Review

JOURNAL OF UNIVERSITY OF BABYLON

For Pure and Applied Sciences (JUBPAS)

Vol.32; No.2.| 2024

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Advancements in Phylogenetic Techniques: Revolutionizing Microbiology and Evolutionary Framework. A Review

Noor Ismeal Nasser[1], Lamees A. Abdul-Lateef [2], Ahmed Abdul Hasan Mohsin[3]

[1] Department of Medical laboratory techniques , Kufa Technical Institute, Al-Furat Al-Awsat Technical University31001, Kufa, Al-Najaf, Iraq. noornasser1984@gmail.com

[2] Department of Microbiology, College of Medicine , Babylon University, Hilla, Iraq. dr.lamees.razzak@gmail.com

[3] Department of prosthetic and orthotic , Kufa Technical Institute, Al-Furat Al-Awsat Technical university ,Kufa, Al-Najaf, Iraq. ahmedabdulhassan84@gmail.com

***Corresponding author email: noornasser1984@gmail.com**

التقدم في تقنيات علم النشوء والتطور: إحداث ثورة في علم الأحياء الدقيقة والإطار التطوري. مراجعة

نور اسماعيل ناصر[1], لميس عبد الرزاق عبد اللطيف[2], احمد عبد الحسن محسن[3]

[1] جامعة الفرات الاوسط التقنية, المعهد التقني كوفه, تقنيات المختبرات الطبية noornasser1984@gmail.com

[2] جامعة بابل, كلية الطب, الاحياء المجهرية dr.lamees.razzak@gmail.com

[3] جامعة الفرات الاوسط التقنية , المعهد التقني كوفه, صناعة الاطراف والمساند ahmedabdulhassan84@gmail.com

Accepted: 29/5/2024 **Published:** 30/6/2024

ABSTRACT

Phylogenetic trees describe the evolutionary relationships among different types of organisms .The selection of methods used in the construction of these trees is not easy because the data used is often incomplete, meaning it comes from studies of both modern organisms and fossils. Also, the morphological data can be misleading ; when closely related organisms may not look the same and distantly related organisms may share certain features, all these reasons make it difficult to determine the true relationships between these organisms. Therefore, the genetic data is more dependable, so it's usually a preferred method used to construct phylogenetic trees. Despite that, the genetic data can be affected by several factors, such as gene transfer, loss, and the choice of genes that are used. It's important to note that phylogenetic trees do not account for the length of time over which evolution happened ; they only reflect the order of divergence among lineages. Despite all these challenges, phylogenetic trees are still a crucial tool in understanding the evolution of life and in identifying new relationships among different organisms. Recently, the use of phylogenetic analysis in microbiological research has increased significantly. This review provides an overview of the concept of phylogeny, the different methods used in the construction of phylogenetic trees, the applications of phylogenetic analysis in microbiology, and the steps involved in building a suitable phylogenetic tree.

Keywords: bacteria; evolution; genome; molecular phylogeny; phylogenomic.

INTRODUCTION

Understanding the evolutionary relationships among prokaryotes seems to be so difficult. In spite of that, the advancements in technology have made a revolution in this field, which allows a deeper understanding of microbial evolution and the effect of that on their relationships. Phylogenetic techniques make it possible to analyze the evolutionary patterns of microbial pathogens, including fungi, viruses, and bacteria, which allow the identification of commonalities, variations, and transmission pathways[1]. This understanding can be used in developing effective vaccines and treatment systems by pinpointing similarities among different strains or species of pathogens. Progress in computational methods has significantly contributed to the field of phylogenetics which, has led to major advances in the microbial diversity perspective in recent decades [2,3]. Phylogenetic trees that are billed on the base of gene sequences act as maps that elucidate the concept of biodiversity. Comparative analyses of a small subunit of rRNA (16S or 18S rRNA) and other gene sequences have determined that life is divided into three major domains (Bacteria, Eucarya, and Archaea). According to the phylogenetic rRNA-based trees, the largest part of Earth's biodiversity is microbial. However, our understanding of the nature of microbial diversity has been limited by our dependence on the study of cultured microorganisms. It is estimated that over 99% of microorganisms found in nature cannot be cultivated using standard methods [4,5] . Recently, the revolution in both recombinant DNA technique and molecular phylogenetic technique has provided several means to identify organisms in microbial communities without the need for cultivation. Applying these methods to diverse environments has revealed that our view of microbial diversity is limited and has uncovered a wealth of novel and environmentally significant diversity yet to be studied. Phylogenetic analysis offers an illuminating glimpse of the evolutionary history of a species [6,7] . This out group, which is typically unrelated to the species under investigation, serves as a benchmark to root the tree, imparting a sense of direction in evolutionary time. The choice of an appropriate out group is of paramount importance. Sometimes, prior knowledge of the group being studied facilitates identifying an out group, or it may become evident during sequence alignment [8] . Choosing a sister group as an out group often gives the most informative results because it shares a recent common ancestor with the investigated group. The construction of phylogenetic trees Faces numerous challenges, including the selection of suitable molecular markers, solving sequence alignment problems, and grappling with potential biases. All these complications indicate the necessity of a thorough comprehension of the methods and data utilized in phylogenetic reconstruction. The selection process for the appropriate species to construct the phylogenetic analysis is vital because it plays an essential role in determining the accuracy of phylogenetic trees. The choice of inappropriate species can cause a significant distortion of the resulting phylogeny, which leads to limiting the analysis to a small number of taxa or choosing the wrong ones [9]. The construction of a phylogenetic tree can be achieved by using various features and characteristics, such as morphological traits, anatomical features, molecular markers, DNA, RNA, and amino acid sequences. The selection of characteristics that are used often depends on factors such as the available data, in addition to the evolutionary distance between species, and the desired level of accuracy. The selection of whether to use DNA or protein sequences is essential when it comes to constructing the molecular phylogeny based on sequencing data [10]. The selection process between DNA sequence or protein sequence in the construction of a phylogenetic tree depends on several factors ,like the evolutionary distance between species, the type of evolutionary changes being examined, and the specific goals of the analysis. DNA sequences have four nucleotides, which gives a large number of possibilities for the characters; this feature provides better resolution for constructing the phylogenetic tree. However, because of increase in the substitution rates, there is a greater chance that two lineages may independently evolve the same nucleotide at the same site, which leads to potential homoplasy. In other words, similar sequences do not result from the same common ancestry, but they rather from convergent evolution or other factors. Nevertheless, it is preferred to use DNA sequences for closely related species because of the finer resolution they can offer, in addition to capturing more detailed evolutionary changes. However, it is preferred to analyze protein sequences in the case of more distantly related species, where changes may have accumulated over longer periods. As they are subject to selective pressures that may result in fewer substitutions and offer a clearer signal of evolutionary relationships [11,12].

