CHAPTER 04: MAIN TYPES of CLASSIFICATION

1. Introduction

Modern bacterial taxonomy brings together different kinds of information: the observable traits of bacteria (phenotypic data), their genetic characteristics (genotypic data), and their evolutionary relationships (phylogenetic data). By combining all these sources of knowledge in what is called a *polyphasic approach*, scientists can create a classification system that is more reliable and stable over time.

2. Phenetic classification

For a very long time, microbial taxonomists relied exclusively on a **phenetic system,** which groups organisms together based on the mutual similarity of their phenotypic characteristics. This classification system succeeded in bringing order to biological diversity and clarified the function of morphological structures. Phenetic studies can reveal possible evolutionary relationships, but *they are not dependent on phylogenetic analysis*.

2.1. Morphological characteristics

Morphological features are important in microbial taxonomy for many reasons. Morphology is easy to study and analyze, particularly in eukaryotic microorganisms and the more complex prokaryotes. In addition, morphological comparisons (colony and cellular morphology) are valuable because structural features depend on the expression of many genes, usually genetically stable, and normally (at least in eucaryotes) do not vary greatly with environmental changes. Thus, morphological similarity is a good indication of phylogenetic relatedness.

Many different morphological features are employed in the classification and identification of microorganisms: cell shape, cell size, cells arrangement (single or in chains, clusters packets, etc), colonial morphology, ultrastructural characteristics, staining behavior, cilia and flagella, flagella arrangement, mechanism of motility, endospore shape and location, spore morphology and location, cellular inclusions and color.

1.2. Staining characteristics

By highlighting the similarities or differences in the composition of the **cell wall** of the studied microorganisms, the staining characteristics provide information about phylogenetic

relationships. The majority of bacteria are either Gram-positive or Gram-negative. Other differential stains, such as acid-alcohol-resistant staining, can be used to identify microorganisms belonging to more specific groups. Remember that this character cannot identify bacteria without a cell wall or Archaebacteria that lack the characteristic cell wall structure.

1.3. Physiological and metabolic (biochemical) characters

Physiological and metabolic characteristics are very useful because they are directly related to the nature and activity of microbial enzymes and transport proteins. Since proteins are gene products, analysis of these characteristics provides an indirect comparison of microbial genomes. Enzymatic activity is often used to differentiate bacteria. These characters can group bacteria into distinct species, and provide information about their ecological niche within an ecosystem (soil bacterium capable of fixing atmospheric nitrogen or oxidizing sulfur).

Gram-negative *Enterobacteria* form a large heterogeneous group of microbes (intestinal tract of humans and animals), and include several species of pathogenic bacteria responsible for diarrheal diseases. Enzymatic activity (biochemical tests) can quickly identify the pathogenic agent (Fig. 01). Thus, all the genera in the family *Enterobacteriaceae* (*Escherichia, Enterobacter, Shigella, Citrobacter,* and *Salmonella*) share the common property of not producing oxidase. *Escherichia, Enterobacter,* and *Citrobacter,* which transform lactose into acid and gas by fermentation, are distinguished from *Salmonella* and *Shigella,* which lack this capacity.

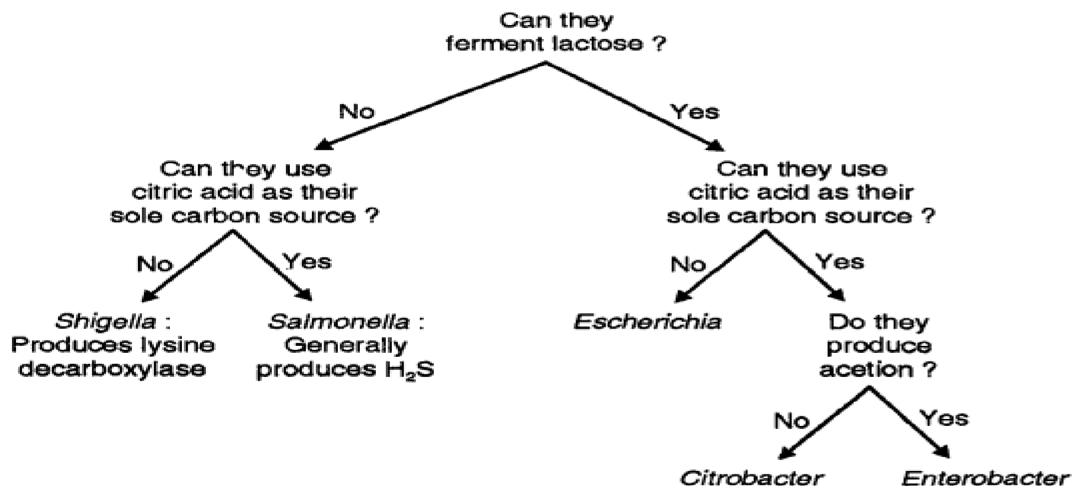

Figure 01: Use of metabolic characteristics to identify selected genera of enteric bacteria.

