UNIT 2

MICROBIAL TAXONOMY

Structure

2.1 Introduction
Expected Learning Outcomes
2.2 Classification
History of Classification
Taxonomic Ranks
Classification System
2.3 Characterization
Phenotypic or Classical Characterization
Molecular Characterization
2.4 Nomenclature
2.5 Bergey's Manual of Systematic Bacteriology
2.6 Numerical Taxonomy
2.7 Summary
2.8 Terminal Questions
2.9 Answers

ignou
THE PEOPLE'S
UNIVERSITY

2.1 INTRODUCTION

With the estimation of 13 million species existing on Earth, only 1.75 million species are described which include a description of 1,56,000 microbial species (Table 2.1). Due to microscopic size, and lack of definite structural features the classification of microorganisms is a great daunting challenge. Additionally, with the advent of modern molecular tools, the dataset of microbial species is going to increase multifold and makes their affiliation more difficult. But the basic question which may come to your mind is that why do we need to classify or identify microbes? The answer to this question lies in the fact that the accurate identification of affiliation of microbe is important regarding both economic, social and health reasons. Therefore, a repository of identified microorganisms is required to track causative organisms of a disease or to find useful microorganisms which have industrial or agricultural importance. Therefore, we need to have proper identification and classification system to understand the microbial diversity as well as to have a thorough reference. This aspect is studied under 'Taxonomy' or 'Systematics.' **Taxonomy** (Greek: taxis- arrangement or order, and nomos- law, or to assign, to distribute or govern) deals with the study of classifying organisms. **Microbial taxonomy** may be defined as the study of the diversity of

microorganisms with the aim of organizing and prioritizing in an orderly manner. The term **systematics** is often used as a synonym of taxonomy. There are three separate but integrated components, i.e., **classification**, **characterization** or **identification**, and **nomenclature** which constitute microbial taxonomy. We shall deal these components separately in further sections.

Table 2.1: Number of described species

Species	Number
Bacteria	4000
Protists (algae, protozoa)	80000
Fungi	72000
Others (Plants, animals)	1594000

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ Understand the significance of microbial taxonomy;
- ❖ Know the components of the microbial taxonomy;
- ❖ Know how to classify organisms and what are the methods available for taxonomic affiliation of the microorganisms;
- ❖ Understand the usefulness of polyphasic taxonomic approaches to identify and classify organisms, and
- ❖ Acknowledge the use of numerical taxonomy in microbial characterization.

2.2 CLASSIFICATION

Classification may be defined as the arrangement of organisms into categories (taxa) by similarities or relationships.

2.2.1 History of Classification

The oldest mention of classification comes with **Aristotle** (384-322 BC) who divided living organisms into plants and animals. After Aristotle, Carl von Linné or **Carolus Linnaeus** (1707-1778), a Swedish botanist, grouped organisms based on their characteristics. This type of method is known as a **natural classification system**. He classified all living organisms in animals and plants as **kingdoms** which were further grouped into five levels, i.e., **class, order, genus, species**, and **variety**. Owing to his contribution in classifying and naming organisms Carl is also called as father of taxonomy.

It was Haeckel (1866) who included microorganisms for the first time in classification system and proposed a separate Kingdom **'Protista'** to include unicellular organisms. Further, the classification system gained much attention

was **five kingdom concept** (Fig. 2.1) proposed by **Robert H. Whittaker (1969).** The five kingdoms namely Monera, Protista, Plantae, Fungi, and Animalia, were propounded based on (i) cell type, i.e., prokaryotic or eukaryotic, (ii) level of organization, i.e., **solitary** and **colonial unicellular** or **multicellular** organization, and (iii) type of nutrition. A brief description of these kingdoms is given below.

