



# Editorial: Methods and Applications in Molecular Phylogenetics

Juan Wang\*

School of Computer Science, Inner Mongolia University, Hohhot, China

**Keywords:** molecular phylogenetics, whole genome sequences, protein, application, disease

## Editorial on the ResearchTopic

### Methods and Applications in Molecular Phylogenetics

The purpose of molecular phylogenetics is to infer the evolutionary history of organisms and gene sequences. In the early stages of research, molecular phylogenetics mainly considers the changes vertically, such as insertion, substitution, and deletion in loci (Siepel and Haussler, 2004). With the development of sequencing technologies, the whole genomes are available for more and more organisms and are used to analyze their phylogenetics (Henz et al., 2005; Birin et al., 2008). The evolutionary history of organisms at this stage is described as a phylogenetic tree (Bruno et al., 2000). Then, genes of genomes are rearranged under horizontal events, such as inversions, duplications, and transpositions, which change the content and order of genes. Many studies introduce computing methods of molecular phylogenetics for whole genomes (Greenman et al., 2012). Phylogenetic networks are used to describe the evolutionary history (Wang and Guo, 2019). Molecular phylogenetics has been applied in many areas, such as the analysis of proteins (Lv et al., 2020).

Traditional methods for molecular phylogenetics need to do the alignment for sequences. It is very time-consuming to process the alignment of whole genome sequences. Therefore, it is a hard issue to do phylogenetic analysis from whole genome sequences of organisms. Wu et al. introduce a metric called information-entropy position-weighted k-mer relative measure (IEPWRMkmer), which combines the position-weighted measure and the information entropy of frequency for k-mers. Accordingly, they denote the whole genomes as feature sequences and then use Manhattan distance to compute the distance between two whole genomes. Finally, they use the Neighbor-Joining method to construct the phylogenetic tree from distance matrices. The IEPWRMkmer is efficient and effective for extracting key information for evolutionary analysis, and it is free to align for whole genomes.

Many studies have been done in applications of molecular phylogenetics. A protein complex contains proteins that interact with each other in function due to the evolutionary relationship. Wang et al. used semantic information of GO terms and the topological information of PPI networks to propose a method called TSSN for constructing a weighted PPI network. They proposed a new algorithm (NNP) for recognizing protein complexes from the weighted PPI network. Experiments showed that the algorithm could identify more protein complexes more accurately. PredMHC, proposed by Chen et al., is used to predict major histocompatibility complex (MHC). The PredMHC extracts information on amino acid composition from proteins, which is different due to the evolution of coding genes. It uses the voting of the SGD, the SMO, and random forest to predict and achieve the best performance on both training and testing datasets than other methods.

Molecular phylogenetics is also applied in predicting disease-related proteins. Anti-inflammatory peptides (AIPs) are important to treat some inflammatory and autoimmune diseases. Zhao et al. introduced a model (called iAIPs) to identify AIPs. iAIPs extract features from AIPs based on the information of sequences changed in evolution and then use the random forest to train.

## OPEN ACCESS

### Edited and reviewed by:

Simon Charles Heath,  
Center for Genomic Regulation (CRG),  
Spain

### \*Correspondence:

Juan Wang  
wangjuan@imu.edu.cn

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 19 April 2022

**Accepted:** 27 May 2022

**Published:** 14 July 2022

### Citation:

Wang J (2022) Editorial: Methods and  
Applications in  
Molecular Phylogenetics.  
Front. Genet. 13:923409.  
doi: 10.3389/fgene.2022.923409

Experimental results show that iAIPs can identify AIPs accurately. Cancer is a serious threat to human health and is one of the main causes of disease death. MultiGATAE, proposed by Zhang et al., can identify the cancer subtypes. It first constructs a similarity graph from multi-omics data (i.e., mRNA, miRNA, and DNA methylation) and then uses a deep learning method to learn embedding representation. It uses the K-means clustering

method to identify cancer subtypes from embedding representation.

## AUTHOR CONTRIBUTIONS

JW wrote the manuscript.

## REFERENCES

- Birin, H., Gal-Or, Z., Elias, I., and Tuller, T. (2008). Inferring Horizontal Transfers in the Presence of Rearrangements by the Minimum Evolution Criterion†. *Bioinformatics* 24 (6), 826–832. doi:10.1093/bioinformatics/btn024
- Bruno, W. J., Succi, N. D., and Halpern, A. L. (2000). Weighted Neighbor Joining: a Likelihood-Based Approach to Distance-Based Phylogeny Reconstruction. *Mol. Biol. Evol.* 17 (1), 189–197. doi:10.1093/oxfordjournals.molbev.a026231
- Greenman, C. D., Pleasance, E. D., Newman, S., Yang, F., Fu, B., Nik-Zainal, S., et al. (2012). Estimation of Rearrangement Phylogeny for Cancer Genomes. *Genome Res.* 22 (2), 346–361. doi:10.1101/gr.118414.110
- Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K., and Schuster, S. C. (2005). Whole-genome Prokaryotic Phylogeny. *Bioinformatics* 21 (10), 2329–2335. doi:10.1093/bioinformatics/bth324
- Lv, Z., Wang, P., Zou, Q., and Jiang, Q. (2020). Identification of Sub-golgi Protein Localization by Use of Deep Representation Learning Features. *Bioinformatics* 36 (24), 5600–5609. doi:10.1093/bioinformatics/btaa1074
- Siepel, A., and Haussler, D. (2004). Phylogenetic Estimation of Context-dependent Substitution Rates by Maximum Likelihood. *Mol. Biol. Evol.* 21, 468–488. doi:10.1093/molbev/msh039
- Wang, J., and Guo, M. (2019). A Review of Metrics Measuring Dissimilarity for Rooted Phylogenetic Networks. *Briefings Bioinforma.* 20 (6), 1972–1980. doi:10.1093/bib/bby062

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.