the methods that scientists use to construct trees. These

methods include phylogenetic analysis, cluster analysis, schematic drawing, distance-

based methods, hierachical clustering, maximum parsimony, maximum likelihood,

Bayesian methods or based on a literature review. Of these methods, distance-based methods, maximum parsimony (MP), and maximum likelihood (ML) are three major basic principles and methods used to build evolutionary trees (Delsuc, Brinkmann, & Philippe, 2005; Hershkovitz & Leipe, 1998).

Data sources: According to our analysis of papers published in *Science* journals from 2012 to 2013, the data sources refer to the data that scientists use to construct trees. The data sources represent the variable properties of the subjects of focus for the specific scientific research. The data collected from research subjects can be molecules, cells, tissue types, or morphological traits.

Deoxyribonucleotides: The smallest unit of DNA. It is a kind of nucleotide that consists of deoxyribose sugar (Freeman, 2008).

Derived trait: An inherited similarity from the most recent common ancestor of the entire group or clade (Raven, Johnson, Losos, Mason, & Singer, 2014). According to Baum and Smith (2012) derived is not a term that should be applied to taxa.

Distance-based method: One of the tree-building methods based on calculated distance between each pair of species to find a tree that predicts the observed distances as closely as possible (Felsenstein, 2004). The actual data is discarded using this method and fixed distances are used in the derivation of trees (Strickler, 2014). Neighbor-joining (NJ) and

Unweighted Pair Group Method with Arithmetic Mean (UPGMA) are examples of

distance-based methods.

Divisive: A "top down" approach in which "all observations start in one cluster, and splits are

performed recursively as one moves down the hierarchy" (Rai & Singh, 2010, p. 3).

DNA (deoxyribonucleic acid): DNA composed of two complementary chains of nucleotides

which wound in a double helix. It contains genetic material of all organisms which can be

transmitted from one generation to the next generation (Raven, Johnson, Losos, Mason,

& Singer, 2014).

Entities on trees: The tips of tree can be labeled with various entities which represent the subject

of focus for specific scientific research. Some entities also represent the variable

properties of the subject. Overall, these entities can represent organisms at molecular

level, cell level, tissue level as well as represent the morphological traits of organisms.

Evolutionary tree (phylogenetic tree): A diagram generated to depict a hypothesis of

evolutionary relationships between populations or species, also called a phylogenetic tree

or cladogram (Halverson, Pires, & Abell, 2011; Meir, Perry, Herron, & Kingsolver, 2007;

Novick & Catley, 2007; Raven, Johnson, Losos, Mason, & Singer, 2014). It provides all

the information relevant to determining the degree of evolutionary relatedness between

taxa (Baum & Smith, 2012). Generally, an evolutionary tree consists of a root, nodes,

branches, and tips. At the tips are sister taxa and an outgroup taxon. Figure 3 depicts a

general evolutionary tree and its components.
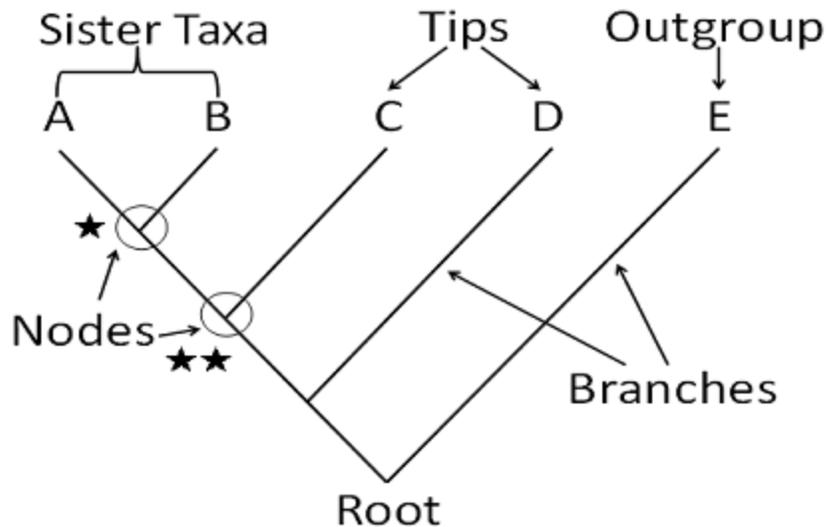
Figure 3 Evolutionary tree

Family tree: A chart that represents the relationships among at least three generations of the

multigenerational family, also called a family diagram. Factual information such as

physical problems can be visually recorded on the tree (Butler, 2008).

