J.Mol.Evol.7,75-86 (1975)

Journal of Molecular Evolution

Evolution of 5sRNA

HIROSHI HORI
Department of Biochemistry and Biophysics, Research Institute for Nuclear Medicine and Biology, Hiroshima University

Received August 19, 1975

Summary. The evolution of 5sRNA of 17 organisms ranging from human to bacteria has been studied using a sequence homology analysis.

The evolutionary rate of 5sRNA genes has been estimated to be 2.2×10^{-10} replacement per one nucleotide site per year. This value is about the same as that of cytochrome C or tRNA's ($\simeq 2\times10^{-10}$).

A phylogenic tree of these organisms including both eukaryotes and prokaryotes has been constructed from the evolutionary distances (the rate of nucleotide substitution per site) data. The time of divergence of prokaryotes and eukaryotes was estimated to be $\gtrsim 1.75\times10^{9}$ years ago and the branching order in eukaryotic kingdoms is consistent with the traditional order. Blue-green algae separated from the bacterial stem $\gtrsim 1.3\times10^{9}$ years ago after eukaryotes had branched.

Key words: 5sRNA Evolution - Rate of 5sRNA Evolution - Phylogenic Tree

INTRODUCTION

Because of its universal occurrence as the sole protein synthetic machinary in all organisms, the ribosome seems to be one of the most useful molecular complexes to study the biologically meaningful phylogeny among organisms. Since RNA, unlike polypeptide, has no codon degeneracy nor masked multihits, evolutionary relationships deduced from base sequences are more reliable than those deduced from polypeptides (Holmquist et al., 1972). Recently, a considerable number of 5sRNA species, a unique constitutent of 50s (60s) ribosomal subunit, in different organisms have been completely or nearly sequenced. These sequences provide us with sufficient information for the study of phylogenic relationships among a wide range of organism as in the cases of cytochrome C and hemoglobin.

In the present study, we compared sequences of seventeen different 5sRNA species including ten prokaryotes and seven eukaryotes, and constructed a phylogenic tree for these organisms.

MATERIALS AND TECHNIQUES

a) 5sRNA Sequences

The 5sRNA sequences used in this study were obtained from the literature listed in Table 1.

b) Alignment of 5sRNA

To obtain the sequence homology between two 5sRNA molecules, the GAAC or GAUC sequences at about position 40-45 were first matched and then the homology was statistically estimated on the remaining sequences. Here, we define Aj and Bi respectively as the jth and ith nucleotide of 5sRNA (A) and 5sRNA (B) from their 5'-terminal, and represent the columns and the rows of the two dimensional array, i.e., matrix (MAT). The cell MATij then represents a pair combination that contains Aj and Bi. A pathway is signified by line which connects cells of the array. All possible pair combinations between A and B can be constructed from this MAT, and every possible sequence comparison is also represented by pathways through the matrix. When MATij is part of the pathway, both i and j must increase in value. Either i or j must increase by only one but the other index may increase by one or more. In this study, MATij was assigned the cell value, one, if Aj is the same kind of nucleotide as Bi. If they are different nucleotides, MATij was assigned zero. A penalty factor, a number subtracted for every gap made, may be assessed as a barrier to allow the gap. No gap would be allowed unless the benefit from allowing the gap would exeed the barrier. The "best matched pathway" is then one for which the sum of the assigned cell values (less than any penalty factors) is largest (Needleman & Wunsh, 1970). We assigned here that one gap is equal to two nucleotide substitutions (Dayhoff & Barker, 1972). With these two primary restrictions (GAAC fixation and assignment of gap penalties) the "best match" was obtained from the matrix with minimum gap insertions. The matrix was punched out by a computer (NEAC 2200 Model 250B). Altogether 150 matrices were made.

