

16s rDNA SEQUENCING TO IDENTIFY UNKNOWN MICROORGANISMS

INTRODUCTION:

The rRNA gene is the most conserved (least variable) DNA 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. Carl Woese, who proposed the three Domain system of classification - Archaea, Bacteria, and Eucarya - based on such sequence information, pioneered this work.

Note on *terminology*: Several pieces of RNA are important for proper ribosome function. This RNA is not translated to protein, the ribosomal RNA is the active component. Thus we can refer to the “rRNA gene” or “rDNA” to designate the DNA in the genome that produces the ribosomal RNA.

Learning Goals:

1. Understand why the DNA of the 16S ribosomal RNA is often used for analysis.
2. Use sequence information from 16S rDNA to get information on phylogeny.

Topic I: RIBOSOMAL RNA

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. Examples of genes in this category are those that define the ribosomal RNAs (rRNAs). 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).

Most prokaryotes have three rRNAs, called the 5S, 16S and 23S rRNA. 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.

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 16S rDNA sequence has **hypervariable regions**, where sequences have diverged over evolutionary time. **Strongly conserved** regions often flank these hypervariable 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) and the Ribosomal Database Project (<http://rdp.cme.msu.edu/>). These sites also provide search algorithms to compare new sequences to their database.

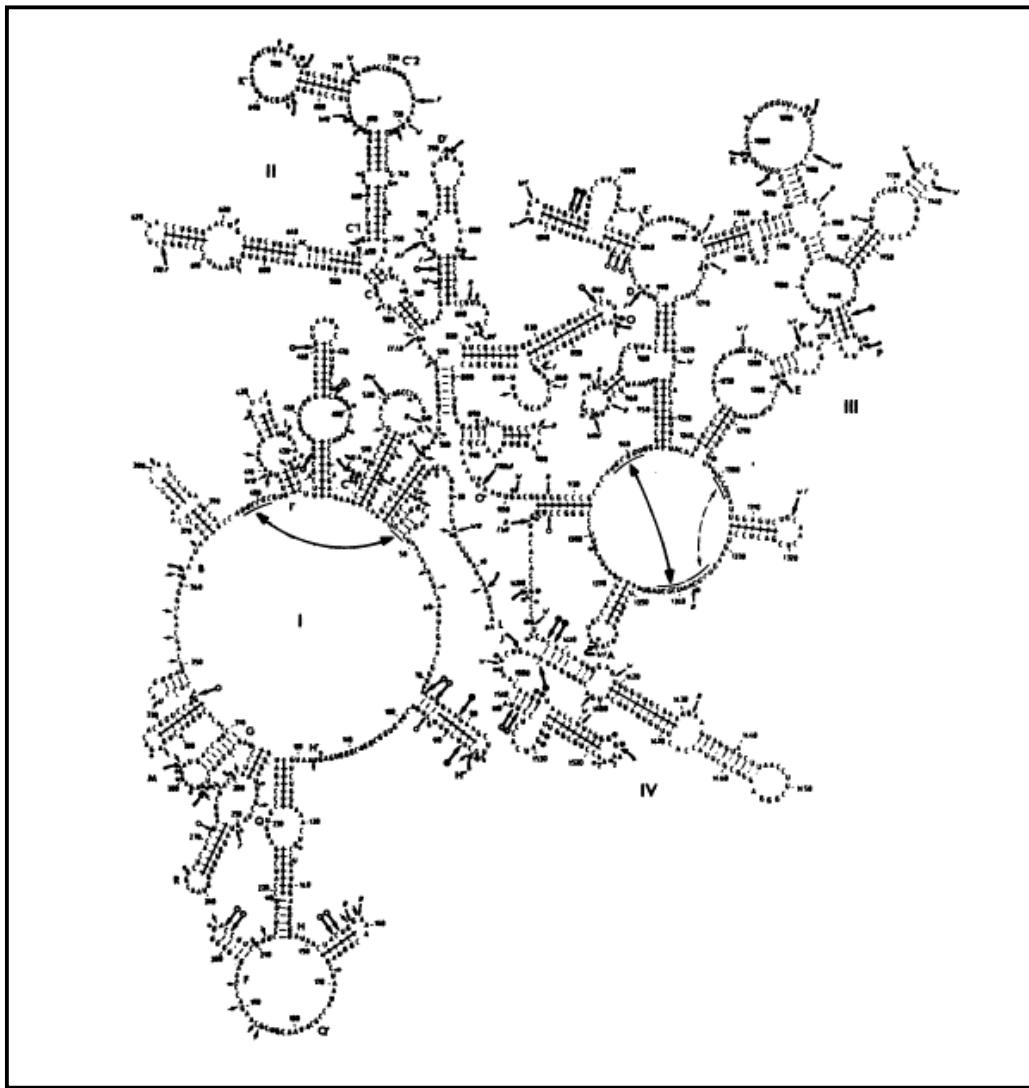


Figure 7.1 from: Structural organization of the 16S ribosomal RNA from *E. coli*. Topography and secondary structure. P Stiegler, P Carbon, M Zuker, J P Ebel, and C Ehresmann. *Nucleic Acids Res.* 1981 May 11; 9(9): 2153–2172.

