

## **Mycoplasma Evolution: A Review of the Use of Ribosomal and Transfer RNA Nucleotide Sequences in the Determination of Phylogenetic Relationships**

R.T. WALKER, D.Sc.

*Department of Chemistry, University of Birmingham, Birmingham, England*

Received January 4, 1983

---

Comparison of the nucleotide sequences of "structural" RNAs (ribosomal and transfer RNA) has enabled the construction of phylogenetic trees to be achieved. Data from 16S rRNA, 5S rRNA, and tRNA from a total of eight Mollicutes (excluding *T. acidophilum*) including representatives of the families *Mycoplasmataceae*, *Spiroplasmataceae*, and *Acholeplasmataceae*, show that these families share a close relationship and a common ancestor with the gram-positive eubacteria. *Thermoplasma acidophilum* is a member of the kingdom Archaeobacteria and has no relationship to the other Mollicutes.

---

The ideal information required for studying the phylogenetic relationship between different species is obviously the complete nucleotide sequence of the genome. While it is now possible to obtain this information and for one or two organisms this may well be available within a year or so (the longest sequence available in April 1983 was 48,502 base pairs) [1], it is unlikely that, in the foreseeable future, this information will be available for even a representative sample of organisms. The work entailed is unlikely to be justified by the results obtained.

Thus as a compromise it is necessary to compare limited pieces of the genome which are related to each other or, alternatively, to look at the gene products. At present, to identify the genome position coding for a specific protein can be a lengthy business, the alternative of comparing protein sequences is also tedious, and differences in genome sequence are often masked by code degeneracy. In any case, there are few, if any, proteins which are present in all species and which can be recognized as having evolved from a common precursor and so, although some valuable information in terms of fine structure has been obtained by comparison of, for example, cytochrome C sequences [2], the method does have its limitations.

There thus remains the possibility of comparing the gene products which are transcribed but not translated—the ribosomal RNAs (rRNA) and transfer RNA (tRNA). Comparison of the genome sequence coding for these "structural" molecules has the advantage that they are present in all cells and had thus evolved into their present form before the precursors of any of the species now present diverged. Thus in principle there should be no barrier to the comparison of phylogenetic relationships of species in all kingdoms.

If we accept that these structural RNAs (as compared to mRNA which contains its information in the form of nucleotide sequence) are suitable candidates for the study

of phylogenetic relationships, we have four candidates in any one species for consideration. These are three ribosomal RNAs—two from the large subunit and one from the smaller subunit—and as these will be the focus of another paper from this Symposium, they will only be briefly mentioned here for the sake of completeness. The large rRNA from the large subunit (23S in the bacterial system) has about 3,700 nucleotides. To sequence this length of RNA is still very difficult as RNA sequencing is not as straightforward as DNA sequencing and, whereas this length of DNA would be a reasonable task to approach for many different organisms, very few RNA sequences of this size have been reported [1]. The alternative of locating the gene for this rRNA species and sequencing that is again possible but tedious if many hundreds of different species need to be compared.

The rRNA from the small subunit (16S in the bacterial system) contains around 1,700 nucleotides. This is still a formidable challenge either by sequencing directly or by locating and sequencing the genome. A compromise has been made with this rRNA by Woese and his co-workers, who have compared sequences of the oligonucleotides produced by T<sub>1</sub> RNase digestion of this rRNA from several hundred species [3,4]. This work led to the discovery of a hitherto unrecognized kingdom—the *Archaeobacteriae*—whose members are as phylogenetically distinct from the *Eubacteriae* (true bacteria) as the latter are from the Eukaryotes. This work, which has been extensively reviewed, has laid the groundwork for our understanding of phylogeny on a molecular basis and has provided detailed phylogenetic trees which appear to be very sensible in areas where there is little controversy and hence demonstrates that this approach to phylogeny via the study of molecular sequences is valid.

The small ribosomal RNA (5S) was originally thought to be far too small (~120 nucleotides) for it to be able to still contain any phylogenetic information of any importance. The advantage of this molecule as a candidate for phylogenetic studies is that it is very easily isolated and sequenced and therefore can be readily compared from a wide range of species. This molecule has also been of considerable interest to molecular biologists for some years, and hence many sequences were available [5] when an attempt to extract phylogenetic information was made. Three groups have shown that this is a very real possibility, and phylogenetic trees have been published [6,7,8].

