

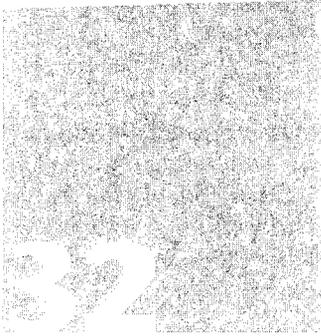
Asembling the Tree of Life

EDITED BY Joel Cracraft

Michael J. Donoghue

OXFORD
UNIVERSITY PRESS

2004



David M. Hillis

The Tree of Life and the Grand Synthesis of Biology

In the 1980s, there was rapid growth of the field of phylogenetics. The developments were so extensive that at the 1988 Nobel symposium titled "The Hierarchy of Life" (Fernholm et al. 1989), one participant wondered aloud if young biologists could be attracted into the field given that "all the big questions have been answered." I doubted that pronouncement; from my view, the field of phylogenetics was still in its nascent stages. I thought most of the big and interesting questions, as well as the major challenges, awaited us in the future. Morris Goodman agreed, and he described his vision for "a new age of exploration that promises to bring to fruition Darwin's dream of reconstructing the true genealogical history of life" (Goodman 1989:43). In many ways, that symposium did represent a turning point for phylogenetics, and the symposium that represents the subject of this book shows just how far we have come since the 1980s. The advances in progress on the Tree of Life have been greater in the 1990s than in all previous years combined, and the prognosis for the future has never been brighter.

A few comparisons between the 1988 Nobel symposium and the present symposium, "Assembling the Tree of Life," demonstrate just how much progress we have made. The description of PCR (polymerase chain reaction) had only been published the year before the Nobel symposium (Mullis and Faloona 1987), and DNA sequencing data were just beginning to have a major impact on the field of phylogenetic analysis. Statistical analysis of phylogenetic trees was

in its infancy in 1988, although several of the papers published in the proceedings of that symposium discussed emerging methods for assessing the strength of support for inferred trees. Even though data sets in 1988 were rather small by today's standards, computational resources (both software and hardware) were already limiting. Maximum likelihood analyses were virtually unmentioned at the 1988 symposium, and the computational constraints of such analyses made their application to large problems impractical. Therefore, systematists were severely limited by lack of data, weakly developed statistical methodology, and computational constraints. However, the stage was set for all of these bottlenecks to be removed or reduced.

In figure 32.1, I show an analysis of papers in the *Science Citation Index* for the past two decades (1982–2001). In 1982, there were 186 papers in the *Science Citation Index* that had the word "phylogeny" (or its derivative "phylogenetic") in the title, abstract, or key words. This means that it was possible to read about one paper every other day, and still read virtually all the literature on phylogenetics published worldwide. As I said above, the growth of the field through the 1980s was impressive: by the end of the decade, there had been more than a doubling of papers on phylogeny (393 papers in 1990), and in that year it would have been necessary to read more than a paper a day to read all the papers in the field. However, the real growth of the field of phylogenetics (at least in terms of number of papers published, and therefore in the number of phylogenetic trees presented) occurred throughout the 1990s.

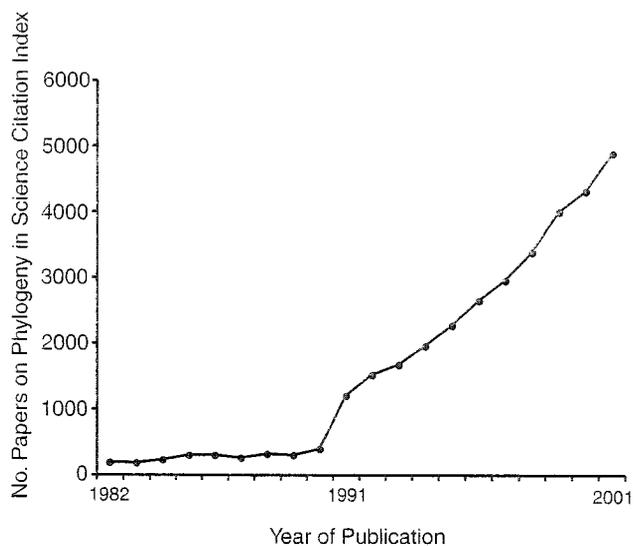


Figure 32.1. Numbers of papers in the *Science Citation Index* that include the words “phylogeny” or “phylogenetic” in the title, abstract, or key words, published from 1982 through 2001.

