Proc. Natl. Acad. Sci. USA
Vol. 76, No. 1, pp. 381–385, January 1979
Evolution

Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species*

(5S RNA models/nucleotide alignments/classification of 5S RNAs)

HIROSHI HORI AND SYOZO OSAWA

Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University, Hiroshima, Japan 734

Communicated by Motoo Kimura, October 23, 1978

ABSTRACT **Secondary structure models of 54 5S RNA species are constructed based on the comparative analyses of their primary structure. All 5S RNAs examined have essentially the same secondary structure. However, there are revealing characteristic differences between eukaryotic and prokaryotic types. The prokaryotic 5S RNAs may be further classified into two types, one having 120 nucleotides (120-N type) and another having 116 (116-N type). A possible mechanism for the conversion of the prokaryotic 116-N type to the 120-N type 5S RNAs (or vice versa) is discussed on the basis of their nucleotide alignments. Finally, by comparing the nucleotide alignments, we propose a phylogenic tree of the 54 5S RNA species.**

Since the publication of phylogenic trees of 5S RNAs (1–4), 5S RNA sequences from a considerable number of organisms have been reported. A total of 54 are known to date. Comparative studies of secondary structures deduced from these sequences suggest that 5S RNA may be classified into three types: (*i*) eukaryotic type, (*ii*) prokaryotic 120 nucleotides (120-N) type, and (*iii*) prokaryotic 116-N type. The nucleotide alignments deduced from the juxtaposed 5S RNA secondary structure have enabled us to construct a phylogenic tree that includes animals, plants, yeasts, a blue-green alga, and a number of bacteria.

MATERIALS AND METHODS

Secondary Structure Models and Alignment of 5S RNAs. The basic method [described by Tinoco *et al.* (5)] for constructing the 5S RNA secondary structure model is the same as was previously used (4, 6). To obtain the alignments of 54 5S RNA sequences (Fig. 2), we first matched the structurally homologous base-paired regions, and then we obtained the "best match alignment," with minimal gap insertions (2), for most of the non-base-paired regions. Two gaps each were inserted into the 5′ and 3′ ends of all *Bacillus* and *Clostridium pasteurianum* sequences. In eukaryotic sequences, one gap each was inserted between positions 4 and 5 and between positions 114 and 115. For the alignment of position 63–77 of eukaryotic vs. prokaryotic sequences, AGU of position 74–76 and GRY (GGU or GAC) of 70–72 eukaryotic 5S RNAs were matched with prokaryotic AGU and RRY (GAU or AAC), respectively, in the corresponding regions. This was followed by manual arrangement between positions 63 and 73. Y (or R) here represents U or C (A or G). The regions of eukaryotic loop (position 78–99) and the corresponding prokaryotic hairpin structure (position 82–94 in *Escherichia coli*) were aligned to obtain the best match. For convenience, the sequence was divided into 15 regions: A, A′, B, B′, C, C′, D, D′, aLb, bLc, cLc′, c′Lb′, b′Ld, dLd′, and d′La′ (Fig. 2). Here, A and A′, B and B′, C and C′, and D and D′ are respectively complementary with each other. The region consisting of A and A′ was named "5′-3′ terminal helix." The aLb is the loop region that connects the base-paired region A with region B.

Construction of Phylogenic Tree. First, the rate of nucleotide substitution, *Knuc*, and the standard error of *Knuc*, σ_k, between sequences i and j were calculated by Eq. 1 (7) and Eq. 2 (8), respectively.

$$Knuc = -(3/4)\ln(1.0 - (4/3)\lambda) \quad [1]$$

$$\sigma_k^2 = \lambda(1.0 - \lambda)/L[(1.0 - (4/3)\lambda)]^2 \quad [2]$$

where λ is the fraction of different sites and L is the number of nucleotide sites to be compared. One gap (represented by - in alignments in Fig. 2) vs. one base, and one blank (represented by blank space) vs. one base, were counted as equal to one, and to one-half nucleotide substitution, respectively. Second, by using the matrix method (2), we constructed a phylogenic tree from *Knuc* values of all possible pairs of the 54 5S RNA sequences. Assuming that *Knuc* is proportional to the number of years that have elapsed since the evolutionary divergence of the two molecules from their common ancestor (7), the value of ½ *Knuc* was taken as the relative time scale in the tree.

