Proc. Int'l Symp. Microorganisms and Health
April 28~29, 2000, Seoul, Korea

A DEEPLY BRANCHED NOVEL PHYLOTYPE FOUND IN PADDY SOIL

Hongik Kim, Kazunori Nakamura and Hiroshi Oyaizu[1]

Microbial Resources and Chemotaxonomy Research Group, National Institute of Bioscience and Human-Technology, Agency of Industrial Science and Technology, Tsukuba, Japan

[1] Department of Global Agricultural Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

SUMMARY

In the course of flora analysis of soil Archaea, we found very strange 16S rDNA clones, which could possibly constitute a sister clade from known two archael, Crenarchaeota and Euryarchaeota, lineages. Overall signature sequences showed that the clones were closely related to domains Archaea and Eucarya. However, at least nine nucleotides distinguished the novel clones from domains Archaea and Eucarya. Phylogenetic trees drawn by maximum parsimony, neighbor joining and maximum likelihood methods also showed unique phylogenetic position of the clones. A very specific primer set was synthesized to detect the presence of the novel group of organisms in terrestrial environments. A specific DNA fragment was amplified from all of paddy soil DNAs, and this fact suggests that the novel organisms inhabit paddy soils.

INTRODUCTION

The diversity of microorganisms is greater than that of any other living organism. However, it has been estimated that only a small portion of living microorganisms on the earth has been cultivated (1-2). Soil microbial communities are one of the important assemblages of organisms in the biosphere, yet little is known about the species that comprise them, due to the limitation of culture-based studies. At best, 2% of the bacterial cells which can be visualized by microscopy are expressed in culture (3). Clearly, the diversity of soil microorganisms is vast and complex. Molecular systematics using the 16S ribosomal RNA gene has been greatly developed in recent years, and also soil microbiologists are applying this molecule to soil samples to detect unknown microorganisms. Unknown microbial flora can be speculated from information retrieved from whole extracted DNAs, which include DNAs of all microorganisms in the examined sample.

In recent years phylogenetic positions of uncultured bacteria have been revealed by 16S rDNA sequence analyses. Archaeal uncultured organisms were detected in marine environments (4,5), hot springs (6, 7), abyssal environments (8) and soil environments (9,10). A crenarchaeotal lineage, which separated very deeply from cultivated crenarchaeota, was revealed to be predominant in marine and abyssal environments (4,8). Euryarchaeotal clones were also detected in marine environments (5) and hot springs (6). In fact, discovering unknown cells in the environmental microbial community is perhaps not as difficult as once thought. Very recently, Barns *et al.* (7) found an ancient lineage of Archaea in a hot spring and proposed a new kingdom "Korarchaeota".

In the course of our study of paddy soil archaea, we found many unusual 16S rDNA clones (11) and here we describe the existence of new group in paddy soil.

MATERIALS AND METHODS

Sample collection and DNA manipulations

Soil DNAs were extracted from four paddy soils collected at Kagoshima-city, Yamaguchi-city, Kumagaya-city, and Yamagata-city in Japan. The DNA was purified by agarose gel electrophoresis and submitted to polymerase chain reaction (PCR). Sampling procedures and PCR conditions are described in detail by Kudo *et al.* (11). The DNA amplified by primer sets of AS564F-SC13 and AS564F-SW42 were cloned into Bluescript vector and ten clones were sequenced with a DNA sequencer.

Primer design for PCR

The forward primer synthesized was AS564F (5'AAC CGT CGA CTG GGC CTA AAG CGY CCG TAG C) which corresponds to positions (in *E.coli*) 564 to 584 of archaeal 16S rRNA Reverse primers (complementary to 16S rRNA) were synthesized to be specific to the unusual 16S rDNA based on the determined partial 16S rDNA sequences (11). The reverse primers, which had sequence complementary to positions 1247 to 1229, were SC13 (5'CGG CGA ATT CTC CAC CAT TGT TGC GCG) and SW42 (5'CGG CGA ATT CCC TAT CAT TGT TGC GCG). These primers had *Eco* RI or *Sal* I restriction sites on the 5' end.

