Volume 9 Number 6 1981 Nucleic Acids Research

Phylogenetic tree derived from bacterial, cytosol and organelle 5S rRNA sequences

Hans Küntzel, Meinhard Heidrich and Birgit Piechulla

Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Str. 3, D-3400 Göttingen, GFR

Received 19 Janaury 1981

ABSTRACT

A phylogenetic tree was constructed by computer analysis of 47 completely determined 5S rRNA sequences.

The wheat mitochondrial sequence is significantly more related to prokaryotic than to eukaryotic sequences, and its affinity to that of the thermophilic Gram-negative bacterium Thermus aquaticus is comparable to the affinity between Anacystis nidulans and chloroplastic sequences. This strongly supports the idea of an endosymbiotic origin of plant mitochondria.

A comparison of the plant cytosol and chloroplast sub-trees suggests a similar rate of nucleotide substitution in nuclear genes and chloroplastic genes. Other features of the tree are a common precursor of protozoa and metazoa, which appears to be more related to the fungal than to the plant protosequence, and an early divergence of the archebacterial sequence (Halobacterium cutirubrum) from the prokaryotic branch.

INTRODUCTION

5S rRNA, a component of the large ribosomal subunit (1), appears to be one of the most suitable macromolecules for the study of molecular evolution, for the following reasons:

(a) it is an almost universally occurring molecule which is apparently absent only from fungal and animal mitochondria (2-5),

(b) it has a rather conservative primary and secondary structure (6-8),

(c) it is a unique RNA of quite informative but still handlable length (116 to 120 nucleotides) which makes its purification and sequence determination relatively easy, and

(d) as a consequence of the latter reason, it is available from a rapidly increasing number of organisms (2).

Several phylogenetic trees derived from 5S rRNA sequences have already been published (9-14). This study was undertaken in order to incorporate new 5S rRNA sequence data from eu-ascomycetes (4), echinodermata, protozoa and chloroplasts (6), and to test the phylogenetic position of the recently determined wheat mitochondrial 5S rRNA sequence (15). In contrast to the most

recently published 5S rRNA tree (14), this analysis is based entirely on completely determined sequences, and the tree is constructed stepwise by considering established phylogenetic affinities.

METHODS

All 47 sequences listed in Table 2 were aligned to maximal sequence homology by introducing a minimal number of gaps. Those nucleotides occurring in only one species at a given position (see the alignment of representative sequences in Fig. 1) were considered as late inserts and, therefore, were eliminated from further analysis. Tree topologies were determined by the matrix method of Fitch and Margoliash (16). For determining differences between pairs of original sequences (difference matrix), a gap versus a nucleotide was counted as one-half the value of a nucleotide substitution (14). Higher values (17) were also tested, but they did not influence tree topologies.

Protosequences were constructed as follows: To each nucleotide or gap of an original sequence the index 1.0 was given. For all positions of a given pair of sequences (which may consist of two original sequences, two protosequences or a combination of both), the "nucleotide composition" (reflecting the probability of a given nucleotide to be at this position in the ancestral sequence) was determined by adding up the one-half indices for each nucleotide at a given position (an example is shown in Table 1).

Table 1: Example for constructing protosequences and calculating differences.

position	(proto)-sequence 1	(proto)-sequence 2	resulting proto-sequence	difference between sequences 1 and 2
a	A 1.0	A 1.0	A 1.0	0
b	G 1.0	C 1.0	G 0.5 C 0.5	1.0
c	U 1.0	gap 1.0	U 0.5 gap 0.5	0.5
d	A - U 0.5 G - C 0.5 gap -	A 0.5 U 0.25 G 0.25 C - gap -	A 0.250 U 0.375 G 0.125 C 0.250 gap -	0.250 (U/A) 0.125 (U/G) 0.250 (C/A) 0.125 (C/U) 0.125 (C/G)

The difference between two (proto)sequences was determined as follows: For each position the product of indices for all possible pairs of different nucleotides was formed (gap versus nucleotide: one-half product), as demonstrated in Table 1, and the sum of all products for a given pair of sequences (= difference, d) was plotted as nodal point in the tree diagram of Fig. 2.