RESULTS AND DISCUSSION

Secondary structure models

The 54 5S RNA species studied here assumed basically the same secondary structure, in accordance with that proposed previously (4). Examinations of these structures and alignments of sequences inspired us to classify the known 5S RNAs into three types. The first, to which all eukaryotic 5S RNAs belong, may be called the eukaryotic type, having 120 nucleotides. The eukaryotic 5S RNA differs from all the prokaryotic 5S RNAs by having a well-conserved loop at position 83–94 and by lacking the hairpin structure that exists in the prokaryotic 5S RNAs. This is a confirmation of the previous result (4), in which only small numbers of 5S RNA species could be compared. An apparent exception is seen in the 5S RNAs from plants, which were reported to have 116–118 nucleotides. However, their secondary structure and alignment clearly show that, qualitatively, they belong to the eukaryotic type. The second type is the prokaryotic120-N type, possessing 120 nucleotides. The 5S RNAs from Gram-negative bacteria belong to this type (nos. 22–40 in Fig. 2). The 5S RNAs from Gram-positive bacteria (*Bacillus* species and *C. pasteurianum*) possess 116 (sometimes 117) nucleotides and form the third type, the prokaryotic 116-N type (nos. 41–54 in Fig. 2). Characteristic differences between the prokaryotic 120-N type and 116-N type may be seen in the regions of A, A′, aLb, and d′La′. The representative secon-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: *Knuc*, rate of nucleotide substitution; R, A or G; Y, U or C.

* This is paper No. 143 of the Department of Biochemistry and Biophysics.

dary-structure models of these three types of 5S RNAs are shown in Fig. 1.

Alignment of Sequences. For the construction of the phylogenic tree from 5S RNA sequences and the evaluation of regional evolutionary stability of the molecule, it is indispensable to have sequence alignments for all 5S RNA species. Indeed, the success of the tree construction will depend largely upon a proper alignment. Because, as mentioned before, all 5S RNAs studied here revealed basically the same secondary structure, it is reasonable first to juxtapose all the 5S RNA secondary structures to obtain the alignments for most parts of the sequence. In this way, the alignments for B, bLc, C, cLc′, C′, c′Lb′, and B′ were easily completed. However, the secondary structures are not exactly the same, revealing some minor variations among 5S RNA species. Alignments for several regions between different models were not possible by the simple juxtaposition technique. Therefore, we aligned the sequences manually for certain regions between 116-N type and 120-N type sequences, and between prokaryotic and eukaryotic sequences. Some explanations for these alignments will be given below together with evolutionary characteristics of these regions.

Two gaps each were inserted into the 5′ and 3′ ends of the 116-N type 5S RNA sequences (Fig. 2). This gave an excellent result for total alignment. No further gap insertions into the remaining regions were necessary both for the 120-N type and 116-N type sequences. Note the great similarity between 120-N type and 116-N type of the "5′-3′-terminal helix" (A and A′) when the gaps were inserted. This suggests that the terminal helix is evolutionally rather stable. On the other hand, a simple alignment (no gaps inserted) gave only 24% identity in the helix of *B. stearothermophilus, B. megaterium, Anacystis nidulans,* and *Pseudomonas fluorescens* when compared with that of *E. coli* (12). Compare this value with 59% identity, on the average, in this region among the above bacteria when aligned by our method. The reasonableness of the two-gap insertion into the 5′ end of the 116-N type prokaryotic sequence is additionally supported by the following fact: the 5′ end of the 5S RNA precursor molecule has a short stretch (spacer) pAUU in *E. coli* (120-N type) (13) and a spacer segment consisting of 21 nucleotides in *B. subtilis* (116-N type) (14). Parts of these sequences are reproduced in Fig. 3. By comparing the 5′-terminal regions of these two precursor molecules, one finds long sequences of great similarity—i.e., *AUU*UGCCUGGCGG in *E. coli* and *AUUUG*UUUGGUGG in *B. subtilis*. (Italic letters represent the precursor stretches; gothic letters show the base-paired region in our alignment.) This strongly suggests that these two segments are the homologous counterparts; UG at the 5′ end of the *E. coli* mature molecule corresponds to the *B. subtilis* precursor *UG* which connects with the 5′ terminus of the mature molecule. Thus, it is reasonable that, in place of *UG*, two gaps are inserted into the mature 5′ end of *B. subtilis* 5S RNA.