PHYLOGENETIC VERSUS PHYLOGENOMIC:

Phylogenomics and phylogenetics: these two terms are related in the field of evolutionary biology, but they refer to different concepts. Phylogenetics represents the study of the evolutionary relationships between a group of organisms based on the construction of a phylogenetic tree, which is built based on either the use of molecular data such as DNA,

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalotbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalotbabylon.com

RNA, or protein sequences or morphological and/or behavioral traits_[13]. This definition reveals that phylogenomics is a subfield of phylogenetics that specifically uses a wide range of genes, or sometimes even whole genomes, to study the evolutionary relationships between different species or other taxonomic units to reconstruct the tree of life with higher accuracy in comparison to traditional phylogenetic methods. Advances in DNA sequencing technologies have made it possible to create a genomic database, which allows for more accurate and detailed reconstructions of evolutionary histories [14].

PHYLOGENETIC TREE:

The evolutionary relationship between organisms is represented by a graph known as the phylogenetic tree, which is typically constructed based on molecular data such as DNA sequences, morphological characteristics, or a combination thereof [15]. this graphic consists of nodes that represent common ancestors, and branches which represent the evolutionary pathways that lead to different species. The evolutionary change that has taken place based on the length of branches and the points at that branches split indicate instances of speciation, where one species is split into two or more descendant species. The relationship between species of organisms is determined by the proximity of their branches to a common node. The closer the branches are, the more closely related the species are. This fact highlights the interconnectedness of all living beings and highlights the importance of preserving biodiversity. It is important to distinguish between homologous and analogous characteristics. When comparing traits between species, homologous characteristics are similar due to shared ancestry, whereas analogous features are similar due to convergent evolution [15,16].

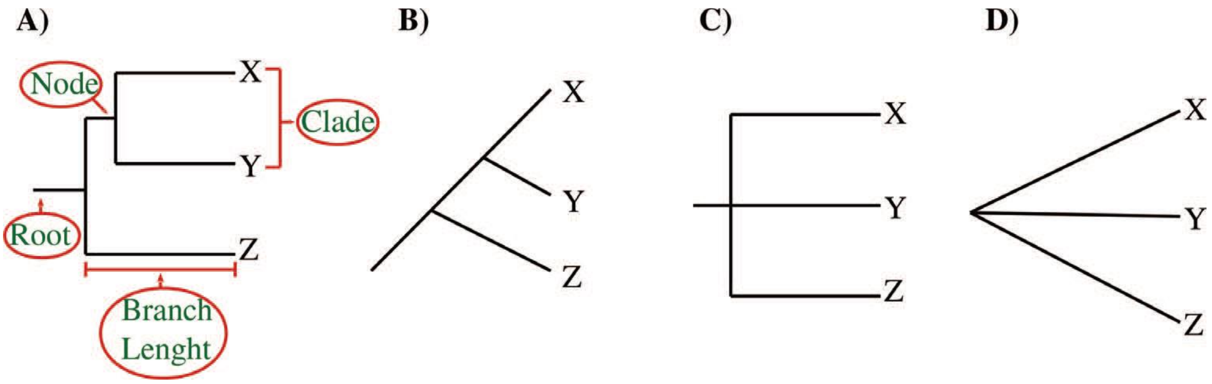

Figure 1. Phylogenetic trees show different representations and topologies. (A) Presentation of the elements of an a phylogenetic tree. (B) Another representation of the phylogenetic tree shown in A. Trees (C) and (D) are like each other and present a topology that differs from trees A and B [17] **.**

GRAPHIC TYPE UTILIZED IN THE DISPLAY OF PHYLOGENETIC TREES:

A phylogenetic tree can be represented using several graphical forms, each of which has a different purpose and could provide varying levels of detail in the context of evolutionary relatedness and divergence. Choosing the appropriate type depends on the type of data that the researcher desires to analyze and how it is presented to provide the best and clearest results[18]. For instance, a cladogram reveals the interconnections between various organisms by pinpointing shared characteristics (synapomorphies) and also highlights branching patterns along with relationships between species or sequences .It shows the common ancestors and points where lines diverge, but the lengths of the branches generally do not represent evolutionary time or genetic change [12]. In the case of the phylogram, branch lengths represent the genetic change between the species or sequences, or the amount of evolutionary divergence [18]. Moreover, a dendrogram usually denotes a branching diagram that illustrates the relationships among species or sequences, which may not necessarily include information about branch lengths or evolutionary changes. However, in the context of phylogenetics, they are usually used interchangeably with cladograms to describe the branching patterns of taxa. This diagram is similar to a tree diagram but is often used more broadly in various fields of study beyond biology [19]. Molecular methods and conventional diagnostic techniques both have advantages and disadvantages. Conventional techniques are sometimes more easy to use, less expensive, and give faster results. However, they may not have the exact sensitivity and specificity as molecular testing. On the other hand

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com
مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

molecular diagnostics ,provide early detection capabilities, but they are more expensive, and difficult, in addition to raising ethical questions. By combining the precision and customization of molecular/genomic techniques with the rapid and widespread availability of older procedures, the integration of both approaches can improve patient care and diagnostic accuracy[18,19].

