# The Troubled Growth of Statistical Phylogenetics

JOSEPH FELSENSTEIN

*Department of Genetics, University of Washington, Seattle, Washington 98919 USA;*
*E-mail: joe@genetics.washington.edu*

Statistical inference of phylogenies almost didn't happen. The story of the origin, growth, and spread of "statistical phylogenetics" needs to be told, because it is so strange. It is not the straightforward story of gradual spread that one might imagine.

It starts with the development of numerical methods in systematics, whose modern proponents were Sokal and Sneath. Their work, embodied in their book *Principles of Numerical Taxonomy*, set off an explosion of work by mathematical clusterers, but did not win many converts in systematics. In the early 1960s two groups started work on numerical inference of phylogenies. Edwards and Cavalli-Sforza, working on trees of human populations and using gene frequencies, invented parsimony and distance matrix methods. In an attempt to choose between these two approaches, they turned to maximum likelihood, which proved to be different from both of the methods. At about the same time, Camin and Sokal began using parsimony on discrete characters, partly to prove how inaccurate it would be. Their paper in *Evolution* in 1965 was widely noticed. It contained detailed discussions of their methods of searching in tree space for the most-parsimonious tree.

In 1966 Eck and Dayhoff used parsimony on molecular sequences for the first time, and in 1967 Walter Fitch and Emanuel Margoliash gave the first detailed description of distance methods and showed their application to protein sequences. When I started writing phylogeny programs and fretting about the statistical properties of these methods, I had read Sokal and Sneath and also Camin and Sokal and had met Sokal. The International Congress of Human Genetics in 1966 brought Edwards, Cavalli-Sforza, and Fitch to Chicago, where I was studying. George Estabrook passed through our department in Chicago soon after, and I met Steve Farris in 1967 at a population biology symposium in Syracuse, New York. I rapidly got the sense of a small band of pioneers exploring new territory.

The stage would now seem set for a gradual spread of statistical methods, but reality was not to be this simple. In 1969 I began to attend the annual Numerical Taxonomy conferences convened by Bob Sokal. In 1971, at that meeting in Ann Arbor, Gareth Nelson advocated Willi Hennig's strictly monophyletic classification. It became clear that some systematists wanted to take a well-defined, almost algorithmic approach. Hennig set forth well-defined methods for inferring phylogenies (provided there was no internal conflict in the data), an approach with enormous appeal to a new generation of morphological systematists.

The difficulty was that although well-defined philosophical principles had great appeal, numerical methods and statistics did not. Nonetheless, some method was needed to resolve conflict among characters. For the new generation of morphological systematists, parsimony was to play this role. They also raised the temperature of debate. Phylogenetic systematists found themselves trying to dislodge a dominant school of evolutionary systematists who would not make their methods explicit and algorithmic. The phylogenetic systematists responded with an increasing stridency.

Pheneticists were algorithmic but not phylogenetic. The rest of the numerical taxonomy crowd were interested in phylogenies but not committed to Hennig's message. That there would be a schism over this message among the numerical types was inevitable, and a few allied themselves with the phylogenetic systematists and moved away from a statistical framework. The polarization was increasingly felt within the Numerical Taxonomy Conference meetings, which were among the only forums for debate on new systematic methods. More and more morphological systematists began to show up, taking the phylogenetic systematic position, wearing with pride Ernst Mayr's epithet for

them, "cladists". The atmosphere grew tense and warlike.

At the same time numerical criteria were gaining ground in the new field of molecular evolution. Molecular evolutionists were not interested in philosophical frameworks or issues of how to classify. They were more pragmatic and eclectic. In the early 1970s major advances were made in the numerical treatment of molecular sequences. Fitch discovered his algorithm for counting the minimum number of steps necessary to evolve a DNA site on a given phylogeny. David Sankoff generalized it, and then went on to propose a parsimony framework that integrated sequence alignment and inference of phylogenies. In this he was about 20 years ahead of his time.

The climax of tensions in the Numerical Taxonomy meetings came in 1979 at Harvard. I recall having to steel myself for about 6 months in advance just to be able to cope with the meeting. The two camps sat on on opposite sides of the auditorium (cladists on the right, others on the left), and the atmosphere was totally partisan. When a speaker from one side made a point, the other groaned, while adherents murmured approval. It was not a relaxing meeting. I had a premonition of an organizational split, having been through similar events 10 years earlier in radical student organizations.

The analogy proved exact. About 6 months later the cladists founded the Willi Hennig Society as a direct reaction to a breakdown of negotiations over the format of the next Numerical Taxonomy meeting. Many of the most important young systematists quickly joined, in a spirit of crusading zeal. Their influence in such journals as *Systematic Zoology* rose rapidly.