The prokaryotic loop region, aLb, which connects the base-paired region A with B, varies in length between the 116-N type and 120-N type. The 120-N type is two (exceptionally, four) bases longer than the 116-N type (see Fig. 2). Similarly, the loop region, d′La′, which connects D′ with A′ is also two (exceptionally, three) bases longer in the 120-N type than in the 116-N type. However, the sequences from positions 11 to 15 (in aLb) or from 97 to 108 (in d′La′) are very similar throughout all prokaryotic 5S RNAs (and also eukaryotic 5S RNAs to a lesser extent). The sequences RUAGC in aLb and AGAGUAGGR in d′La′, may be seen in both the 116-N type and the 120-N type. Therefore, the alignments in the main parts of these two regions posed no problem. The two pairs of nucleotides (positions 9–10 in aLb and 109–110 in d′La′) of the 120-N type, which are included in the loop structures, would correspond to positions 9–10 in A and 109–110 in A′, respectively, of the 116-N type terminal helix.

The alignments of the sequences in the vicinity of the 5′-3′-terminal helix in *E. coli* and *B. subtilis* precursor 5S RNAs immediately suggest a possible conversion mechanism of the 116-N type to the 120-N type (or vice versa) during evolution. It has been known that during maturation enzymatic cleavages occur at the points indicated by arrows in Fig. 3 at both the 5′ and 3′ ends. Suppose that the cutting points of the 116-N type are shifted further by the change in enzyme specificity so that two bases of the spacer sequences are added to both ends and then the complementarity of two pairs of nucleotides in the A and A′ regions (positions 9–10 vs. 107–108) is lost by mutations. Then the lengths of the aLb and d′La′ regions will become the same as those of the 120-N type, keeping constant the length of the terminal helix. Thus the 116-N type may easily be converted to the 120-N type. Conversion in the other direction—i.e., from the 120-N type to the 116-N type—is also possible by the reverse process. Note that the 5′ and 3′ spacer regions of the 116-N type are complementary with each other and exhibit a great homology with the corresponding regions of the 120-N type. The order of appearance in these two types of 5S RNA

