KOR. JOUR. MICROBIOL. December, 1986 p. 400-405 Vol. 24, No. 4

Construction of a Phylogenetic Tree from tRNA Sequences

Lee, B.J.*, D.H. Lee, Y.J. Kim and H.S. Kang
Department of Microbiology, College of Natural Sciences, Seoul National University

tRNA 염기 순서를 이용한 계통학적 연구

이병재*·이동훈·김영준·강현삼
서울대학교 자연과학대학 미생물학과

Abstract: We have constructed a phylogenetic tree for eleven species by comparing their tRNA sequences. The tree suggests that prokaryotes diverged very early before the emergence of animals. The fact that *H. volcano*, an archaebacterium, clusters with eukaryotes implies that eukaryotes did not diverge directly from their common ancestor with eubacteria. The branching order of phage T_4 and phage T_5 indicates that they have diverged separately from their hosts and they might have evolved independently. A correlation between nucleotide substitution in tRNAs and paleontological record was observed. We verified that our phylogenetic tree fits very well with traditional ones very well by imposing the molecular clock on the tree.
Key words: tRNA sequences/computer analysis/phylogenetic tree

Early studies in evolution were derived largely from paleontology, anthropology and systematic biology. However, with the progress of biochemistry and molecular biology new approaches to understanding evolution at the molecular level have been developed. Examples of their approaches include studies on the degree of interspecific hybridization of DNA, the degree of cross reactivity of antisera to purified proteins, the number of differences in peptides from enzymatic digests of purified homologous proteins and the number of amino acid replacements between homologous proteins whose complete primary structures have been determined (for review, see Ferguson, 1980; Wilson *et al.*, 1977; Wiley, 1981, Nei and Koehn, 1983). Many phylogenetic trees using each of these approaches have been constructed. However, it should be noted that none of these methods are completely satisfactory because the portion of the genome examined is very restricted and the variable measured does not reflect the mutation distance between the genes examined with sufficient accuracy. But the advent of various nucleotide sequencing methods has made it possible to detect all mutations (Maxam and Gilbert, 1977; Sanger *et al.*, 1977; Simoncsits *et al.*, 1977; Doniskeller *et al.*, 1977). Some evolutionist have constructed phylogenetic trees by comparing 5S rRNA sequences (Hori *et al.*, 1985; Hori and Osawa, 1979; Deiko *et al.*, 1984; Rogers *et al.*, 1985; Walker, 1985). However, the study with 5S rRNA

* Present address; Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, MD. 20892, USA.

has several disadvantages. There is no consensus on the secondary structue of 5S rRNA. Sequence alignment is, therefore, somewhat arbitrary. Most of the 5S rRNA sequences have been derived from RNA sequencing data which are less accurate than DNA sequencing data. There is less 5S rRNA sequencing data compared to those of tRNA. In this report we constructed phylogenetic tree by comparing the nucleotide sequences of tRNA.

MATERIALS AND METHODS

Sequence alignment of tRNA

The sequences used in this report were adapted from the compiled tRNA sequences of Sprinzl and Gauss (1984). To make it possible to compare both DNA and RNA sequences, RNA was transformed into DNA, i.e., modified bases were changed into their premodified form and the base 'U' was designated as 'T'. Organelle tRNA sequences were excluded and tRNAs which have the same pattern of recognition were compared (see Results).

Construction of phylogenetic tree

The evolutionary distance, Knuc., and standard error of Knuc., σ k, between two sequences compared were calculated using the equations previously described by Kimura (1980) and Deiko *et al.* (1984). Knuc. corresponds to the number of base substitutions per nucleotide site that have occurred in the course of evolution extending over T years.

$$\text{Knuc.} = -(1/2)\log_e[(1-2P-Q)(1-2Q)^{1/2}]$$

Where P and Q are fractions of nucleotides showing transition- and transversion-type differences, respectively.

$$\sigma k^2 = [(a^2P + b^2Q) - (aP + bQ)^2]/n$$

Where n is the number of nucleotide sites to be compared a and b are calculated by the following equations:

$$a = 1/(1-2P-Q)$$

$$b = [(1-2P-Q) + 1/(1-2Q)]/2$$

The mismatching bases were regarded as transversion type mutations.