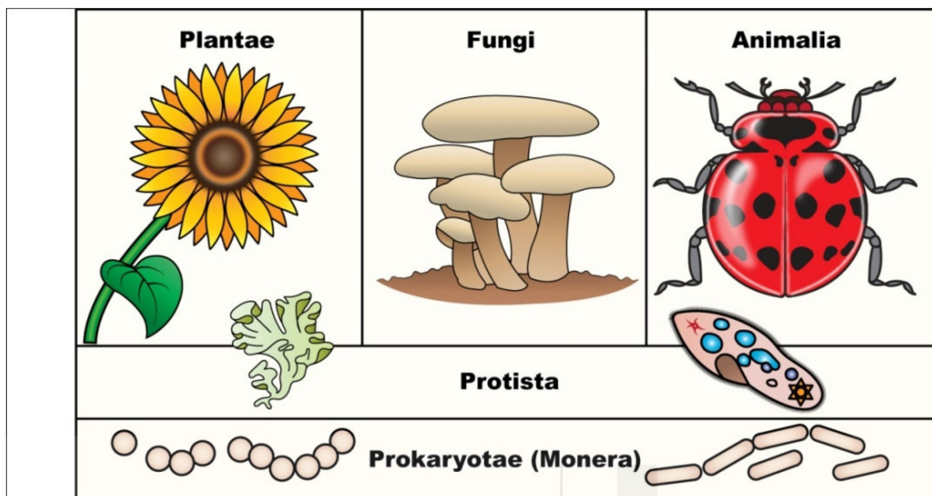


Fig. 2.1: Representation of five kingdoms as per classification system of Whitakker. Figure adopted from web page.

Monera: (i) Members of monera are most abundant microorganisms which include prokaryotic organisms such as bacteria, cyanobacteria, and mycoplasma.

(ii) They are primarily unicellular and have cell walls (except mycoplasma which lacks a cell wall) and are of varied shapes (spherical, rod-shaped, comma-shaped, spiral, etc.) and size.

(iii) They include both autotrophic (phototrophic and chemotrophic), and heterotrophic microorganisms. Some of them are parasites and cause deadly diseases.

Protista: (i) It includes single cellular eukaryotic organisms represented by Chrysophytes, Dinoflagellates, Euglenoids, Slime molds, and Protozoans.

(ii) Majority of them are aquatic and have an autotrophic or heterotrophic mode of nutrition.

(iii) Protists exhibit both sexual and asexual mode of reproduction.

Fungi: (i) They are heterotrophic eukaryotic microorganisms which include unicellular yeast, multicellular but microscopic molds, macroscopic mushrooms, etc.

(ii) Most of them exhibit an absorptive mode of nutrition and live on dead organisms and thus, termed as **saprophytes**

(iii) They reproduce by vegetative and sexual mode of reproduction. Some of them form fruiting bodies.

(iv) Many fungal species live in symbiotic relationship with algae and called as **lichens**.

Plantae: (i) These includes all plants which are multicellular organisms having cells wall made of cellulose.

(ii) They are autotrophic (photosynthetic) and synthesize their food. However, some of them such as some epiphytes and insectivorous plants (Venus flytrap, bladderwort) are partially heterotrophic.

(iii) They show alternation of generation, i.e., the diploid sporophytic and the haploid gametophytic stage– that alternate with each other.

Animalia: (i) It includes all multicellular eukaryotic organisms which lack a cell wall.

(ii) They are heterotrophs and depend directly or indirectly on plants. They show the ingestive type of nutrition.

Despite the systematic classification system, five kingdom concept is not accepted by many biologists as it lacks distinction between **archaea** and bacteria. Therefore, it followed proposal of **six kingdom** concept which included archaea.

The most recent classification has been proposed by **Carl R. Woese** (1998) who grouped all living life forms in **three domains**, i.e., **Eukarya, Bacteria,** and **Archea** based on the sequence of rRNA genes, a component of ribosomal structure (Fig. 2.2). In this classification system, Eukarya includes three kingdoms namely plants, animals, and fungi. The kingdom Bacteria has several kingdoms which include phototrophic, chemotrophic, pathogenic (disease-causing) as well as non-pathogenic bacteria. The Archea includes prokaryotes that lack peptidogly can in their cell walls. They are usually extremophiles which can survive the extreme conditions such as high temperature (**thermophiles**), high salt concentration (**halophiles**), and pH (very low or very high), etc. Methane-producing (**methanogenic**) anaerobic bacteria are also the member of Archea.

