

## Lab 5: IDENTIFICATION OF UNKNOWN MICROORGANISMS BASED ON 16s rDNA

### SEQUENCE ANALYSIS

#### INTRODUCTION:

The rRNA is the most conserved (least variable) gene in all cells. Portions of the rDNA sequence from distantly-related organisms are remarkably similar. This means that sequences from distantly related organisms can be precisely aligned, making the true differences easy to measure. For this reason, genes that encode the rRNA (rDNA) have been used extensively to determine **taxonomy**, **phylogeny** (evolutionary relationships), and to estimate **rates of species divergence** among bacteria. Thus the comparison of 16s rDNA sequence can show evolutionary relatedness among microorganisms. This work was pioneered by Carl Woese, who proposed the **three Domain system** of classification - Archaea, Bacteria, and Eucarya - based on such sequence information.

#### The Ribosomal RNAs

In Bacteria, Archaea, Mitochondria, and Chloroplasts the small ribosomal subunit contains the 16S rRNA (where the S in 16S represents Svedberg units). The large ribosomal subunit contains two rRNA species (the 5S and 23S rRNAs). Bacterial 16S, 23S, and 5S rRNA genes are typically organized as a co-transcribed **operon**. There may be one or more copies of the operon dispersed in the genome (for example, *E coli* has seven). The Archaea contains either a single rDNA operon or multiple copies of the operon.

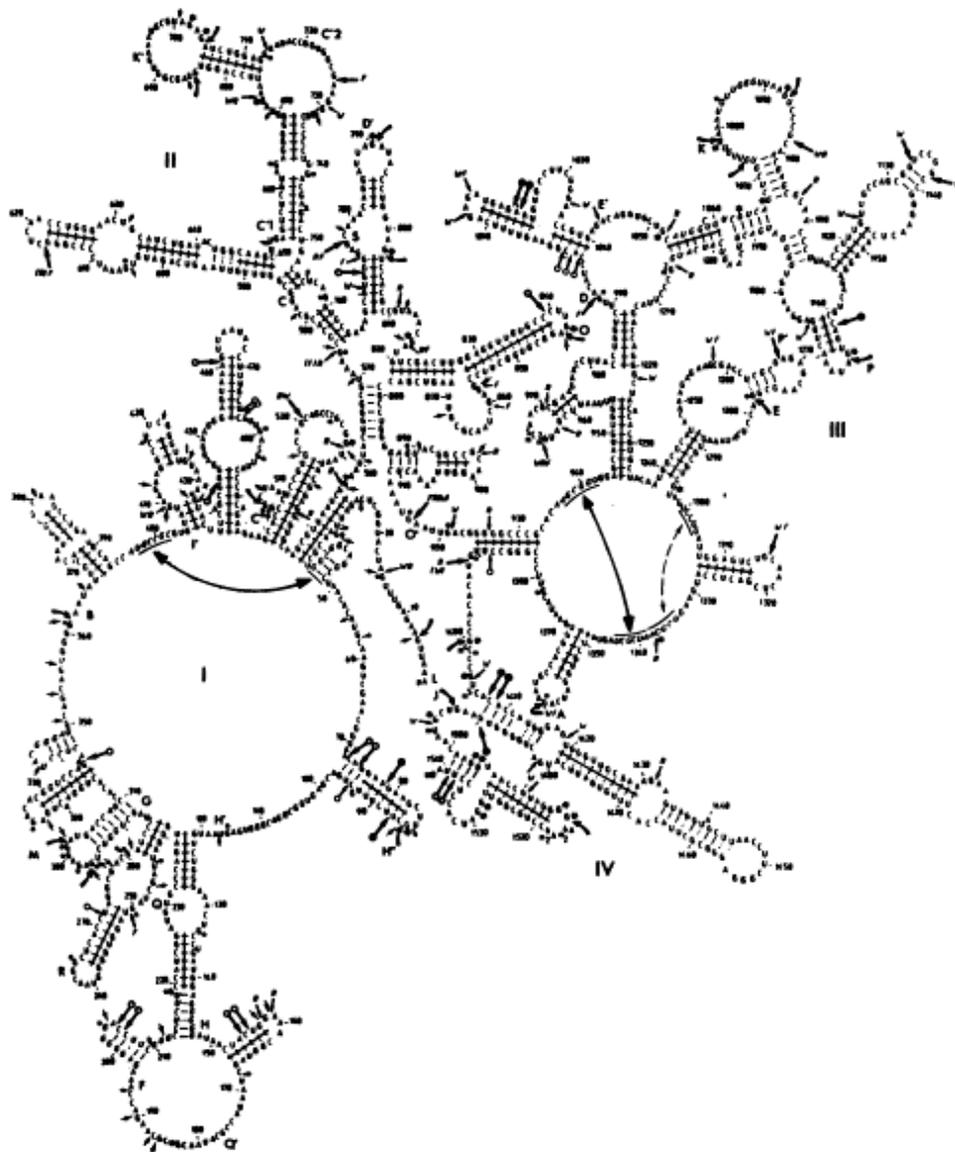
To infer relationships that span the diversity of known life, it is necessary to look at genes conserved through the billions of years of evolutionary divergence. An example of genes in this category are those that define the ribosomal RNAs (rRNAs). Most prokaryotes have three rRNAs, called the 5S, 16S and 23S rRNA.