TECHNIQUES EMPLOYED IN THE PHYLOGENETIC TREE CONSTRUCTION.

There are two main primary methods for constructing a phylogenetic tree, which, in turn, includes several sub-methods for building the tree. The distance-based and character-based methods, each of which uses distinct algorithms and criteria for analysis. Character-Based Methods, also known as cladistic or phylogenetic inference methods , depend on the analysis of shared derived characters among taxa to reveal their evolutionary history. These methods are based on the concept of parsimony, which needs an uncomplicated tree and the fewest evolutionary changes to reveal the observed character data. When compared with the distance-based methods, character-based methods are considered more robust and accurate for concluding phylogenetic relationships. These are favorable when dealing with complex evolutionary scenarios and large datasets. Despite that, it can be computationally intensive and may require large amounts of computational resources. Maximum Likelihood (ML) techniques create a model that shows how sequences evolve,which makes this technique a powerful and widely used method for constructing phylogenetic trees. It depends on advanced methods that consider various factors, such as the variability of substitution rates and the probability of observed data. These techniques are highly preferred because they give researchers the opportunity to gain deeper insights into the mechanisms that drive the evolution of genetic sequences. ML trees can give informative details about internal node branch lengths and support values. Maximum Likelihood techniques use statistical models to estimate the likelihood of different tree topologies given the character data, and they select the tree with the highest probability [20-23].

The Maximum Parsimony (MP) technique represents a valuable tool for researchers in various fields because it minimizes the evolutionary changes required to explain the observed sequence data. which leads to the accurate representation of the evolutionary relationships between different organisms [20-25]. Moreover, the Bayesian Inference (BI) techniques are used when dealing with evolutionary model uncertainty and help estimate clades' posterior probabilities. These methods can accommodate complex models of sequence evolution. Also, it involves the use of statistical models to calculate posterior probabilities for different tree topologies based on character data [22,23,26-28]. As well, distance-based methods (phenetic methods) rely on calculating genetic or morphological distances between taxa to infer their evolutionary relationships. It's based on the idea that more closely related organisms will have shorter genetic or morphological distances. These methods are relatively simple and computationally efficient. Still, they can be sensitive to the choice of distance metric. They may not always accurately represent authentic evolutionary relationships, especially in cases of long-branch attraction or other tree-building artifacts [29-31] . This method includes several techniques with different purposes, Unweighted Pair Group Method with Arithmetic Mean, in this technique, the constructing process of a tree occurs by grouping taxa based on their pairwise distances and using an arithmetic mean to determine branch lengths. It is expected to construct trees based on genetic or sequence dissimilarity. Microbiology often uses these methods to explore relationships between different microbial strains or species [31, 26]. The neighbor-joining (NJ) technique is widely used in microbiology, especially for constructing phylogenetic trees based on iteratively joining the closest neighbors (taxa) based on their pairwise distances. It is computationally efficient and suitable for large datasets. NJ trees are often used when the evolutionary distances between sequences are small [30, 26, 32,33].Minimum Evolution (ME) technique is a phylogenetic tree-building algorithm that aims to construct the tree with the minimum overall branch length, assuming that the observed differences among sequences or taxa have evolved under an evolutionary model. Seek the tree with the minimum total branch length across all possible tree topologies. This methods Relatively efficient in handling moderate-sized datasets [34]. The Fitch-Margoliash Method (FM) technique is a classic phylogenetic approach used to construct phylogenetic trees based on genetic distance matrices. It works with a matrix of pairwise genetic distances between taxa or sequences. It aims to estimate branch lengths and infer the tree topology that minimizes the discrepancy between the observed distances and the expected distances computed under a given model of evolution [35], each employing distinct algorithms and criteria for analysis. Each method has its advantages, so consider