The various sub-trees of metazoa, protozoa, mycophyta, plants, Gram-negative and Gram-positive bacteria were constructed stepwise and separately by using similar criteria as described for constructing tRNA trees (18), and considering biologically supported affinities, previous 5S tree data (13,14) and tree data derived from RNase T_1 catalogues of bacterial 16S rRNAs (19).

Calculations were performed with a UNIVAC 1100/82 computer, using a Fortran program.

RESULTS AND DISCUSSION

Table 2 summarizes the origin of forty seven 5S rRNA sequences used in this analysis. This list does not contain some bacterial species included in the 5S tree of Hori and Osawa (14), because they were reconstructed from RNase-T_1-generated oligonucleotides. Although such RNase-T_1 fragment "catalogues" of ribosomal RNAs have yielded valuable information about phylogenetic relationships (19), they could possibly be misleading, as in the case of the human mitochondrial 12S rRNA gene sequence, where several regions of significant homology to *E. coli* 16S rRNA sequences occur in a different order (21).

The sequence matrix of Fig. 1 contains only 14 out of 47 sequences, but most of these 14 species represent groups of similar sequences showing a group-specific pattern of deletions or insertions.

The alignment of Fig. 1 is similar to that of Hori and Osawa (14), exept for positions 80 to 110, where the correlation between the two groups of eukaryotic and prokaryotic sequences has been altered in such a way to increase the overall homologies of both primary structure (Table 3) and potential secondary structure (boxed regions of Fig. 1). The helical regions of *E. coli* 5S rRNA as shown in Fig. 1 have recently been determined by several independent methods (V.A. Erdmann, personal communication), and there is a general agreement (7,14) that other 5S rRNAs are similarly backfolded, except for the stem region S4 which can be formed only in eukaryotic molecules and, interestingly, in the archebacterial species *H. cutirubrum*.

The alignment of the recently determined wheat mitochondrial 5S rRNA sequence deserves some comments: It is obvious that this sequence fits well the general frame of invariant and semi-invariant nucleotides (Table 3)

Table 2: Sources of 5S rRNAs.

sequence number	organism	reference	sequence number	organism	reference
	eukaryotes, cytosol			bacteria and organelles	
1	Homo sapiens (HeLa cells)	6	28	Bacillus megaterium	6
2	Gallus gallus (chicken embryo)	"	29	Bacillus licheniformis	"
3	" " "	"	30	Bacillus subtilis	"
4	Iguana iguana (leguan)	"	31	Bacillus stearothermophilus	"
5	Terrapene carolina (turtle)	"	32	Lactobacillus vividescens	"
6	Xenopus laevis (toad, somatic cells)	"	33	Clostridium pasteurianum	"
7	" " (" , oocytes)	"	34	Escherichia coli	"
8	Xenopus mulleri (" , somatic cells)	"	35	Proteus vulgaris	"
9	" " (" , oocytes)	"	36	Photobacter	"
10	Salmo gairdneri (trout)	"	37	Pseudomonas fluorescens	"
11	Lytechinus variegatus (sea urchin)	"	38	Anacystis nidulans	"
12	Drosophila melanogaster (fly)	"	39	Lemna minor, chloroplast	"
13	Crithidia fasciculata (trypanosome)	19	40	Vicia faber, chloroplast	"
14	Tetrahymena thermophila (ciliate)	19	41	Phaseolus vulgaris, chloroplast	"
15	Aspergillus nidulans (mold)	4	42	Nicotiana tabacum, chloroplast	"
16	Neurospora crassa (mold)	4	43	Thermus aquaticus	"
17	Saccharomyces cerevisiae (yeast)	6	44	Triticum aestivum, mitochondria	15
	Kluyveromyces lactis (yeast)	"	45	Mycobacterium smegmatis	6
18	Torulopsis utilis (yeast)	"	46	Halobacterium cutirubrum	"
19	Pichia membranaefaciens (yeast)	"			
20	Secale cereale (rye)	"			
21	Triticum aestivum (wheat embryo)	"			
22	Lemna minor (duckweed)	"			
23	Vicia faber (broad bean)	"			
24	Phaseolus vulgaris (dwarf bean)	"			
25	Helianthus anuus (sun flower)	"			
26	Lycopersicum esculentum (tomato)	"			
27	Chlorella pyrenoidosa (green alga)	"			