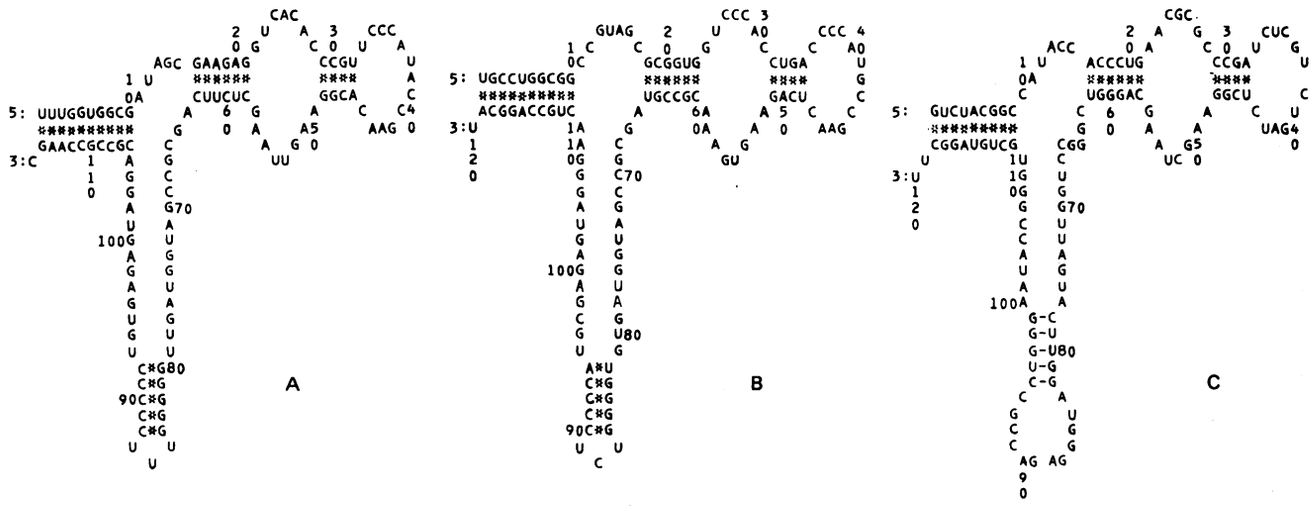

FIG. 1. Models of the secondary structures of the three types of 5S RNA. (*A*) *E. coli* prokaryotic 116-N type; (*B*) *B. subtilis* prokaryotic 120-N type; (*C*) human KB cell eukaryotic type.