There now remains tRNA as the final molecule to be considered, which is the main subject of this review. tRNAs have one advantage, namely that they are interesting molecules in terms of structure and function in protein biosynthesis and they have therefore attracted a good deal of attention and were indeed the first nucleic acid to be sequenced [9]. Since then, several hundred sequences have become available [10], determined for the most part by people who had no or little interest in phylogeny. However, there are several disadvantages in sequencing tRNA molecules for phylogenetic purposes. Although the molecules are relatively small (~70–90 nucleotides), any cell normally contains around 50 or more different tRNAs which are used to translate the 61 codons. To sequence a tRNA, a unique species has to be isolated, and this can involve a lengthy process of separating the one molecular type required from some 50 or more almost identical molecules. RNA sequencing is not easy and tRNA sequencing is even more difficult because of the tight tertiary structure of the molecule and the presence of many modified bases.

There then arises the question as to whether the tRNA molecule contains any phylogenetic information anyway. Certainly, until recently the general opinion was

that it did not [11]. The molecule is small, there are many “invariant” bases (those which are present in the same position in nearly all tRNAs from all species), and within the double-stranded regions, if one base changed, in order to retain the base pair, the partner automatically has to change; thus the number of independent separate mutations which are allowable is much lower than the number of nucleotides in the sequence. Indeed, in the normal 76 nucleotide-containing tRNA, Eigen has shown that only 30 nucleotides can be accepted as being able to vary with complete independence [12]. Is this number of only 30 nucleotides sufficient for phylogenetic information to be extracted? The answer, quite clearly demonstrated by elegant analyses of the data by the independent groups of Cedergren et al. [13,14] and Eigen and Winkler-Oswatitsch [12] is that indeed sufficient information remains and that the constraint upon base mutation has been such that tRNAs do not contain random sequences.

Thus to summarize this theoretical introduction to the use of nucleotide sequence as a basis for phylogenetic relationships, the position is that the “structural” RNAs are very suitable candidates for this type of analysis. The larger ribosomal RNAs