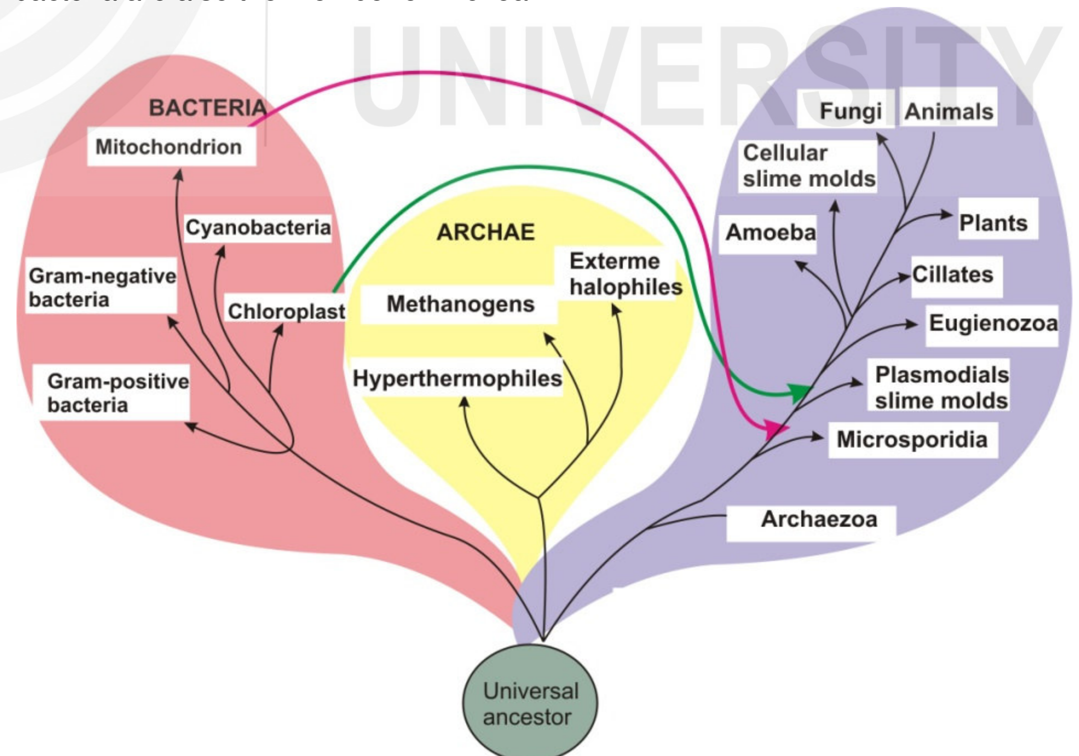


Fig. 2.2: Three domain system as proposed by Carl R. Woese based on rRNA gene sequence. Figure adapted from Benjamin Cummings, an imprint of Addison Wesley Longman, Inc.

A brief account of different characteristics of these three classes in shown in Table 2.2.