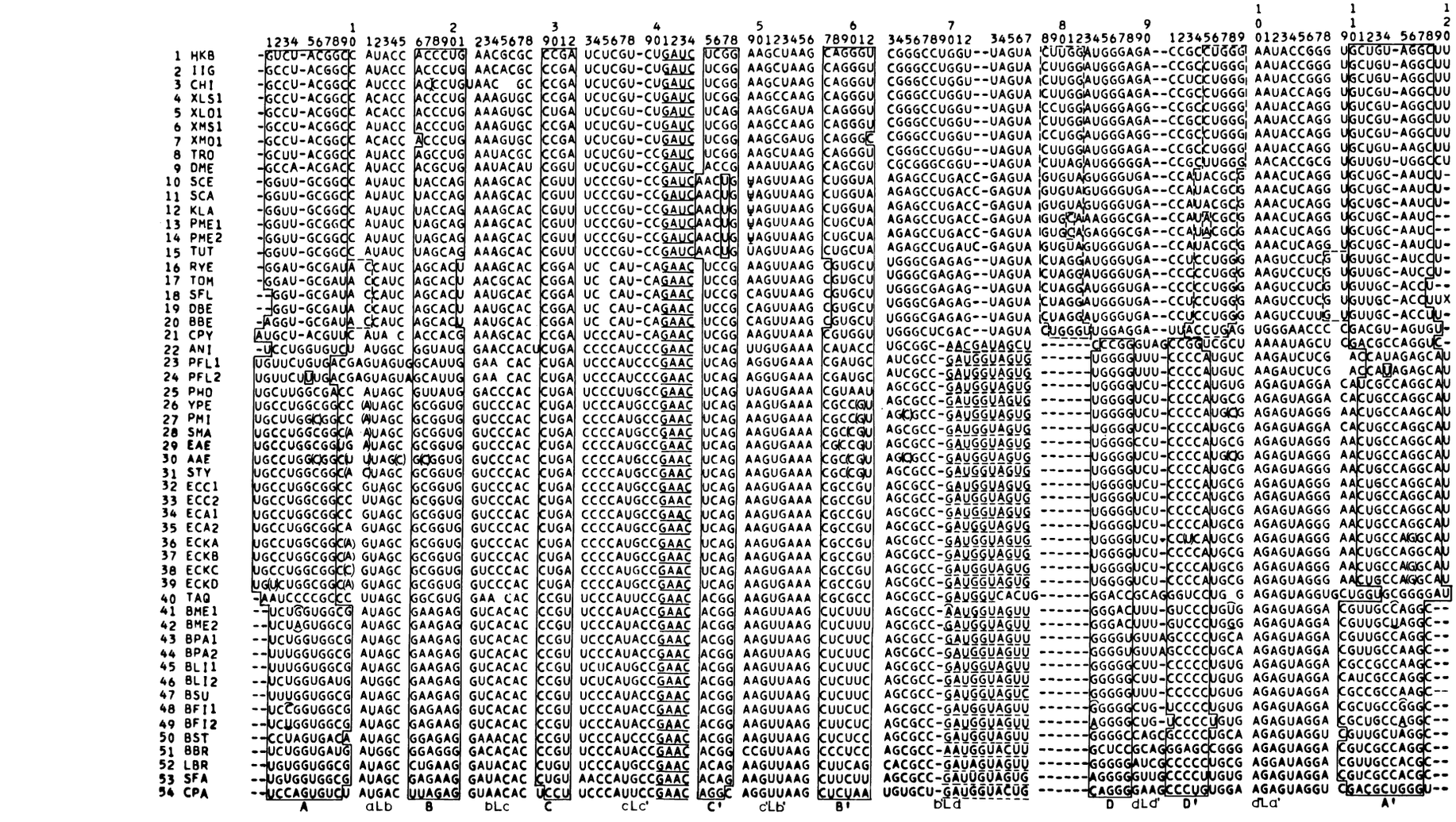

FIG. 2. Alignment of 5S RNAs. The number ahead of the abbreviated species name corresponds to the number after the name of the organism in Fig. 4. The squared off sequences correspond to the helix region in secondary structures. The underlined sequences probably interact with the GTψC loop of tRNA (9). The sequences with broken underlines probably interact with 23S RNA (10). A, A′, B, B′, C, C′, and D, D′ in the lowest line are symbols for helix regions. aLb, bLc, cLc′, c′Lb′, b′Ld, dLd′, and d′La′ in the lowest line are symbols for loop regions. ⟩, Sequence of CHI (no. 3) that was originally reported had C in this place. The sequences in parentheses have not been determined definitively. Sequences of PMI (no. 27), SMA (no. 28), EAE (no. 29), AAE (no. 30), and STY (no. 31) were not completely determined, and, therefore, they were reconstituted from the oligonucleotide maps in the literature. For 5S RNA primary sequences, see refs. 2, 4, and 11.

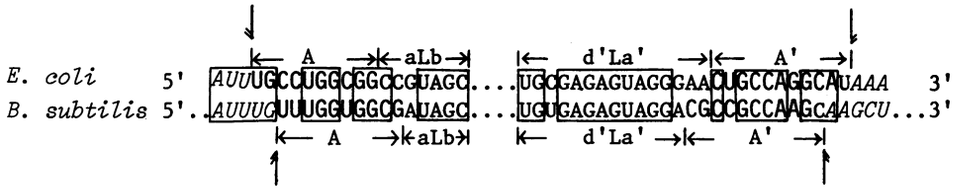

FIG. 3. Partial alignment of the *E. coli* and *B. subtilis* precursor 5S RNA sequences. Italic and gothic letters represent the precursor regions and the base-paired regions, respectively.

cannot be decided with certainty, because no direct evidence is available. We tentatively propose that the 116-N type is more primitive, because *C. pasteurianum*, known as one of the primitive bacteria, has this type of 5S RNA.

For the alignment of the A and A′ regions between prokaryotic and eukaryotic sequences, one gap each was inserted between positions 4 and 5 and between positions 114 and 115 in eukaryotic sequences, because a greater similarity of these regions was obtained in this way (41% identity, on the average). The alignments of b′Ld, D, dLd′, D′, and d′La′ between prokaryotes and eukaryotes were tentatively done simply to obtain the best matching. It is difficult to give a concrete basis for these alignments because of a considerable discontinuity in these regions between the prokaryotic and eukaryotic secondary-structure models.

Phylogenic tree

Fig. 4 shows a phylogenic tree derived from 54 5S RNA sequences of animals, plants, yeasts, a blue-green alga, and bacteria. The time of divergence of prokaryotes and eukaryotes was 1.5 times as great as that of human and yeast. Assuming that the human and yeast divergence occurred 1200 ± 75 million years before (point A in Fig. 4; refs. 1 and 15), then the divergence time of prokaryotes and eukaryotes goes back to about