We constructed a phylogenetic tree by employing the unweighted pair-group method using arithmatic averages (UPGMA) from the evolutionary distance values calculated by the above equations (Hori, 1976; Sneath and Sokal, 1973; Jardin and Sibson, 1970).

RESULTS

We examined the sequence homology of $tRNA^{Arg}$ sequences in *Aspergillus nidulans* and other species to observe whether there is any correlation between the sequence homology of tRNA and the phylogenetic development. It was found that the sequence homology shows good agreement with traditional evolutionary distance (Lee, 1986). Fig.1 shows the dendrogram for tRNA constructed from homology matrix using the arithmatic average linkage cluster method. This figure strongly supports the idea that organelle tRNA genes have evolved separately from the nuclear tRNA genes. Rodriguez-Vargas *et al* (1984) reached the same conclusion, even when they changed the clustering methods. Thus, to simplify the system, the organelle tRNA data were discarded.

In addition, it appears that tRNAs that recognize the same codon usage i.e. the same patterns of codon recognition, were clustered together (data not shown). Therefore, we only compared the sequences of tRNAs that have the same wobble codon among tRNAs within the same amino acid family.

Fig.2 shows the dendrogram constructed by Kimura's distance matrix of $tRNA^{Arg}$ using UPGMA. This shows relatively good consistency with the traditional evolutionary tree for amino acid sequence except in the branching order of fungi. We thought that this inconsistancy may be due to the large contribution of each nucleotide due to the short sequence of tRNA. To counteract this, we chose eleven model species in which many tRNA sequence data have been reported. The dendrogram for these model species were obtained from Kimura's

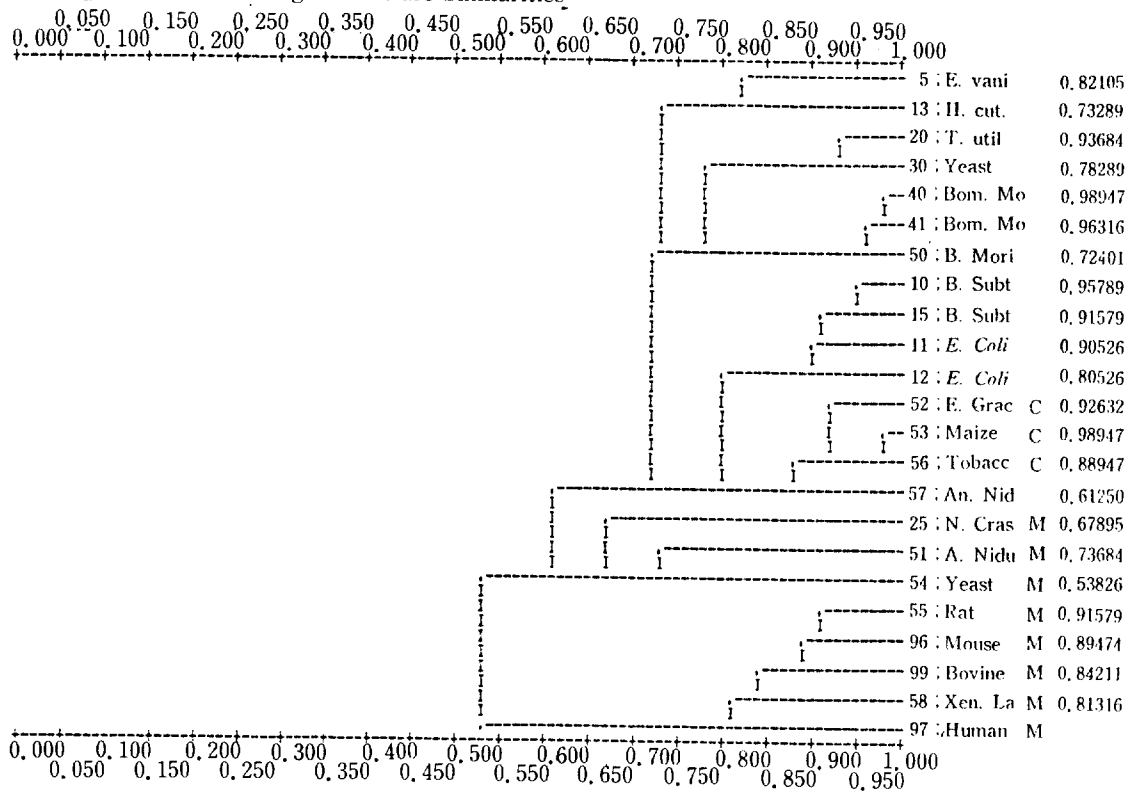

Fig.1. *Dendrogram of tRNA*^{Ala}

After constructing the homology matrix the branching order was determined using the average linkage cluster method. This figure is a copy of the original picture drawn by computer. M represents mitochondria tRNA^{Ala} and C represents chloroplast tRNA^{Ala}