Many biochemical and physiological characters used to classify bacteria are based on conditions that support growth as: carbon and nitrogen sources, cell wall constituents, energy

sources, fermentation products, general nutritional type, luminescence, mechanisms of energy conversion, motility, osmotic tolerance, oxygen relationships, pH optimum and growth range, photosynthetic pigments, salt requirements and tolerance, secondary metabolites formed, sensitivity to metabolic inhibitors and antibiotics and storage inclusions.

2.4. Ecological characteristics

The ability of a microorganism to colonize a specific environment is of taxonomic value. Some microbes may be very similar in many other respects but inhabit different ecological niches, suggesting they may not be as closely related as first suspected. Some examples of taxonomically important ecological properties are life cycle patterns; the nature of symbiotic relationships; the ability to cause disease in a particular host; and habitat preferences such as requirements for temperature, pH, oxygen, and osmotic concentration. Many growth requirements are considered physiological characteristics as well (section 2.3).

2.5. Serological characteristics

Serology is the science that studies serum, the noncellular fraction of blood, and the immune responses revealed by examining serum. Microorganisms are antigenic, meaning that their presence in an animal's body induces the production of antibodies. Antibodies are proteins produced by the infected organism that circulate in the blood and bind very specifically to the bacteria that triggered their production (Fig. 02) (e.g. the immune system of a rabbit injected with dead typhoid bacteria will react by producing antibodies against this type of bacteria).

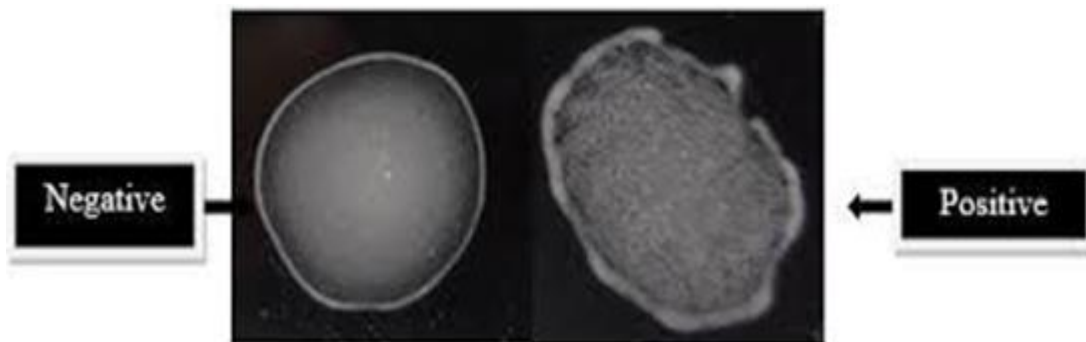

Figure 02: Slide agglutination test.

Sera that inactivate particular bacteria are called antisera (*s.* **antiserum** or **immune serum)**, they are commercial kits available on the market containing antibody solutions designed to identify various microorganisms of medical importance (agglutination test).

2.6. Chemotaxonomy

Chemotaxonomic characterization relies on the analysis of cellular constituents in prokaryotes. This approach uses the chemical composition of organisms as a basis for classification.

2.6.1. Cell wall composition

Given that peptidoglycan is ubiquitous in the bacterial domain (excluding Tenericutes, *Chlamydiae*, and Planctomycetes), the composition of the cell wall serves as a reliable criterion for taxonomic classification. Archaeal wall structure and chemistry differ from those of the *Bacteria*. Before they were distinguished as a unique domain of life, the *Archaea* were characterized as being either Gram-positive or Gram-negative. Archaeal cell walls lack peptidoglycan and also exhibit considerable variety in terms of their chemical make-up. For example, *Methanobacterium* and some other methane-generating archaea (Methanogens) have walls containing **pseudomurein** (Fig. 03), a peptidoglycan-like polymer that has L-amino acids instead of D-amino acids in its cross-links, *N*-acetyltalosaminuronic acid instead of *N*-acetylmuramic acid, and β (1→3) glycosidic bonds instead of β (1→4) glycosidic bonds.