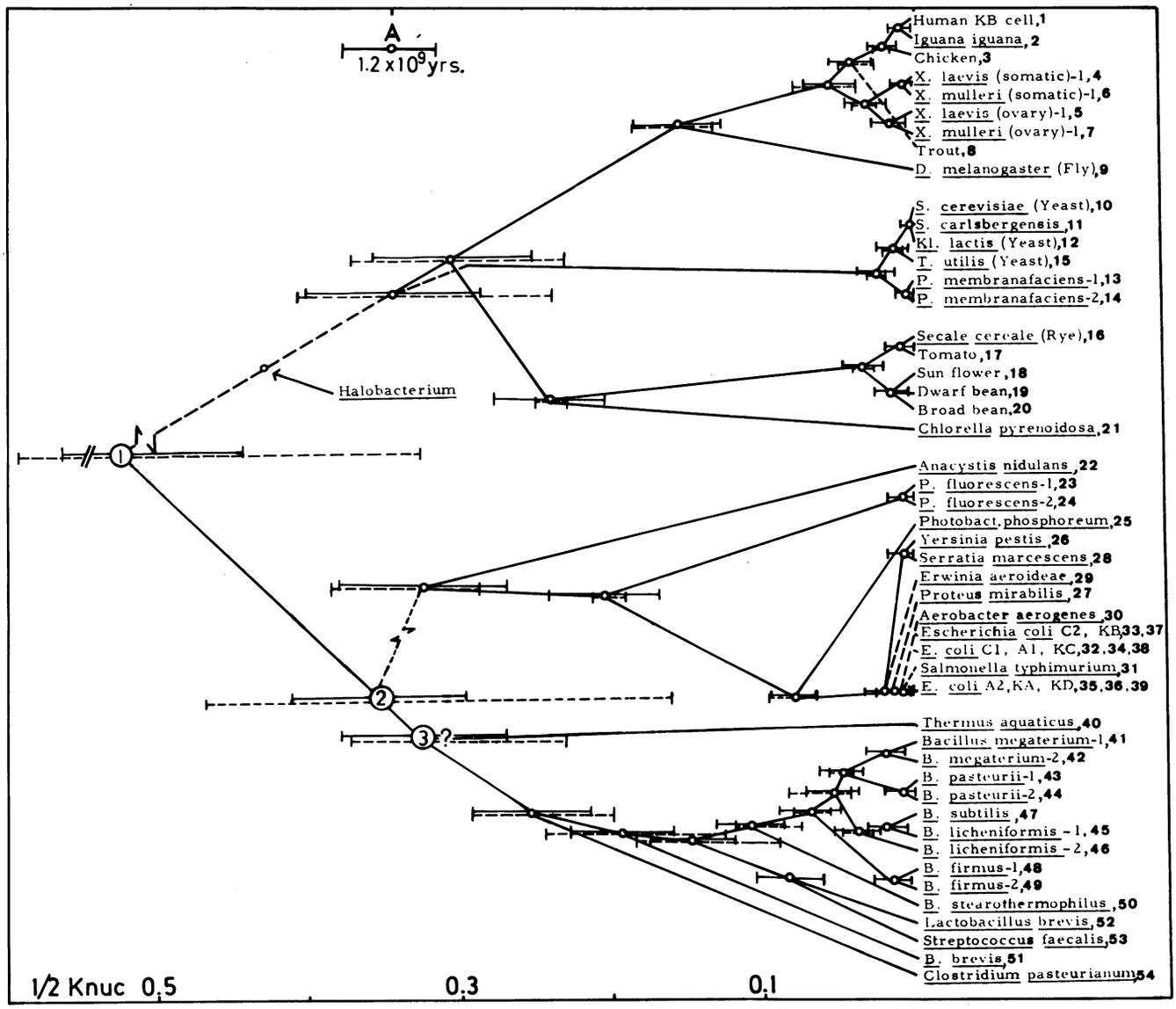

FIG. 4. Phylogenic tree. |—o—|, Range of standard error from [2]; |- - o - -|, range of ½ *Knuc* values of all pairs at the branching point.

1.8×10^9 yr. This value agrees well with that of Kimura and Ohta (1). However, due to certain discontinuities between eukaryotic and prokaryotic sequences or between the prokaryotic 120-N type and 116-N type, some uncertainties could not be avoided in constructing the tree and in determining the divergence time of the early branches (e.g., points 1 and 2 in Fig. 4).

The tree shows that first the fungi diverged from the animals and plants, and slightly after that, plants and animals separated. However, plants could have first diverged from fungi and animals as reported before (4), because (*i*) *Knuc* of animals/plants and animals/fungi are practically the same (0.61 vs. 0.63, both with a standard error of about 17%; *Knuc* of plants/fungi = 0.75); (*ii*) all 5S RNAs of animals and fungi have a GAUC sequence at about position 41–44, whereas the corresponding sequence in plants is GAAC, which is also common to all the prokaryotic 5S RNAs; and (*iii*) human 5.8S RNA reveals 75% identity with yeast 5.8S RNA, whereas it is considerably less similar to plant 5.8S RNA (16).

