

## The Molecular Tangled Bank: Not Seeing the Phylogenies for the Trees

LYNN MARGULIS<sup>1</sup>, MICHAEL DOLAN<sup>2</sup>, AND RICARDO GUERRERO<sup>3</sup>

<sup>1</sup> *Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003-5820;*

<sup>2</sup> *Organismic and Evolutionary Biology Graduate Program, University of Massachusetts, Amherst, Massachusetts 01003-5810; and* <sup>3</sup> *Department of Microbiology, University of Barcelona, Av. Diagonal, 645, Barcelona, Spain 08028*

About 40 years ago, in a gesture of civic enlightenment, the United States Congress passed legislation that converted the National Advisory Committee for Aeronautics (NACA; in existence since 1915) to NASA (National Aeronautics and Space Administration). The Space Act, no doubt in partial response to the launching of the first near-Earth satellite, Sputnik, by the Soviet Union, established NASA in 1958. This legislation is arguably the most well-written and enlightened government document to be created since the 18th century Jeffersonian Declaration of Independence.

Many of the stated goals of the 1958 Space Act were meant to generate new knowledge about the universe at large, but the statement that has most directly concerned molecular evolution is the following (1, p. 4):

A major goal of the National Aeronautics and Space Administration is to understand the origin and evolution of life as a phenomenon in the solar system [or wherever in the universe it may reside].

Three objectives are seen as fundamental to planetary biology and chemical evolution (1, p. 5):

- (1) Understanding the origin and evolution of life;
- (2) understanding the cycles that sustain life—the interactions between the physical, chemical and biological phenomena on the surface of the planet Earth; and
- (3) understanding the effects of life, both past and present, on the planet.

This paper was submitted in conjunction with a workshop titled *Evolution: A Molecular Point of View*. The workshop, which was held at the Marine Biological Laboratory, Woods Hole, Massachusetts, from 24–26 October 1997, was sponsored by the Center for Advanced Studies in the Space Life Sciences at MBL and funded by the National Aeronautics and Space Administration under Cooperative Agreement NCC 2-896.

Clearly the stated goals of planetary biology and chemical evolution of the National Academy's Space Science Board are congruent with the evolutionary ones of the NASA-sponsored Center for Advanced Studies in the Space Life Sciences at the Marine Biological Laboratory. Only one specific goal was stated: the Center was meant to provide a forum "for scientists to think and discuss" about *Evolution: A Molecular Point of View*.

This meeting at the MBL examined and explored in a new way evidence for ancestral relationships among the multitudinous extant organisms on Earth. By examining the sequences of amino acids in proteins, and the sequences of nucleotides that constitute nucleic acids, the community of scholars attempts to reconstruct from extremely recent data the history of the lineages of major forms of extant life. Macromolecular sequencing techniques, high-speed computing, international telephone lines that confer access to, and connect, biochemical databases have unified the study of evolution as never before in the history of life science.

The advantages of molecular evolution in reconstructing the evolutionary history of life are indisputable. Macromolecular sequences of proteins or nucleic acids taken from living organisms provide a large quantity of detailed data hitherto unavailable. High-speed computers permit alignment and assessment of vastly more data than ever before in unique ways not possible until recently.

However, here we recognize some problems and goals.

Macromolecules are always extracted from single individuals or a number of like individuals that were living in the same place at the same time (a population) and are taken to represent the taxon in question (*e.g.*, *Spongomonas* as Cercozoa, Cavalier-Smith, pp. 393–396, these proceedings). Clearly, there will be individual examples of named organ-

isms that, in fact, are not proper representatives of the taxa placed on the branches of phylogenetic drawings.