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.

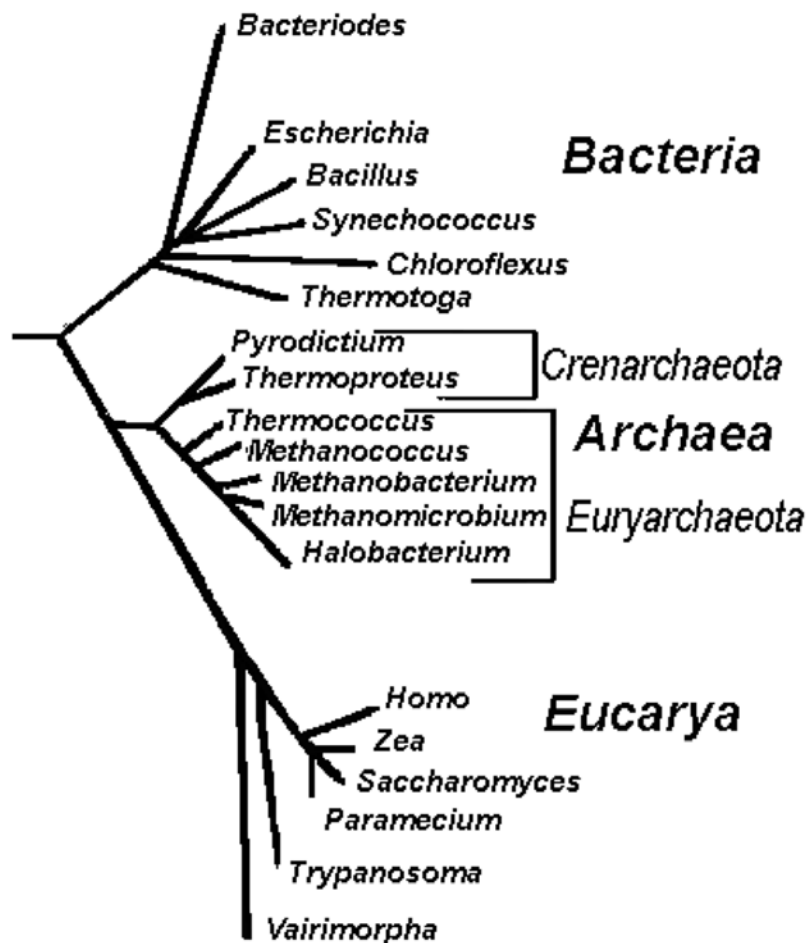
The extraordinary conservation of rRNA genes can be seen in these fragments of the small subunit (16S) rRNA gene sequences from organisms spanning the known diversity of life. Note several areas of identity among these diverse organisms in the Table below:

| | |
|--------------------------------|---|
| Human | GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGT ATATTAAGTTGCTGCAGTTAAAAAG |
| Yeast | GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGT ATATTAAGTTGTTGCAGTTAAAAAG |
| Corn | GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGT ATATTTAAGTTGTTGCAGTTAAAAAG |
| <i>E. coli</i> | GTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG TTAATCGGAATTACTGGGCGTAAAGCG |
| <i>Anacystis nidulans</i> | GTGCCAGCAGCCGCGGTAATACGGGAGAGGCAAGCG TTATCCGGAATTATTGGGCGTAAAGCG |
| <i>Thermotoga maratima</i> | GTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCG TTACCCGGATTTACTGGGCGTAAAGGG |
| <i>Methanococcus vannielii</i> | GTGCCAGCAGCCGCGGTAATACCGACGGCCCGAGTG GTAGCCACTCTTATTGGGCCTAAAGCG |
| <i>Thermococcus celer</i> | GTGGCAGCCGCCGCGGTAATACCGGCGGCCCGAGTG GTGGCCGCTATTATTGGGCCTAAAGCG |
| <i>Sulfolobus sulfotaricus</i> | GTGTCAGCCGCCGCGGTAATACCAGCTCCGCGAGTGG TCGGGGTGATTACTGGGCCTAAAGCG |

Topic II: 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 7.2: 16s rDNA Phylogenetic Tree from: <http://lecturer.ukdw.ac.id/dhira/ClassAndPhylo/molPhylogeny.html>



REFERENCES

"Identification of bacteria using two degenerate 16s rDNA sequencing primers"; K. Boye, E. Høgdall, M. Borre; *Microbiol Res.* 1999 May;154(1):23-6.

"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 "Microbiology Laboratory Manual" by J. D. Newman, Lycoming College
(<http://srv2.lycoming.edu/~newman/>)

Dr Dag Harmstad, University of Muenster, Germany (personal communication).
<http://lecturer.ukdw.ac.id/dhira/ClassAndPhylo/molPhylogeny.html>