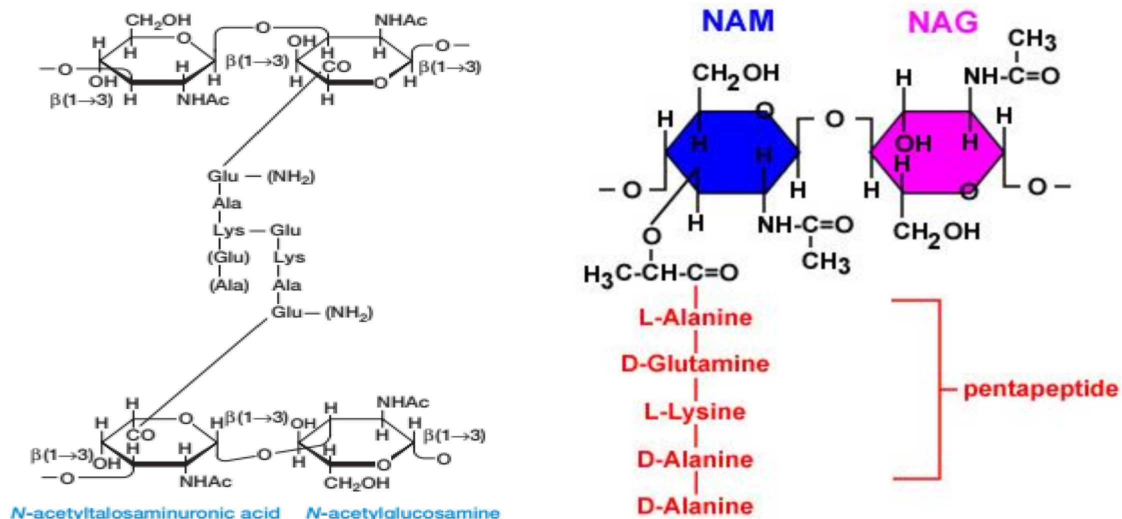

Figure 03: Pseudomurein (in left) and peptidoglycan (in right) structure.

Also, determining its composition is mainly used for Gram-positive bacteria, in which it can be characteristic of a genus or of a particular bacterial species. The type of peptidoglycan in Gram-negative bacteria is rather uniform and has little taxonomic value. Teichoic acids embedded in this peptidoglycan can also be used as phenotypic markers after extraction, purification, and analysis

by chromatography. Their absence in *Micrococcus* spp. makes it possible to distinguish these bacteria from *Staphylococcus* spp. Within this genus, the type of teichoic acid makes it possible to differentiate *Staphylococcus aureus* from other staphylococcal species.

2.6.2. Fatty acid analysis

Bacteria differ in the type and reactive quantity of fatty acids that make up their membranes, thus the cellular fatty acids composition can be used as an identifying marker. Bacterial membranes differ from eukaryotic membranes in lacking sterols (steroid-containing lipids) such as cholesterol. However, many bacterial membranes contain sterol-like molecules called **hopanoids**, that are synthesized from the same precursors as steroids (like the sterols in eukaryotic membranes, they probably stabilize the membrane).

One of the most distinctive features of the *Archaea* is the nature of their membrane lipids. They differ from both *Bacteria* and *Eucarya* in having branched chain hydrocarbons attached to glycerol by ether links rather than fatty acids connected by ester links (Fig. 04). Sometimes two glycerol groups are linked to form an extremely long tetraether. Usually the diether hydrocarbon chains are 20 carbons in length, and the tetraether chains are 40 carbons. Cells can adjust the overall length of the tetraethers by cyclizing the chains to form pentacyclic rings (Fig. 04). Archaeal lipids are derivatives of isopranyl glycerol ethers rather than the glycerol fatty acid esters in *Bacteria*.

