A Novel Ribosomal-based Method for Studying the Microbial Ecology of Environmental Engineering Systems

Tao Yuan, Asst/Prof. Stephen Tiong-Lee Tay and Dr Volodymyr Ivanov
School of Civil and Environmental Engineering,
Nanyang Technological University, Nanyang Avenue, Singapore 639798

Abstract
A novel ribosomal-based method was applied to examine genetic relationships among microorganisms in the phylogenetic domains of Archaea and Bacteria. Representative unaligned nucleotide sequences of 16s rRNA genes were retrieved from the RDP database. In the domain of Archaea, based on the numbers of certain tri-nucleotides or triplets, the two phyla, Euryarchaeota and Crenarchaeota could be separated. The numbers of certain triplets, especially those containing high guanine and cytosine content such as GGG or CCC, also correlate with the optimum temperatures for growth of the species. In the domain of Bacteria, it was noticed that no single set of triplets or triplet ratios could separate all the phyla at once. However, in certain phylum (Proteobacteria in this study), there were selected triplets whose occurrences could help categorise the sequences into their classes. Although this ribosomal-based method presently requires a substantial amount of data processing and is not standardised in the way that it includes some ad-hoc manual observations and selections of triplets, triplet analysis appears to be a very quick and efficient method in studying the phylogenetic characteristics of microorganisms. These characteristics may be potentially determined in the experiments and then used to for study the microbial ecology of environmental engineering systems.

INTRODUCTION
The prokaryotes are classified based on the 16s rRNA sequences analysis to reflect their phylogenetic features [1, 2]. Previous studies have observed that the numbers of certain triplets in 16s rRNA sequences are species-specific [3]. Therefore it would be reasonable to assume that this trait of triplets applies to a wider scope, i.e. the number of triplets in 16s rRNA sequences could be phylum-specific, class-specific, etc. It could in turn give rise to a simpler way to identify microorganisms: by means of triplets rather than complete 16s rRNA sequences. Additionally, some genetic information in 16S rRNA gene sequence is lost in conventional 16s rRNA sequence analysis because of the requirements to align sequences and compare unambiguous sequences. This study aims to assess the feasibility of the triplet-based method in investigating the genetic relationships among the microorganisms in Archaea and Bacteria domains.

It was also noted in previous studies that the triplets in 16s rRNA sequences reveal some of the phenotypic characteristics [4, 5]. Hence the triplets might reflect some other characteristics of microorganisms as well. In this study, Archaea microorganisms with higher optimum temperatures were found to possess a higher GC (guanine and cytosine) content. In addition, the occurrences of selected triplets were observed to correlate with optimum growth temperatures.
The domain of Bacteria contains numerous phyla; therefore the range was narrowed down to Proteobacteria in this study. Characteristic triplets were found to help categorise sequences into each of the five classes of Proteobacteria.

METHODS

Representative unaligned nucleotide sequences of 16s rRNA genes for Archaea and Bacteria were retrieved from the Ribosomal Database Project II home page [6]. Characteristics (e.g. number of occurrences, average, standard deviation, coefficient of variation, etc.) of a total 64 of triplets were obtained by software “DNA Sequence”. The triplet characteristics were compared to similarities and differences among different microorganisms in the domains of Archaea and Bacteria as described in Bergey’s Manual [1-2]. Scatter plots were produced to reveal the results.

RESULTS

Archaea

The Phyla of Euryarchaeota and Crenarchaeota

The two domains of Archaea, namely Euryarchaeota and Crenarchaeota, could be separated by a few sets of triplets or selected triplet ratios. One example is shown in the figure below, the CCC/GGU values range from 0.366 to 1.143 for Euryarchaeota, and 1.290 to 2.000 for Crenarchaeota; whereas the GGG/GGU values range from 0.902 to 1.829 for Euryarchaeota and 1.903 to 2.783 for Crenarchaeota (Figure 1.).