Table 2.2: Comparison of Bacteria, Archaea, and Eucarya

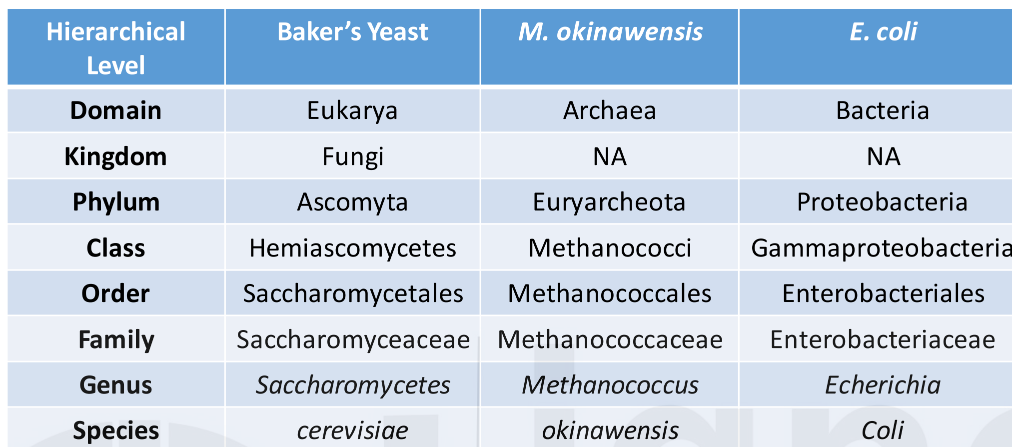
Property	*Bacteria*	*Archaea*	*Eucarya*
Membraene-Enclosed Nucleus with Nucleolus	Absent	Absent	Absent
Complex Internal Membranous Organelles	Absent	Absent	Absent
Cell Wall	Almost always have peptidoglycan containing muramic acid	Variety of types, no muramic acid	No muramic acid
Membrane Lipid	Have ester-linked, straight-chained fatty acids	Have ether-linked, branched aliphatic chains	Have ester-linked, straight-chained fatty acids
Gas Vesicles	Present	Present	Absent
Transfer RNA	Thymine present in most tRNAs *N*-formylmethionine carried by initiator tRNA	No thymine in T or TΨC arm of tRNA Methionine carried by initiator tRNA	Thymine present Methionine carried by initiator tRNA
Polycistronic RNA	Present	Present	Absent
mRNA Introns	Absent	Absent	Present
mRNA Splicing, Capping, and Poly A Tailing	Absent	Absent	Present
Ribosomes			
Size	70S	70S	80S (ctoplasmic ribosomes)
Elongation factor 2	Does not react with diphtheria toxin	Reacts	Reacts
Senstivity to chloramphenicol and kanamycin	Sensitive	Insensitive	Insensitive
Sensitivity to anisomycin	Insensitive	Sensitive	Sensitive

DNA-Dependent RNA Polymerase			
Number of enzymes	One	Several	Three
Structure	Simple subunit pattern (4 subunits)	Complex subunit pattern similar to eukaryotic enzymes (8-12 subunits)	Complex subunit pattern (12-14 subunits)
Rifampicin sensitivity	Sensitive	Insensitive	Insensitive
Polymerase II Type Promoters	Absent	Present	Present
Metabolism			
Similar ATPase	No	Yes	Yes
Methanogenesis	Absent	Present	Absent
Nitrogen fixation	Present	Present	Absent
Chlorophyll-based photosynthesis	Present	Absent	Present[a]
Chemolithotrophy	Present	Present	Absent
[a]Present in chloroplasts (of bacterial origin)			

2.2.2 Taxonomic Ranks

The microorganisms are arranged in different hierarchical or taxonomic level based on the similarity of traits. In a taxonomic hierarchy, the lowest level is considered to be **species**. The species definition of microbes (bacteria in particular) is debatable and different from the eukaryotic species (higher organisms) the latter of which are referred to the members of interbreeding population producing fertile progeny. The bacterial species is defined as a collection of strains that share stable properties and differ significantly from another group of strains. The **strain** is descendant of a single pure microbial culture of species. The strains may further be categorized as **biovar** (variant based on biochemical or biophysical properties), **morphovars** (morphological variant), phagovars (variants with different phage susceptibility) and **serovars** (variants differing in antigenic properties). Often, a **type strain** is used for identification or characterization of the new microbe. The type strain is one of the first and fully characterized strains which can be used as a reference point of the given species. However, the type strain may not necessarily be a representative strain of a species.