Table 1
Source of 5sRNA sequences used in this paper

Species	Abbreviation	Nucleotide length of 5sRNA	Reference
Human KB cell	KB	120	Forget & Weissman (1967)
Xenopus laevis (African toad), kidney cell	XK	120	Brownlee et al.(1972); Ford et al.(1973)
ovary cell	XO	120	Wegnez et al.(1972); Ford et al.(1973)
Chlorella (cytoplasmic)	CC	119	Jordan & Galling(1973)
Saccharomyces cerevisiae (yeast)	SC	121	Miyazaki (1974)
Saccharomyces carlsbergensis (yeast)	SB	121	Hindley & Page (1972)
Torulopsis utilis (yeast)	TU	121	Nishikawa & Takemura(1974); Miyazaki(1974)
Anacystis nidulans (blue-green alga)	AN	120	Corry et al. (1974)
Bacillus stearothermophilus	BS	119	Marotta et al.(1973)
Bacillus megaterium	BM	116	Pribula et al.(1974)
Pseudomonas fluorescens	PF	120	DuBuy & Weissman (1971)
Proteus milabilis	PM	120	Sogin et al.(1972)
Serratia marscescens	SM	120	Sogin et al.(1972)
Erwinia aeroideae	EA	120	Sogin et al.(1972)
Aerobacter aerogenes	AA	120	Sogin et al.(1972)
Salmonella typhimurium	ST	120	Sogin et al.(1972)
Escherichia coli C	EC	120	Brownlee et al.(1967); Jarry & Rosset(1971)

c) The Rate of Nucleotide Substitution

Since the rate of nucleotide substitution during evolution has been reported to be constant per site per year in most of the informational macromolecules, we assumed the same for 5sRNA. All equations shown below were essentially the same as those used by Kimura & Ohta (1971, 1972, 1973).

The rate K(nu), the average number of nucleotide substitutions per site, may be shown as follows:

$$(1) \qquad K(nu) = -3/4 \ln (1-(4/3)\lambda)$$

where λ is the fraction of corresponding sites which differ from each other.

The standard error (σ_K) of the K(nu) is given as

$$(2) \qquad \sigma_K = \sqrt{\frac{\lambda}{(1-\lambda)n}}$$

where n is the number of nucleotide sites for which comparison of nucleotides can be made. The rate of substitution per nucleotide per site per year (k) may be calculated as

$$(3) \qquad k = K(nu)/2T$$

where T is the number of years that have elapsed since the evolutionary divergence of the two polynucleotides from their common ancestor.

d) Construction of the Phylogenic Tree

The K(nu) value, conventionally called "evolutionary distance" in this study, was used for the construction of the phylogenic tree. All of the pairs of organisms were rearranged in the order of increasing K(nu) values from Table 4, and the pair to be formed first was decided simply by choosing the pair with the smallest K(nu) values. The value of 1/2K(nu) of the pair was taken to settle the branching point between them. The branching points between two or more pairs were determined from the average numbers of 1/2K(nu) between the pairs. In Fig.1, the time scale, which had been calculated from K(nu) values from Eq.(3), was used in the abscissa.

RESULTS AND DISCUSSION

a) Alignment of 5sRNA

There are a number of methods to compare more than two homologous molecules. The visual comparison is sometimes useful. However, as one molecule of 5sRNA is composed of 116 to 121 nucleotides, this method is quite tedious and moreover a certain rationalisation is usually requested for the significance of the results obtained.

The ideal way would be a comparison of the primary structures as aligned according to the constructed secondary or tertiary structure of the molecules. However, possible secondary structure models have been suggested for several 5sRNA by Forget & Weissman (1969), Brownlee et al. (1972), Nishikawa & Takemura (1974), Jordan et al. (1974), Kearns & Wong (1974) and others, without establishing their common structure. It is therefore difficult at present to compare the 5sRNA's with this sort of information.

The third way to obtain the "best matches" between two sequences would be the purely statistical approach. Given two definite sequences, Needleman & Wunsch (1970) devised an algorithm for finding the longest common sequences without considering gap constraints. If the length of the nucleotide chain is relatively short, this method might be useful. However in the case of 5sRNA, we could not detect any significant difference in homology among widely diverged 5sRNA species and/or random sequences by this method (data not shown).

Sankoff (1972) gave an algorithm for constructing "best matches" between two sequences under constraints on the number of gaps allowed. The probability distributions for the tests of significance were calculated using a Monte-Carlo method. However, this method was too complex and needed too much time for computation. We have therefore decided to use the following method.