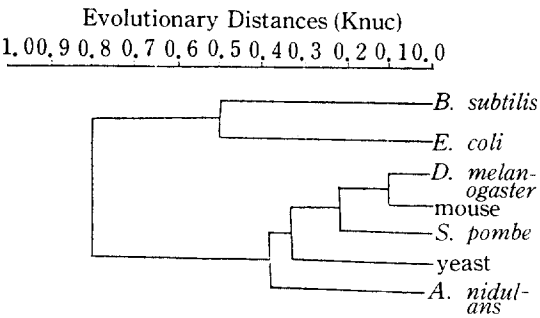

Fig.2. *Dendrogram of tRNA*^{Arg}

Only the arginine tRNAs that have ACG anticodon were analyzed and those arginine tRNAs from organelles were excluded. After computing the evolutionary distance values, dendrogram was constructed using UPGMA. Except for the sequence of *A. nidulance* (Lee and Kang, 1986), all data were adapted from Gauss and Sprinzl (1984).

distance value matrix by using UPGMA (Fig. 4). At least 2 tRNA sequences were compared between two nearest species in this figure. The clustering pattern in this dendrogram shows very good agreement with the traditional phylogenetic tree.

Fig.3 denotes the relationship between nucleotide substitution and paleontological time. The fit of points to the line is fairly good, i.e., every point meets the line within the error range.

Finally, to verify that our dendrogram fits the traditional phylogenetic tree constructed by comparing amino acid divergence, we plotted the paleontological time on the dendrogram (Fig.4). Our phylogenetic tree shows good agreement with that obtained from amino acid divergence in relation to the molecular clock.

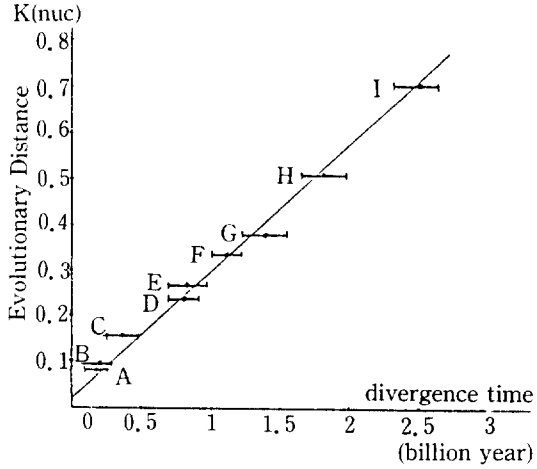

Fig.3. *Relationship between nucleotide substitution and paleontological time.*

The total nucleotide substitution for tRNAs are plotted for pairs of organisms: A: rat-human; B: mouse-human; C: Xenopus-mouse; D: Drosophila-human; E: rat-Caenorhabditis; F: rat-Euglena; G: Bacillus-*E. coli*; H: human-yeast; and I: human-*E. coli*. Calculations for determining divergence time of each pair, were taken from Wilson *et al.* (1977), Ferguson (1980), Hori *et al.* (1985) and Hori and Osawa (1979). The extended bar from each dot (⊢●⊣) represents the error range.

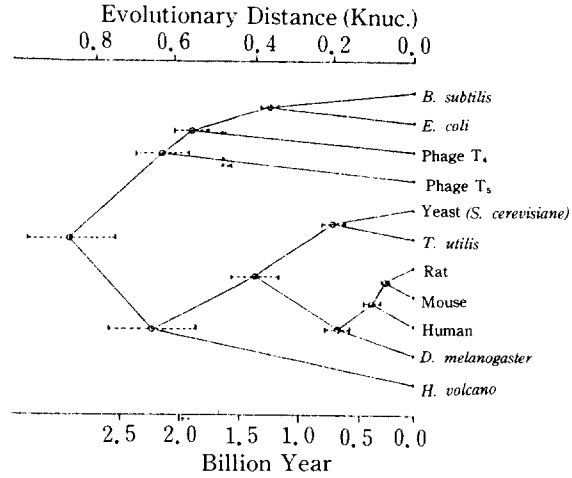

Fig.4. *Phylogenetic tree of eleven model species.*

The tree was constructed by the same methods used in Fig.2, except that tRNAs which can recognize the same wobble codon were regarded as the same tRNA. In addition, the paleontological time was plotted and the stadard deviations (σ k, designated as ⊢--o--⊣) were calculated (see Materials and Methods).

