

What is and why do we have to know the phylogenetic tree?

Editorial

Sherif M Abaza

Medical Parasitology Department, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

All creatures are related to each other in a branching evolutionary tree with a common ancestry, and according to their genetic diversity, they separate as leaves on the external branches of the tree. Phylogeny is defined as the history of descent from the common ancestry, whereas the phylogenetic tree shows the lines of evolutionary descent of different species from the common ancestor. In other words, it is the visual representation of the relationship between different organisms or species in relation to their evolutionary time. The phylogenetic tree, also called a dendrogram, demonstrates the distance of one group of creatures from other groups indicating the degree of relationship. That is to say closely related groups, also known as clades, are located on branches close to each other. In spite of being intellectual, *i.e.* depending on analytical hypotheses, it provides a convenient simple method for studying phylogenetic relationships. Accordingly, phylogenetic analysis is the ability to read and interpret phylogenetic trees and to apply these data to accurately represent the evolutionary process.

Keywords: epidemiological studies, dendrogram, gene sequencing, genetic diversity, species identification.

Received: 15 July, 2020, **Accepted:** 21 August, 2020.

Corresponding Author: Tel. +20 1005243428, E-mail: smabaza@hotmail.com

Print ISSN: 1687-7942, **Online ISSN:** 2090-2646, **Vol. 13, No. 2, August, 2020.**

What we have to know?

The term phylogenetic is derived from the two ancient Greek words (*phûlon*) and (*gênesis*). The first means "race or lineage", and the latter means "origin or source". Parasitologists have to know that phylogeny of a group of species, called a clade, is based upon similarities and/or differences in their biological, morphological, and genetic characteristics. To comprehend a phylogenetic tree, one must keep in mind that it is composed of internal branches that connect two nodes or it is formed of just one node with one external branch. The node is a branching diverging point, while the crown node, *i.e.* the tree root, represents the last common ancestor of the clade of interest. The clade is a monophyletic group of creatures that can be separated from the rest of the tree, including its root, by cutting a single branch. In other words, monophyletic species are all the descendants of a particular ancestor, included in one clade. The root is the earliest node in the phylogenetic tree that represents the last common ancestor of all tree tips or

leaves. The latter are also known as the entities (*e.g.*, taxa, genes) whose relationships are depicted using the tree diagram. Therefore, taxa are a named group of organisms that are represented by tree tips or leaves (Figure 1 A-C).

A cladogram represents the branching pattern of taxa and provides all the information relevant to determine the degree of relatedness between the known group of organisms. This means that the cladogram indicates that the more closely related organisms are those which share more recent common ancestors. It also provides time distance between evolutionary lineage splitting which is the division of an ancestral lineage into two or more descendant lineages. It is worth mentioning that the history of lineage splitting, depicted by a rooted tree including the list of all its clades, is termed the tree's topology. During the last two decades, several algorithmic programs were developed for phylogenetic analyses for estimation of the lineage divergence time between species.

