

Phylogenetic Analyses: Past, Present, and Future

Keynote Address

Bernard M.E. Moret
School of Computer and Communication Sciences
EPFL (Swiss Federal Institute of Technology)
Lausanne, Switzerland
<http://lcbp.epfl.ch>

Abstract

Phylogenies are simplified histories of the evolution of a group of taxa (organisms, genes, biological networks, computer malware, artistic styles, etc.) These phylogenies are inferred from modern-day specimens, in a process that starts by collecting comparable data about the taxa (such as the sequences of a few genes), then devising an appropriate model of evolution for the data, and finally running an inference procedure (machine-learning) to obtain a tree and some parameter values about that tree. Each year, thousands of citations are made to existing phylogenetic inference packages, mostly in the life sciences, but also in computer science, linguistics, forensics, and art history. As enounced by Th. Dobzhansky in the title of one of his papers, “biology makes no sense except in the light of evolution” and phylogenetic analyses are our spotlights. Yet in this talk I will argue that phylogenetic analyses are underused and in need of generalization.

For the last 80 years, phylogenies have used sequence data as the basis for inference; at first these sequences coded for morphological characteristics or simple genomic characteristics such as chromosomal banding; for the last 40 years, they have been RNA or DNA sequences. Phylogenetic analyses of languages, artistic styles, criminal activities, biological networks, or entire genomes have had to use tools developed to analyze relatively short sequences with very simple evolutionary models: the complexity of evolutionary models for other data, along with the relative paucity of studies based on such data, prevented the development of analysis techniques better adapted to the data. We thus need to enlarge and generalize existing techniques to improve the quality of phylogenetic analyses of data other than genome sequence data and to enable phylogenetic analyses for entirely new types of data. In particular we need new models, sophisticated preprocessing, and reasonable optimization criteria.

Most importantly, we need an enlightened view of phylogenetic analyses in science. We are all familiar with comparative methods, but a comparison between two taxa or a collection of pairwise comparisons among a collection of taxa is just a *degenerate* phylogenetic analysis, one that makes no (or minimal) use of evolution and models. In any area where the objects of study are subject to some form of evolution, phylogenetic analyses will yield much better results than simple comparative studies.

Bibliography

Bernard M.E. Moret grew up in Switzerland, received baccalauréats in Latin-Greek and Latin-Mathematics, then did a Diploma in Electrical Engineering at EPFL. After working for 2 years for Omega and Swiss Timing on the development of real-time OS for sports applications, he left for the US. He received his PhD

from the U. of Tennessee in 1980 and joined the University of New Mexico (UNM), where he served as chairman of the department from 1991 to 1993. In 2006 he accepted the chair of Computational Biology and Bioinformatics in the School of Computer and Communication Sciences at EPFL.

He founded the ACM Journal of Experimental Algorithmics (JEA) and served as its Editor-in-Chief for 7 years; he also helped found the IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), where he served as Associate Editor until 2008. He founded the annual Workshop on Algorithms in Bioinformatics (WABI) and chairs its steering committee. Until summer 2008, he chaired the Biodata Management and Analysis (BDMA) study section of the US National Institutes of Health (NIH). He led a team of over 50 biologists, computer scientists, and mathematicians in the CIPRES (Cyber Infrastructure for Phylogenetic Research) project, funded by the US National Science Foundation (NSF). He has published over 100 papers in computational biology in the last 15 years, under funding from the US NSF, the Alfred P. Sloan foundation, the IBM Corporation, the US NIH, the Swiss NSF, and SystemsX.ch.

His main research interests are modelling and algorithmic questions in evolutionary biology, particularly in phylogenetic analysis; for the last 15 years, he has worked mostly on problems in the evolution of whole genomes, but he has also published papers on problems in biological network evolution, the evolution of alternative splicing, and more theoretical papers on measures and methods in phylogenetic analysis.