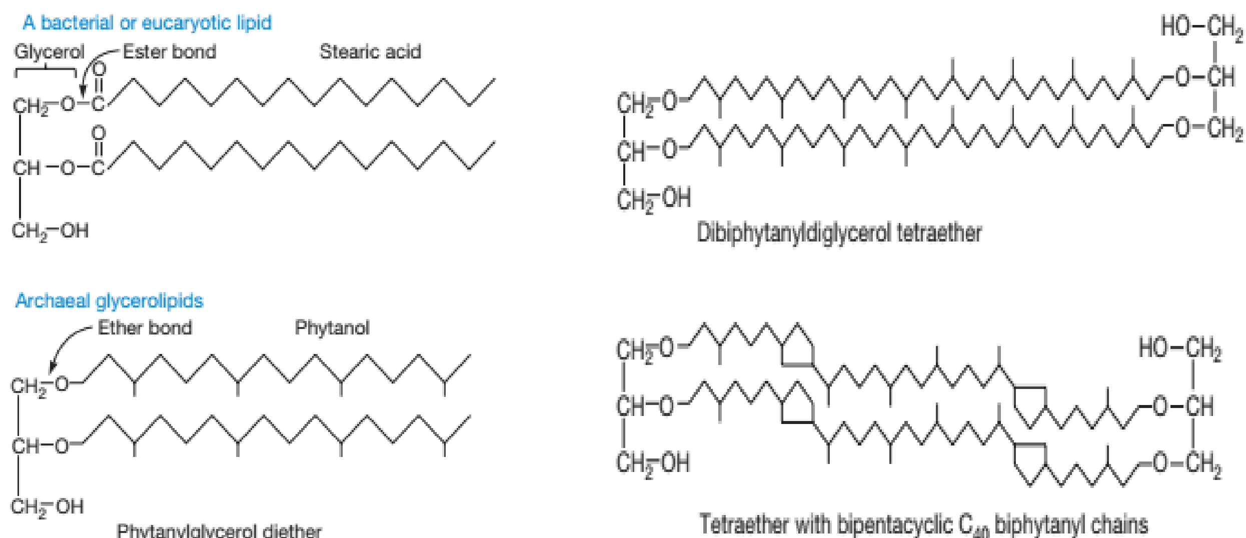

Figure 04: Archaeal membrane lipids.
(Isoprene units, C_5H_8; Phytanol is derived from an isoprene unit, $C_{20}H_{42}O$).

In Gram-negative bacteria, the fatty acids are present in both the cytoplasmic and the outer membranes. Whereas in Gram-positive bacteria (lacking an outer membrane), the cytoplasmic

membrane is the source of fatty acids. To analyze their fatty acid composition, bacterial cells are grown under standardized conditions. After this the cells are chemically treated with sodium hydroxide and methanol to release the fatty acids and to convert those acids to their more volatile methyl ester form (FAME = Fatty acid methyl ester). The resulting fatty acids (methylated esters) can then be separated and analyzed using gas chromatography. By comparing the pattern of peaks, or chromatogram, to those of known species and isolate can be identified. A fatty acid profile can therefore be characteristic of a bacterial genus or species.

It has also been demonstrated that the lipid composition of bacterial membranes varies with environmental temperature in such a way that the membrane remains fluid during growth. For example, bacteria growing at lower temperatures will have fatty acids with lower melting points in their membrane phospholipids.

2.6.3. Comparison of protein and amino acid sequencing

Amino acid sequences of proteins are direct reflections of mRNA sequences and therefore closely related to the structures of the genes coding for their synthesis. Comparisons of proteins (determine and compare the amino acid sequences) of different microorganisms is very useful taxonomically. If the sequences of proteins with the same function are similar, the organisms possessing them are probably closely related. Nevertheless, the sequences of proteins with dissimilar functions often change at different rates; some sequences change quite rapidly whereas others are very stable. The sequences of cytochromes and other electron transport proteins, histones, heat shock proteins, transcription and translation proteins and a variety of metabolic enzymes have been used in taxonomic studies.

Because protein sequencing is slow and expensive, more indirect methods of comparing proteins frequently have been employed. The electrophoretic mobility (electrophoretic profile) of proteins is useful in studying relationships at the species and subspecies level. Antibodies can discriminate between very similar proteins, and immunologic techniques are used to compare proteins from different microorganisms.

2.7. Numerical Taxonomy

The earliest system of biological classification was based on arbitrarily chosen criteria (artificial classification). Later when the fact of biological evolution was recognized, another dimension was immediately added to the concept of natural or phenetic classification. An

alternative approach for the classification is empirical one where the taxonomic arrangement is based on quantification of the similarities and differences among organism. This was first suggested by a French biologist Michal Adansona and the classification is also named as **Adansonian** (or **numerical**) **taxonomy**.