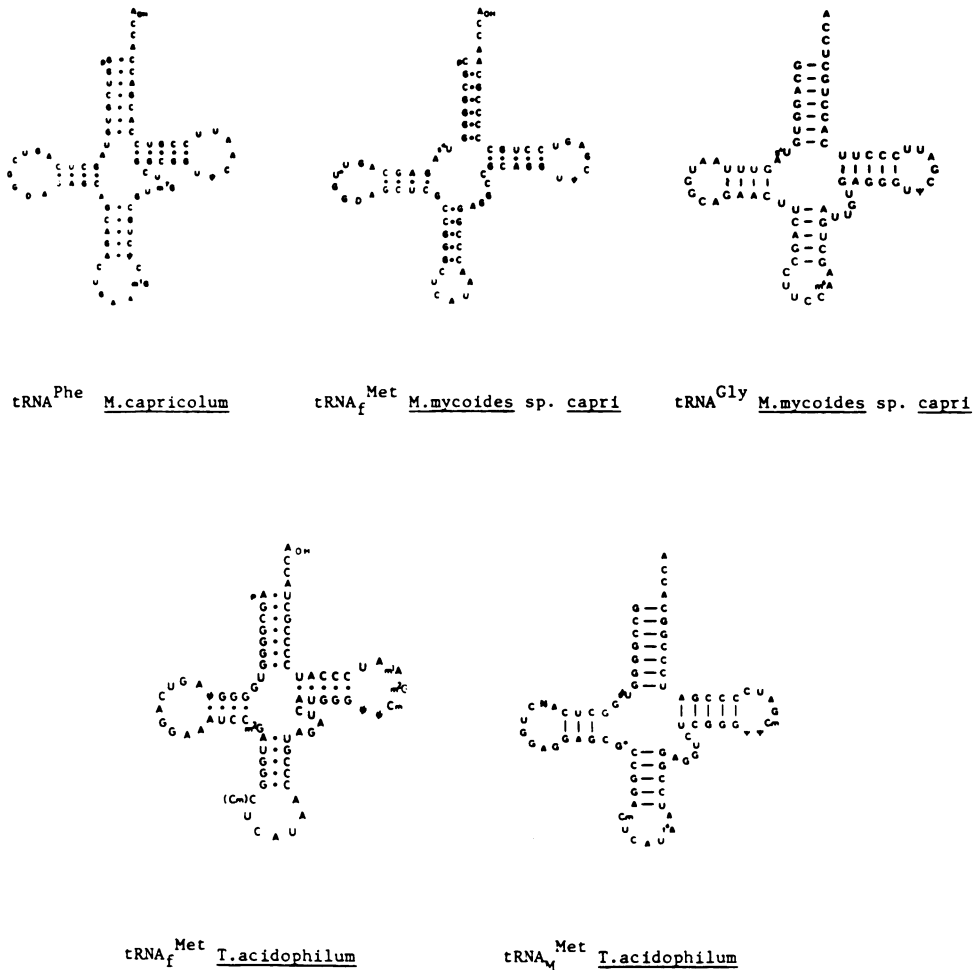


FIG. 1. Cloverleaf structures of all tRNAs sequenced from Mollicutes.

have been too large for complete analyses to be performed but the compromise of comparison of oligonucleotides found by specific enzymatic digestion does give very valuable, and so far the most detailed, information. 5S rRNA and tRNA sequences can also be used for this purpose; they have their disadvantages but, as long as the data analysis is interpreted with caution, general phylogenetic relationships can be deduced although the fine detail may need to be established by other means.

#### *Phylogeny of Thermoplasma acidophilum*

It is reasonable to deal with the phylogeny of *T. acidophilum* separately as, once its correct base analysis became available [15], it was clear that it could not possibly have any close relationship to the other members of the Mollicutes. Indeed, it has several distinctive properties which are now recognized to be characteristic of archaeobacteria, and the work of Woese and his co-workers clearly establishes from 16S rRNA data that this cell wall-less organism is definitely an archaeobacterium [4]; thus the loss of a cell wall has occurred at least twice in evolution. The results from 5S rRNA [16] and the tRNA<sub>f</sub><sup>Met</sup> [18] and tRNA<sub>M</sub><sup>Met</sup> [17] species (see Fig. 1) clearly confirm that *T. acidophilum* is an archaeobacterium, but at present there is not really sufficient information for one to be able to locate its closest relative with any certainty, although it does seem to have more in common with the other archaeobacterial acidothermophile *Sulpholobus acidocaldarius* than with the halophiles or methanogens [4]. The exact relationship between the ancestor of the present eukaryotes and the ancestral eubacterial and archaeobacterial species is not yet finally settled [19]. *T. acidophilum* tRNA<sub>f</sub><sup>Met</sup> sequence is very similar to that of yeast, whereas the tRNA<sub>M</sub><sup>Met</sup> sequence shows nearly 90 percent homology with the tRNA ancestral quasi-species deduced by Eigen and has little homology with anything else [12].

#### PHYLOGENY OF MEMBERS OF THE ORDER MYCOPLASMATALES

Only three tRNA sequences from mycoplasma species have been published. These are tRNA<sup>Phe</sup> from *M. capricolum* [20] and tRNA<sub>f</sub><sup>Met</sup> [21] and tRNA<sup>Gly</sup> [22] from *M. mycoides* sp. *capri* PG3 (see Fig. 1), 5S rRNA sequences from *M. capricolum* [23], *M. mycoides* sp. *capri*, Spiroplasma BC3 [24], Acholeplasma laidlawii [Woese, unpublished results], *M. pneumoniae* and *A. modicum* [Leach, Rogers, Walker, unpublished results] are also known. 16S rRNA Catalogue data is available for *S. citri*, *M. capricolum*, *M. gallisepticum*, and *A. laidlawii* [4]. Thus the data is very limited and rather fragmented and one must be careful not to generalize too much as there is only data available on these few species.