In all 5sRNA's, so far sequenced, there is a long nucleotide sequence between positions 20 and 60 which has been highly conserved (Corry et al., 1974). All 5sRNA's of prokaryotes have a GAAC sequence within this conserved region at about position 40-45. This sequence is complementary to GTψC in all prokaryotic tRNA molecules (Forget & Weissman, 1967), which suggests that the sequence of 5sRNA probably interacts with the common GTψC loop of tRNA (Ofengand & Henes, 1969; Shimizu et al., 1973; Erdmann et al., 1973; Richter et al., 1973). Eukaryotic initiator tRNA's (of yeast and rabbit cytoplasm) have the sequence GAUC in place of

Table 2

Number of nucleotide residues which differs from each other (left lower half) and number of gaps inserted (right upper half)

	KB	XK	XO	CC	SC	SB	TU	AN	BM	BS	PF	PM	SM	EA	AA	ST	EC
KB		0	0	3	6	6	6	6	6	5	10	8	8	8	8	8	8
XK	9		0	3	6	6	6	6	6	5	10	8	8	8	8	8	8
XO	15	6		3	6	6	6	6	6	5	10	8	8	8	8	8	8
CC	40	43	46		6	6	6	7	6	4	7	5	5	5	5	5	5
SC	46	48	52	57		0	0	9	7	8	9	9	9	9	9	9	9
SB	47	49	53	58	4		0	9	7	8	9	9	9	9	9	9	9
TU	48	50	54	57	2	6		9	7	8	9	9	9	9	9	9	9
AN	67	68	67	57	56	56	57		4	7	4	4	4	4	4	4	4
BM	56	59	64	50	50	50	49	48		3	4	4	4	4	4	4	4
BS	59	61	65	59	49	49	47	40	24		3	7	7	7	7	7	7
PF	52	53	52	54	55	54	56	49	44	53		2	2	2	2	2	2
PM	45	50	51	52	46	47	44	45	32	39	34		0	0	0	0	0
SM	47	51	52	54	46	47	44	41	30	37	39	7		0	0	0	0
EA	46	50	50	53	45	46	43	43	33	37	38	8	6		0	0	0
AA	47	50	51	54	44	45	42	43	30	37	36	5	5	5		0	0
ST	47	50	51	55	45	46	43	41	32	37	37	5	3	4	3		0
EC	46	49	50	55	45	46	43	43	32	35	36	5	4	4	3	2	

GTψC which is common in all the known eukaryotic tRNA's (Simsek et al., 1973). The sequence GAUC which is complementary to GAUC is found also at about position 40-45 5sRNA of human KB cell, *Xenopus laevis*, *T.utilis* and *S.cerevisiae*. (Two exceptions have however been reported: GAUA in *S.carlsbergensis* and GAAC in *Chlorella*). The occurrence of similar sequences having functionally the same meaning in this particular region of the molecules suggests their common origin. Therefore when two 5sRNA molecules are compared to derive the sequence homology, it is reasonable first to match this GAAC or GAUC region between the two and then the "best match alignment" is statistically obtained on the rest of the sequences with minimum gap insertions.

By the procedure discussed above and described in Materials and Techniques, the number of sites which differ from each other and the number of gaps to be inserted were obtained (Table 2). Based on these data, "best match alignments" were constructed and the representative ones are shown in Table 3.

b) Rate of Nucleotide Substitution in 5sRNA

Two homologous molecules were compared only on the non-gapped nucleotide sites to obtain the rate of substitutions in the course of evolution. Deletions and insertions (=gaps) are usually not taken into account even if they exist (see Kimura & Ohta, 1971). Therefore, K(nu) (=evolutionary distance) was computed without considering gaps using Eq.(1) (Table 4).

Table 3

Representative 5sRNA alignments. Parenthesis and underline: undetermined. The squared-off sequences (GAAC or GAUC): sequence first fixed. Sequences of PM, SM, EA, AA and ST were not completely determined and therefore reconstructed from oligonucleotide maps (Sogin et al., 1974). Not all sequence comparisons were shown here, since the sequences in vertebrate groups (KB, XK, XO), yeast groups (SC, SB, TU) and entero-bacteria groups (EC, ST, AA, EA, SM, PM) are about the same in each group