Genus lies just above the species in hierarchical level which is a collection of similar species. Similarly, members of similar genera are grouped in a **Family**, similar family in **Order**, the similar order in **Class**, and so on (Fig. 2.3). With the flooding sequence data of rRNA gene and protein-coding genes, the inclusion of **superphylum** just below the kingdom has been recommended recently. There are many sequences which do not match with the current database of known sequences and indicate the possibility of new taxonomic levels in the future.



Hierarchical Level	Baker's Yeast	*M. okinawensis*	*E. coli*
Domain	Eukarya	Archaea	Bacteria
Kingdom	Fungi	NA	NA
Phylum	Ascomyta	Euryarcheota	Proteobacteria
Class	Hemiascomycetes	Methanococci	Gammaproteobacteria
Order	Saccharomycetales	Methanococcales	Enterobacteriales
Family	Saccharomyceaceae	Methanococcaceae	Enterobacteriaceae
Genus	*Saccharomycetes*	*Methanococcus*	*Echerichia*
Species	*cerevisiae*	*okinawensis*	*Coli*

Fig. 2.3: The hierarchy of taxonomy: taxonomic representation of the representative organisms belonging to three domains proposed by Carl R. Woese.

2.2.3 Classification System

Microorganisms, primarily bacteria, can be grouped or assigned to taxa based on different classification systems which include phenotypic features, genotypic character, and/or phylogenetic relationship. Currently, the taxonomic assignments are done employing **polyphasic taxonomy** which employs all three, i.e., phenotypic (phenetic), genotypic, and phylogenetic characteristics. To better understand how these data are useful in taxonomic criteria, we discuss these components individually in brief.

Phenetic Classification: This is one of the most common and old methods to identify and classify organisms. This classification system employs screening of phenotypically observable characteristics such as the morphological structure of the cell, colony morphology, biochemical and physiological traits which can be assayed or observed. This includes analysis of as many as possible traits (more than 50) for robust classification. Methods used in phenetic classification are described in section 2.3.1. Though this method can be used to establish an evolutionary relationship, this is not always the case. For instance, not all flagellated bacteria are evolutionarily related.

Genotypic Classification: As the name suggests, this system is based on similarity in the sequence of evolutionarily conserved genes or whole genomes. Majority of genotypic classification is based on small subunit rRNA gene sequence. The genetic classification is well suited for studying evolutionary relatedness among the organisms.

ignou THE PEOPLE'S UNIVERSITY

Phylogenetic Classification: The phylogenetic classification system compares organisms based on their evolutionary relationship. The term ‘phylogeny’ is originated from Latin word ‘*phylon,’* i.e., tribe or race, and ‘*genesis,’* i.e., generation or origin. Most of the evolutionary studies are based on the fossil record which is not available for the majority of microbes. Therefore, the phylogenetic classification could be employed only after the development of microbial identification based on 16S rRNA gene sequence.

2.3 CHARACTERIZATION

Another component of taxonomy is **characterization** which means a description of detailed features of given organism. The comprehensive characterization can enable the accurate **identification** of an organism. Several approaches are there for characterizing the microbes. All the approaches can be broadly divided into phenotypic and molecular characterization. One or more of these approaches can be employed to characterize the microbe. These methods differ in their resolution. Greater the resolution, better identification on different taxonomic rank. The method which can differentiate microbe at strain level will be considered to have the highest resolution. We shall have a brief description of various methods used to characterize microbe at phenotypic or molecular level. Majority of the descriptions are focused to prokaryotes, and the description specific to eukaryotic microbe will be mentioned wherever necessary.

2.3.1 Phenotypic or Classical Characterization

This includes methods which analyze morphological, biochemical, physiological, and ecological features of an organism. These methods form the basis of phenetic classification. The phenotypic characteristics are useful and sometimes essential for assigning to the different taxonomic levels such as phylum, class, order, family, or even species. Some of the important phenotypical properties which are analyzed for characterization are given below. The resolution of phenotypic methods is shown in Fig. 2.4.