Peter H. A. Sneath and Robert Sokal have defined numerical taxonomy as "*the grouping by numerical methods of taxonomic units into taxa on the basis of their character states*." Here each phenotypic character is given equal weighting, it should express numerically the taxonomic distances between organisms in terms of the number of characters they share, relative to the total number of characters examined. This approach was not feasible before the advent of computers because of the large number of calculations involved.

The process begins with a determination of the presence or absence of selected characters in the group of organisms under study. Many characters, at least 50 and preferably several hundred, should be compared for an accurate and reliable classification (morphological, biochemical, physiological). After character analysis, an association coefficient is calculated for each pair of organisms in the group, the ***simple matching coefficient*** (**S_{SM}**) or the ***Jaccard coefficient*** (**S_J**) (Tab. 01). Both coefficients increase linearly in value from 0.0 (no matches) to 1.0 (100% matches).

Table 01: Calculation of association coefficients for two organisms

In this example, organisms A and B are compared in terms of the characters they do and do not share. The terms in the association coefficient equations are defined as follows:

		Organism B 1	Organism B 0
Organism A	1	*a*	*b*
	0	*c*	*d*

a = number of characters coded as present (1) for both organisms
b and c = numbers of characters differing (1,0 or 0,1) between the two organisms
d = number of characters absent (0) in both organisms
Total number of characters compared = $a + b + c + d$

The simple matching coefficient $(S_{SM}) = \dfrac{a + d}{a + b + c + d}$

The Jaccard coefficient $(S_J) = \dfrac{a}{a + b + c}$

The simple matching coefficients, or other association coefficients, are then arranged to form a **similarity matrix**. This is a matrix in which the rows and columns represent organisms, and each value is an association coefficient measuring the similarity of two different organisms; so that each organism is compared to every other one in the table (Fig. 05*a*). Organisms with great similarity are grouped together and separated from dissimilar organisms (Fig. 05*b*); such groups of organisms are called **phenons** (sometimes called phenoms).

The results of numerical taxonomic analysis are often summarized with a treelike diagram called a **dendrogram** (Fig. 05*c*). Each branch point is at the similarity value relating the two branches. The organisms in the two branches share so many characteristics that the two groups are seen to be separate and below the branch point value, the two groups appear to be one.

Sometimes phenoms are preceded by a number showing the similarity level above which they appear (e.g., a 70-phenon is a phenom with 70% or greater similarity among its constituents). Phenoms formed at about 80% similarity often are equivalent to species.

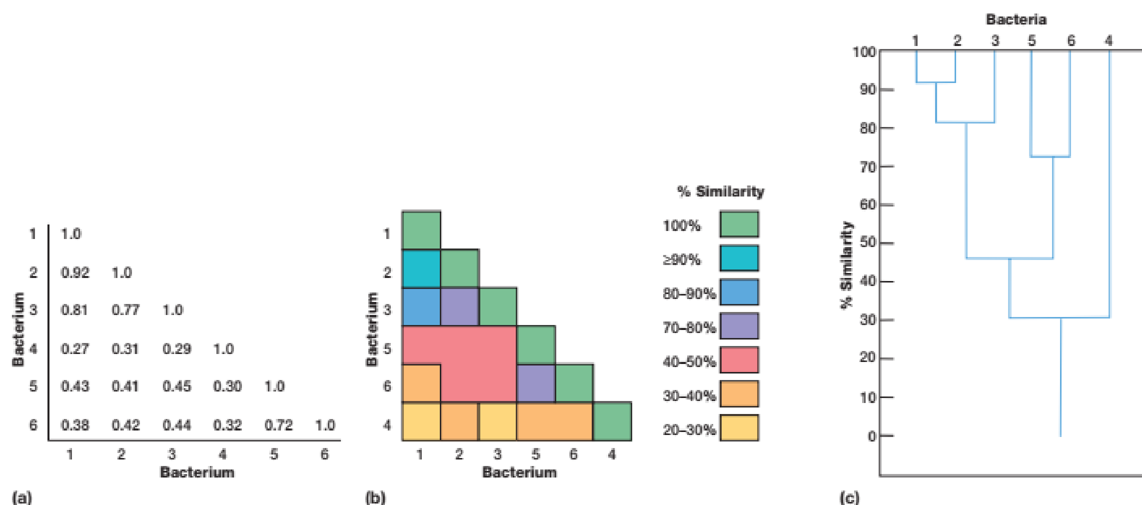

Figure 05: Clustering and dendrograms in numerical taxonomy.