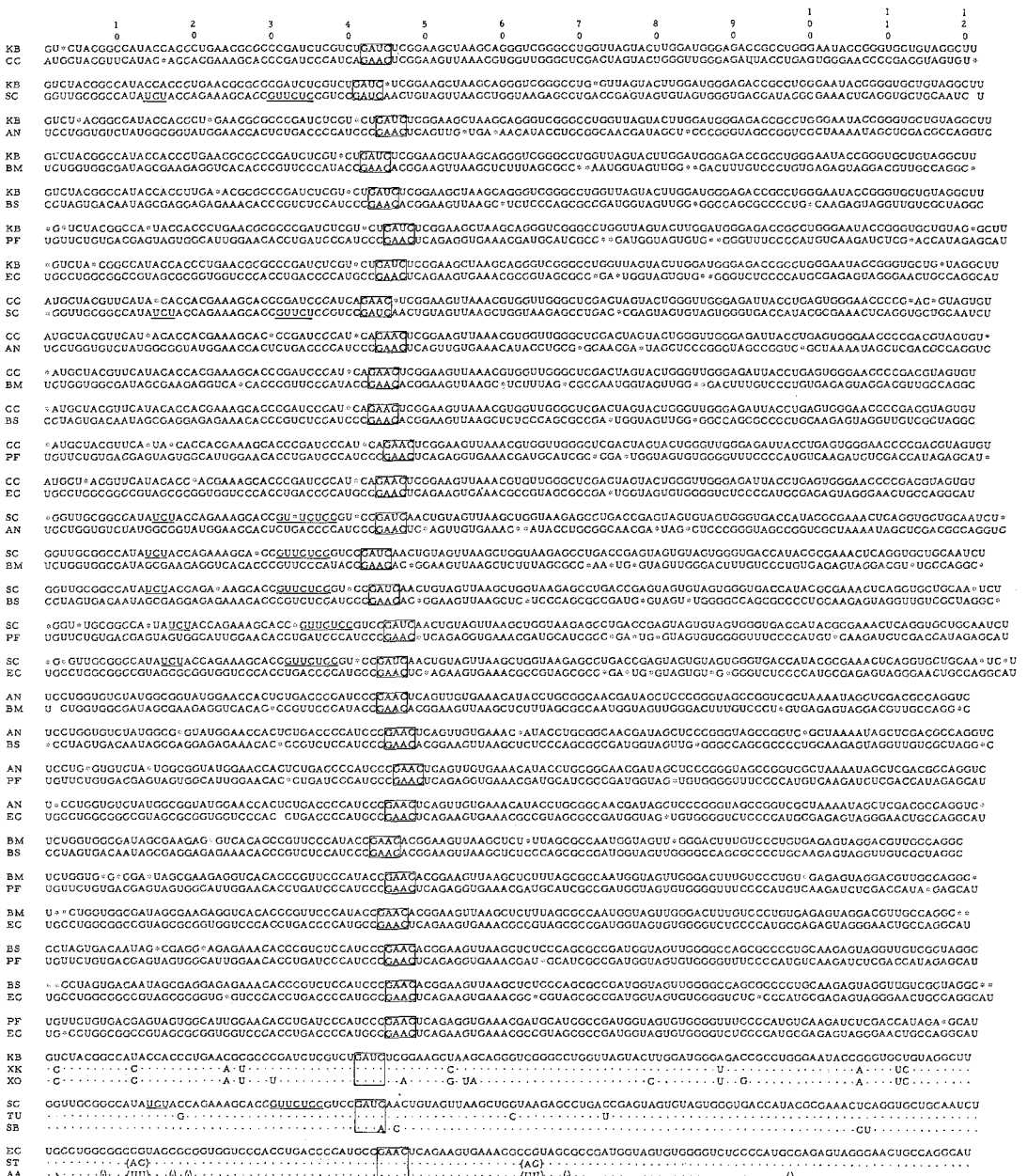

Table 4

Fraction of different sites between two 5sRNA's (λ; lower left half) and evolutionary distance (K(nu); upper right half)

