Ribosomal RNA Phylogeny and the Primary Lines of Evolutionary Descent

Norman R. Pace,* Gary J. Olsen,* and Carl R. Woese†

* Department of Biology
and Institute for Molecular and Cellular Biology
Indiana University
Bloomington, Indiana 47405
† Department of Genetics and Development
University of Illinois
Urbana, Illinois 61801

Over the past decade, developments in molecular phylogeny using ribosomal RNA (rRNA) sequences have revealed the outlines of the master phylogenetic tree relating all known life-forms. This serves as a framework within which we can begin to understand evolutionary diversity, the evolution of intermediary metabolism, and a host of other biological principles. We provide here a brief overview of some of the findings, which at times offer a different perspective than textbook versions of the evolutionary history of life.

As pointed out by Zuckerkandl and Pauling (J. Theor. Biol. 8, 357–366, 1965), macromolecular sequence comparisons are the most accurate and reliable basis from which to infer phylogenetic relationships. Sequence data are preferable to other molecular methods for assessing evolutionary relatedness (e.g., nucleic acid hybridizations and immunological tests) because they permit straightforward, quantitative interpretation and, importantly, because they form a growing data base for subsequent reference. Moreover, use of sufficiently lengthy sequences avoids mistaking relationships because of "convergence" of phenotypes. There is not necessarily a "best" molecule for comparative sequence studies. The only requirements are that the molecule's sequence does not undergo genetic transfer between species, that its function is as strictly constant as possible, and that it contains a sufficient number of residues which change at a rate commensurate with the evolutionary distance considered. The 16S-like and 23S-like rRNAs meet these criteria and are sufficiently conservative in structure that they can be used to establish the most ancient relationships among all organisms. The 5S rRNAs and tRNAs are also conservative in structure, but they contain too few independently varying nucleotide positions to establish the more distant relationships accurately.

About four hundred full and partial 16S-like rRNA sequences, mostly prokaryotic, have been determined. The figure shows an unrooted phylogenetic tree (Fitch and Margoliash, Science 155, 279–284, 1967) based on approximately one thousand unambiguously homologous characters (nucleotides) in 16S-like rRNAs. Extant organisms are seen to fall into three, phylogenetically coherent groups, the eukaryotes, the eubacteria, and the archaeabacteria (Pace et al., Science 209, 457–463, 1980). These are the primary lines of evolutionary descent—the three "primary kingdoms." The exact root of this tree cannot be placed because that requires a more deeply branching sequence, which cannot exist in a universal phylogeny. It is noteworthy that the line segments connecting the various modern organisms to their common ancestors are not the same length. This means that the lines of descent represented have not evolved at the same rate, a fact evident from comparisons of protein sequences, as well. I therefore, an assumption of isochronicity cannot be used to establish the root of the tree.

It often has been supposed that the eukaryotic lineage arose relatively recently, 1–2 billion years ago, from a prokaryotic ancestry. This age for the eukaryotes is loosely based on an absence of earlier, eukaryote-like fossils and on extended extrapolations of protein sequence comparisons. However, the figure shows that the eukaryotic nuclear line of descent did not arise from within either of the bacterial lines; rather, it seems as ancient as the bacterial lines. Regardless of the root position in the tree, deep branchings among the eukaryotes attest to this (Sogin et al., PNAS 83, 1383–1387, 1986). On the other hand, the molecular phylogenetic data (not only from rRNA) prove beyond reasonable doubt that the mitochondria and chloroplasts are of eubacterial origins (below).

The organisms most familiar to us, the multicellular plants and animals, occupy a rather shallow domain within the eukaryotic line of descent. The developmental programs of the multicellular forms engender incredible
diversity in form and in function. However, both bacteria and unicellular eukaryotes span far greater evolutionary histories.

Bacterial evolution was essentially uncharacterized until rRNA sequences were analyzed. Of the prokaryotes, it is the eubacteria that are most familiar to us. A sufficient variety of 16S rRNA oligonucleotide catalogs have been inspected that it is possible to define about ten large phylogenetic groups of eubacteria (Woese et al., System. Appl. Microbiol. 6, 143–151, 1985). Although representatives of only four of these “phyla” are shown in the figure, all diverge from the main line of descent at about the same depth. Each of the well-inspected groups, although phylogenetically coherent, contains representatives with greatly different phenotypes. For instance, the group of “purple bacteria and relatives” (represented in the figure by E. coli, P. testosteroni, and A. tumefaciens) contains all of the major nutritional themes — heterotrophs (e.g., E. coli), chemotrophs (e.g., Thiobacillus, a sulfur oxidizer), and phototrophs (e.g., Chromatium). Thus, the traditional classification of bacteria, using nutritional and morphological aspects, often gives little indication of phylogenetic affiliations. The group of purple bacteria and relatives also includes many organisms that enter symbiotic or parasitic relationships with eukaryotes (e.g., rhizobacteria, agrobacteria, and rickettsias, all of which would group specifically with A. tumefaciens in the figure). The mitochondria, as well, were derived from the purple bacteria and relatives (Yang et al., PNAS 82, 4443–4447, 1985). Organelle genesis has not been unique to this group, however. The cyanobacteria (represented in the figure by A. nidulans), another of the ten eubacterial phyla, also enter symbiotic relationships with eukaryotes (e.g., lichens), and their lineages gave rise to the chloroplasts. It remains to be established whether the mitochondria and the chloroplasts each are monophyletic or polyphyletic; i.e., had singular origins or arose several times.