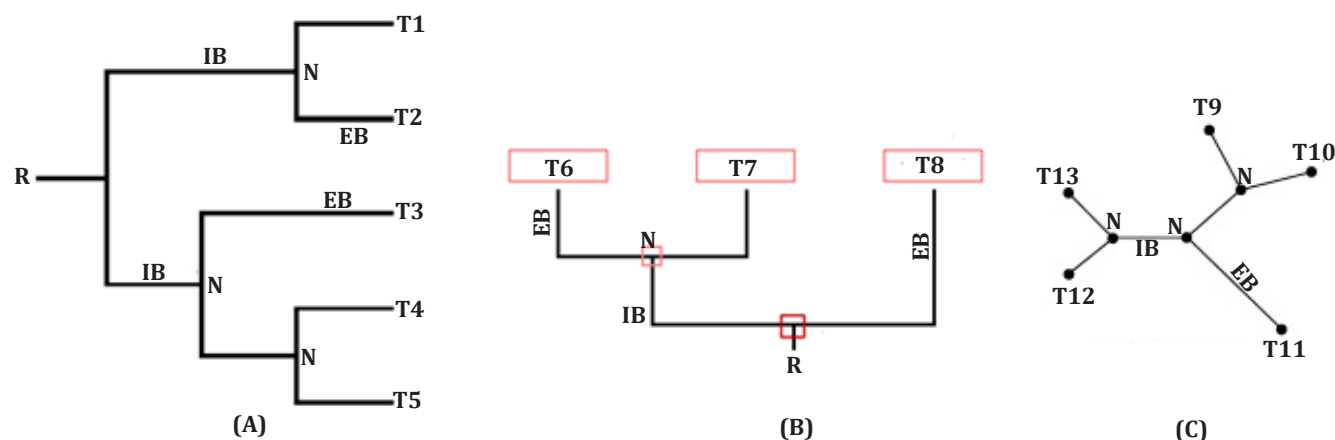


Fig. 1. Dendrogram representing rooted tree (A, B) and unrooted tree (C). It is observed that monophyletic clade includes taxa 1/2, 3, 4/5, 6/7, 8, 9/10, 11, and 12/13. In the rooted tree, Ts 1-8 are children (one ancestor), while Ts 9-13 are neighbors in unrooted tree. Note that T3 and T8 are outgroups for Ts 4/5, and 6/7, respectively.

EB: external branch; **IB:** internal branch; **N:** node; **R:** root; **T:** taxon.

Depending on the research question being addressed, phylogenetic tree is either rooted or unrooted (Figure 1 A-C), *i.e.*, with or without a specified time axis, respectively. The first is a direct tree with an unique node (root) representing the most recent common ancestor of all tree entities. On the other hand, an unrooted tree illustrates the relatedness of the leaf nodes without making assumptions about ancestry. Entities or taxa of an unrooted tree do not require the identification of the ancestral root. Both types of trees can be divided into two (binary) or more (multi) branches, and in these instances, branches of rooted and unrooted trees are known as children and neighbors, respectively. Accurate root identification is crucial in order to show the evolutionary path of the biological species. Analysis of rooted phylogenetic trees is much required by research biologists because the unrooted tree shows only the relationships among species without definite evolutionary path. However, when the research question is focused only on relationships among species, rather than the evolutionary change, analysis of the unrooted phylogenetic tree is required.

Before the amazing evolution of informatics technology, organisms were classified into classes, orders and families based on morphological criteria (traits). However, over the evolutionary history of a species, these traits may be gained or derived and then lost to revert back to the ancestral form. Therefore, counting on morphological traits is not feasible to construct the phylogenetic tree of the required species. Nowadays, DNA sequencing of a gene or certain protein (genetic traits) shows few differences in amino acids sequence in closely related species rather than in species of other clades. Genetic data include mitochondrial DNA sequences, ribosomal RNA gene sequences, and any genomic data of interest. To analyze sequence data and identify the most probable phylogenetic tree, computer programs and statistical algorithms are utilized. The most common method for constructing a rooted tree is the use of an uncontroversial outgroup. Outgroup organisms are taxa that are not initially assumed to be within the clade of interest (Figure 1 A-C). On the other hand, the unrooted tree can be generated from rooted ones by simply omitting the root node.

Several parameters are evaluated to assess methods used in construction of the phylogenetic tree including 1) efficient time and cost; 2) powerful in utilizing given data; 3) consistent answers in hypothesized data; 4) robustness, *i.e.* it overcomes the assumptions of the underlying hypothesis; and above all its ability to alert us when it is not good to use, *i.e.* falsifiability. It should be close enough to allow inference from trait data or molecular sequencing, but far enough to be a clear outgroup.

Why we have to know?

Previously, taxonomic classification of parasites relied on morphological similarity between families

to be assigned in the same order of each phylum. Now, after recent revolutionary molecular techniques, parasites are classified according to their molecular biology simultaneously with their morphological characters. In other words, species classification has become less static and more dynamic, *i.e.* phylogenetic tree construction aims to figure out the path of species descent from their ancestor, if we look at the modification pattern of the present-day organisms.

Taxonomic classification of eukaryotic protozoa remains a topic of debate, including problems in taxonomy of some protozoa such as species of *Blastocystis* and *Microsporidium*, and previously *Pneumocystis carinii*, now classified as fungi. In addition, several protozoal species are known by their high genetic diversity such as *Leishmania*, *Blastocystis*, and *Microsporidium* as well as *T. cruzi*. Also, it was reported that variable clinical presentations observed in several diseases caused by these protozoal species and others such as *E. histolytica*, *G. lamblia*, *T. vaginalis* and *Cryptosporidium* species are attributed to genetic diversity. In contrast, only few helminthes are characterized by high genetic diversity such as species of *Echinococcus*, *Taenia* and *Fasciola*. Therefore, phylogenetic analysis might help in species identification, genotyping and subtyping as well as taxonomic classification.