	KB	XK	XO	CC	SC	SB	TU	AN	BM	BS	PF	PM	SM	EA	AA	ST	EC
KB		0.079	0.137	0.451	0.564	0.582	0.601	1.081	0.786	0.837	0.693	0.546	0.582	0.564	0.582	0.582	0.564
XK	0.075		0.052	0.499	0.602	0.621	0.641	1.120	0.864	0.891	0.715	0.641	0.662	0.641	0.641	0.641	0.621
XO	0.125	0.050		0.550	0.683	0.705	0.727	1.081	1.016	1.016	0.692	0.662	0.682	0.641	0.661	0.661	0.641
CC	0.339	0.364	0.390		0.786	0.811	0.786	0.799	0.659	0.838	0.727	0.673	0.716	0.694	0.716	0.739	0.739
SC	0.397	0.414	0.448	0.487		0.033	0.017	0.774	0.650	0.621	0.750	0.572	0.572	0.553	0.535	0.553	0.553
SB	0.405	0.422	0.457	0.496	0.033		0.051	0.774	0.650	0.621	0.727	0.590	0.590	0.571	0.553	0.571	0.571
TU	0.414	0.431	0.465	0.487	0.017	0.050		0.798	0.629	0.582	0.773	0.535	0.535	0.517	0.500	0.517	0.517
AN	0.573	0.581	0.573	0.491	0.483	0.483	0.491		0.602	0.462	0.605	0.532	0.467	0.467	0.499	0.467	0.499
BM	0.487	0.513	0.557	0.436	0.435	0.435	0.426	0.414		0.242	0.529	0.343	0.317	0.358	0.317	0.344	0.344
BS	0.504	0.521	0.556	0.504	0.422	0.422	0.405	0.345	0.207		0.705	0.446	0.415	0.415	0.415	0.415	0.386
PF	0.452	0.461	0.452	0.466	0.474	0.466	0.483	0.415	0.379	0.456		0.360	0.431	0.416	0.387	0.402	0.387
PM	0.388	0.431	0.440	0.444	0.400	0.409	0.383	0.381	0.276	0.33	0.286		0.061	0.070	0.043	0.043	0.043
SM	0.405	0.440	0.448	0.462	0.400	0.409	0.383	0.347	0.259	0.319	0.328	0.058		0.052	0.043	0.025	0.034
EA	0.397	0.431	0.431	0.453	0.391	0.400	0.374	0.364	0.284	0.319	0.319	0.067	0.050		0.043	0.034	0.034
AA	0.405	0.431	0.440	0.461	0.383	0.391	0.365	0.364	0.259	0.319	0.303	0.042	0.042	0.042		0.025	0.025
ST	0.405	0.431	0.440	0.470	0.391	0.400	0.374	0.347	0.276	0.319	0.311	0.042	0.025	0.033	0.025		0.017
EC	0.397	0.422	0.431	0.470	0.391	0.400	0.374	0.364	0.276	0.302	0.303	0.042	0.033	0.033	0.025	0.017	

Assuming that the rate of nucleotide substitution in 5sRNA of all organisms is constant, the substitution rate was directly obtained from the 5sRNA difference between human (h) and *Xenopus* kidney (xk) or ovary (xo) 5sRNA. The evolutionary distance between human and *Xenopus* $K(nu)_{h-xk}$ or $K(nu)_{h-xo}$ was 0.079 ± 0.026 or 0.137 ± 0.034 where the error was computed using Eq.(2). It has been known from paleontological studies that the common ancestor of the amphibians and the mammals appeared about 250 million years ago (see Simpson, 1950). It then follows that k_{h-xk} is 1.58×10^{-10} and k_{h-xo} is 2.74×10^{-10}. In this study, the mean of k_{h-xk} and k_{h-xo} was used for the evolutionary clock. The mean value $k_{h-x} = 2.2\times10^{-10}$ is almost the same as the one (2.3×10^{-10}) obtained by Kimura & Ohta (1973) based on the time scale estimated by cytochrome C. Thus the rate of nucleotide substitution in 5sRNA is considerably slow during evolution. It is about the same rate as that of tRNA ($2.2\times10^{-10} \pm 0.5\times10^{-10}$) or of cytochrome C ($2.0\times10^{-10} \pm 0.8\times10^{-10}$) (Holmquist et al., 1973).

c) The Phylogenic Tree

The phylogenic tree constructed from K(nu) values (evolutionary distances) of 5sRNA for different organisms is shown in Fig.1. The lowest node represents the divergence of two populations, one of which developed into the bacterial and blue-green algal groups, and the other eventually evolved into an ancestor of the eukaryotes.