#### **Ribosomal RNAs in Prokaryotes**

Name	Size (nucleotides)	Location
5S	120	Large subunit of ribosome
16S	1500	Small subunit of ribosome
23S	2900	Large subunit of ribosome

The 5S has been extensively studied, but it is usually too small for reliable phylogenetic inference. The 16S and 23S rRNAs are sufficiently large to be useful.

The 16s rDNA sequence has **hypervariable regions**, where sequences have diverged over evolutionary time. These are often flanked by strongly-conserved regions. Primers are designed to bind to **conserved** regions and amplify variable regions. The DNA sequence of the 16S rDNA gene has been determined for an extremely large number of species. In fact, there is no other gene that has been as well characterized in as many species. Sequences from tens of thousands of clinical and environmental isolates are available over the internet through the **National Center for Biotechnology Information** ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and the **Ribosomal Database Project** ([www.cme.msu.edu/RDP/html/index.html](http://www.cme.msu.edu/RDP/html/index.html)). These sites also provide search algorithms to compare new sequences to their database.



**Figure 1** from: Structural organization of the 16S ribosomal RNA from *E. coli*. Topography and secondary structure. P Stiegler, P Carbon, M Zuker, J P Ebel, and C Ehresmann. *Nucleic Acids Res.* 1981 May 11; 9(9): 2153–2172.

The extraordinary conservation of rRNA genes can be seen in these fragments of the small subunit (16S) rRNA gene sequences from organisms spanning the known diversity of life. Note several areas of identity among these diverse organisms:

human	GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAGTTGCTGCAGTTAAAAAG
yeast	GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAGTTGTTGCAGTTAAAAAG
corn	GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAGTTGTTGCAGTTAAAAAG
Escherichia coli	GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCG
Anacystis nidulans	GTGCCAGCAGCCGCGGTAATACGGGAGAGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCG
Thermotoga maritima	GTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTACCCGGATTTACTGGGCGTAAAGGG
Methanococcus vannieli	GTGCCAGCAGCCGCGGTAATACCGACGGCCCGAGTGGTAGCCACTCTTATTGGGCCTAAAGCG
Thermococcus celer	GTGGCAGCCGCCGCGGTAATACCGGCGGCCCGAGTGGTGGCCGCTATTATTGGGCCTAAAGCG
Sulfolobus sulfotaricus	GTGTCAGCCGCCGCGGTAATACCAGCTCCGCGAGTGGTCGGGGTGATTACTGGGCCTAAAGCG