DISCUSSION

The goal of this study is to investigate the possibility of estabilishing the methods by which one can construct a phylogenetic tree using tRNA sequence data. As discussed in the introduction, many methods have been developed and applied to constructing a phylogenetic tree (Ferguson, 1980; Wilson *et al.*, 1977; Fitch and Margolish, 1967; Sneath and Sokal, 1973). However, construction of the phylogenetic tree by direct comparison of the nucleotide sequence of gene(s) may be more accurate and convenient than comparision of other biochemical parameters (such as nucleic acid hybridization, antigenic cross hybridization, isozymic pattern analysis etc.). In this respect, tRNA provides a very useful system because a lot of tRNA sequence data have been reported (Sprinzl and Gauss, 1984) and the structure of tRNA is strict (Schimmel *et al.*, 1979). It is, therefore, very easy to align and compare interspecies sequences. Another advantage of this system is that tRNA genes are relatively easier to isolate compared to other genes.

Previous efforts to construct phylogenetic trees from an analysis of tRNA sequences have provided some problems. One approach involved finding ancester of tRNA, which is both very difficult and somewhat abstract (Dillon, 1974). Another approach is very similar to ours except that only one tRNA sequence per species in the tree was compared, and Kimura's equation for distance value was not used (Rodriguez-Vargas *et al*, 1984). In this latter approach, there is a large contribution by each nucleotide; thus, a single base difference might change the branching order. Actually, in both this latter study and in our study, the branching order is slightly different according to the kind of tRNA analyzed. Therefore, in this report we

chose eleven model species and analyzed them as described in the Materials and Methods. We finally decided to use Kimura's equation to calculate the distance value and UPGMA as the clustering tool. This approach has two advantages; first, it is generally regarded as a good method for construction of a phylogenetic tree. Secondly, most of the phylogenetic trees of 5 S rRNA were constructed using this procedure, it is, therefore, convenient to compare our results with those of 5S rRNA.

As a result, we found that our approach to constructing a phylogenetic tree is consistent with traditional approaches. For example, the evolutionary distance value (Knuc.) shows a linear relationship with paleontological time (Fig.4). When we plotted the paleontological time on our phylogenetic tree, the divergence time of mammals from Drosophila, is about seven hundred million years. This result is consistent with the studies on cytochrome c (Ferguson, 1980).

Another significant fact emerging from this study is that *Halobacterium volcano,* an archaebacterium, clusters with eukaryotes. This is not really surprising; two other lines of evidence supporting this idea have been reported. Hori *et al.* (1979 ; 1985) reached the same conclusion using 5S rRNA sequence analysis. Moreovor, the base modification pattern of Halobacterium $tRNA^{Arg}$ is more similar to that of eukaryotes than prokaryotes (Lee, 1986). The importance of this fact cannot be as yet accurately assessed. However, it seems obvious that eukaryotes did not diverge directly from a common ancestor with the eubacteria. Additionaly, there is a possibility that eukaryotes, eubacteria and archaebacteria have diverged differently from a common ancestor (Pace *et al.*, 1986). Phage tRNAs are not closely related to their host tRNA, a feature that is very similar to the relationship between nuclear tRNA and organelle tRNA. Thus, we can say that phages have evolved separately from their host.

The best method for constructing the phylogenetic tree would be to compare all the nucleotide sequences between organisms, but this is impossible. Transfer RNA is a very good system for the construction of a phylogenetic tree. We have shown the validity of tRNA sequence data for the construction of phylogenetic tree. The phylogenetic position of a certain species can be determined using this eleven model species.

ACKNOWLEDGEMENTS

We thank Drs. D. Hatfield and M. Falzon for their valuable comments on manuscript. Part of this research were supported by a grant from the Korea Science and Engineering Foundation.