Limitations

Although a phylogenetic tree constructed on the basis of gene sequencing or genomic data in different species can provide evolutionary insights, its analysis has important limitations. Phylogenetic analyses may not accurately represent the evolutionary history of the included taxa because results are subjected to falsification by further studies, *e.g.* gathering of additional data, or analyzing the existing data with improved construction methods. Moreover, there are problems in basing phylogenetic analysis on a single gene sequencing because such trees constructed from one data source may differ from that constructed from another source. Therefore, great care is needed in inferring phylogenetic relationships among species. For this reason, recent phylogenetic analyses use a combination of genes that come from different genomic sources, or genes that would be expected to evolve under different selective regimes, so that homoplasy (false homology) would be unlikely to result from natural selection. A third limitation encountered in phylogenetic analyses is the absence of particular types of genes or any insertion and deletion, or any other event thought to contain an evolutionary signal such as gene mutation.

For further reading

- Baum D. Reading a phylogenetic tree: The meaning of monophyletic group. Nature Education 2008; 1(1): 190.
- Martínez-Aquino A. Phylogenetic framework for coevolutionary studies: a compass for exploring

jungles of tangled trees. *Curr Zool* 2016; 62(4):393-403.

- Kinene T, Wainaina J, Maina S, Boykin LM. Rooting trees, methods for. *Encyclopedia of Evolutionary Biology* 2016; 3:489-493.

Applications

1. Taxonomic position: In 2002, a group of investigators from Japan and USA conducted two phylogenetic analyses. Nucleotide sequences of the genes encoding small subunit ribosomal RNA, and cytosolic heat shock protein-70 (HSP70) of a *B. hominis* strain positioned it within stramenopiles, the closest relatives of alveolates where *P. falciparum* belongs^[1]. The second study was conducted utilizing the genes encoding mitochondrial HSP70 from amitochondriate protozoa, *G. lamblia*, *E. histolytica*, and *Encephalitozoon hellem*. They demonstrated that all sequences were located within the eukaryotic mitochondrial clade, suggesting that none of these protozoa are primitively amitochondriate, but they lost their mitochondria secondarily in their evolutionary past^[2]. Recently, and after the widespread application of phylogenomics and achievement of several discoveries of the major lineages of eukaryotes, the new eukaryotic tree was discussed. The reviewers claimed that eukaryotes tree is exclusively constructed on molecular phylogenetic evidences. Neither the biological nor morphological data are involved in the phylogenetic construction of almost all the new supergroups^[3].

2. Epidemiological studies: In the last decade, several phylogenetic analyses were performed and all demonstrated genetic diversity in *F. hepatica*^[4], *L. tropica*^[5], *T. saginata*^[6], *E. multilocularis*^[7], *G. lamblia*^[8], *Blastocystis* spp.^[9], *E. granulosus*^[10]. The most common genes utilized in reconstruction of the phylogenetic trees were those encoding cytochrome c oxidase subunit 1 (*cox1*), small-subunit ribosomal DNA (SSU rDNA), and internal transcribed spacer 1 (ITS1), a region between the genes encoding 18S rRNA and 5.8S rRNA. Other specific genes were also utilized in the phylogenetic studies to identify genotypes of *Cryptosporidium* spp.^[11], and *G. lamblia* assemblages^[12] such as those encoding *Cryptosporidium* HSP70, and beta-giardin (*bg*), respectively. It is worth noting that multilocus sequence typing (MLST) is utilized in investigating the genetic diversity of *T. vaginalis* Egyptian isolates. Three housekeeping genes were used; tryptophanase (*p1*), alanyl tRNA synthetase (*p8*) and DNA mismatch repair protein (*p13*)^[13]. All these studies demonstrated and evidenced the usefulness of phylogenetic analyses in conducting the epidemiological genotyping studies in parasites with high genetic diversity isolated from the same geographic area.

3. Species identification: By construction of phylogenetic trees, the investigators were able to identify *T. cruzi* isolates and discriminated it from other *Trypanosoma* spp., utilizing *cox1* gene^[14]. Also, in a phylogenetic tree constructed by utilizing gene encoding ITS1, the investigators grouped isolates of *L. major* and *L. tropica* in two separate clades^[15].

4. Isolates subtyping: It was reported that *T. cruzi* has seven discrete typing units (*Tc* I-VI and *Tc*bat), however, there is much controversy concerning the origin of its hybrid lineages. Also, *Blastocystis* spp. is one of the most common protozoa with genetic diversity that might be associated with variable clinical presentations. Phylogenetic analysis enabled the investigators to 1) clarify the evolutionary history of *T. cruzi* lineages utilizing its satellite DNA (SatDNA) sequencing^[16], and 2) subtype *Blastocystis* isolates from Thailand and the results revealed that subtype 3 was the most prevalent subtype^[9].