Phylogenetic analysis

For phylogenetic analysis, sequences were manually aligned to remove some unalignable regions and identified homologous regions were realigned using the program Clustal W ver. 1.7 (12). A maximum parsimony (MP) tree reconstruction was performed using the PAUP software ver. 3.1.1 (13). The applied ratio of transition and transversion was 2:1. A heuristic search was

used with a random stepwise addition sequence of 100 replicates, tree-bisection-reconnection branch swapping and the MULPARS option. A further analysis was run with 100 bootstrap replicates, each consisting of 10 random additional replicates. A maximum likelihood analysis (ML) was carried out using a program package MOLPHY ver. 2.3b3 (14). A ML distance matrix was calculated using NucML and the initial neighbor joining tree was reconstructed by NJdist in the MOLPHY. A ML tree was obtained using NucML with R (local rearrangement search) option based on the HKY model (15). Local bootstrap probabilities (LBPs) were estimated by the RELL method. A Neighbor joining (NJ) analysis was performed by using the PHYLIP software ver. 3.572 (16). DNADIST from this program package was used to create a distance matrix based on the two parameter method of Kimura (Kimura, 1980). This distance matrix was used to construct a NJ tree using the NJ algorithm from NEIGHBOR (PHYLIP ver. 3.572) and bootstrap analyses utilized 1000 replicate data sets.

RESULTS AND DISCUSSION

Sequences of ten cloned DNAs were determined for further analyses, and sequences of each clone were 82 to 98% similar to each other. Each sequence formed secondary structure of 16S rRNA positions from 585 to 1228, and these sequences seemed not to be chimeras (data not shown). Phylogenetic analysis by MP, NJ and ML all resulted with similar topologies and novel soil clones formed a monophyletic clade supported by high bootstrap proportions (BPs), to which they formed a sister group with Euryarchaeota and Crenarchaeota. MP topology was not bifurcated. This branching was probably due to limited number of informational sites which might be reduced when manually realigned. The monophyletic relationships of Euryarchaeota and Crenarchaeota is supported by quite low BP. This may be a result of systematic problems due to a rate of nucleotide substitution or base compositions. The ML tree of small data set which removed problematic 6 OTUs from 43 OTUs showed high BPs (90%) in this relationship but NJ tree was not changed (data not shown). The ML method is available for the phylogenetical analysis among the organisms which have the multiple substitution. On the other hand, the NJ method underestimated the multiple substitutions, so it is possible that its internal branch lengths are shorter than those of ML method. In this case, the ML tree possibly estimates the right internal branch length and true evolutional tree.

The topology of phylogenetic trees was not reliable when the bootstrap show moderate to lower level. Deep branchings, such as those between bacterial phylum, usually show relatively low bootstrap, and the different branching topology was presented in various published papers,

even though the same molecule was used for these analyses (17). Signature sequences, by contrast, seem to be more informative than the unreliable trees. For example, the bacterial phylums, which are not easily separated by the tree construction, are very clearly distinguished by signature sequences. The signature sequences shown in Table1 are very conservative

Table 1. 16S rRNA sequence signature distinguishing Bacteria, Eucarya, Archaea, planktonic and abyssal clones, and soil novel clones

Position	Novel soil clones	Archaea*	Eucarya	Bacteria
688	C	G	G	G
693	A	R	G	mostly G
699	G	Y	C	C
716	C	C	Y	A
756	G	G	mostly A	mostly C
912	U	mostly U	U	mostly C
931	G	G	mostly G	C
946	G	A	A	A
952	C	C	C	U
962	Y	G	U	mostly C
966	U	U	U	G
973	R	C	G	mostly G
975	mostly A	G	G	mostly A
1050	A	G	G	G
1086	C	mostly C	C	mostly U
1087	C	C	mostly U	mostly G
1109	A	A	A	C
1110	G	G	G	A
1114	U	C	Y	C
1186	A	G	mostly A	G
1208	mostly U	C	mostly C	C
1235	A	U	U	U

*including planktonic abyssal clones and "Korarchaeota".

sequences and the ten novel soil clones exhibited non-universal sequences strongly supports

their connection to an ancient lineage of life, and possibly argues for their distinction as a new lineage.