Gene expression profile: A unique pattern or inventory of gene expression from certain types of

cell, organs, or tissues, under specified physiological or disease conditions (Smith, Datta,

Smith, Campbell, Bentley, & McKenzie, 2000).

Gene expression: The process of selectively reading and using genetic information to make

proteins required for a cell to operate (Clark, 2005; Lodish et al., 2008).

Gene: The fundamental molecular unit of heredity of a living organism. Genes are made up of several hundred to many thousand nucleotides in a sequence (Raven, Johnson, Losos, Mason, & Singer, 2014).

Heat map: A heat map is a graphical representation of data where rectangular color coded arrays represent the intensity of signals as the dependent variable (Pleil, Stiegel, Madden, & Sobus, 2011). However, the purpose for a heat map is not to portray a chronology. Instead a heat map is used to show patterns in variable functions such as to characterize responses to conditions such as variations in temperature or nutrients, or to track changes over time.

Hierarchical clustering: This method of cluster analysis is aimed at building a hierarchy of clusters. There are two types of strategies for hierarchical clustering: agglomerative and divisive (Steinbach, Karypis, & Kumar, 2000). Most methods of hierarchical clustering are distanced-based methods in which clusters are merged based on distances between clusters. The commonly used linkage criteria are: complete linkage clustering, single-linkage clustering, and average linkage clustering (Olson, 1995).

Homology: In biology, homology refers to any trait of organisms that is derived from a common ancestor (Hall, 2012). Tree-thinking attempts to show how targeted data sources share common (homologous) features inherited from a recent common ancestor on the tree.

Homoplasy: In biology, homoplasy refers to the presence of any similar trait of organisms that is not derived from a common ancestor. Homoplasy occurs for some reason other than inheritance from common or shared ancestry (Lieberman, Wood, & Pilbeam, 1996).

Labeled tree: Labeled trees have distinct labels on each tip (Felsenstein, 2004).

Linkage criterion: Linkage criteria (plural) specify "the dissimilarity of sets as a function of the pairwise distances of observations in the sets" (Sasirekha & Baby, 2013, p. 1).

Literature: According to our analysis of papers published in *Science* journals from 2012 to 2013, some evolutionary trees were constructed based on literature that depicted an evolutionary relationship with a particular tree shape and with entities on the tree.

Maximum Likelihood (ML): ML is a kind of standard statistical method used to compute moderate numbers of data (e.g., nucleotide sequences) (Felsenstein, 2004). It "finds the topology and branch lengths that have the highest probability of producing the observed multiple sequence alignment" (Hershkovitz & Leipe, 1998, p. 191). Maximum likelihood attempts to estimate the actual amount of change according to the evolutionary model (Hershkovitz & Leipe, 1998; Strickler, 2014; Swofford, 1996).

Maximum Parsimony (MP): MP is a kind of standard statistical method that attempts to find "the tree that explains with the fewest number of discrete steps"(Hershkovitz & Leipe, 1998, p. 191) to explain an origin for all the differences. It uses the original data set for

inference which falls under the philosophy of "the simpler hypotheses are preferable to the complicated ones" (Swofford, 1996). This method is widely used for morphological data (Lamboy, 1994).

Molecular data: These data are identified as the sequence of nucleotides, amino acids, and gene expression profiles that are characteristic of the organisms.

Molecular level entities: These entities are identified as molecular which include gene, chromosome, the sequences of nucleotide，gene expression profile, the sequences of amino acid, and proteins of organisms.

Morphological data: The data are identified as morphological traits of organism. Morphological traits are the structural traits an organism expresses. For example, the leg feathers of dinosaurs is a kind of morphological trait.