The tRNA data is interesting from many viewpoints but here we are only concerned with its phylogenetic significance. This shows quite clearly that all the sequences determined are characteristic of those expected for gram-positive bacteria [10]. As it happens, phylogenetic trees are available for all these three tRNA species (Phe [13,14], Met<sub>f</sub> [12], and Gly [13,14]), and in each case the mycoplasma species concerned is placed in such a position that it and the gram-positive bacteria share a common ancestor. It is not possible, nor is it ever likely to be possible, to be more specific than this; particularly when one bears in mind that with a genome size of N 10<sup>6</sup> bp, it is likely that 75 percent of the genome has been lost in the process of evolution from its ancestor.

Again one must emphasize the paucity of the data here. Strictly, all that has been shown is that *M. capricolum* and *M. mycoides* sp. *capri* have a common ancestor

with present gram-positive bacteria. Thus one can assume (and the 5S rRNA data supports this), that these two species are related to each other but, although it is tempting (and probably correct) to assume that all mycoplasma (including spiroplasma, ureaplasma, and acholeplasma) species are all related, there is as yet no evidence from tRNA sequences to support or refute this possibility. As we know that the cell wall has been lost on more than one occasion (*vide* thermoplasma), it is still possible that the remaining members of the Mollicutes have arisen independently from closely related ancestors. It is also not possible to be more specific concerning the identity of the gram-positive bacterium which most closely resembles any one mycoplasma, and this data will need to come from other sources—presumably serological.

Some 5S rRNA sequence data supports and extends these conclusions. *M. capricolum* [23] and *M. mycoides* sp. *capri* [Walker et al., unpublished results] differ in only three nucleotides and are thus probably as closely related to each other as two *E. coli* strains. They are phylogenetically related to the gram-positive bacteria but the unusually small size of mycoplasma 5S rRNAs (107 or 108 nucleotides), spiroplasma (107 nucleotides), and acholeplasma (112 and 108 nucleotides) makes it very difficult to be more specific than this. The 5S rRNA sequences of the bee Spiroplasma BC3 and *A. laidlawii* show that representatives of these two families are related to each other and to the mycoplasma. The 16S rRNA data of Woese [4] shows that *S. citri* and the other three Mollicutes investigated (*M. capricolum*, *M. gallisepticum*, and *A. laidlawii*) are related to each other.

The exact positioning of the different families in the order Mollicutes both to each other and with respect to other gram-positive bacteria is not straightforward because of the differences in mutation rates and/or genome sizes; thus the present clustering of mycoplasma and spiroplasma and a close but distinct clustering of acholeplasma with *Clostridium ramosum* and *C. innocuum* [3] must be regarded as provisional. The mycoplasma undoubtedly fall within the bacillus-lactobacillus-streptococcus cluster but even then the loss of a cell wall and loss of a major fraction of the genome may have occurred more than once. For a more detailed analysis of the phylogeny of these organisms, a recent review [25] should be consulted.

#### ACKNOWLEDGEMENTS

I would like to thank all those scientists who have been associated with my work with Mollicute RNA sequence determination. This started in 1973 with the award of a Leverhulme Fellowship to work at M.I.T. with Professor U.L. RajBhandary. This collaboration continued for some years with the award of a NATO research grant to myself and Professor H.G. Khorana and more recently with grants from the MRC, SRC, and SERC.

Several of the sequences described have been determined by Dr. M.W. Kilpatrick and more recently by M.J. Rogers and J. Simmons. The *T. acidophilum* 5S rRNA was sequenced in the laboratory of Professor H.J. Gross. Much of this work has been done in association with the Mycoplasma Reference Laboratory of the PHLS and I gratefully acknowledge the help and advice of Dr. R.H. Leach. Dr. Dennis Searcy has provided cells of the "pseudoplasma" *T. acidophilum* and the cells of *M. mycoides* sp. *capri*, and the bee Spiroplasma BC3 have been grown with great skill and devotion over many years by Mr. E.T.J. Chelton, M.I.Biol., F.I.M.L.T., to whom this paper is dedicated on the occasion of his retirement.