((a) A small similarity matrix that compares six strains of bacteria. The degree of similarity ranges from none (0.0) to complete similarity (1.0). (b) The bacteria have been rearranged and joined to form clusters of similar strains. For example, strains 1 and 2 are the most similar. The cluster of 1 plus 2 is fairly similar to strain 3, but not at all to strain 4. (c) A dendrogram showing the results of the analysis in part (*b*). Strains 1 and 2 are members of a 90-phenon, and strains 1–3 form an 80-phenon. While strains 1–3 may be members of a single species, it is quite unlikely that strains 4–6 belong to the same species as 1–3).

Numerical taxonomy has proved to be a powerful tool in microbial taxonomy. Although it often has simply reconfirmed already existing classification schemes, sometimes accepted classifications are found wanting. Numerical taxonomic methods also can be used to compare sequences of macromolecules such as RNA and proteins.

3. Genotypic Classification

There are currently many ways in which the genotype of a microbe can be evaluated in taxonomic terms. In general, **genotypic classification**, using molecular characteristics such as DNA, RNA and proteins, seeks to compare the genetic similarity between organisms and advances our understanding of microbial evolution and taxonomy. Molecular analysis is the only feasible means of collecting a large and accurate data set from a number of microbes. Individual genes or whole genomes can be compared.

3.1. Nucleic acid base composition

It is, possibly, the simplest technique to be employed in the determination of DNA base composition by calculating (G + C)/(A + T) ratio or **G + C content** (or **CHARGAFF Coefficient**) as follow:

$$\text{Mol\% G + C} = \frac{G + C}{G + C + A + T} \times 100$$

- The G + C content can be ascertained after hydrolysis of DNA and analysis of its bases with high-performance liquid chromatography (HPLC).

- The G + C content often is determined from the **melting temperature (*Tm*)** of DNA.

- The DNA with a greater G + C content have more hydrogen bonds, and its strands separate at higher temperatures, that is, it has a higher melting point.

- DNA melting can be easily followed spectrophotometrically because the absorbance of DNA at 260 nm (UV light) increases during strand separation (Fig. 06). The midpoint of the rising curve gives the melting temperature, a direct measure of the G + C content.

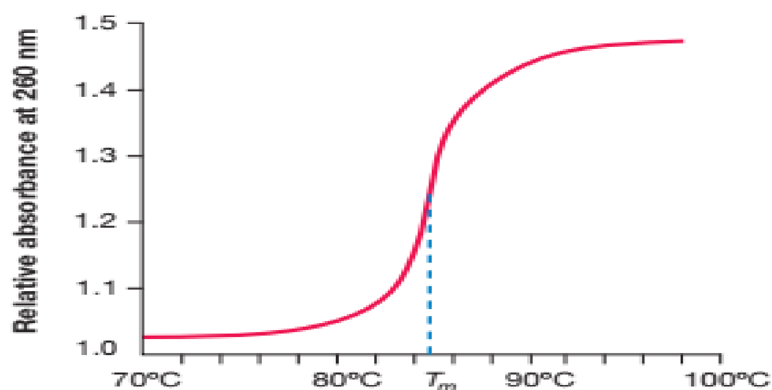

Figure 06: A DNA melting curve. The *Tm* is indicated

Taxonomically these G + C data are valuable for at least two reasons:

- They can confirm a taxonomic scheme developed using other data. If organisms in the same taxon are too dissimilar in G + C content, the taxon probably should be divided.
- The G + C content appears to be useful in characterizing prokaryotic genera since the variation within a genus is usually less than 10% even though the content may vary greatly between genera (ranging from around 25 to almost 80%).

3.2. Nucleic acid hybridization

The similarity between genomes can be compared more directly by use of nucleic acid hybridization studies, this includes DNA-DNA homology (degree of sequence homology). If a mixture of single stranded DNA (formed by heating of dsDNA) is cooled and held at a temperature about 25°C below the T_m, strands with complementary base sequences will reassociate to form stable dsDNA, and the non-complementary strands will remain single (the incubation temperature determines the degree of homologous sequences to form a stable hybrid).

Two strains whose DNAs show at least 70% relatedness under optimal hybridization conditions and less than a 5% difference in Tm often, but not always, are considered members of the same species. However, DNA preparation from two unrelated bacteria, could not hybridize or they will not form a *stable detectable hybrid*.