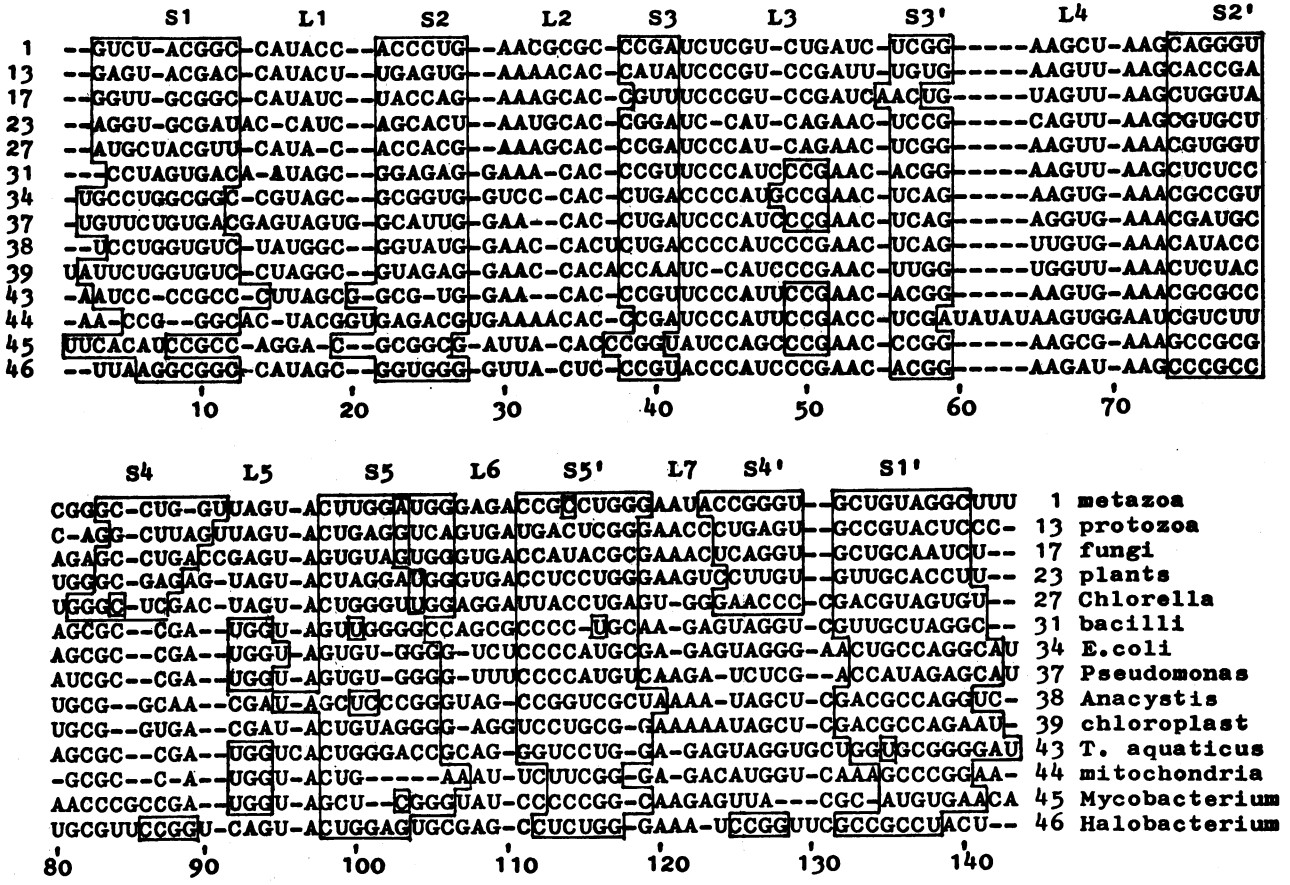

Fig. 1: Alignment of 14 representative 5S rRNA sequences. The polarity is 5'⟶3'. Boxed regions are possibly involved in base pairing (stems S1 to S5).

Table 3: Invariant and semi-invariant nucleotides of 46 5S rRNA sequences.

position nucleotide	exceptions (sequence #)	position nucleotide	exceptions (sequence #)
10 G	-	59 G	16,33
12 Y	-	66 A	33,37-42
19 C	13-15,37,44	67 G	12,15
27 G	1,2,20-26	71 A	-
34 C	6-9,11	72 A	7-9
36 C	12,15,40	74 C	45
38 C	15,16,33	81 G	13,15,37,42,45
42 U	18,34,38,45,46	83 G	45
43 C	18,45	94 G	35,38-42
45 C	-	95 U	-
47 U	45	97 A	-
49 C	14	99 U	7,9,38,45
51 G	-	112 C	13,16,27,32,43
52 A	15,16	117 G	-
54 C	13	119 G	31,33,37,38,45
		132 G	34-37,43,44

and of possible base-paired regions, if the alternating U and A residues at positions 60 to 65 and the nucleotides at positions 21, 28 and 70 are interpreted as inserts, and if other positions are deleted. The mitochondrial inserts and the unique C residue of T. aquaticus at position 96 were considered as late additions in evolution and, therefore, were disregarded for further analysis.

It is interesting to note that the mitochondrial sequence shares some properties with most or all prokaryotic sequences:

(a) The "eukaryotic stem" S4 cannot be formed. Instead, the two loop regions L3 and L5 could interact similarly as suggested for E. coli 5S rRNA (V.A. Erdmann, personal communication) by forming three or four adjacent base pairs (boxed regions of Fig. 1).