Morphology and Growth behavior: The primary information about the test microbe can be obtained from morphological data which is based on cell or colony morphology and growth behavior. Morphological information is valuable as it is stable and guided by genetic functions.

Growth conditions: Bacteria are analyzed based on their requirement of the oxygenic or anoxygenic condition, CO_2, and the composition of media, upon which organisms can grow.

Colony and cell morphology: Information related to colony morphology such as colour, edge, texture, elevation, opacity, motility on a solid surface, and production of extra-colonial **pigments** are used as primary visual data. However, there is a limited number of characteristics which can be analyzed with unaided eyes. Further cellular structure, the formation of **prosthecae**, **branching**, **endospores**, presence and insertion of **flagellae**, etc., are determined by light and electron microscopy, which provide valuable data for morphological description.

Staining characteristics of cells: Staining methods such as Gram staining, acid-fast staining, etc. are used to differentiate prokaryotes based on differential cell wall or membrane composition.

Phenotypic Methods	Family	Genus	Species	Sub species	Strain
Serotyping					
Raman spectroscopy					
SDS-PAGE					
MALDI-ToF					
Chemotaxonomy (e.g. FAME profiling)					
Phenotyping (growth, morpholoigy, API)					
BIOLOG, Omnilog, Vitek					
Genotypic Methods					
Genome sequencing					
16S rRNA, rpoB gene sequencing					
Mol%GC					
DNA-DNA Hybridization					
MLSA/MLST					
Whole cell protein profiling					
DNA fingerprinting (BOX-, ERIC-, REP-PCR					

Fig. 2.4: Phenotypic and genotypic methods for the characterization of prokaryotes and the approximate respective taxonomic levels of resolution. Light grey region denotes limited resolution of given method at given taxonomic level. Some of the methods such as serotyping, SDS-PAGE, and whole cell protein profiling is not discussed in the text but they are used for typing and identification of prokaryotes. MLSA and MLST are expanded as Multilocus sequence alignment and Multilocus sequence typing respectively.

ignou THE PEOPLE'S UNIVERSITY

Chemical characterization ("chemotaxonomy"): The difference in a structural component of the cell wall, cell membrane or cytoplasm can provide useful information for identification and classification. Differences in peptidoglycan, teichoic acids, mycolic acids, fatty acids, polar lipids, respiratory lipoquinones, pigments and polyamines (Tindall et al., 2010) are used in microbial identification. Such cellular features offer systematic identification often shows the evolutionary relationship. However, it may differ regarding resolution. For instance, **FAME (fatty acid methyl ester) analysis** is one of the most common methods used which is used to identify bacteria based on type and proportion of fatty acid present in the cell membrane. The membrane lipid of bacteria varies in species to species, and they differ in chain length, presence or absence of a double bond, rings, branched rings, etc. In this method, the fatty acids are extracted from bacteria grown under standard conditions, chemically treated to generate methyl derivative, and analyzed by gas chromatography. The fatty acid profile is matched with known database and identified. FAME profiling is one of the most common methods. However, its application is limited with the fact that the fatty acid profile changes with environmental and growth factors (Fig. 2.5).