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

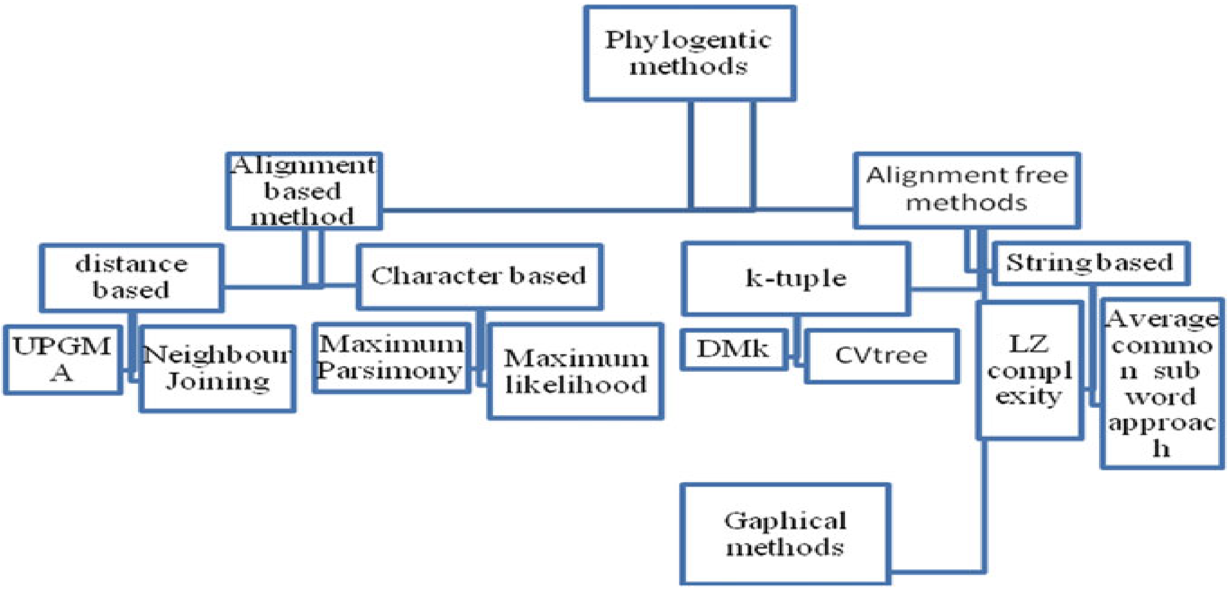

Fig. 2: Hierarchical view of phylogenetic methods [17].

TYPES OF PHYLOGENETIC TREE:

Phylogenetic trees can be classified into various types based on their representation and the specific relationships they depict among organisms ; theycan also be classified based on the type of data used to construct them. The data sources often determine the methods and approaches employed in creating these trees [37]. The rooted phylogenetic tree has a designated point, the root, from which all branches emanate. This root represents the common ancestor of all the species included in the tree. Also, it indicates evolutionary direction, with the root representing the oldest ancestor and the tips (leaves) representing the present-day species. Branch lengths can represent evolutionary distance or time. This type of phylogenetic tree is used to visualize the evolutionary history and relationships among different species in a hierarchical manner [38]. while the unrooted tree has no designated root point. It only displays the relationships between species without specifying an ancestral lineage. It focuses on the relationships between species rather than their historical connection to a common ancestor. Emphasizes branching patterns and relative distances between species rather than absolute evolutionary timelines. This does not imply an evolutionary direction or lineage from a single root [39]. Moreover, the term gene phylogenetic tree means that a tree is reconstructed from comparisons between the sequences of orthologous genes (those derived from the same ancestral sequence) to make inferences about the evolutionary history of the species from which the genes are obtained. It is distinct

مجلة جـامعة بـابل للعلـــوم الصـرفة والتطـبيقية مجلة جـامعة بـابل للعلـوم الصـرفة والتطـبيقية مجلة جـامعة بـابل للعلـوم الصـرفة والتطـــبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

from species trees or organismal phylogenetic trees, which depict the evolutionary relationships between whole organisms or species. Gene trees focus on the evolutionary history of a particular gene or sequence. It is particularly useful in various biological contexts, including molecular biology, genetics, and comparative genomics. They provide insights into the evolution of specific genes and can help understand how genes have diverged, duplicated, or been horizontally transferred across different species. These trees can also be essential for studying gene function and regulation within an evolutionary context [40]. while the species phylogenetic tree provides insights into evolutionary history and relationships among various organisms at the species level. It plays a crucial role in evolutionary biology, systematics, and comparative genomics, providing a framework for understanding the history of life on Earth, evolutionary connections between different species, and the timing of divergence events. They are often used to answer questions related to the classification and taxonomy of species as well as to study the patterns of speciation, extinction, and adaptation over time. It is commonly assumed that the gene and species trees are identical, but this is not necessarily the case. Internal nodes in the gene and species trees must be precisely equivalent for this assumption to hold true. However, they are not equivalent [41].

APPLICATION OF PHYLOGENETIC TREES IN MICROBIOLOGY.

Phylogenetic analysis:

The history of life is a big mystery. Throughout the ages, scientists have tried to understand this mystery, and one of the most important methods used to understand this fact and how species are related was phylogenetic analysis, which is characterized by its reliability that stems from its comprehensive nature that provides a detailed examination of evolutionary histories, givesinsights into speciation events, genetic diversification, and the interrelationships among different organisms or genes that help in understanding the shared ancestry and divergence among them. The simpler the representation of phylogenetic analysis, the more accessible, understandable, and visually grasp complex evolutionary patterns. On the other hand, this visual aid helps in communicating findings to a broader audience as well. The applications of phylogenetics extent to various fields in biology, including evolutionary biology, microbiology, conservation biology, and even fields like medicine and forensics.This reflects its importance as a fundamental tool for understanding biological diversity and tracing the origins of diseases [17]. For constructing the phylogenetic tree several steps should be taken into consideration, starting with the type of data and its quality [22,42-44] which could be genetic data, typically DNA or RNA sequences. This data may be derived from specific genes or regions of the genome that are used for phylogenetic analysis [45,46]. After that, the collected sequences are aligned to identify homologous regions or positions within the gene, genome, or genetic markers to ensure that comparable characters are used for tree construction [23]. the application of phylogenetic methods, such as maximum likelihood, maximum parsimony, bayesian inference, or neighbor-joining, to construct the tree. These methods aim to find the most likely branching pattern that explains the observed sequence data [45].To perform tree construction, several software packages could be used, like PhyML, RAxML, MrBayes, or MEGA. The construction started with Input the alignment data, selected method, and model parameters into the chosen software, then run the analysis, and finally execute the software to estimate the tree. The software calculates branch lengths and tree topology based on the alignment data and the chosen method [47]. It is possible to use a tree visualization software such as FigTree, Dendroscope, or Archaeopteryx to visualize and explore the phylogenetic tree. Also its Analyze the tree's branching patterns and consider the evolutionary relationships it represents [48]. Currently, it is possibleto trackthe sources and transmission pathways of infectious diseases, identify the origins of outbreaks , study microbial resistance to antibiotics, guide drug discovery efforts, and understand the spread of resistance genes in bacterial populations by using phylogenetic analysis in epidemiology [44,49-51]. Also ,molecular phylogeneticsis applied to study the evolutionary relationships of viruses. This is important for understanding the origin and transmission of viral diseases and for developing strategies to combat them [50]. This application is crucialin understanding the diversity and ecological roles of microorganisms in various ecosystems, including soil, water, and extreme environments. in addition to its importance in biotechnology and industrial microbiology by facilitating [50], [52]. The selection of strains with desirable genetic traits for various applications, such as biofuel production, bioremediation, and biopharmaceutical production [53].