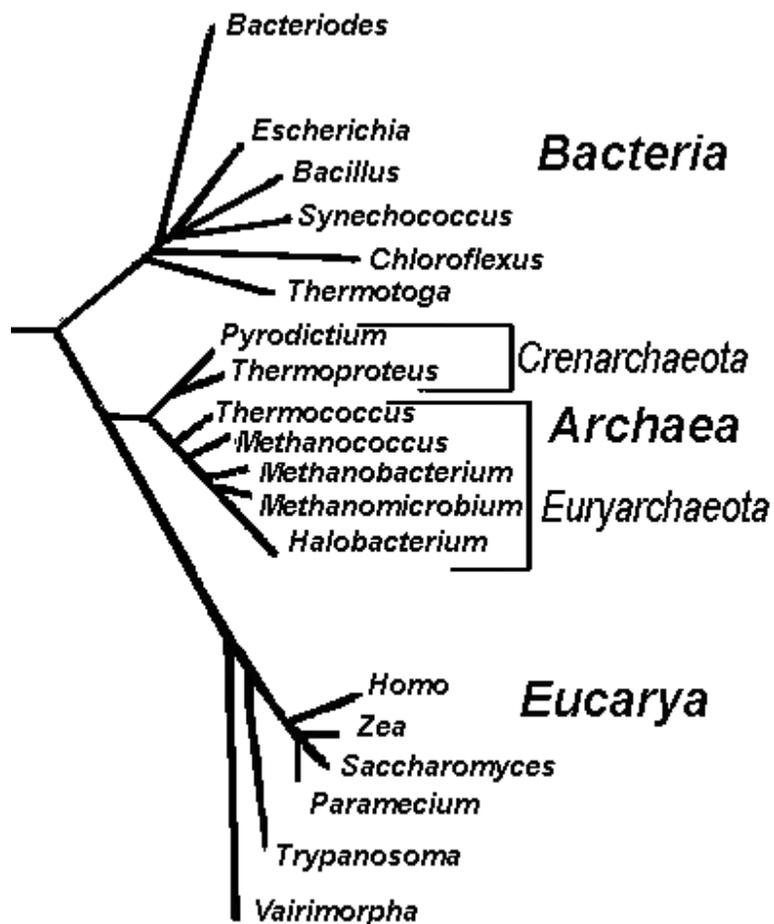
As a graduate student at the University of Illinois, Bernadette Pace used the annealing of rRNA with genomic DNA to measure the similarity of rRNAs in various species. These experiments demonstrated that rRNA-based methods are applicable to directly comparing a broader range of organisms (i.e., spanning greater phylogenetic distances) than is **whole genome DNA-DNA hybridization**. However, as with DNA-DNA measurements, it was necessary to have DNA and/or RNA from each species of interest.

If relationships were analyzed by comparing sequence data, rather than hybridizing the molecules, one could infer relationships without having all of the molecules in hand (only the sequence data from previous studies are necessary). This was already being done with protein sequences.

Carl Woese recognized the full potential of rRNA sequences as a measure of phylogenetic relatedness. He initially used an RNA sequencing method that determined about 1/4 of the nucleotides in the 16S rRNA (the best technology available at the time). This amount of data greatly exceeded anything else then available. Using newer methods, it is now routine to determine the sequence of the entire 16S rRNA molecule. Today, the accumulated 16S rRNA sequences (about 10,000) constitute the largest body of data available for inferring relationships among organisms.

### Molecular Phylogenies can Reflect Genealogy and Amount of Change

By comparing the inferred rRNA sequences (or those of any other appropriate molecule) it is possible to estimate the historical branching order of the species, and also the total amount of sequence change. An example of a 16S rRNA-based **phylogenetic tree** showing the three (identified) Domains of life - **Bacteria**, **Archaea** and **Eucarya** - is below. In this tree, lineages diverge from a common ancestral lineage on the far left. The lengths of the individual lines reflect the amount of sequence change (note that some lineages have modified the gene sequence substantially more than others, and thus have accumulated longer total branch lengths).