![Figure 1. The phyla of Euryarchaeota and Crenarchaeota on the triplet ratio dot scatter plot](image)

The Phylum of Euryarchaeota

There are seven classes as indicated by Bergey’s Manual [1-2] in the phylum of Euryarchaeota. No single set of triplets or triplet ratios were found to separate the seven classes all at once. But it is shown that there are certain triplets or triplet ratios featuring each of these classes, e.g. Class 2 Methanococci clusters in a plot with ACG and GGC as the coordinates and Class 3 Halobacteria clusters in a plot with ACG and CCG as the coordinates (Figure 2.).
Mesophiles, Thermophiles and Hyperthermophiles in the domain of Archaea

Study on optimum growth temperatures of Archaea species shows a remarkable relationship between triplet numbers/triplet ratios and the optimum temperatures [7, 8, 9] of these Archaea sequences. Species with higher optimum temperatures (Thermophiles and Hyperthermophiles) tend to have a higher GC content in the 16s rRNA gene sequence. In fact the G+C mol% increases almost linearly with the optimum temperature (Figure 3a). Similarly, Thermophiles and Hyperthermophiles have generally more triplets containing guanine and cytosine, and less triplets containing adenine or uracil.

It could also be seen that the species range from mesophiles to thermophiles and hyperthermophiles, and from Euryarchaeota to Crenarchaeota, was associated with increasing numbers of GGG and CCC triplets, in practically a linear relationship (Figure 3b).
**Bacteria**

The Bacteria domain contains a huge collection of 16s rRNA sequences and there was no single set of triplets or triplet ratios to distinguish the numerous phyla all at once. The range is thus narrowed down to the phylum level (Proteobacteria in this study).

**Phylum of Proteobacteria**

There are five classes under the phylum of Proteobacteria: Alpha, Beta, Gamma, Delta and Epsilon, namely [1, 2]. Again there is no single set of triplets or triplet ratios that could separate the five classes all at once although combinations of triplet ratios can be employed to separate one class from the rest (Figure 4).

![Figure 4. Triplet ratios CUA/CCG and CUA/GGG featuring Epsilonproteobacteria](image)

**DISCUSSION**

By analysis of triplet numbers and triplet ratios in 16s rRNA sequences, it could be seen that the triplet method functions analogously as the whole 16s rRNA sequences to categorise the microorganisms into phyla and classes in each phylum. With scatter plots, Euryarchaeota and Crenarchaeota were separated. Each of the seven classes in Euryarchaeota possessed unique triplet characteristics that allowed individual classes to be separated from the others.

Although G+C mol% content in bacterial chromosomes is one of the recommended characteristics for the standard description of bacterial species, this is the first study, to the best of our knowledge, that demonstrates a positive correlation between G+C mol% in the 16S rRNA gene and optimum growth temperature. Since GC bonds are stronger than AT bonds, we postulate that thermophilic and hyperthermophilic bacteria require high G+C mol% in the 16S rRNA sequence to prevent the ribosome from melting *in vivo* at the high temperatures encountered. This is also evidenced in the higher occurrences of triplets with higher GC content.

In the domain of Bacteria, so far as this study goes, there was no single set of triplets or triplet ratios found to distinguish different phyla and different classes in each phylum. Yet there were signature triplets for each of the phyla or class. Therefore by combining a few triplet plots, each featuring different phyla or classes, it would be possible to categorise a sequence into specifically one of them.
There are some very recently discovered genera that were not included in this study. Further research would be done to take account of them.

In this study, there were occasions where manual observations or ad-hoc selections of triplets were applied. This ribosomal-based method might need further refinement or standardisation, e.g. devising a logical strategy to search for suitable triplet ratios to resolve target microbial groups.

REFERENCES

[6] Online resource RDP - The Ribosomal Database Project
   http://rdp.cme.msu.edu/html
[9] Online resource DSMZ - German Collection of Microorganisms and Cell Cultures
   http://www.dsmz.de/dsmzhome.htm