REFERENCES

1. Arisue N, Hashimoto T, Yoshikawa H, Nakamura Y, Nakamura G, Nakamura F, *et al.* Phylogenetic position of *Blastocystis hominis* and of stramenopiles inferred from multiple molecular sequence data. *J Eukaryot Microbiol*; 2002, 49(1):42-53.
2. Arisue N, Sánchez LB, Weiss LM, Müller M, Hashimoto T. Mitochondrial-type HSP70 genes of the amitochondriate protists, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Parasitol Int* 2002; 51(1): 9–16.
3. Burki F, Roger AJ, Brown MW, Simpson AGP. The new tree of eukaryotes. *Trends Ecol Evol* 2020; 35(1):43-55.
4. Reaghi S, Haghighi A, Harandi MF, Spotin A, Arzamani K, Rouhani S. Molecular characterization of *Fasciola hepatica* and phylogenetic analysis based on mitochondrial (nicotiamide adenine dinucleotide dehydrogenase subunit I and cytochrome oxidase subunit I) genes from the North-East of Iran. *Vet World* 2016; 9(9):1034-1038.
5. Ekşi F, Özgözaşı O, Karşılıgil T, Sağlam M. Genotyping *Leishmania* promastigotes isolated from patients with cutaneous leishmaniasis in south-eastern Turkey. *J Int Med Res* 2017; 45(1):114-122.
6. Sanpool O, Rodpai R, Intapan PM, Sadaow L, Thanchomnang T, Laymanivong S, *et al.* Genetic diversity of *Taenia saginata* (Cestoda: Cyclophyllidae) from Lao People's Democratic Republic and northeastern Thailand based on mitochondrial DNA. *Parasit Vectors* 2017; 10(1):141.
7. Li JQ, Li L, Fan YL, Fu BQ, Zhu XQ, Yan HB, *et al.* Genetic diversity in *Echinococcus multilocularis* from the Plateau vole and Plateau pika in Jiuzhi

- County, Qinghai Province, China. *Front Microbiol* 2018; 9:2632.
8. Rayani M, Hatam G, Unyah NZ, Ashrafmansori A, Abdullah WO, Hamat RA. Phylogenetic analysis of *Giardia lamblia* human genotypes in Fars Province, Southern Iran. *Iran J Parasitol* 2017; 12(4):522-533.
 9. Srichaipon N, Nuchprayoon S, Charuchaibovorn S, Sukkapan P, Sanprasert V. A Simple genotyping method for rapid differentiation of *Blastocystis* subtypes and subtype distribution of *Blastocystis* spp. in Thailand. *Pathogens* 2019; 8(1).
 10. Barazesh A, Sarkari B, Shahabi S, Halidi AG, Ekici A, Aydemir S, *et al.* Genetic diversity of *Echinococcus granulosus* isolated from Humans: A comparative study in two cystic echinococcosis endemic areas, Turkey and Iran. *Biomed Res Int* 2020; 2020:3054195.
 11. Hong SH, Anu D, Jeong YI *et al.* Molecular characterization of *Giardia duodenalis* and *Cryptosporidium parvum* in fecal samples of individuals in Mongolia. *Am J Trop Med Hyg*; 2014, 90(1):43-47.
 12. Dan J, Zhang X, Ren Z, Wang L, Cao S, Shen L, *et al.* Occurrence and multilocus genotyping of *Giardia duodenalis* from post-weaned dairy calves in Sichuan province, China. *PLoS One* 2019; 14(11): e0224627.
 13. 13. Mohamed BO, Elleboudy NA, Hussein HM, Khalifa KE, Azab ME. Genotyping of *Trichomonas vaginalis* isolates from Egypt. *PUJ* 2019; 12(3):209-220.
 14. Rodrigues MS, Morelli KA, Jansen AM. Cytochrome c oxidase subunit 1 gene as a DNA barcode for discriminating *Trypanosoma cruzi* DTUs and closely related species. *Parasit Vectors* 2017; 10(1):488.
 15. Mohammadiha A, Dalimi A, Mohebbali M, Sharifi I, Mahmoudi M, Mirzaei A, *et al.* Molecular identification and phylogenetic classification of *Leishmania* spp. isolated from human cutaneous leishmaniasis in Iran: A cross-sectional study. *Iran J Parasitol* 2018; 13(3):351-361.
 16. Ramírez JC, Torres C, Curto MLA, Schijman AG. New insights into *Trypanosoma cruzi* evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis. *PLoS Negl Trop Dis* 2017; 11(12):e0006139.