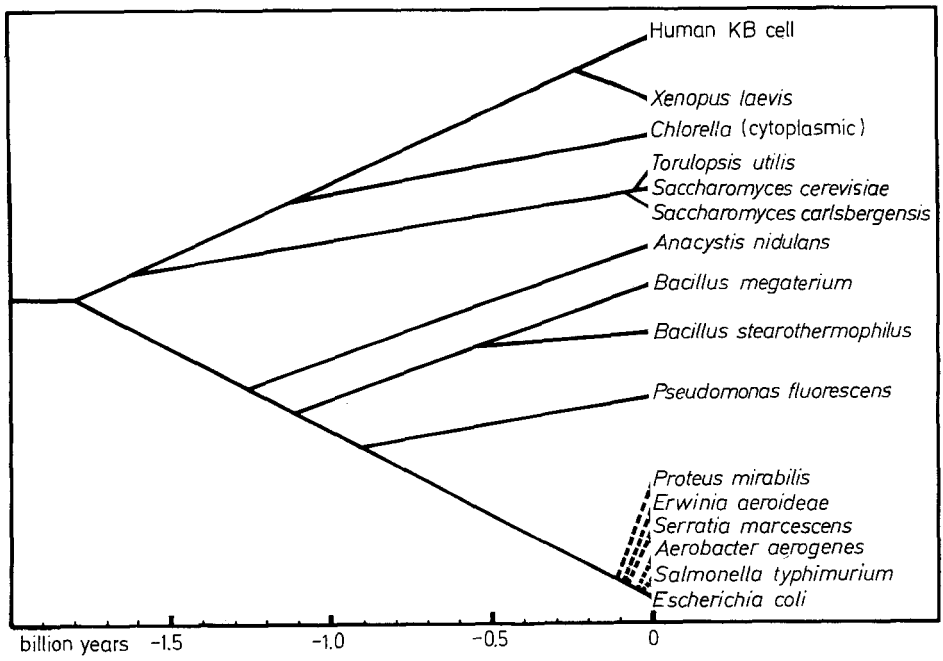

Fig.1
Phylogenic tree constructed from 5sRNA sequence comparisons

The time of divergence of prokaryotes and eukaryotes was estimated to be $\gtrsim 1.75\times10^9$ years ago. This is in almost exact agreement with the value of 1.8×10^9 years obtained by Kimura & Ohta (1973) who compared 5sRNA sequences of four different organisms. These values are also in good accordance with that calculated from the Precambrian fossil records (between 1×10^9 and 2×10^9 years ago) (Barghoorn, 1971). Since the alignments of 5sRNA's were based on the "best (maximum) matches" between two molecules, the actual time of divergence could be somewhat greater than that calculated here.

A possible branching order deduced from the 5sRNA data of eukaryotic kingdoms is Fungi, Planta and Animalia. The Fungi group diverged from the *Chlorella*-animal stem $\gtrsim 1.7\times10^9$ years ago, and the different fungi species dealt with here diverged very recently in evolution. More 5sRNA data, especially of Planta and Protista, may be required to elaborate on these branching orders. Even so, the order presented here is consistent with the traditional branching order, but not, however, with that from cytochrome C (McLauglin & Dayhoff, 1973).

According to Fig.1, blue-green algae separated $\gtrsim 1.3\times10^9$ years ago from *Bacilli-E.coli* stem after eukaryotes had branched. However, the amino acid sequence of ferredoxin or cytochrome f of blue-green algae is more similar to that of

eukaryotes than to that of photosynthetic bacteria (Wada et al., 1974; Ambler & Bartsch, 1975), and therefore, according to these data, blue-green algae should have been separated from the bacterial stem before eukaryotes had branched. It has been speculated that the eukaryotic chloroplasts were derived from symbiotic prokaryotes related to blue-green algae (see Margulis, 1970), and the remainder, including cytoplasmic ribosomes, were of genuine eukaryotic origin. If this is true, the phylogenic tree should be made from the molecules of genuine eukaryotic origin and not from those of possible parasitic nature. Then the discrepancy between the tree constructed from cytoplasmic ribosomal 5sRNA and the one from photosynthetic components may easily be understood.

In *Enterobacteriaceae*, the six species studied here branched quite recently, and the order of branching nearly reflects the taxonomic relatedness (Osawa et al., 1971; Pace & Campbell, 1971). In bacterilogy, enterobacteria, *Pseudomonas* and blue-green algae are classified into gram-negative organisms and *Bacilli* into gram-positive. Since blue-green algae differentiated from the others $\gtrsim 1.3\times10^9$ years ago, the ancestor of these bacteria could be gram-negative; some time after the branching between *Bacilli* and enterobacteria, *Bacilli* became gram-positive, and other bacteria have remained gram-negative. Thus the gram-positive bacteria branched from the gram-negative stem $\gtrsim 1.2\times10^9$ years ago.

Acknowledgements. This investigation was aided by a grant from the Scientific Research Funds of the Ministry of Education, Japan (No.94815). I wish to express my sincere thanks to Prof. S. Osawa for continuous encouragement and critically reading the manuscript, and to Drs. M. Kimura and T. Ohta, National Institute of Genetics, Mishima, Japan, for valuable comments and suggestions. I am also greatly indebted to Drs. U. Yoshida of the Faculty of Science, Hiroshima University, H. Ueoka of the Department of Biometrics of this Institute, and K. Higo of this Department, for discussion and advice on the computer programming, and Mr. N. Sumida for help in the course of computer operations.