3.3. Nucleic acid sequencing

Despite the usefulness of G + C content determination and nucleic acid hybridization studies, rRNAs from small ribosomal subunits (5S and 16SrRNA's isolated from 50S and 30S subunits respectively of prokaryotic ribosomes) have become the molecules of choice for inferring microbial phylogenies and making taxonomic assignments at the genus level. Thus, on identical sequence means identical organisms and a similar sequence means closely related organisms.

The rRNA are almost ideal for studies of microbial evolution and relatedness, their functional role is same in all ribosomes. Furthermore, their structure changes very slowly with time. Because the structure of ribosome cannot tolerate much change and still remain functional, rRNA is highly conserved.

Comparative analysis of 16S rRNA sequences from thousands of organisms has demonstrated the presence of ***oligonucleotide signature sequences***. These are short, conserved nucleotide sequences that are specific for a phylogenetically defined group of organisms.

3.4. Genomic Fingerprinting

A group of techniques called **genomic fingerprinting** can also be used to classify microbes and help determine phylogenetic relationships. Unlike the molecular analyses so far discussed, genomic fingerprinting does not involve nucleotide sequencing. Instead, it employs the capacity of restriction endonucleases (restriction enzymes) to recognize specific nucleotide sequences, and cut DNA molecule at these points. Thus, the pattern of DNA fragments generated by endonuclease cleavage (called restriction fragments) is a direct representation of nucleotide sequence. For example, the enzyme *EcoRI* cuts DNA at the point indicated by the arrow in each sequence (Fig. 07). The comparison of restriction fragments between species and strains is the basis of ***r*estriction *f*ragment *l*ength *p*olymorphism** (***RFLP***) analysis.

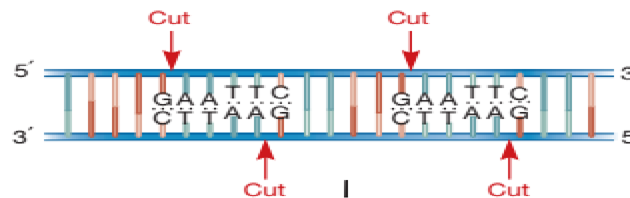

Figure 07: Restriction endonuclease *Eco*Rl action.

Because DNA fingerprinting enables identification to the level of species, subspecies, and often strains, it is valuable not only in the study of microbial diversity, but in the identification of human, animal, and plant pathogens as well.

Figure 08 shows the taxonomic utility of several kinds of molecular analyses including protein profiling; with the exception of genome sequencing, it is clear that a combination of approaches is best for identification at the species level or lower.

Family	Genus	Species	Subspecies	Strain
Genome sequencing	Genome sequencing	Genome sequencing	Genome sequencing	Genome sequencing
16S rDNA sequencing	16S rDNA sequencing			
	Mol% G+C	Mol% G+C	Mol% G+C	
	DNA-DNA hybridization	DNA-DNA hybridization	DNA-DNA hybridization	
		Multilocus sequence typing	Multilocus sequence typing	Multilocus sequence typing
		Whole cell protein profiling	Whole cell protein profiling	Whole cell protein profiling
		Genomic fingerprinting	Genomic fingerprinting	Genomic fingerprinting

Figure 08: Relative taxonomic resolution of various molecular techniques.

4. Phylogenetic classification

The term **phylogeny** (Greek *Phylon,* tribe or race, and *Genesis,* generation or origin) refers to the evolutionary development of a species. Scientists realized that when they observed differences and similarities between organisms as a result of evolutionary processes, they also gained insight into the history of life on Earth. However, for much of the twentieth century, microbiologists could not effectively employ phylogenetic classification systems, primarily because of the lack of a good fossil record. When Woese and Fox proposed using rRNA nucleotide sequences to assess evolutionary relationships among microorganisms, the door opened to the resolution of long-standing inquiries regarding the origin and evolution of the majority of life forms on Earth, the microbes.

The sequence of nucleic acids and proteins change with time and are considered to be ***molecular chronometers***. This term is important in the use of molecular sequences in determining phylogenetic relationship. It is assumed that there is an evolutionary clock where the sequences of many rRNA and protein gradually change over time without destroying or severely altering their functions. Using molecular chronometer for phylogenetic analysis is complex because the rate of sequence change can vary because some periods are recognized by especially rapid change. Different molecules and various parts of the same molecule can change at different rates. Highly conserved molecules like rRNA are used to follow large scale evolutionary changes.