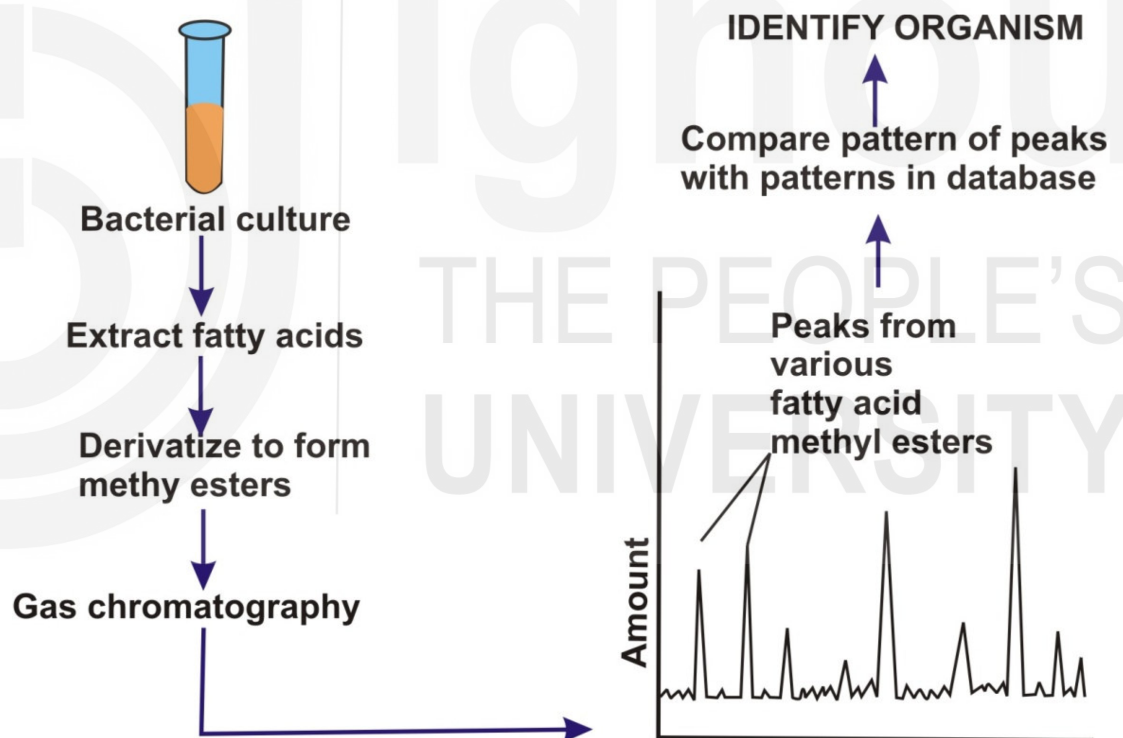


Fig. 2.5: Workflow of microbial identification by FAME analysis. Figure adapted from Brock Biology of Microorganism 13th edition, Pearson.

With the development of **mass spectrometric** methods, fast and accurate identification is possible by analyzing microbial molecules. A spectrometric method known as **MALDI-ToF** (Matrix-assisted laser desorption/ionization-time of flight) identifies a microbe by analyzing masses of microbial proteins after comparing with the protein database available (Fig. 2.6). In this method, microbes are grown in specific/standard conditions, transferred to a sample target, dried, and analyzed. This method is applied even to those organisms which are difficult to culture.