Comparative analysis of rRNA sequences first established the existence of two fundamentally distinct types of prokaryotes: eubacteria and archaeabacteria. Profound physiological and structural differences between the archaebacteria and the eubacteria confirm their deep evolutionary separation (The Bacteria, Vol. VIII. Archaebacteria, eds. C. R. Woese and R. S. Wolfe, Academic Press, 1985). Depending on the position of the root of the tree in the figure, archaebacteria are closer to the common ancestor of all the kingdoms than are one or both of the other primary kingdoms. This suggests that the archaebacterial phenotype is more primitive than one or both of the other lines.

The archaebacteria so far inspected exhibit one of three phenotypes: “sulfur-dependent” (formerly thermoacidophilic, represented in the figure by S. solfataricus and T. tengae), methanogenic, or extremely halophilic. Two thermoacidophilic genera, Thermoplasma and Thermococcus, are aligned by rRNA relatedness with the methanogen/halophile domain of the archaebacteria. This is an interesting finding because it suggests that the common ancestor to the main archaebacterial lineages also was thermophilic and sulfur-metabolizing, the most ancient phenotype that we so far can infer.

In light of the phylogenetic relationships that emerge from rRNA sequence comparisons, it is of interest to consider other sets of “kingdom” definitions. Two conspicuous ones are the five kingdoms of Whittaker (a popular textbook version), and the recent proposals of Lake and his colleagues.

The five kingdoms defined by Whittaker (Science 163, 150–159, 1969) are the animals, the plants, the fungi, the protists (unicellular eukaryotes), and the monera (prokaryotes). These definitions focus attention on the morphological diversity of the “higher” eukaryotes without recognizing the molecular and biochemical diversity now evident in the eukaryotic and prokaryotic microbes. Although evolutionary relationships between the multicellular forms and the microbes were suggested on the basis of modes of nutrition, themes which we now know can obscure true genealogy, the five kingdoms of Whittaker were considered fundamental units: no interkingdom affiliations could credibly be made. With current phylogenetic perspective, however, the natural relationships among the eukaryotes can be properly defined.

More recently, Lake et al. (PNAS 82, 3716–3720, 1985) have argued that the tree topology shown in the figure is incorrect. These authors would pool the extreme halophiles with the eubacteria, to form the “photocyte” kingdom; the sulfur-dependent archaebacteria would be raised to kingdom status (“ecocyte”), with a close relationship to the eukaryotes. The methanogens would remain the sole members of the archaebacteria. These conclusions were drawn in large part from assumed evolutionary progressions of certain features of ribosome morphology and from selected supportive correlations. We feel, however, that the proposals are based on too few characteristics to justify the conclusions (for a detailed critique, see Woese and Olsen, Syst. Appl. Microbiol. 8, in press), and that the best measures of kingdom-level relationships at this time are the unambiguously homologous sequence characters in the 16S-like rRNAs.

Although the outlines of a universal phylogeny are emerging, the survey of organisms has not been comprehensive. Many major groups have received little attention and analyses have largely been limited to laboratory cultures. Yet, organisms in natural microbial populations often are refractory to cultivation, as are many symbionts and obligate parasites. However, such organisms are amenable to rRNA-based phylogenetic analyses using cloned rRNA genes. In this manner, previously uninvestigated life-forms can be characterized phylogenetically, relative to other organisms of known rRNA sequence.

The phylogenetic characterization of organisms is more than an exercise in taxonomy, since evolutionary relationships are established in a credible and quantitative way. Closely related organisms are expected to be similar in their general biochemical properties; conversely, diversity in rRNA sequences indicates potential biochemical differences. Therefore, knowledge of evolutionary relationships points to organisms where novelties are to be found. Truly homologous traits common to evolutionarily diverse organisms must also have belonged to their common ancestor. Commonalities of the kingdoms must have resided in the “progenote,” the common ancestor of all life on this planet (Woese and Fox, J. Mol. Evol. 10, 1–6, 1977).