Morphological traits entities: These entities are identified as morphological traits of organism. The morphological traits mean the structural traits an organism expresses. For example, the leg feather of dinosaurs is a kind of morphological trait.

mRNA (messenger RNA): An intermediate form of the information in DNA that is transported out of the eukaryotic nucleus to the cytoplasm for ribosomal processing to translate genetic information from DNA into proteins (Raven, Johnson, Losos, Mason, & Singer, 2014).

Multifurcating trees: In multifurcating trees, interior nodes connect to two or more branches. A

bifurcating tree is a type of multifurcating tree. Figure 4 and Figure 5 show

multifurcating trees that are not bifurcating trees because some nodes are connected to
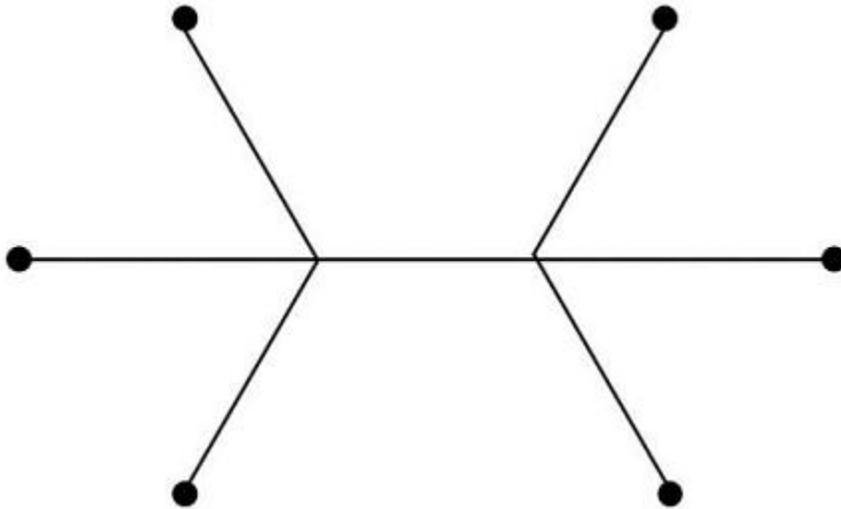
three or more branches.

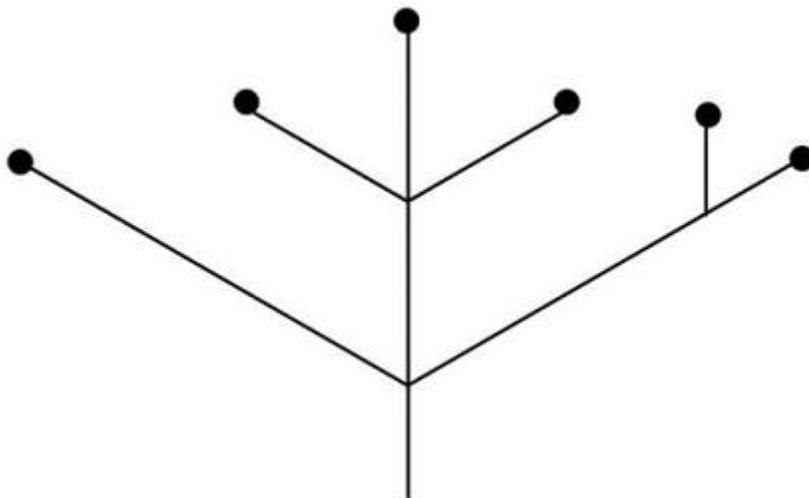Figure 4 Unrooted multifurcating tree

Figure 5 Rooted Multifurcating tree

Neighbor-Joining (NJ): This method is aimed at reconstructing phylogenetic trees from

        evolutionary distance data (Baum & Smith, 2012; Saitou & Nei, 1987). The tree is

        constructed by linking the least distant pairs of nodes as defined by a modified matrix.

        The modified distance matrix is constructed by adjusting the separation between each

        pair of nodes on the basis of the average divergence from all other nodes (Hershkovitz

        1998; Strickler, 2014; Swofford, 1996).

Node: Each node represents the divergence of organisms or species. Two species shared their

        most recent common ancestor at the nearest node. In Figure 1, A and B shared their most

        recent common ancestor at the node marked with a single asterisk, while A, B, and C

        shared a more distant common ancestor at the node with two asterisks.

