

Computational Techniques in Phylogenetics: Harnessing Big Data for Tree Inference

Rosset Lemey*

Department of Human Genetics and Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Introduction

Phylogenetics, the study of evolutionary relationships among species, has undergone a revolution in recent years, thanks to the advent of big data and advanced computational techniques. Traditional methods of inferring evolutionary trees, also known as phylogenetic trees, were limited by the amount of genetic data that could be analyzed at any given time. However, the availability of vast amounts of genomic data and the development of sophisticated algorithms have enabled researchers to reconstruct more accurate and detailed evolutionary histories. This article delves into the computational techniques driving modern phylogenetics, highlighting how big data is harnessed to infer intricate tree structures. One of the key drivers behind the advancement of phylogenetic techniques is the exponential growth of biological data. The rapid development of high-throughput sequencing technologies has led to the generation of vast genomic datasets encompassing diverse species. This wealth of genetic information provides researchers with the opportunity to explore relationships between species at unprecedented resolutions.

Homology, the presence of shared ancestry, is a fundamental concept in phylogenetics. Identifying homologous regions within sequences allows researchers to discern conserved genes and motifs, providing insights into evolutionary relationships. Big data aids in the identification of subtle homologous features, contributing to the accuracy of tree reconstruction. Maximum Likelihood (ML) and Bayesian inference, both methods aim to find the most probable evolutionary tree given the data, but they differ in their approach. Maximum Likelihood seeks to find the tree that maximizes the likelihood of the observed data under a given evolutionary model. This approach involves exploring a vast tree space and optimizing model parameters to fit the data. Advances in computing power have made it possible to efficiently search through large tree spaces, enabling the analysis of extensive datasets.

Bayesian Inference, on the other hand, treats the tree as a random variable and uses Bayes' theorem to update the probability of different trees based on the observed data. It involves sampling from the posterior distribution of trees, which requires significant computational resources. Big data has enabled the application of more complex Bayesian models, leading to better handling of uncertainties in tree inference. Phylogenomics is an emerging field that leverages genomic data to reconstruct evolutionary relationships. Rather than focusing on a single gene, researchers use information from thousands of genes to create a more comprehensive picture of evolution. This approach enables the resolution of previously unresolved branches in the tree and improves the accuracy of tree inference [1].

***Address for Correspondence:** Rosset Lemey, Department of Human Genetics and Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; E-mail: rosset@le.ucla.edu

Copyright: © 2023 Lemey R. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 06 July, 2023, Manuscript No. [ijbdt-23-111955](#); **Editor assigned:** 08 July, 2023, Pre QC No. [P-111955](#); **Reviewed:** 22 July, 2023, QC No. [Q-111955](#); **Revised:** 27 July, 2023, Manuscript No. [R-111955](#); **Published:** 04 August, 2023, DOI: [10.37421/2376-0214.2023.9.55](#)

Description

Supermatrix approaches involve concatenating sequences from different genes into a single matrix, resulting in a larger dataset for analysis. While this technique increases the computational demands, it also harnesses the power of big data to provide a more robust and detailed evolutionary history. While computational techniques in phylogenetics have advanced significantly, challenges remain. Handling the massive amounts of data generated by high-throughput sequencing platforms requires efficient algorithms and scalable computing resources. Additionally, the choice of evolutionary models, treatment of missing data and accounting for potential biases pose ongoing challenges [2].

The future of computational phylogenetics holds exciting prospects. As computational power continues to increase and algorithms become more sophisticated, researchers will be able to analyze even larger datasets with improved accuracy. Integrating additional sources of data, such as morphological traits and fossil records, will lead to more comprehensive and reliable phylogenetic trees. The field of phylogenetics has been propelled into a new era by the confluence of big data and advanced computational techniques. As genetic data continues to accumulate and computational tools evolve, researchers are poised to uncover more nuanced and accurate evolutionary relationships among species. The ability to harness vast datasets and employ powerful algorithms marks a transformative moment in our understanding of the tree of life [3].

With the immense growth in genomic data and the collaborative nature of scientific research, ethical considerations and data sharing have become paramount. As researchers leverage big data for phylogenetic analysis, they must ensure that they have the necessary permissions and adhere to data privacy regulations. Genomic data often contains sensitive information that could potentially be traced back to individuals or populations. Open data sharing practices can greatly accelerate scientific progress, enabling researchers to build upon each other's work and validate findings. However, striking a balance between data accessibility and protecting individual privacy remains a challenge. Machine learning techniques have begun to intersect with phylogenetics, offering new avenues for analysis and prediction. Deep learning models, such as convolutional neural networks and recurrent neural networks, have demonstrated promising results in tasks like sequence alignment and predicting protein structures. Integrating machine learning algorithms with traditional phylogenetic methods could lead to improved accuracy and efficiency in tree inference [4].

As the field of computational phylogenetics evolves, the demand for skilled researchers who can harness the power of big data continues to grow. Educational initiatives and training programs play a crucial role in equipping scientists with the computational and analytical skills needed for modern phylogenetics. By fostering interdisciplinary collaborations between biologists, computer scientists and statisticians, the scientific community can collectively address the challenges posed by big data analysis. The insights gained from computational phylogenetics have far-reaching implications beyond academia. Understanding the evolutionary relationships among species has practical applications in fields such as conservation biology, drug discovery and understanding the spread of infectious diseases. By translating research findings into actionable strategies, computational phylogenetics can contribute to solving real-world problems [5].

Conclusion

Computational techniques in phylogenetics have ushered in an era of unprecedented insight into the evolutionary history of life on Earth. The synergy between big data and advanced algorithms has empowered researchers to uncover intricate relationships among species, reshaping our understanding of biodiversity. As technology continues to evolve, computational phylogenetics will remain at the forefront of biological research, driving innovation and discovery in ways that were once unimaginable. By embracing ethical considerations, fostering collaboration and continuing to push the boundaries of computational methods, scientists are poised to unravel the mysteries of evolution and illuminate the interconnectedness of all living organisms.

Acknowledgement

We thank the anonymous reviewers for their constructive criticisms of the manuscript.

Conflict of Interest

The author declares there is no conflict of interest associated with this manuscript.

References

1. Azouri, Dana, Shiran Abadi, Yishay Mansour and Itay Mayrose, et al. "Harnessing machine learning to guide phylogenetic-tree search algorithms." *Nat Commun* 12 (2021): 1983.
2. Dornburg, Alex, Jeffrey P. Townsend and Zheng Wang. "Maximizing power in phylogenetics and phylogenomics: A perspective illuminated by fungal big data." *Adv Genet* 100 (2017): 1-47.
3. Schwartz, Russell and Alejandro A. Schäffer. "The evolution of tumour phylogenetics: Principles and practice." *Nat Rev Genet* 18 (2017): 213-229.
4. Suchard, Marc A. and Andrew Rambaut. "Many-core algorithms for statistical phylogenetics." *Bioinform* 25 (2009): 1370-1376.
5. Philippe, Hervé, Henner Brinkmann, Dennis V. Lavrov and D. Timothy J. Littlewood, et al. "Resolving difficult phylogenetic questions: Why more sequences are not enough." *PLoS Biol* 9 (2011): e1000602.

How to cite this article: Lemey, Rosset. "Computational Techniques in Phylogenetics: Harnessing Big Data for Tree Inference." *J Biodivers Biopros Dev* 9 (2023): 55.