In 2001, almost 5000 papers were published on phylogeny. The total number of papers in the *Science Citation Index*, across all fields of science, was 999,618 in 2001. That means that a staggering 1 paper out of every 200 published in all fields of science was on phylogeny! Today, when I pick up a journal in almost any biological field, I expect to see some kind of phylogenetic analysis in at least one of the articles. If one wanted to attempt to read all the papers on phylogeny, that would require reading about 100 papers a week.

Recent progress on the Tree of Life has not resulted just because phylogenies are so much easier to infer now than they were a decade or so ago. The importance of understanding the relationships among the subjects of their studies finally became widely accepted (by biologists, of all fields) in the 1990s, as well. As phylogenies for many groups (as well as genes) became widely available, the power of comparative analyses became apparent in all areas of biology. Until phylogenies were widely available, biologists were likely to view objects of study in biology much as a chemist would view atoms in a chemical equation. Every hydrogen atom (of the same isotope) can be treated like all others. However, virtually nothing in biology is like hydrogen atoms. Every gene, every individual, every species, and every clade is more closely related (and more similar) to some genes, individuals, species, and clades than it is to others. This makes biology difficult, but not impossible. However, it does mean that every biologist must think at some level about phylogeny to put his or her work in the context of the rest of biology.

As I watched the presentations in this symposium, I was awed in two ways. First, the progress on reconstructing the Tree of Life has been nothing short of phenomenal. Our annual progress on understanding new relationships within

the Tree of Life is now much greater than all the accumulated knowledge on relationships that we had in the late 1980s. The applications of the Tree of Life to problems as diverse as forensics, origins of new diseases, ecology, behavior, development, molecular evolution, and assessment of global biodiversity is astonishing, and it is hard to keep up with all the new developments. Second, and despite all the recent progress, I was struck with the view that we are on the brink of yet another turning point: as the Tree of Life becomes more complete, its applications are also expanding exponentially. A complete Tree of Life would allow analyses that we would never contemplate today. Even the goal of discovering all the species on Earth is much more likely to be achieved if we have a complete Tree of Life for all the known species. A complete Tree of Life would allow us to catalog and organize all the species we know about, greatly increasing the potential to automate the discovery and description of the remaining unknown species. Fields such as ecology could move from treating communities as unknown “black boxes” to understanding their complexity and differences, perhaps allowing ecology to emerge as a truly predictive science. With phylogeny as a framework, molecular biology could move from a largely descriptive science to a field of explanation and prediction. The Tree of Life would also allow us to organize, connect, and synthesize all the information on all the species of Earth. A grand, web-based “encyclopedia of life” would result, and the field of biology would be immediately transformed. After that point, any information that anyone collected on any species would contribute to the understanding of all of life. In short, the Tree of Life represents the first (and most critical) step in the Grand Synthesis of biology.

Will someone writing an overview of the 2022 Tree of Life Symposium see the trend shown in figure 32.1 continue? My guess is that the trend will continue for at least a few years, but perhaps not decades, if the phylogenetic revolution is to be truly successful. The term “phylogeny” is now emphasized in papers that use phylogenetic methods in part because the approach is still considered innovative in many fields. However, in the future, if the Tree of Life initiative is truly successful, people will not think to distinguish their papers in this way. If all of biology is connected through a Tree of Life, then studying biology in a phylogenetic context should become almost transparent. People will include phylogenetic analyses as a matter of ordinary operating procedure. So, the best measure of the success of the phylogenetic revolution will come when analyzing biological data in a phylogenetic context merits as much of an emphasis in a paper as using a computer to analyze data does today, namely, something that virtually everyone does as a matter of necessity. And as with computers, new students in biology won't even be able to imagine how we ever got along without phylogenetic analysis.

Literature Cited

Fernholm, B., K. Bremer, and H. Jörnvall (eds.). 1989. The hierarchy of life: molecules and morphology in phylogenetic analysis. Excerpta Medica (Elsevier Science), Amsterdam.

Goodman, M. 1989. Emerging alliance of phylogenetic systematics and molecular biology: a new age of explora-

tion. Pp. 43-61 in The hierarchy of life: molecules and morphology in phylogenetic analysis (B. Fernholm, K. Bremer, and H. Jörnvall, eds.). Excerpta Medica (Elsevier Science), Amsterdam.

Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction. Methods Enzymol. 155:335-350.