We know nothing about these organisms other than their 16S rDNA sequences, so it might be difficult to predict their phenotypic characteristics. However, it is expected that sequences derived from the paddy fields are from mesophilic organisms. The high G+C content of the Crenarchaeota is correlated to its thermophilic characteristic ; thermophiles usually have rDNA G+C contents of > 60% whereas mesophiles generally have rDNA G+C contents of 55% or less. Judging from the sampling environments and the G+C contents of novel soil clones, the sequences probably come from mesophilic and anaerobic microorganisms.

REFERENCES

1. **Ward, D. M., M. M. Bateson, R. Weller and A. L. Ruff-Roberts (1992)** Ribosomal RNA analysis of microorganisms as they occur in nature. pp. 219-286. in K. C.Marshall, ed. Advances in Microbial Ecology vol. 12, Plenum Press, New York.

2. **Hawksworth, D. L. and L. A. Mound (1991)** Biodiversity database: The crucial significance of collections. In The Biodiversity of microorganisms and invertebrates : Its role in sustainable agriculture. D. L. Hawksworth (ed.). CAB international, Redwood Press Ltd., Melksham, UK, pp17-29.

3. **Bakken, L. R. (1985)** Separation and purification of bacteria from soil. Appl. Environ. Microbiol. **49**: 1482-1487.

4. **Fuhrman, J. A., K, McCallum and A. A. Davis (1992).** Novel major archaebacterial group from marine plankton. Nature **356**: 148-149.

5. **DeLong, E. F. (1992)** Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA **89**: 5685-5689.

6. **Barns, S. M., R. E. Fundyga, M. W. Jeffries and N. R. Pace (1994)** Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA **91**: 1609-1613.

7. **Barns, S. M., C. F. Delwiche, J. D. Palmer and N. R. Pace (1996)** Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA **93**: 9188-9193.

8. **Mcinerney, J. O., M. Wilkinson, J. W. Patching, T. M. Embley and R. Powell (1995)** Recovery and phylogenetic analysis of novel archaeal rRNA sequences from deep-sea deposit feeder. Appl. Environ. Microbiol. **61**: 1646-1648.

9. **Bintrim, S. B., T. J. Donohue, J. Handelsman, G. P. Roberts and R. M. Goodman (1997).** Molecular phylogeny of Archaea from soil. Proc Natl Acad Sci USA 94, 277-282.

10. Großkopf , R., Stubner, S. & Liesack W. (1998). Novel euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. Appl Environ Microbiol **64**: 4983-4989.

11. **Kudo, Y., T. Nakajima, T. Miyaki and H. Oyaizu (1996)** Methanogen flora of paddy soils in Japan. FEMS Microbiol. Ecol. **22**: 39-48.

12.**Thompson, J. D., D. G. Higgins and T. J. Gibson (1994)** Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673-4680.

13.**Swofford, D. L (1991)** PAUP: Phylogenetic analysis using parsimony. Version 3.1. Illinois Natural History Survey Champaign, Illinois.

14.**Adachi, J. and M. Hasegawa (1996)** Computer Science Monographs, No. 28, MOLPHY: Version 2.3. programs for molecular phylogenetics based on maximum likelihood. Institute of Statistical Mathematics, Tokyo.

15.**Hasegawa, M., H. Kishino and T. Yano (1985)** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**: 160-174.

16. **Felsenstein, J. (1995)** PHYLIP (Phylogeny Inference Package) Version 3.57c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

17. **Olsen, G. J., C. R. Woese and R. Overbeek (1994)** The winds of (evolutionary) changes : Breathing new life into microbiology. J. Bacteriol. **176**:1-6.