The significance of the out-group in the phylogenetic tree's construction

Phylogenetic analysis sometimes needs to build a rooted tree to determine the direction of evolution, indicating which branch represents the common ancestor of all the in group (target) species. To do that, it should be used as an out-group which is a species or group of

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

species used as a reference point to root the tree. The outgroup is critical and increases the overall reliability and informativeness of the tree. The choice of an appropriate outgroup is essential, and it should be selected based on its evolutionary relationship to the ingroup species while satisfying the goals of the analysis [54,55].

STRATEGIES TO ASSESS THE ACCURACY OF A PHYLOGENETIC TREE.

There are various strategies aimed at evaluating the robustness, reliability, and consistency in addition to the accuracy of a phylogenetic tree

Bootstrap Analysis:

This method is based on the generation of multiple pseudo-data through random resampling by replacement characters from the original alignment. Then construct phylogenetic trees for each bootstrap data. After that, a consensus tree is computed from the multiple trees, where the branches that appear in the majority of the bootstrap set of replicates are considered well-supported [56].

Cross-Validation:

This method is based on splitting the original alignment into subsets such as 80/20 or 70/30 to create training and test data. After that, using the training data to reconstruct the tree and evaluate its accuracy is done by comparing it to the true relationships known from the test data [57].

COMPARATIVE METHODS:

This method is based on the use of the tree to make predictions about characteristics that are known to correlate with phylogeny and compare these predictions to empirical data [58].

SIMULATED DATA:

This method is based on generating simulated data that has known evolutionary relationships. After that, the simulated data have been analyzed with the same methods used for real data and assessed for their ability to recover the true tree [59].

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

DATA QUALITY AND SENSITIVITY ANALYSIS:

This method is based on the use of sensitivity analyses to examine the quality of sequence alignment or molecular markers by varying alignment parameters or models to assess how these choices impact the tree because low-quality or unreliable data can lead to inaccurate trees [60].

CONCLUSION

Development in phylogenetic techniques have revolutionized the science of microbiology and the larger field of evolutionary biology, providing profound insights into the evolutionary relationships among organisms. These techniques have allowed a more accurate and detailed understanding of the evolutionary processes shaping life on Earth, particularly among microbial species.

Conflict of interests.

There are non-conflicts of interest.

References

1- R. S. Gupta, “The Natural Evolutionary Relationships among Prokaryotes,” *Critical Reviews in Microbiology*, vol. 26, no. 2, pp. 111–131, Jan. 2000, doi: https://doi.org/10.1080/10408410091154219.

2- Z. J. C. TOBIAS, A. K. YADAV, A. SCHMIDT-RHAESA, and R. POULIN, “Intra- and interspecific genetic diversity of New Zealand hairworms (Nematomorpha),” *Parasitology*, vol. 144, no. 8, pp. 1026–1040, 2017. doi:10.1017/S0031182017000233

3- A. Sharma *et al.*, “Advanced strategies for development of vaccines against human bacterial pathogens,” *World Journal of Microbiology and Biotechnology*, vol. 37, no. 4, Mar. 2021, doi: https://doi.org/10.1007/s11274-021-03021-6.

4- M. Höhl and M. A. Ragan, “Is Multiple-Sequence Alignment Required for Accurate Inference of Phylogeny?,” *Systematic Biology*, vol. 56, no. 2, pp. 206–221, Apr. 2007, doi: https://doi.org/10.1080/10635150701294741.

5- S. R. Amit Roy, “Molecular Markers in Phylogenetic Studies-A Review,” *Journal of Phylogenetics & Evolutionary Biology*, vol. 02, no. 02, 2014, doi: https://doi.org/10.4172/2329-9002.1000131.

6- P. Hugenholtz, B. M. Goebel, and N. R. Pace, “Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity,” *Journal of Bacteriology*, vol. 180, no. 18, pp. 4765–4774, Sep. 1998, doi: https://doi.org/10.1128/jb.180.18.4765-4774.1998.

7- S. Ghazanfar, A. Azim, M. Suleri, M. I. Anjum, and I. Begum, “Metagenomics and its application in soil microbial community studies: biotechnological prospects.,” vol. 6, no. 2, pp. 611–622, Jan. 2010.

8- D. H. Parks *et al.*, “Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life,” *Nature Microbiology*, vol. 2, no. 11, pp. 1533–1542, Sep. 2017, doi: https://doi.org/10.1038/s41564-017-0012-7.