Thanks are also due to Mr. Kelvin Lee of the Radiation Effects Research Foundation (RERF) for reading the manuscript and for making several editorial suggestions.

REFERENCES

Ambler, R.P., Bartsch, R.G. (1975). Nature 253, 285
Barghoorn, E.S. (1971). Sci.Am. 224, 30
Brownlee, G.G., Sanger, F., Bonell, B.G. (1967). Nature 215, 735

Brownlee, G.G., Cartwright, E., McShane, T., Williamson, R. (1972). FEBS Letters 25, 8
Corry, M.J., Payne, P.I., Dyer, T.A. (1974). FEBS Letters 46, 63
Corry, M.J., Payne, P.I., Dyer, T.A. (1974). FEBS Letters 46, 67
Dayhoff, M.O., Barker, W.C. (1972). In: Atlas of protein sequence and structure, M.O. Dayhoff, ed., pp. 41-45. Washington, D.C.: National Biomedical Research Foundation
DuBuy, B., Weissman, S.M. (1971). J.Biol.Chem. 246, 747
Erdmann, V.A., Springe, M., Pongs, O. (1973). Biochem.Biophys.Res.Comm. 54, 942
Ford, P.J., Southern, E.M. (1973). Nature, N.Biol. 241, 7
Forget, B.G., Weissman, S.M. (1967). Sci. 158, 1695
Hindley, J., Page, S.M. (1972). FEBS Letters 26, 157
Holmquist, R., Cantor, C., Jukes, T.H. (1972). J.Mol.Biol. 64, 145
Holmquist, R., Jukes, T.H., Pangrburm, S. (1973). J.Mol.Biol. 78, 91
Jordan, B.R., Galling, G. (1973). FEBS Letters 37, 333
Jordan, B.R., Galling, G., Jourdan, R. (1974). J.Mol.Biol. 87, 205
Kearns, D.R., Wong, Y.P. (1974). J.Mol.Biol. 87, 755
Kimura, M., Ohta, T. (1971). Theoretical aspects of population genetics, p. 16-32. Princeton University Press
Kimura, M., Ohta, T. (1971). J.Mol.Evol. 1, 1
Kimura, M., Ohta, T. (1972). J.Mol.Evol. 2, 87
Kimura, M., Ohta, T. (1973). Nature New Biol. 243, 199
Marotta, C.A., Levy, C.C., Weissman, S.M. (1973). Biochem. 12, 2901
Margulis, L. (1970). Origin of eukaryotic cells. New Haven and London: Yale Univ.Press
McLaughlin, P.J., Dayhoff, M.O. (1973). J.Mol.Evol. 2, 99
Miyazaki, M. (1974). J.Biochem.(Tokyo) 75, 1407
Needleman, S.B., Wunsch, C.D. (1970). J.Mol.Biol. 48, 443
Nishikawa, K., Takemura, S. (1974). FEBS Letters 40, 106
Ofengand, J., Henes, C. (1969). J.Biol.Chem. 244, 6241
Osawa, S., Itoh, T., Otaka, E. (1971). J.Bacteriol. 107, 168
Pace, B., Campbell, L.L. (1971). J.Bacteriol. 107, 543
Pribula, C.D., Fox, G.E., Woese, C.R., Sogin, M., Pace, N. (1974). FEBS Letters 44, 322
Richter, D., Erdmann, V.A., Sprinzl, M. (1973). Nature, New Biol. 246, 132
Sankoff, D. (1972). Proc.Nat.Acad.Sci.USA 69, 4
Shimizu, N., Hayashi, H., Miura, K. (1970). J.Biochem.(Tokyo) 67, 373
Simsek, M., Petrissant, G., Rajbhandary, U.L. (1973). Proc.Nat.Acad.Sci. USA 70, 2600
Simpson, G.G. (1950). The meaning of evolution: A study of the history of life and its significance for man. Yale University Press
Sogin, S.J., Sogin, M.L., Woese, C.R. (1972). J.Mol.Evol. 1, 173

Wada, K., Kagamiyama, H., Shin, M., Matsubara, H. (1974). J.Biochem. (Tokyo) 76, 1217
Wegnez, M., Monier, R., Denis, H. (1972). FEBS Letters 25, 13

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