Nucleic acids: Nucleic acids are made up of nucleotides. DNA and RNA are the two main

        varieties of nucleic acids (Freeman, 2008).

Nucleotide: A kind of molecule that consists of three components: a phosphate group, a sugar,

        and a nitrogenous base. There are two types of nucleotides: ribonucleotides and

        deoxyribonucleotides (Freeman, 2008).

Origin of species: An evolutionary tree traces the origin of features identified in the targeted data

        sources that are selected to characterize each species.

Outgroup (taxon): An outgroup is a species or group of organisms that is closely related to the group under study but not a member of the group under study (Raven, Johnson, Losos, Mason, & Singer, 2014). In Figure 1, E is related to other branches only at the root, so it is called the outgroup taxon.

Parsimony: Parsimony means extreme unwillingness to use resources. The principle of parsimony is an approach that connects things in the simplest or most economical way when there are two equally adequate models (Parsimony, 2014).

Peptides: Peptides are polymers made up of a linear arrangement of amino acids (Raven, Johnson, Losos, Mason, & Singer, 2014).

Phylogenetic analysis: The process of phylogenetic analysis can be summarized in steps. The first step involves alignment of traits such as the nucleotide or amino acid sequences for the taxa of interest to determine the presence of a phylogenetic signal. A next step is to decide the most appropriate tree-building method for a specified data set, and then to choose a strategy to find the best tree under the selected optimality criterion. Finally, the tree obtained must be scrutinized to determine the level of confidence that can be placed on the results (Hillis, 1993; Strickler, 2014). Phylogenetic analysis is done by various methods of cluster analysis.

Phylogenetic tree: also called evolutionary tree, it is a model for the evolutionary history generated to depict a hypothesis of evolutionary relationships between populations or

species. Although some scientists may use tree-shaped diagrams to describe relationships among individual entities with a focus on their characters (Morrison, 2013), most of these diagrams are unable to show evolutionary history. For example, some heat maps depict relationships among various modified genes. These diagrams are not phylogenetic trees because phylogenetic tree refers to a tree diagram that can depict evolutionary history in which the entities on the tree are usually a group of organisms but not individual organisms (Baum & Smith, 2012).

Proteins: Proteins are made up of one of more peptides folded into tertiary and quaternary (three-dimensional) structure. Some proteins form structures in organisms and proteins also provide functional regulation of organisms (Raven, Johnson, Losos, Mason, & Singer, 2014).

Reasoning with trees: Generally speaking, reasoning refers to the ability of thinking and understanding things in a logical way (Hacker, 2011). In our study, reasoning with trees refers to the reasoning that scientists do with an evolutionary tree. According to our analysis of papers published in *Science* journals from 2012 to 2013, scientists have seven types of reasons for publishing a tree: cladistics, homology, homoplasy, chronology, trace the origin of traits, absolute time scale, and compare with published trees.

Representation of trees: According to our analysis of papers published in *Science* journals from 2012 to 2013, the evolutionary trees as represented in scientific articles consisted of

different types of tree shapes as well as the entities on the tree. Root, node, tip, branch, and taxon are components of an evolutionary tree.

Ribonucleotides: A kind of nucleotide in which the sugar is ribose (Freeman, 2008).

RNA (ribonucleic acid): The polymer of ribonucleotides is called RNA. There are various RNAs in organisms: mRNA, rRNA, tRNA.

Root: The root of an evolutionary tree indicates that all organisms on the tree share a common ancestor. In Figure 1, all the species (A, B, C, D, and E) share a common ancestor. Root is the earliest node in a tree from which the descendant lineages arise (Baum & Smith, 2012).

Rooted bifurcating tree: A tree can be rooted and bifurcating. Figures 3 is one example of a rooted bifurcating tree.

Rooted multifurcating tree: A tree can be rooted and multifurcating. Figure 4 is one example of a rooted multifurcating tree.