5. Polyphasic taxonomy

As we come to know by previous paragraph that all phylogenetic results vary with data used in analysis, many taxonomist believe that all possible valid data should be employed in determining phylogeny, in the approach called ***polyphasic taxonomy***, taxonomic schemes are developed using a wide range of phenotypic and genotypic information ranging from molecular properties to ecological characteristics. The criteria for the selection of techniques depend upon the level of taxonomic resolution needed. Like serological technique is used to identify strains, but not for genera or species. Protein electrophoretic pattern is useful in determining species but not for genera or families. DNA hybridization and percentage G + C content can be used to study species and genera other characteristic like chemical composition, DNA probe result, rRNA sequences, DNA sequences can be used to define species, genera and families

6. Phylogenetic trees

Phylogenetic relationships are illustrated in the form of branched diagrams or trees. A ***phylogenetic tree*** is a graph made of branches that connect nodes (Fig. 09). The nodes represent taxonomic units such as species or genes; the external nodes at the end of the branches represent living (extant) organisms. As in the universal phylogenetic tree (Fig. 10), the length of the branches represents the number of molecular changes that have taken place between the two nodes. A tree may be unrooted or rooted. An unrooted tree (Fig. 09*a*) simply represents phylogenetic relationships but does not provide an evolutionary path. Figure 09*a* shows that A is more closely related to C than it is to either B or D, but does not specify the common ancestor for the four species or the direction of change. In contrast, the rooted tree (Fig. 09*b*) gives a node that serves as the common ancestor and shows the development of the four species from this root.

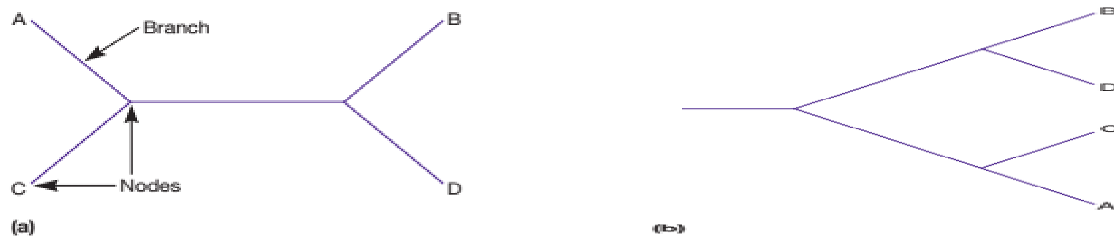

Figure 09: Examples of Phylogenetic trees.
((a) Unrooted tree joining four taxonomic units. (b) Rooted tree).

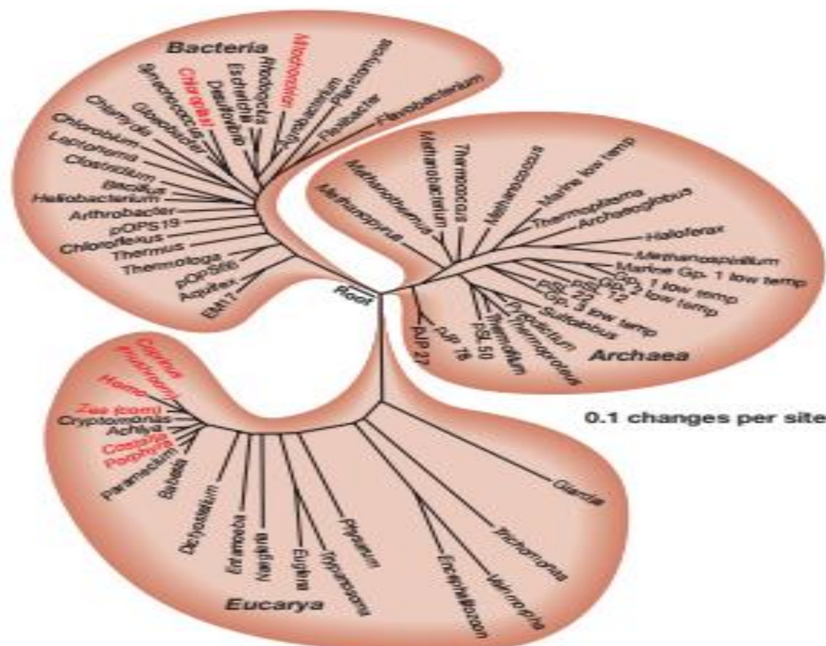

Figure 10: Universal Phylogenetic Tree.
(These evolutionary relationships are based on rRNA sequence comparisons. Length of branches indicates evolutionary relationships between organisms but not time. Microbes are printed in black).