(b) The mitochondrial sequence shares with all eubacterial sequences the nucleotides G-22, G-29, N-48 and the deletion at position 90, and it has in common with most bacterial species the deletions at positions 86, 110 and 118.

(c) Table 4 shows that the mitochondrial sequence is significantly more related to prokaryotic than to eukaryotic sequences, and that the affinity between the Thermus aquaticus and plant mitochondrial sequences is almost as high as between the cyanobacterial and chloroplastic sequences.

Table 4: Sequence differences between 5S rRNAs from organelles and cytosol or bacteria (PS, protosequence)

	wheat mitochondria	Lemna minor chloroplast
Thermus aquaticus	41	55
Anacystis nidulans	57	39
Gram-negative bacteria, PS	51	47
Gram-positive bacteria, PS	59	49
wheat cytosol	55	68
Lemna minor cytosol	55	66
eukaryotes, PS	61	66

Thermus aquaticus is a thermophilic, aerobic, Gram-negative eubacterium (22) which apparently has diverged rather early from other Gram-negative bacteria (Fig. 2). The surprisingly high affinity between the plant mitochondrial and the T. aquaticus sequence could be explained by convergence, but might also point to a phylogenetic relationship. This interesting possibility could be tested by sequence analysis of a truly universal molecule, such as the small ribosomal subunit RNA (or its gene) from different mitochondria and bacteria, including T. aquaticus and members of the Rhodospirillaceae. The latter group of photosynthetic bacteria has been reported to be the closest relative to the "bacterial precursor" of mitochondria, on the basis of cytochrome c sequence analysis (13).

