



Jeffrey L. Thorne, Ph.D.

Professor of Genetics and Statistics

North Carolina State University

Born in 1963 in Evanston, Ill., USA

Studied Mathematics and Molecular Biology at the University of Wisconsin and Genetics at the University of Washington and at Cornell University

FOCUS

PROJECT

Statistical Techniques for Studying DNA and Protein Sequence Evolution

My research focus is on developing statistical techniques for extracting evolutionary information from DNA sequence data. During the last several years, I have mostly concentrated on connecting genotypic change (i. e., DNA sequence evolution) to phenotype (i. e., the characteristics of an organism). The relationship between genotype and phenotype is central to biology, but statistical techniques for studying how this relationship affects evolution are primitive.

A wide variety of computational techniques aim to predict aspects of phenotype (e. g., protein structure, RNA structure, antigenicity) from DNA sequences. We are working to improve probabilistic models of molecular evolution by incorporating these computational biology prediction systems. Because phenotype can induce dependence among changes that occur at different positions within a gene sequence, conventional procedures for making evolutionary inferences are not computationally feasible and we have developed new procedures.

On a separate topic, I am also interested in combining fossil and DNA evidence to estimate divergence times (i. e., the times since most recent common ancestry of related evolutionary lineages). Within this topic, another emphasis is to characterize the pattern by which chronological rates of sequence evolution change over time. This is important because the pattern of evolutionary rate

change has great potential to illuminate the process of biological evolution.

Recommended Reading

Thorne, J. L., H. Kishino, and J. Felsenstein. "An evolutionary model for the maximum likelihood alignment of DNA sequences." *J Mol Evol* 33: (1991): 114-124.

Thorne, J. L., H. Kishino, and I. S. Painter "Estimating the rate of evolution of the rate of molecular evolution." *Mol Bio Evol* 15: (1998): 1647-1657.

Robinson, D. M., D. T. Jones, H. Kishino, N. Goldman, and J. L. Thorne. "Protein evolution with dependence among codons due to tertiary structure." *Mol Biol Evol* 20, 10 (2003): 1692-1704.

DNA as a historical tool

The majority of the most important events in the history of life occurred millions or even billions of years prior to the advent of humans who could record them. As is the case for other historians, evolutionary biologists want to know as much as possible about what events occurred and they also want to characterize the forces that gave rise to those events. Some indispensable information about these events can be extracted from the fossil record. However, the fossil record is sparse and can be difficult to interpret.

The other major source of information about biological history is genetic information from extant organisms. Recent technological improvements have dramatically improved the ability to collect genetic material such as DNA sequence data. As a result, DNA sequence data now have a fundamental role in the study of biological history.

The phenotype of an organism is its appearance and behavior whereas the genotype of an organism is its DNA. The impact of the phenotype on perpetuation of the genotype is a central issue in evolutionary biology. When assessing natural selection, one challenge that needs to be overcome is that the relationship between genotype and phenotype is often poorly understood. Increasingly, computational methods are being developed to predict aspects of phenotype from DNA sequences. Our strategy is to incorporate these phenotypic prediction schemes into probabilistic models of DNA sequence change. We then base statistical techniques upon the resulting evolutionary models. Our techniques infer the effects of phenotypic traits on rates of DNA sequence change.

Most statistical procedures for assessing natural selection are designed to analyze data sets consisting of DNA from different individuals of the same species. In contrast, our procedures are designed to analyze data sets consisting of DNA sequence data from different species. This is important because the within-species genetic variation for some aspects of phenotype may be negligible or uninteresting, whereas it is much more likely that important genetic variation has accrued subsequent to the divergence of different species.

This talk will provide an overview of our statistical approach and some of its strengths and weaknesses. A focus will be on how we have applied the approach to quantify the influences of protein three-dimensional structure and RNA secondary structure on DNA sequence change. The talk will conclude with a discussion of potential future directions for this line of research.

PUBLICATIONS FROM THE FELLOW LIBRARY

Thorne, Jeffrey L. (2007)

Population genetics without intraspecific data

<https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1047193876>

Thorne, Jeffrey L. (2007)

Quantifying the impact of protein tertiary structure on molecular evolution

<https://kxp.k10plus.de/DB=9.663/PPNSET?PPN=1047193353>