#### REFERENCES

1. EMBL: Nucleotide Sequence Data. Library News (April): 1983
2. Fitch WM, Margoliash E: Construction of phylogenetic trees: A method based on mutation distances as estimated from cytochrome C sequences is of general applicability. *Science* 155:279-284, 1967

3. Fox GE, Stackebrandt E, Hespell RB, et al: The phylogeny of prokaryotes. *Science* 209:457-463, 1980
4. Woese CR, Maniloff J, Zablén LB: Phylogenetic analysis of mycoplasma. *Proc Nat Acad Sci USA* 77:494-498, 1980
5. Erdmann VA, Huysmans E, Vandenberghe A, et al: Collection of published 5S and 5.8S ribosomal RNA sequences. *Nucleic Acids Research* 11:r105-r133, 1983
6. De Wachter R, Chen M-W, Vandenberghe A: Conservation of secondary structure in 5S ribosomal RNA: A uniform model for eukaryotic, eubacterial, archaeobacterial and organelle sequences is energetically favourable. *Biochimie* 64:311-329, 1982
7. Hori H, Osawa S: Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 54 5S rRNA species. *Proc Nat Acad Sci USA* 76:381-385, 1979
8. Küntzel H, Heidrich M, Piechulla B: Phylogenetic tree derived from bacterial, cytosol and organelle 5S rRNA sequences. *Nucleic Acids Research* 9:1451-1461, 1981
9. Holley RW, Appgar J, Everett GA, et al: Structure of a ribonucleic acid. *Science* 147:1462-1466, 1965
10. Gauss DH, Sprinzl M: Compilation of tRNA sequences. *Nucleic Acids Research* 11:r1-r53, 1983
11. Holmquist R, Jukes TH, Pangburn S: Evolution of transfer RNA. *J Mol Biol* 78:91-116, 1973
12. Eigen M, Winkler-Oswatitsch R: Transfer RNA: The early adaptor. *Die Naturwissenschaften* 68:217-228, 1981
13. Cedergren RJ, LaRue B, Sankoff D, et al: Convergence and minimal mutation criteria for evaluating early events in tRNA evolution. *Proc Nat Acad Sci USA* 77:2791-2795, 1980
14. Cedergren RJ, Sankoff D, LaRue B, et al: The evolving tRNA molecule. In *CRC Critical Reviews in Biochemistry, Vol II*. Edited by GD Fasman. Cleveland, CRC Press, 1981, pp 35-104
15. Searcy DG, Doyle EK: Characterization of *Rernoplasma acidophilum*. *Int J Systematic Bacteriol* 25:286-289, 1975
16. Luehrsen KR, Fox GE, Kilpatrick MW, et al: The nucleotide sequence of the 5S rRNA from the archaeobacterium *Rernoplasma acidophilum*. *Nucleic Acids Research* 9:965-970, 1981
17. Kilpatrick MW, Walker RT: The nucleotide sequence of the tRNA<sub>M<sup>Met</sup></sub> from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Research* 9:5407-6410, 1981
18. Kuchino Y, Ihara M, Yabusaki Y, et al: Initiator tRNAs from archaeobacteria show common unique sequence characteristics. *Nature* 298:684-685, 1982
19. Van Valen LM, Maiorana VC: The archaeobacteria and eukaryotic origins. *Nature (London)* 287:248-250, 1980
20. Kimball ME, Szeto KS, Soll D: The nucleotide sequence of phenylalanine tRNA from *Mycoplasma* sp. (Kid). *Nucleic Acids Research* 1:1721-1732, 1974
21. Walker RT, RajBhandary UL: The nucleotide sequence of formylmethionine tRNA from *Mycoplasma mycoides* sp. capri. *Nucleic Acids Research* 5:57-70, 1978
22. Kilpatrick MW, Walker RT: The nucleotide sequence of glycine tRNA from *Mycoplasma mycoides* sp. capri. *Nucleic Acids Research* 8:2783-2786, 1980
23. Hori H, Sawada M, Osawa S, et al: The nucleotide sequence of 5S rRNA from *Mycoplasma capricolum*. *Nucleic Acids Research* 9:5407-6410, 1981
24. Walker RT, Chelton ETJ, Kilpatrick MW, et al: The nucleotide sequence of the 5S rRNA from *Spiroplasma* species BC3 and *Mycoplasma mycoides* sp. capri. PG3. *Nucleic Acids Research* 10:6363-6367, 1982
25. Maniloff J: Evolution of wall-less prokaryotes. *Ann Rev Microbiol* 37:477-500, 1983