**Figure 2** 16s rDNA Phylogenetic Tree from: <http://lecturer.ukdw.ac.id/dhira/ClassAndPhylo/molPhylogeny.html>

## PROCEDURE:

In this laboratory, we will amplify a region of the 16s rDNA gene (based on the *E. coli* 16s rDNA sequence) using Polymerase Chain Reaction (PCR). The DNA will be sequenced using the capillary DNA sequencer at the Grice Marine Lab. We will then compare the results obtained with the public databases (NCBI and RDP) and determine the identity of the unknown bacteria.

## EXPERIMENTAL PROTOCOL:

1. Each student will obtain one liquid culture of bacteria. Note the code number. Also note the characteristics of the liquid culture such as color or consistency.
2. Every unknown culture should be treated as though it were a human pathogen. Wear gloves while handling the bacteria. Be careful not to contaminate anything, especially yourself.
3. Carefully resuspend the culture in the test tube. Remove 500 ul of culture into a special locking eppendorf

tube. Pellet in the microfuge for 2 minutes at 10,000 x g. Carefully draw off the medium using a P200. Some pellets are not firm and are easily disturbed.

4. Thoroughly resuspend cells in 400 ul dH<sub>2</sub>O. Freeze cells in a dry ice bath for 2 minutes. Place in the heating block (70 - 80' C) for 7-10 minutes. Open the cap *carefully* to release pressure. Add 100 ul of resuspended Chelex to the bacterial cells. Resuspend the Chelex-100 solution *each time* before use. Mix thoroughly using a vortex or pipettor.
5. Again freeze cells in the dry ice bath for 2 minutes, followed by 7-10 minutes in the heat block. Open cap *carefully* to release any built-up pressure. Then close cap tightly and spin in the microfuge for 2 minutes at 10,000 x g. Remove 200 ul of the clear supernatant to a new microfuge tube and keep the tube on ice.
6. Set up PCR reactions. Each student will add 8 ul of dH<sub>2</sub>O and 2 ul of the DNA supernatant to the small (0.5 ml) PCR tube. You should have a total of 10 ul in the tube. Write your initials on the TOP of the tube.
7. The instructor will prepare Master Mix plus primers, as shown in the table below. The Master Mix contains all the common components for a set of reactions. It improves consistency among the reactions and reduces pipetting error.
8. We will set up an "assembly line" to add Master Mix and mineral oil before putting the tubes into the PCR machine. Add 40 ul of Master Mix plus the 10 ul already added to the PCR tube = 50 ul final reaction volume.

Master Mix

Reactant	per Reaction	x (number of students) + 1 extra =
Taq Polymerase 5 U/ul	0.1 ul	_____ ul
PCR buffer 10x	5 ul	_____ ul
MgCl <sub>2</sub> 1.5 mM	3 ul	_____ ul
each NTP 10 mM	1 ul	_____ ul
forward primer 10 mM	1 ul (1.0 mM)	_____ ul
reverse primer 10 mM	1 ul (1.0 mM)	_____ ul
dH <sub>2</sub> O	15.4 ul	_____ ul
<b>TOTAL</b>	<b>40 ul</b>	_____ ul

## PCR Parameters:

30 cycles of:

Initial Denaturation:	94 'C - 1 min	Linked to :	Linked to:
95'C - 3 min	55 'C - 1 min	72 'C - 10 min	5 'C - indefinitely
	72 'C - 1 min		

**Questions to consider:**

1. What features of 16s rDNA make it suitable for phylogenetic analysis?
2. Describe the process of PCR. What happens during one PCR cycle?

Adapted from:

"Identification of bacteria using two degenerate 16s rDNA sequencing primers" by Boye, *et al.*, the "Microbiology Laboratory Manual" by J. D. Newman, Lycoming College ([srv2.lycoming.edu/~newman](http://srv2.lycoming.edu/~newman)), and Dr Dag Harmstad, University of Muenster, Germany (personal communication).