Rooted tree: A rooted tree has one vertex on the tree designated as the root, which represents the earliest node, and the tips and branches have an orientation with more recent lineages farther away from the root (Baum & Smith, 2012). According to Felsenstein (2004), an unrooted tree can be rooted at one of its nodes.

rRNA (ribosomal RNA): A class of RNA found in the ribosomes of cells (Raven, Johnson,

Losos, Mason, & Singer, 2014).

Schematic drawing: According to our analysis of papers published in *Science* journals from 2012

to 2013, some evolutionary trees were biologists' schematic drawings to represent their

ideas about evolutionary relationships of the entities on the tree.

Single-linkage clustering: This commonly used linkage criteria is for merging clusters based on

the minimum distance between elements of each cluster (Krebs, 1999).

Sister Taxa: Two groups that share a more recent common ancestor that is not shared with other

species. In Figure 1, A and B are called sister taxa.

Taxon (plural: taxa): In biology, a taxon is adjudged by taxonomists to be a group of one or more

populations of organisms that are considered to be a unit (Futuyma, 2013). Taxa are

represented by the tips of a phylogenetic tree (Baum & Smith, 2012).

Tips: In an evolutionary tree, the entities whose relationships are studied and depicted are called

tips (Baum & Smith, 2012). In Figure 1, A, B, C, D, and E represent organisms that

currently exist.

Tissue: In biology, a tissue is a group of similar cells that act as a function unit in organisms

(Raven, Johnson, Losos, Mason, & Singer, 2014).

Tissue data: The data are identified as tissue types in organisms.

Tissue level entities: These entities are identified as tissue types in organisms.

Trace the origin of traits: The origin of the traits of the entities on a tree explain the history of

how and why the traits disappeared or appeared by depicting splitting of lineages in a tree

(Baum & Smith, 2012).

Tree thinking: A technique to chronicle the sequence of events when new heritable traits

emerged (Brooks & McLennan, 1991; O'Hara, 1988). Evolution explanations depend on

this chronicle (O'Hara, 1988). In Figure 6, the evolution process shows species having

evolved from the oldest to the youngest represented as the vertical axis, with progress of

time from the bottom to the top. Figures b, c, and d draw the clades of B and C and/or the

clade of A, B, and C flipped with respect to each other. Although these four Figures (a, b,

c, and d) look different, they represent the same evolutionary explanations: species B and

species C are closely related and they share a more recent common ancestor when

compared with species A, D, E, and F. Species A, B and C shared a more common

ancestor when compared with species D, E, and F. Tree thinking ability refers to "the

ability to read and interpret phylogenetic trees and use trees to accurately represent the

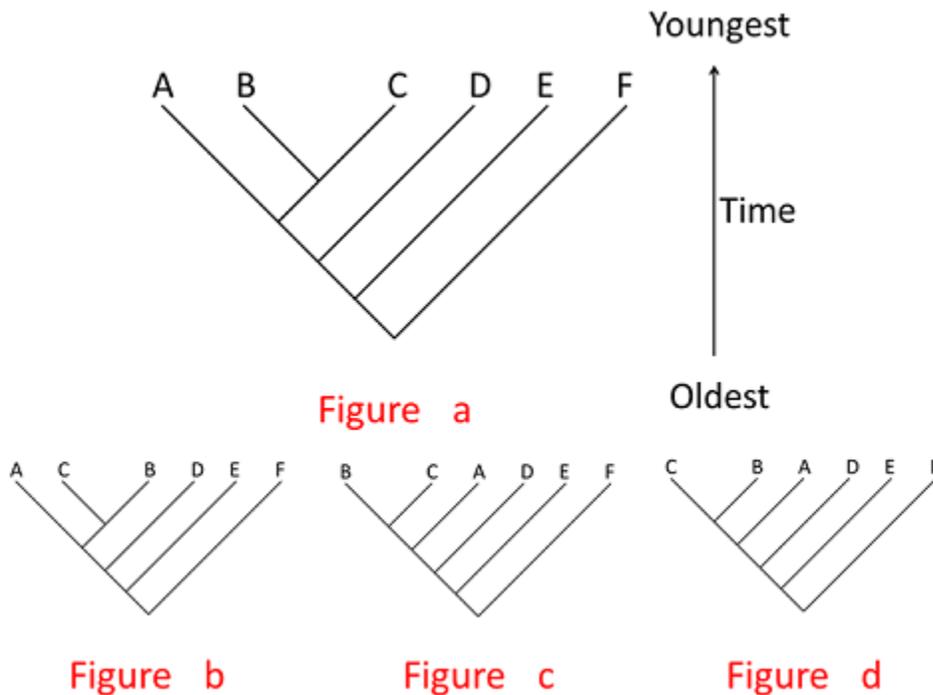evolutionary process" (Baum & Smith, 2012, p.457).