9- Y. Tian and L. Kubatko, “Rooting phylogenetic trees under the coalescent model using site pattern probabilities,” *BMC Evolutionary Biology*, vol. 17, no. 1, Dec. 2017, doi: https://doi.org/10.1186/s12862-017-1108-7.

10- R. Godini and H. Fallahi, “A brief overview of the concepts, methods and computational tools used in phylogenetic tree construction and gene prediction,” *Meta Gene*, vol. 21, p. 100586, Sep. 2019, doi: https://doi.org/10.1016/j.mgene.2019.100586.

11- F. Delsuc, H. Brinkmann, D. Chourrout, and H. Philippe, “Tunicates and not cephalochordates are the closest living relatives of vertebrates,” *Nature*, vol. 439, no. 7079, pp. 965–968, Feb. 2006, doi: https://doi.org/10.1038/nature04336.

12- L. R. Novick, K. M. Catley, and D. J. Funk, “Characters Are Key: The Effect of Synapomorphies on Cladogram Comprehension,” *Evolution: Education and Outreach*, vol. 3, no. 4, pp. 539–547, Jun. 2010, doi: https://doi.org/10.1007/s12052-010-0243-z.

13- P. Hugenholtz, B. M. Goebel, and N. R. Pace, “Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity,” *Journal of Bacteriology*, vol. 180, no. 18, pp. 4765–4774, Sep. 1998, doi: https://doi.org/10.1128/jb.180.18.4765-4774.1998.

14- I. Borozan, S. Watt, and V. Ferretti, “Integrating alignment-based and alignment-free sequence similarity measures for biological sequence classification,” *Bioinformatics*, vol. 31, no. 9, pp. 1396–1404, Jan. 2015, doi: https://doi.org/10.1093/bioinformatics/btv006.

15- M. Sardaraz, M. Tahir, T. Aziz Ikram, and H. Bajwa, “Applications and Algorithms for Inference of Huge Phylogenetic Trees: a Review,” *American Journal of Bioinformatics Research*, vol. 2, no. 1, pp. 21–26, Aug. 2012, doi: https://doi.org/10.5923/j.bioinformatics.20120201.04.

16- G. Alves-Silva, E. R. Drechsler-Santos, and R. M. B. da Silveira, “Bambusicolous Fomitiporia revisited: multilocus phylogeny reveals a clade of host-exclusive species,” *Mycologia*, vol. 112, no. 3, pp. 633–648, May 2020, doi: https://doi.org/10.1080/00275514.2020.1741316.

17- S. Manel, M. K. Schwartz, G. Luikart, and P. Taberlet, “Landscape genetics: combining landscape ecology and population genetics,” *Trends in Ecology & Evolution*, vol. 18, no. 4, pp. 189–197, Apr. 2003, doi: https://doi.org/10.1016/s0169-5347(03)00008-9.

18- J.-H. . Choi, H.-Y. . Jung, H.-S. . Kim, and H.-G. . Cho, “PhyloDraw: a phylogenetic tree drawing system,” *Bioinformatics*, vol. 16, no. 11, pp. 1056–1058, Nov. 2000, doi: https://doi.org/10.1093/bioinformatics/16.11.1056.

19- D. H. Huson, D. C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp, “Dendroscope: An interactive viewer for large phylogenetic trees,” *BMC Bioinformatics*, vol. 8, no. 1, Nov. 2007, doi: https://doi.org/10.1186/1471-2105-8-460.

20- N. Alon, B. Chor, F. Pardi, and A. Rapoport, “Approximate Maximum Parsimony and Ancestral Maximum Likelihood,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 7, no. 1, pp. 183–187, Jan. 2010, doi: https://doi.org/10.1109/tcbb.2008.13.

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

21- G. Talavera and J. Castresana, “Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments,” Systematic Biology, vol. 56, no. 4, pp. 564–577, Aug. 2007, doi: https://doi.org/10.1080/10635150701472164.

22- M. Talianová, “Survey of molecular phylogenetics,” Plant, Soil and Environment, vol. 53, no. No. 9, pp. 413–416, Jan. 2008, doi: https://doi.org/10.17221/2290-pse.

23- A. Cherkasov, “Bioinformatics: A practical guide to the analysis of genes and proteins,” American Journal of Human Biology, vol. 17, no. 3, pp. 387–389, 2005, doi: https://doi.org/10.1002/ajhb.20118.

24- S. L. Robson and B. Wood, “Hominin life history: reconstruction and evolution,” Journal of Anatomy, vol. 212, no. 4, pp. 394–425, Apr. 2008, doi: https://doi.org/10.1111/j.1469-7580.2008.00867.x.

25- L. L. Cavalli-Sforza and A. W. F. Edwards, “PHYLOGENETIC ANALYSIS: MODELS AND ESTIMATION PROCEDURES,” Evolution, vol. 21, no. 3, pp. 550–570, Sep. 1967, doi: https://doi.org/10.1111/j.1558-5646.1967.tb03411.x.

26- A. Stamatakis, “Distributed and Parallel Algorithms and Systems for Inference of Huge Phylogenetic Trees based on the Maximum Likelihood Method,” mediatum.ub.tum.de, 2004. https://mediatum.ub.tum.de/?id=601760 (accessed May 28, 2024).

27- T. L. Williams and B. M. E. Moret, “An investigation of phylogenetic likelihood methods,” Infoscience (Ecole Polytechnique Fédérale de Lausanne), Nov. 2003, doi: https://doi.org/10.1109/bibe.2003.1188932.