Fig. 2 shows a phylogenetic tree derived from the 5S rRNA sequences listed in Table 2 and constructed as described in Methods. The topology of the tree for sequences 1,4 and 5 (mammals and reptils) was only slightly superior to the two alternative topologies, and the same is true for the topology of the deuterostomian branch (fish sequence and protosequences of amniota and amphibia).

The sea urchin sequence is significantly more related to the protostomian (insect) sequence (d = 16) than to the deuterostomian protosequence (d = 22). This is unexpected since the echinodermata are generally classified as early deuterostomians (23). On the other hand, the phylogenetic significance of gastrulation and deuterostomy has been questioned (24), and it might be worthwhile to study 5S rRNA sequences of molluscs and tunicates, in order to test alternative phylogenetic pathways leading to the separation of vertebrates and invertebrates.

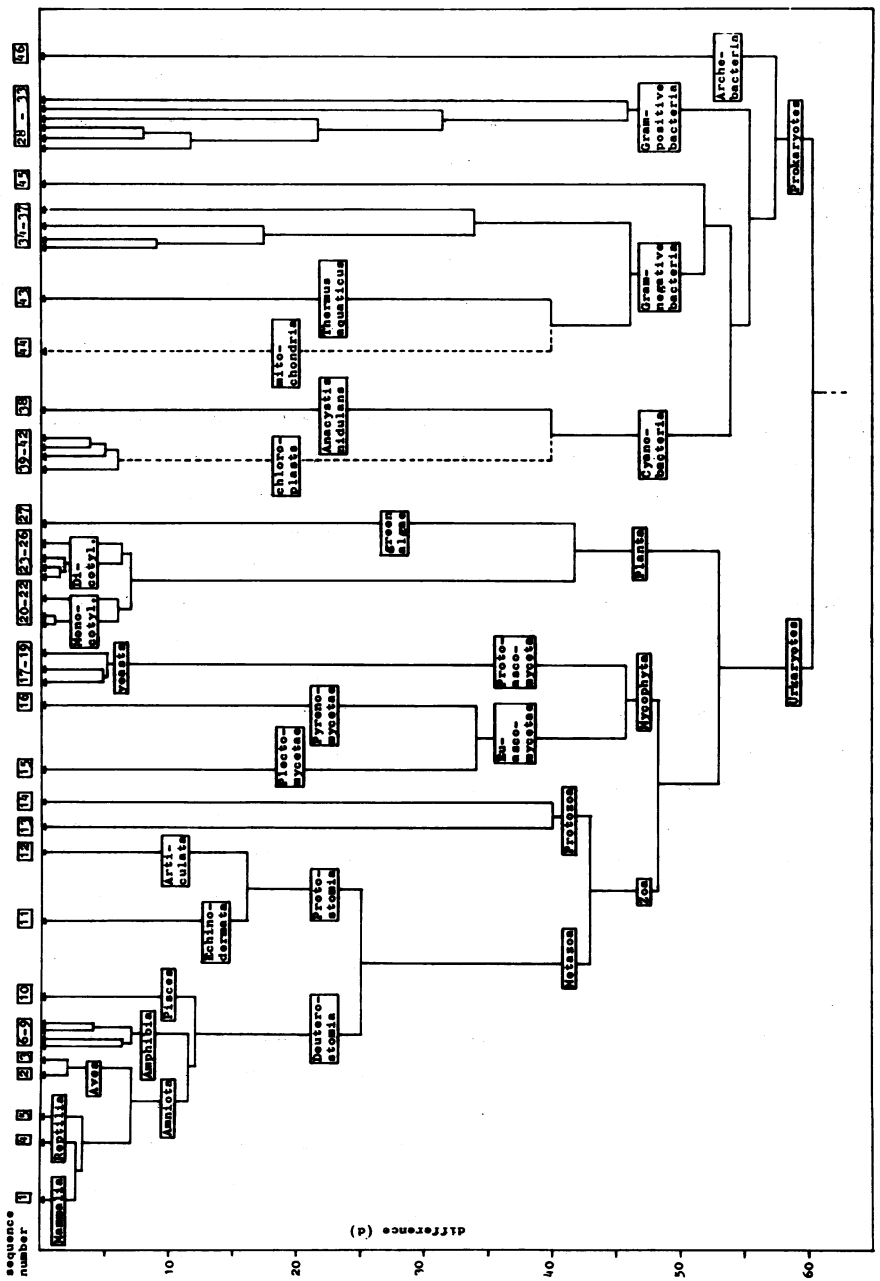

Fig. 2: Phylogenetic tree derived from 46 5S rRNA sequences (Table 2).

The protosequence of the two protozoa (trypanosome and ciliate) is more related to the protosequence of metazoa (d = 43) than to that of fungi (d = 50) or plants (d = 53). Therefore, a common precursor of metazoa and protozoa ("zoa") was constructed and compared to that of fungi (4) and plants (including the green alga Chlorella pyrenoidosa which is clearly related to higher plants) (14).

Fig. 3 shows the three possible tree topologies. The third alternative is the most unlikely one, and the first tree is a slightly better solution than the second one, since the fungal protosequence is more related to that of "zoa" than to that of plants.