Historians like David Hull have tended to see this as a struggle over classification, phylogenetic versus phenetic or evolutionary. It was that in part, but in addition many of the people who had been developing numerical and statistical methods for inferring phylogenies found themselves outside the Hennig Society. And outside versus inside was important. The cladists of that era had accepted a number of points as an intellectual package. At one point in the mid-1980s I tried to summarize the package and came up with these points, in order of importance:

- Use Hennig's terminology—autapomorphy, symplesiomorphy, and so forth—rather than terms like ancestral or derived.
- Classify cladistically; use only monophyletic groups.
- Do biogeography by vicariance (*pace* Hennig).
- Use only computer programs written by leaders in the Hennig Society, all others are fundamentally flawed.
- Use only parsimony methods. Compatibility methods are evil.
- Do not weight characters.
- Be hostile to molecular data.
- Consider your methods to be hypothetico-deductive.
- Fossils are to be treated the same as living species.
- Parasites always have exactly the same phylogenies as their hosts.
- It is important to go around saying that one cannot infer ancestor–descendant relationships.
- It is important to go around saying that species are individuals, not classes.
- Be skeptical of the reality of the species as nonoperational.
- History: William of Ockham told Popper to tell Hennig to use parsimony.

For those trying to survive outside the fold, the early 1980s were scary. We felt that the systematics journals were becoming closed to us. In 1981 I sent a paper on a statistical approach to weighting overseas, to the *Biological Journal of the Linnean Society*, rather than see it handled by *Systematic Zoology*. It was touch and go as to whether people outside the Hennig Society could continue to work on phylogenies. There was the recurrent fear that the National Science Foundation (NSF) Systematic Biology program would come to be controlled by reviewers who would automatically reject our grants. In fact, this never happened (though I have heard rumors that NSF Systematic Biology took formal steps to ensure that the strife would not affect their review of proposals).

The fighting in systematics grew more intense than in any field I have known. I used to think that we fought a lot in when I worked in population genetics, but in that field we used to sit side by side at meetings without

growing red-faced, hissing at each other, or spreading scurrilous rumors. In systematics, however, the controversy attracted extreme personalities, mostly to the other side of the issue (probably both sides felt this way; I still do).

Systematics and molecular evolution passed through this Dark Ages in the 1980s. By the early 1990s the mood was beginning to lighten. Conflicts within the Hennig Society drove a large fraction of young systematists to desert that organization, though without abandoning their commitment to phylogenetic systematics. Ed Wiley, Joel Cracraft, and Vicki Funk took the important step of taking over the Society of Systematic Zoology, playing a role in moving it out of the American Society of Zoologists and into meeting with the Society for the Study of Evolution and the American Society of Naturalists. These "evolution meetings" grew rapidly, providing a forum where open and friendly exchange of views could take place.

At the same time, statistical tools for phylogenies were becoming available, including maximum likelihood, the bootstrap, invariants, the KHT test, and parametric bootstrapping. This was reflected in available software such as PAUP, MacClade, and my own package, PHYLIP. Computers continued to get faster and could more easily carry out statistical methods. The prestige of parsimony as the only philosophically sound method of inferring phylogenies was tarnished by the realization that it sometimes would be statistically inconsistent. Computer simulation tests by John Huelsenbeck and others showed that distance and likelihood methods work at least as well as parsimony.

There have thus emerged two schools of thought, statistical phylogenetics and cladistic parsimony. Many younger systematists think of their inferences of phylogeny as being basically statistical, and their choices of methods are driven by pragmatism rather than by prior philosophical commitment. Within the Hennig Society, an operationalist strain of thought emerged in the 1980s as "pattern cladistics"; it shared with phenetics the view that the patterns in the data should be used whether or not they reflected history. This view was heavily criticized and effectively silenced within the society, but its very critics then adopted a quite similar view when they said that parsimony should be used even when it is statistically inconsistent.

Although the cladistic parsimony school in the Hennig Society has renewed its criticism of statistical approaches, these have become established as an important part of the standard repertory. In the era of Markov Chain Monte Carlo methods, Hidden Markov Models, and mathematical genomics, little sign is left of the harrowing conflicts of the 1980s. We old survivors of the systematics wars sit around nursing our post-traumatic stress disorder and telling our war stories to the few who will listen. The late-1990s generation wonders what these old folks are grumbling about.

Historians and philosophers of science have scarcely noticed these developements. While they looked elsewhere, debate shifted from cladistics-versus-phenetics to whether inference of phylogenies should be statistical. The focus of systematics has shifted massively away from classification: It is the phylogenies that are central, and it is nearly irrelevant how they are then used in taxonomy. Once historians awake to these developments they will find a fascinating story, and one that happens to have a reasonably happy ending.