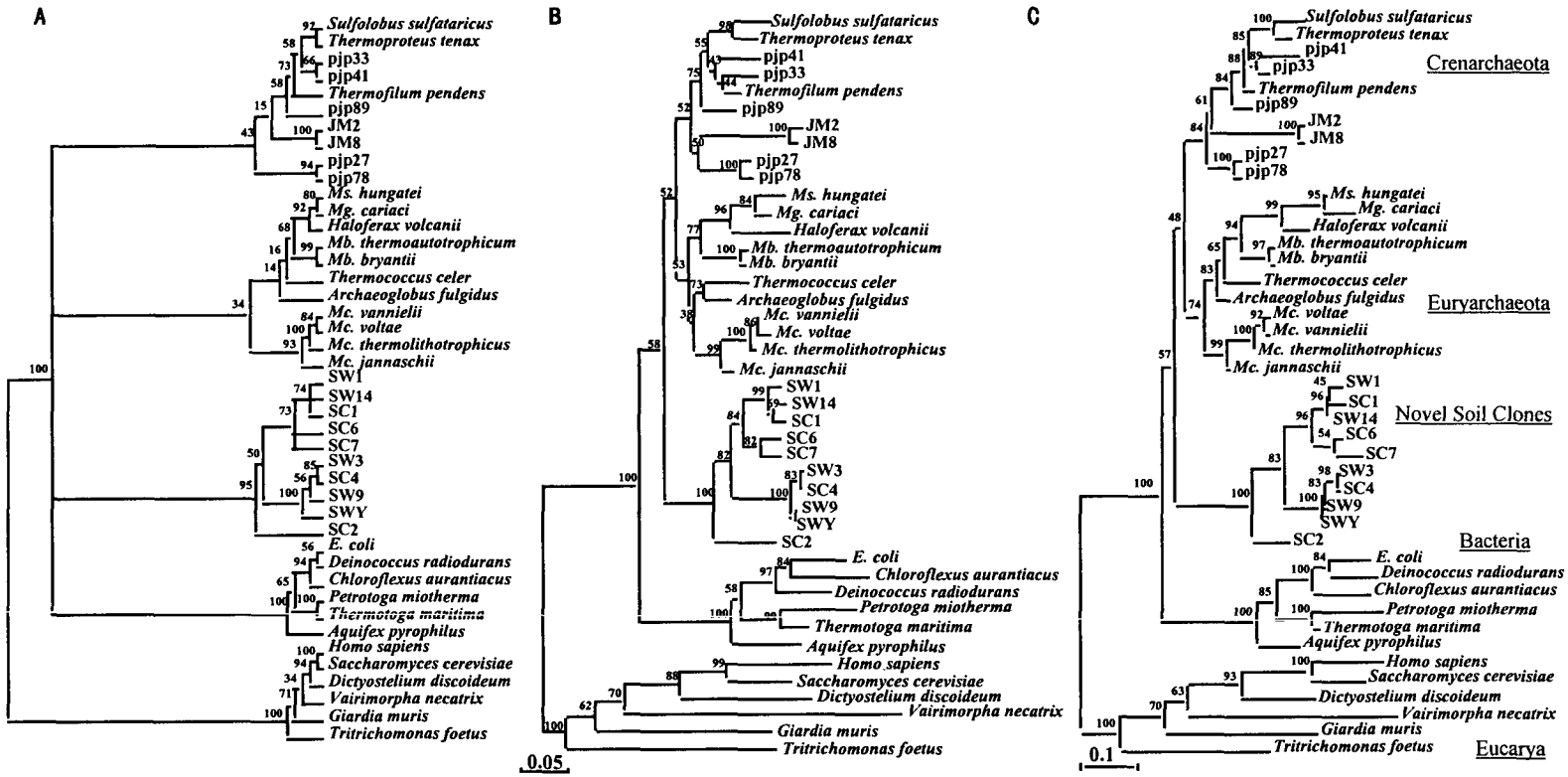

Fig.1. Phylogenetic relationships of the novel soil clones inferred from the three analytical methods. (A) Strict consensus tree of 3 most parsimonious trees (B) NJ tree (C) ML tree (a / b = 2.713). Numbers in MP trees, NJ trees, and ML trees indicate bootstrap percentage and local bootstrap probabilities respectively. Each scale bar below NJ trees and ML trees means substitutions per site.