The affinities between metazoa, protozoa and fungi are still increased if a recently published slime mold 5S rRNA sequence (Dictyostelium discoideum) (28) is included in the tree analysis (data not shown here): the closest relative to this sequence is that of the protozoa C. fasciculata (d = 40). Another independent argument favouring an early separation of green algae and plants from all other eukaryotes (fungi and zoa) comes from the analysis of mitochondrial genomes: plant mitochondria contain a heterogenous genome of very high complexity (up to 300 kb) (25) possessing a 5S rRNA gene and gene product (26,27), whereas fungi and animal mitochondria have a homogenous genome of much smaller size (75 to 15 kb) and apparently have lost a 5S rRNA gene (2-4).

The tree data of Fig. 2 suggest an extremely early separation of "urkaryotic" precursors to nucleated cells (19) from the prokaryotic branch (d = 61), followed by an early divergence of the archebaterium H. cutirubrum from eubacteria (d = 57). Alternatively, H. cutirubrum could have diverged from the urkaryotic branch (d = 59), but the alignment of Fig. 1 suggests a more prokaryotic character (presence of G-22, G-29, N-48 and deletions at positions 110 and 118 in the archebacterial sequence).

The eubacteria appear to diverge into Gram-positive bacteria (bacilli

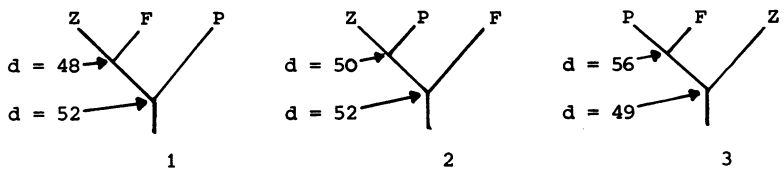

Fig. 3: Alternative tree topologies of the eukaryotic kingdoms "zoa" (metazoa + protozoa, Z), fungi (F) and plants (P).

and relatives) and a common precursor to cyanobacteria and Gram-negative bacteria. Although the organelle sequences are significantly related to members of the latter two groups (dotted lines of tree diagram) they were not included for constructing protosequences.

The difference scale of Fig. 2 is equivalent to the "mutational distance" between 5S rRNA sequences (17), but not necessarily equivalent to the elapsed time, because possible differences in the rate of nucleotide substitution in various organisms have not been taken into account. It should be noted, however, that the striking similarity of mutational distances between chloroplast sequences and between cytosol sequences of higher plants suggests a rather similar mutational rate for nuclear and chloroplast 5S rRNA genes.