적 요

이미 발표된 각 시료들의 $tRNA^{Arg}$ 염기 서열을 이용하여 계통학적 연구를 하였다. archaebacterium인 *H. volcano* 가 진핵생물과 연계된 결과는 진핵생물이 eubacteria와의 공통적 조상에서 분화되지 않았음을 제시하며 Phage T_4와 T_5의 연계 순서는 그들이 각각 독립적으로 숙주로부터 분화되었음을 나타낸다. tRNA의 염기 순서의 상관관계를 이용한 연구 결과가 기존의 다른 연구 결과 및 고생물학적 기록들과도 일치함을 알 수 있었다.

REFERENCES

1. Deiko, S., R. Yamasaki, J. Jidoi, H. Hori and S. Osawa, 1984. Secondary structure and the phylogeny of *Staphylococcus* and *Micrococcus* 5S rRNAs. *J. Bacteriol.* **159**: 233-237.
2. Donis-Keller, H., A.M. Maxam and W. Gilbert, 1977. Mapping adenines, guanines and pyrimidines in RNA. *Nucl. Acids Res.* **4**: 2527-2538.
3. Ferguson. A. 1980. Biochemical system-

atics and evolution. Halsted Press.
4. Fitch, W.M. and E. Margoliash, 1967. Construction of phylogenic trees. *Science.* **155**: 279-284.
5. Hori, H. and S. Osawa, 1979. Evolutioary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species. *Proc. Natl. Acad. Sci. USA.* **76**: 381-385.
6. Hori, H., B.L. Lim and S. Osawa, 1985. Evolution of green plants as deduced from 5S rRNA sequences. *Proc. Natl. Acad. Sci. USA.* **82**: 820-823.
7. Jardin, N. and R. Sibson, 1970. Mathematical Taxonomy. John Wiley and Sons Ltd. London.
8. Lee, B.J., 1986. Ph.D. theses. Seoul National University. Seoul Korea.
9. Lee, B.J. and H.S. Kang, 1986. Studies on the Organization and Expression of *Aspergillus nidulans* tRNA Genes. (V) The molecular structure of *A. nidulans* $tRAN^{Arg}$ gene. Manuscript submitted to Kor. *J. Microbiol.*
10. Maxam, A.M. and W. Gilbert, 1977. A new methods for sequencing DNA. *Pro. Natl. Acad. Sci. USA.* **74**: 560-564.
11. Nei, M and R.K., Koehn, 1983. Evolution of genes and proteins, Sinauer Associates Inc. Publishers. Massachusetts.
12. Pace, N.R., G.J. Olsen and C.R. Woese, 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell. **45**: 325-326.
13. Rodriguez-Vargas, A.M., J.E. Fajardo and B.C. Ramirez, 1984. Phylogeny of transfer RNA. Origins of Life. **14**: 547-555.
14. Rogers, M.J., J. Simmons, R.T. Walker, W.G. Weisburg, C.R. Woese, R.S. Tanner, I.M. Robinson, D.A. Stahl, G. Olsen, R.H. Leach and J. Maniloff. 1985. Construction of the mycoplasma evolutionaly tree from 5S rRNA. **82**: 1160-1164.
15. Sanger, G., S. Niklen and A.R. Coulson 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**: 5463-5467.
16. Schimmel, P.R., D. Soll and J.N. Abelson, 1979. Transfer RNA Structure Properties and Recognition. Cold Spring Harbor Laboratory.
17. Simoncsits, A., G.G. Brownlee, R.S. Brown, J.R. Rubin and M. Guilley, 1977. New rapid gel sequencing method for RNA. *Nature* **269**: 833-836.
18. Sneath, P.H.A. and R.R. Sokal, 1973. Numerical Taxonomy. Freeman and Company, San Fransisco.
19. Sprinzl, M. and D.H. Gauss, 1984. Compilation of tRNA(gene) sequences. *Nucl. Acids Res.* **12**: r1-r131.
20. Walker, W., 1985. 5S and 5.8S ribosomal RNA sequences and protist phylogenetics. Bio Systems. **18**: 269-278.
21. Wiley, E.O., 1981. Phylogenetics. John Wiley and Sons. New York.
22. Wilson, A.C., S.S. Caulson and T.J. White, 1977. Biochemical evolution. *Ann. Rev. Biochem.* **46**: 573-639.

(Received Dec. 6, 1986)