28- D. L. Swofford, “PAUP^* : Phylogenetic Analysis Using Parsimony (and other methods) 4.0b8,” Jan. 2001.

29- F. Bai, J. Xu, and L. Liu, “Weighted relative entropy for phylogenetic tree based on 2-step Markov Model,” Mathematical biosciences, vol. 246, no. 1, pp. 8–13, Nov. 2013, doi: https://doi.org/10.1016/j.mbs.2013.09.004.

30- R. R. Sokal and C. D. Michener, “A statistical method for evaluating systematic relationships,” vol. 38, no. 6, pp. 1409–1438, Jan. 1958.

31- N. Saitou and M. Nei, “The neighbor-joining method: a new method for reconstructing phylogenetic trees.,” Molecular Biology and Evolution, vol. 4, no. 4, pp. 406–425, Jul. 1987, doi: https://doi.org/10.1093/oxfordjournals.molbev.a040454.

32- P. Crescenzi, D. Goldman, C. H. Papadimitriou, A. Piccolboni, and Mihalis Yannakakis, “On the Complexity of Protein Folding,” Journal of Computational Biology, vol. 5, no. 3, pp. 423–465, Jan. 1998, doi: https://doi.org/10.1089/cmb.1998.5.423.

33- D. Bryant, “Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks,” Molecular Biology and Evolution, vol. 21, no. 2, pp. 255–265, Aug. 2003, doi: https://doi.org/10.1093/molbev/msh018.

34- R. Desper, “Theoretical Foundation of the Balanced Minimum Evolution Method of Phylogenetic Inference and Its Relationship to Weighted Least-Squares Tree Fitting,” Molecular Biology and Evolution, vol. 21, no. 3, pp. 587–598, Dec. 2003, doi: https://doi.org/10.1093/molbev/msh049.

35- W. M. Fitch and E. Margoliash, “Construction of Phylogenetic Trees,” Science, vol. 155, no. 3760, pp. 279–284, Jan. 1967, doi: https://doi.org/10.1126/science.155.3760.279.

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

36- G. Munjal, M. Hanmandlu, and S. Srivastava, “Phylogenetics Algorithms and Applications,” Advances in Intelligent Systems and Computing, vol. 904, pp. 187–194, 2019, doi: https://doi.org/10.1007/978-981-13-5934-7_17.

37- E. Barbosa Evanovich dos Santos, “Phylogenetics,” Recent Advances in Phylogenetics, Apr. 2019, doi: https://doi.org/10.5772/intechopen.79422.

38- D. H. Huson and C. Scornavacca, “Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks,” Systematic Biology, vol. 61, no. 6, pp. 1061–1067, Sep. 2012, doi: https://doi.org/10.1093/sysbio/sys062.

39- A. Francis, K. T. Huber, and V. Moulton, “Tree-Based Unrooted Phylogenetic Networks,” Bulletin of Mathematical Biology, vol. 80, no. 2, pp. 404–416, Dec. 2017, doi: https://doi.org/10.1007/s11538-017-0381-3.

40- Y. Nakano, Y. Domon, and K. Yamagishi, “Phylogenetic trees of closely related bacterial species and subspecies based on frequencies of short nucleotide sequences,” vol. 18, no. 4, pp. e0268847–e0268847, Apr. 2023, doi: https://doi.org/10.1371/journal.pone.0268847.

41- C. R. Woese, “Interpreting the universal phylogenetic tree,” Proceedings of the National Academy of Sciences, vol. 97, no. 15, pp. 8392–8396, Jul. 2000, doi: https://doi.org/10.1073/pnas.97.15.8392.

42- B. G. Hall, “Building Phylogenetic Trees from Molecular Data with MEGA,” Molecular Biology and Evolution, vol. 30, no. 5, pp. 1229–1235, Mar. 2013, doi: https://doi.org/10.1093/molbev/mst012.

43- A. Stamatakis, P. Hoover, and J. Rougemont, “A Rapid Bootstrap Algorithm for the RAxML Web Servers,” Systematic Biology, vol. 57, no. 5, pp. 758–771, Oct. 2008, doi: https://doi.org/10.1080/10635150802429642.

44- X. Didelot, R. Bowden, D. J. Wilson, T. E. A. Peto, and D. W. Crook, “Transforming clinical microbiology with bacterial genome sequencing,” Nature Reviews Genetics, vol. 13, no. 9, pp. 601–612, Aug. 2012, doi: https://doi.org/10.1038/nrg3226.

45- R. D. M. Page, “Tree View: An application to display phylogenetic trees on personal computers,” Bioinformatics, vol. 12, no. 4, pp. 357–358, 1996, doi: https://doi.org/10.1093/bioinformatics/12.4.357.

46- B. G. Hall, “Building Phylogenetic Trees from Molecular Data with MEGA,” Molecular Biology and Evolution, vol. 30, no. 5, pp. 1229–1235, Mar. 2013, doi: https://doi.org/10.1093/molbev/mst012.

47- P. Kapli, Z. Yang, and M. J. Telford, “Phylogenetic tree building in the genomic age,” Nature Reviews Genetics, vol. 21, no. 7, pp. 428–444, Jul. 2020, doi: https://doi.org/10.1038/s41576-020-0233-0.

48- R. D. M. Page, “Visualizing Phylogenetic Trees Using TreeView,” Current Protocols in Bioinformatics, vol. 0, no. 1, pp. 6.2.1–6.2.15, Aug. 2002, doi: https://doi.org/10.1002/0471250953.bi0602s01.