The branching point of *Thermus aquaticus* is not definite at present. The sequence similarity of this species is greatest with *B. stearothermophilus* (70% similarity) (17), while it is about 64% with *E. coli* group. However, *T. aquaticus* is a Gram-negative bacterium and its 5S RNA has 120 nucleotides possessing AU at the 3′ end, which is a characteristic of the 120-N type. Thus this bacterium may be more related to the 120-N type than to the 116-N type, even though the homology percentage revealed the reverse relationship.

All bacteria treated in this paper belong to one stem which originated at point 1 in Fig. 4. However, recent studies on the sequence of *Halobacterium cutirubrum* 5S RNA (18) indicate that certain aspects of its secondary structure (19) and its *Knuc* value would suggest that *H. cutirubrum* is more related to eukaryotes than to prokaryotes and that the emergence of this organism occurred around the point indicated by the arrow in Fig. 4. Furthermore, amino acid sequence data for the ribosomal protein HL20 (equivalent to L7/L12 from *E. coli*) and the 5S RNA binding proteins, HL13 and HL19, from *H. cutirubrum* indicate considerable sequence homology with the equivalent proteins from eukaryotes (ref. 20; M. Yaguchi, personal communication).

We thank Drs. S. Takemura, M. Miyazaki, and H. Komiya for their permission to use trout and yeast 5S RNA sequences, Drs. A. T. Matheson and M. Yaguchi for sending us their unpublished data on *Halobacterium*, and Mrs. S. W. Mirsky for her reading of the manuscript. This work was supported in part by Grant 220623 from the Ministry of Education of Japan.

1. Kimura, M. & Ohta, T. (1973) *Nature (London) New Biol.* **243,** 199–200.
2. Hori, H. (1975) *J. Mol. Evol.* **7,** 75–86.
3. Schwartz, R. M. & Dayhoff, M. O. (1976) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Georgetown Univ. Medical Center, Washington, DC), Suppl. 2, Vol. 5, pp. 293–300.
4. Hori, H. (1976) *Mol. Gen. Genet.* **145,** 119–123.
5. Tinoco, I., Jr., Uhlenbeck, O. C. & Levine, M. D. (1971) *Nature (London)* **230,** 362–367.
6. Fox, G. E. & Woese, C. R. (1975) *Nature (London)* **256,** 505–507.
7. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
8. Kimura, M. & Ohta, T. (1972) *J. Mol. Evol.* **2,** 87–90.
9. Erdmann, V. A., Sprinzl, M. & Pongs, O. (1973) *Biochem. Biophys. Res. Commun.* **54,** 942–948.
10. Herr, W. & Noller, H. F. (1975) *FEBS Lett.* **53,** 248–252.
11. Erdmann, V. A. (1978) *Nucleic Acids Res.* **5,** r1–r13.
12. Erdmann, V. A. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **18,** 45–90.
13. Jordan, B. R., Forget, B. G. & Monier, R. (1971) *J. Mol. Biol.* **55,** 407–421.
14. Sogin, M. L., Pace, N. R., Rosenberg, M. & Weissman, S. M. (1976) *J. Biol. Chem.* **251,** 3480–3488.
15. Dickerson, R. E. (1971) *J. Mol. Evol.* **1,** 26–45.
16. Nazar, R. N., Sitz, T. O. & Busch, H. (1975) *Biochem. Biophys. Res. Commun.* **62,** 736–743.
17. Nazar, R. N. & Matheson, A. T. (1977) *J. Biol. Chem.* **252,** 4256–4261.
18. Nazar, R. N., Matheson, A. T. & Bellemare, G. (1978) *J. Biol. Chem.* **253,** 5464–5469.
19. Matheson, A. T., Yaguchi, M., Nazar, R. N., Visentin, L. P. & Willick, G. E. (1978) in *Energetics and Structure of Halophilic Microorganisms*, eds. Caplan, S. R. & Ginzburg, M. (Elsevier/North-Holland, Amsterdam), in press.
20. Amons, R., van Agthoven, A., Pluijms, W., Möller, W., Higo, K., Itoh, T. & Osawa, S. (1977) *FEBS Lett.* **81,** 308–310.