Figure 6 Evolutionary trees

tRNA (transfer RNA): A class of RNA that recognizes information in mRNA to help position the

amino acid on the ribosome. It helps bring about the linear arrangement of amino acids

during translation of an mRNA into a peptide sequence (Raven, Johnson, Losos, Mason,

& Singer, 2014).

Types of tree shapes: Trees can be sorted as labeled trees and unlabeled trees. Trees can be also

sorted as rooted trees and unrooted trees as well as bifurcating trees and multifurcating

trees. Thus, there are four basic types of tree shapes: rooted bifurcating tree shape (Figure

3); rooted multifurcating tree shape (Figure 4); unrooted bifurcating tree shape (Figure 5);

and unrooted multifurcating tree shape (Figure 6).

Unrooted bifurcating trees: A tree can be unrooted and bifurcating. Figure 5 is one example of an unrooted bifurcating tree.

Unrooted multifurcating trees: A tree can be unrooted and multifurcating. Figure 6 is one example of an unrooted multifurcating tree.

Unrooted tree: A tree that has no component designated as the root has no time axis or earliest (oldest) ancestor identified.

Unrooted tree and rooted tree: Figure 7 shows an unrooted tree as well as two rooted versions of the same tree. In Figure 7, the unrooted tree does not show complete chronology because any point within this tree can be indicated as a root. For example, if the ancestor at dot 5 were to be indicated as the root, meaning it existed first, then the group at dot 6 would be an ancestor for tips 3 and 4, but not for tips 1 and 2 that were derived from node A which would have diverged before the existence of the group at dot 6. In contrast, if the ancestor at dot 6 were to be indicated as the root, meaning it existed first, then dot 6 would be an ancestor for all of the tips 1, 2, 3 and 4. Placing dot 6 at the root means it existed before dot 5 which would then only be an ancestor for tip 1.

Figure 7 An unrooted tree and two examples of rooted possibilities for that unrooted tree

Unweighted Pair Group Method with Arithmetic Mean (UPGMA): The UPGMA method is used

to infer phylogenies if evolutionary rates were assumed to be the same in all lineages

(Felsenstein, 2004). This method joins the shortest distance for two nodes at each stage of

the clustering (Hollich, Milchert, Arvestad, & Sonnhammer, 2005). This method of

hierarchical clustering is based on the mean distance between elements of each cluster as

a linkage criterion.

References

Baum, D. A., & Smith, S. D. (2012). *Tree thinking: An introduction to phylogenetic biology.*

Englewood, CO: Roberts and Company Publishers.

Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40).

Indianapolis, IN: John Wiley & Sons.

Brooks, D. R., & McLennan, D. A. (1991). *Phylogeny, ecology, and behavior: A research program in comparative biology.* Chicago, IL: University of Chicago Press.

Burns, R. P., & Burns, R. (2008). *Business research methods and statistics using SPSS*. London: Sage.

Butler, J. F. (2008). The family diagram and genogram: Comparisons and contrasts. *The American Journal of Family Therapy, 36,* 169-180.

Clark, D. P. (2005). *Molecular biology: Understanding the genetic revolution*. UK: Elsevier Academic Press.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics, 6,*361-375.

Felsenstein, J. (2004). *Inferring phylogenies* (Vol. 2). Sunderland, MA: Sinauer Associates.

Freeman, S. (2008). *Biological science* (3rd ed.). San Francisco, CA: Pearson.

Gordon, A. (1999). *Classification* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Hacker. (2011). In Merriam-Webster.com. Retrieved May 8, 2011, from http://www.merriam-webster.com/dictionary/hacker

Hall, B. K. (Ed.). (2012). *Homology: The hierarchial basis of comparative biology*. San Diego, CA: Academic Press.

Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Science Education, 95,* 794-823.

Hershkovitz, M. A., & Leipe, D. D. (1998). Phylogenetic analysis. In Baxevanis, A.D. & Ouellette, B.F.F. (Eds.), *Bioinformatics: A practical guide to the analysis of genes and proteins* (pp. 189-230). New York: Wiley Interscience.

Hollich, V., Milchert, L., Arvestad, L., & Sonnhammer, E. L. (2005). Assessment of protein

distance measures and tree-building methods for phylogenetic tree reconstruction.

Molecular Biology and Evolution, 22, 2257-2264.

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic

trees. *Bioinformatics, 17,* 754-755.

Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic networks: concepts, algorithms

and applications*. New York, NY: Cambridge University Press.

Janvier, P. (1984). Cladistics: Theory, purpose and evolutionary implications. *Evolutionary

Theory: Path into the Future, 39-75.*

Krebs, C. J. (1999). *Ecological methodology* (Vol. 620). Menlo Park, CA: Benjamin/Cummings.

Lamboy, W. F. (1994). The accuracy of the maximum parsimony method for phylogeny

reconstruction with morphological characters. *Systematic Botany*, 19.

Lieberman, D. E., Wood, B. A., & Pilbeam, D. R. (1996). Homoplasy and early Homo: An

analysis of the evolutionary relationships of *H. habilis sensu stricto* and *H. rudolfensis*.

Journal of Human Evolution, 30, 97-120.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M. Scott, M. P., Bretscher, A., Ploegh, H., &

Matsudaira, P. (2008). *Molecular Cell Biology*. New York, NY: Macmillan Press.

Meir, E., Perry, J., Herron, J., & Kingsolver, J. (2007). College students' misconceptions about

evolutionary trees. *The American Biology Teacher, 69,* 71-76.

Morris, J., Hartl, D., Knoll, A., & Lue, R. (2013). *Biology: How life works*. New York, NY:

Macmillan Press.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: the influence of a

Gestalt Perceptual Principle. *Journal of Experimental Psychology: Applied*, *13,* 197-223.

O'Hara, R. J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary

biology. *Systematic Zoology, 37,* 142-155.

Olson, C. F. (1995). Parallel algorithms for hierarchical clustering. *Parallel Computing, 21,*

1313-1325.

Parsimony. (2014). In Oxford Dictionaries' online dictionary. Retrieved

fromhttp://www.oxforddictionaries.com/us/definition/american_english/parsimony

Pleil, J. D., Stiegel, M. A., Madden, M. C., & Sobus, J. R. (2011). Heat map visualization of

complex environmental and biomarker measurements. *Chemosphere, 84,* 716-723.

Rai, P., & Singh, S. (2010). A survey of clustering techniques.*International Journal of Computer

Applications, 7,* 156-162.

Raven, P., Johnson, G., Losos, J., Mason, K., & Singer, S. (2014). *Biology* (10th ed.). New York,

NY: McGraw-Hill.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing

phylogenetic trees. *Molecular Biology and Evolution, 4,* 406-425.

Sasirekha, K., & Baby, P. (2013). Agglomerative hierarchical clustering algorithm − A Review.

International Journal of Scientific and Research Publications, 3, 1-3.

Smith, A., Datta, S. P., Smith, G. H., Campbell, P. N., Bentley, R., & McKenzie, H. A. (2000).

Oxford dictionary of biochemistry and molecular biology. New York, NY: Oxford

University Press.

Steinbach, M., Karypis, G., & Kumar, V. (2000, August). *A comparison of document clustering

techniques.* In *KDD workshop on Text Mining* (Vol. 400, No. 1, pp. 525 – 526).

Strickler, P. (2014, January 13). A brief review of the common tree-building methods used in

phylogenetic inference. Retrieved

from http://bioinfo.mbb.yale.edu/mbb452a/projects/Patricia-M-Strickler.html

Swofford, D.J. & G.J. Olsen, G.J. (1996). *Molecular systematics* (D.M. Hillis and C. Moritz,

eds.). Sunderland, MA: Sinauer Associates.