Since just after the publication of *Origins of Species* by Charles Darwin, 19th century naturalists, especially Ernst Haeckel, recognized that drawings with branches (dubbed "phylogenetic trees") can represent in two dimensions our concepts of the evolutionary history of life. In the era immediately prior to the molecular biological revolution, a lengthy review of evolutionary relationships among the smaller eukaryotes was published by R. Klein (University of Vermont) and A. Cronquist (director of the Bronx Botanical Garden in New York) (2). In an era where all organisms were either animals or plants, Klein and Cronquist were oriented to the living world as botanists: and they called the unruly set of beings that they analyzed "Thallophytes" ("lower plants" including "plant-like protozoa"). The problem of partial phylogeny was well analyzed by these authors: branching diagrams for relations between the parts of organisms (partial phylogenies) can be depicted independent of any full phylogeny that attempts to show the evolutionary history of the entire population or species of organism in question. The points they made are still useful. Partial phylogenies, even when they are fully valid for the characteristics (genes, pigments, organs, etc.) in question, are not, nor should they ever claim to be, full phylogenies. All molecular phylogenies are partial. Investigators in the new science of molecular evolution often contend that phylogenies based on ribosomal RNA are superior (or inferior) to those derived from protein sequences. Single protein sequences are less effective than more than one in positioning organisms on phylogenetic diagrams. Yet even multiple protein phylogenies fall short of descriptions of the evolutionary history of a species since all organisms, except a few bacteria, have thousands of proteins that are required at all times for self-maintaining metabolism. This implies that, to date, all published phylogenetic diagrams based solely on data from nucleic acid sequences, protein sequences, or a combination of both are seriously deficient.

Because they failed to incorporate ideas of plastid and mitochondrial origins by symbiogenesis, Klein and Cronquist published many incongruent phylogenetic diagrams. Similar inconsistent and even mutually exclusive molecular phylogenies have been published and reviewed at this meeting. In part, the problems stem from differing alignments and computing methods. But probably far more relevant to both the molecular and taxonomic contentiousness is the problem of unstated or inexplicit assumptions about the organisms, their names, and the extent to which the phylogenies are partial.

The major egregious unstated hypothesis is that the amount of sequence divergence in the extant organisms studied directly reflects their evolutionary divergence from

common ancestors. In fact, macromolecular sequence might easily be a more sensitive function of precise nuclease cleavage—highly selected genetic recombination leading to rearrangement of protein domains and consistent selective environmental constraints. Consider the case of all eukaryotes, from the time that they evolved through symbiosis (3), the rates of evolution of their component symbionts (bionts) have differed from the rates of the entire integrated organism (holobiont). Differences in the extent of genetic integration and the elimination of redundancy in multiple symbionts (e.g., like the seven genomic components that constitute *Mesodinium rubrum* (as reviewed by G. Sellers, Univ. of Maryland, unpubl. data), or the five biotic components of *Mixotricha paradoxa*) exacerbate the problem.

We have mentioned here but a few of the oversights that have led to overinterpretation of molecular phylogenies as reflective of past history. We feel that no substitute exists for knowing the live organism in question, its life history and habitat. Natural history, life cycle details, morphology including ultrastructure, the history of environmental change as documented by the rock record, and good sense are all crucial elements in reconstructing the ancient lineages of life on Earth.

Perusing this set of papers makes clear that both the problems and prospects are enormous: no doubt molecular insights will be crucial in resolving the genealogies of our remote ancestors.

The value of this first meeting must be emphasized. Our colleagues raise awareness of the relation between the goals of NASA and the construction of molecular phylogenies, state problems of extrapolation and overinterpretation from limited data, and point out the immense range of microbes and their descendants worthy of study. Future scientific gatherings under this aegis, we hope, will make explicit, through open discussion, some of the unstated assumptions and will promote better-evaluated molecular data. The amazing revelation of macromolecular sequences will then be integrated into other ways of learning about the evolutionary past history of life and its environment on planet Earth.

### Literature Cited

1. Space Science Board, Planetary Biology and Chemical Evolution, 1980, National Academy of Sciences, *National Academy of Sciences Press*, Washington DC.
2. Klein, R., and A. Cronquist. 1967. A consideration of the evolutionary and taxonomic significance of some biochemical, micromorphological and physiological characteristics in the Thallophytes. *Q. Rev. Biol.* 42: 105–296.
3. Margulis, L. 1996. Archaeal-eubacterial merges in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl. Acad. Sci. USA* 93: 1071–1076.