49- W. Alsanie, E. Felemban, M. Farid, M. Hassan, A. Sabry, and A. Gaber, “Molecular Identification and Phylogenetic Analysis of Multidrug-resistant Bacteria using 16S rDNA Sequencing,” Journal of Pure and Applied Microbiology, vol. 12, no. 2, pp. 489–496, Jun. 2018, doi: https://doi.org/10.22207/jpam.12.2.07.

50- Y.-J. Lee, H.-W. Sung, J.-G. Choi, J.-H. Kim, and C.-S. Song, "Molecular epidemiology of Newcastle disease viruses isolated in South Korea using sequencing of the fusion protein cleavage site region and phylogenetic relationships," Avian Pathology, vol. 33, no. 5, pp. 482–491, Oct. 2004, doi: https://doi.org/10.1080/03079450400003700.

51- G. P. S. Raghava, R. J. Solanki, V. Soni, and P. Agrawal, "Fingerprinting Method for Phylogenetic Classification and Identification of Microorganisms Based on Variation in 16S rRNA Gene Sequences," BioTechniques, vol. 29, no. 1, pp. 108–116, Jul. 2000, doi: https://doi.org/10.2144/00291bc02.

52- J. Borneman and E. W. Triplett, "Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation.," Applied and environmental microbiology, vol. 63, no. 7, pp. 2647–2653, 1997, doi: https://doi.org/10.1128/aem.63.7.2647-2653.1997.

53- R. S. Gupta, "Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification," FEMS Microbiology Reviews, vol. 40, no. 4, pp. 520–553, Jun. 2016, doi: https://doi.org/10.1093/femsre/fuw011.

54- M. S. Rosenberg and S. Kumar, "Taxon Sampling, Bioinformatics, and Phylogenomics," Systematic Biology, vol. 52, no. 1, pp. 119–124, Feb. 2003, doi: https://doi.org/10.1080/10635150390132894.

55- M. Kellis et al., "Defining functional DNA elements in the human genome," Proceedings of the National Academy of Sciences, vol. 111, no. 17, pp. 6131–6138, Apr. 2014, doi: https://doi.org/10.1073/pnas.1318948111.

56- D. E. Soltis and P. S. Soltis, "Applying the Bootstrap in Phylogeny Reconstruction," Statistical Science, vol. 18, no. 2, pp. 256–267, May 2003, doi: https://doi.org/10.1214/ss/1063994980.

57- D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography, vol. 40, no. 8, pp. 913–929, Mar. 2017, doi: https://doi.org/10.1111/ecog.02881.

58- "An Improved Phylogenetic Tree Comparison Method," cdr.lib.unc.edu, Mar. 18, 2013. https://cdr.lib.unc.edu/concern/dissertations/ms35t9783 (accessed May 28, 2024).

59- "Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice," Choice Reviews Online, vol. 52, no. 06, pp. 52–308952–3089, Jan. 2015, doi: https://doi.org/10.5860/choice.187276.

60- B. Misof, K. Meusemann, B. M. von Reumont, P. Kück, S. J. Prohaska, and P. F. Stadler, "A priori assessment of data quality in molecular phylogenetics," Algorithms for Molecular Biology, vol. 9, no. 1, Jan. 2014, doi: https://doi.org/10.1186/s13015-014-0022-4.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

الخلاصة

ان فهم العملية التطورية للحياة أمر صعب، ومن أكثر الأدوات المستخدمة لتمثيل وفهم هذه العملية هو التحليل التطوري، من خلال رسم أشجار التطور. في الواقع، هذه الأشجار هي فرضيات، تستخدم لوصف العلاقات التطورية بين الأنواع المختلفة من الكائنات الحية. إن اختيار الطرق المستخدمة في بناء هذه الأشجار ليس بالأمر السهل، لأن البيانات المستخدمة غالبًا ما تكون غير مكتملة، بمعنى آخر أنها تأتي من دراسات كل من الكائنات الحية الحديثة والحفريات. كما أن البيانات المورفولوجية يمكن أن تكون مضللة، فعندما قد لا يبدو الكائن الحي المرتبط ارتباطًا وثيقًا بنفس الشكل، وقد تشترك الكائنات البعيدة في سمات معينة، كل هذه الأسباب تجعل من الصعب تحديد العلاقات الحقيقية بين هذه الكائنات الحية. ولذلك، فإن البيانات الوراثية أكثر موثوقية، لذلك عادة ما تكون الطريقة المفضلة المستخدمة لبناء أشجار النشوء والتطور. وعلى الرغم من ذلك فإن البيانات الوراثية يمكن أن تتأثر بعدة عوامل مثل نقل الجينات وفقدانها واختيار الجينات المستخدمة. من المهم أن نلاحظ أن أشجار النشوء والتطور لا تأخذ في الاعتبار طول الفترة الزمنية التي حدث خلالها التطور، ولكنها تعكس فقط ترتيب الاختلاف بين الأنساب. على الرغم من كل هذه التحديات، لا تزال أشجار النشوء والتطور أداة حاسمة في فهم تطور الحياة وفي تحديد العلاقات الجديدة بين الكائنات الحية المختلفة. في الآونة الأخيرة، زاد استخدام التحليل الوراثي في الأبحاث الميكروبيولوجية بشكل ملحوظ. تقدم هذه المراجعة نظرة عامة على مفهوم التطور والتطور، والطرق المختلفة المستخدمة في بناء أشجار التطور، وتطبيقات التحليل التطوري في علم الأحياء الدقيقة، والخطوات المتبعة في بناء شجرة التطور والتطور المناسبة.

الكلمات المفتاحية: البكتيريا، التطور، الجينوم، السلالة الجزيئية، النشوء والتطور.خلاصة.