Furthermore, the similarity of the chloroplast sequences and their similar mutational distance from the cyanobacterial sequence points to a common endosymbiotic origin of higher plant chloroplasts. This may not be so in the case of mitochondria. Our tree data leave open the possibility that the early eukaryotes (containing "urkaryotic" nuclei but no organelles) have differentiated into different cellular prototypes before the invasion of protomitochondria, and that the different types of mitochondrial genomes (and different eukaryotic kingdoms) have originated from several independent cellular fusion events.

REFERENCES

1. Erdmann, V.A. (1976) in Progress in Nucleic Acids Research and Molecular Biology, Cohn, W.E., ed., Vol. 18, pp. 45-90, Academic Press, New York.
2. Lizardi, P.M. and Luck, D.J.C. (1971) Nature New Biol. 229, 140-142.
3. Borst, P. and Grivell, L.A. (1978) Cell 15, 705-723.
4. Piechulla, B., Hahn, U., McLaughlin, L.W. and Küntzel, H. (1981) Nucl. Acids Res. 9,1445-1450
5. Dubin, D.T. and Friend, D.A. (1972) J. Mol. Biol. 71, 163-175.
6. Erdmann, V.A. (1981) Nucl. Acids Res.,
7. Fox, G.E. and Woese, C.R. (1975) J. Mol. Evol. 6, 61-76.
8. Benhamov, J., Jourdan, L. and Jordan, B.R. (1977) J. Mol. Evol. 9, 279-298.
9. Kimura, M. and Ohta, T. (1973) Nature New Biol. 243, 199-200.
10. Hori, H. (1975) J. Mol. Evol. 7, 75-86.
11. Schwartz, R.M. and Dayhoff, M.O. (1976) in Atlas of Protein Sequences and Structure, Dayhoff, M.O., ed., Suppl. 2, Vol. 5, pp. 293-300, Natl. Biomedical Res. Foundation, Washington DC.
12. Hori, H. (1976) Mol. Gen. Genet. 145, 119-123.
13. Schwartz, R.M. and Dayhoff, M.O. (1978) Science 199, 395-403.
14. Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. USA 76, 381-385.
15. Spencer, D.F., Bonen, L. and Gray, M.W. (1980) in International Bari Conference on the Organization and Expression of the Mitochondrial Genome, 12th, Martina Franca, Italy, Abstr. 36.
16. Fitch, W.M. and Margoliash, E. (1967) Science 155, 279-284.

17. Sankoff, D., Cedergren, R.J. and Lapalme, G. (1976) J. Mol. Evol. 7, 133-149.
18. Cedergren, R.J., LaRue, B., Sankoff, D., Lapalme, G. and Grosjean, H. (1980) Proc. Natl. Acad. Sci. USA 77, 2791-2795.
19. Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980) Science 209, 457-463.
20. MacKay, R.M., Gray, M.W. and Doolittle, W.F. (1980) Nucl. Acids Res. 8, 4911-4917.
21. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) Nature 286, 460-467.
22. Zeikas, J.G., Taylor, M.W. and Brock, T.D. (1970) Biochim. Biophys. Acta 204, 512-520.
23. Ubaghs, G. (1969) in Chemical Zoology, Florkin, M. and Scheer, B.T., eds., Vol. III, pp. 3-45, Academic Press, New York.
24. Brien, P. (1974) in Chemical Zoology, Florkin, M. and Scheer, B.T., eds., Vol. VIII, pp. 99-146, Academic Press, New York.
25. Bonen, L. and Gray, M.W. (1980) Nucl. Acids Res. 8, 319-335.
26. Leaver, C.J. and Harmey, M.A. (1972) Biochem. J. 129, 37P.
27. Cunningham, R.S., Bonen, C., Doolittle, W.F. and Gray, M.W. (1976) FEBS-Lett. 69, 116-122.
28. Hori, H., Osawa, S. and Iwabuchi, M. (1980) Nucl. Acids Res. 8, 5535-5539.