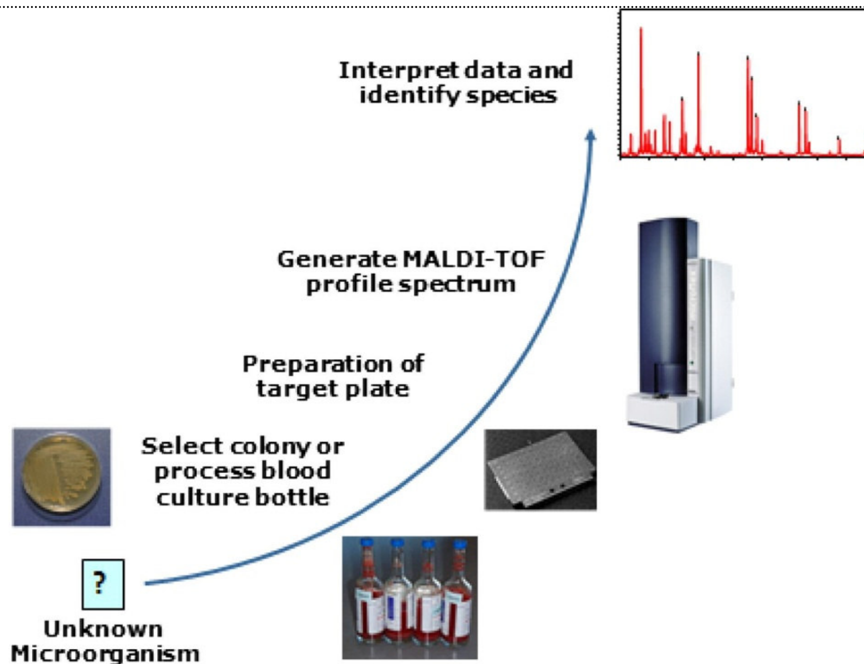


Fig. 2.6: Schematics showing methodogy for bacterial identification using MALDI-ToF Mass spectrometry. Figure adapted from web source.

Physiological and Metabolic characteristics

Physiological testing is based on the growth properties, the ability to utilize certain substrates or to show features that provide information about the basic metabolic activities of prokaryotes. The physiological and metabolic properties provide useful information which is directly related to the nature and activity of microbial proteins. There are many commercial products available that claims reliable identification of microorganisms which includes BiOLOG's carbon utilization test. Conventionally, BiOLOG plates contain 96 wells each containing different carbon sources and a redox dye. One or more wells do not have any carbon source and act as a control. After inoculation and incubation of test microbe under standard conditions, the wells are analyzed for the development of pink colour which indicates the ability to utilize given carbon source (Fig. 2.7). The profile is then matched with the microbial database, and the bacterium is identified using **numerical taxonomic** approach (discussed in section 2.6). In this system, different plates are used for Gram-positive and Gram-negative bacteria.

More recently, BiOLOG has introduced highly automated system with updated microbial database called Omnilog which enables identification of microbes including bacteria, yeast, and filamentous fungi based on ability to metabolize all major classes of biochemicals, in addition to determining other important physiological properties such as pH, salt, and lactic acid tolerance, reducing power, and chemical sensitivity. This system allows profiling of more than 50 different microbial sample at a time and provide data in a minute. Similar to BiOLOG, there are other fast microbial identification system such as **Vitek** (manufactured by BIOMÉRIEUX) and **API (Analytical Profile Index)** system which exploits physiological properties to identify microbe.

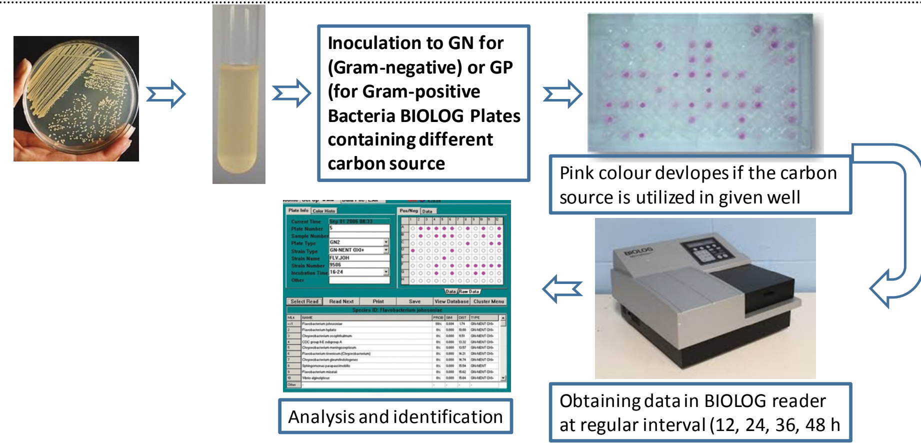


Fig. 2.7: Schematics showing identification of bacteria using BIOLOG carbon utilization test. BIOLOG is commercially supplied.

Ecological characteristics

The ability of a microorganism to colonize certain ecological conditions, to establish symbiotic relationship with specific host, or having requirement of specific pH, temperature, or oxygen can have taxonomic value.

> ***Why should molecular methods be used for identification?***
>
> The phenotypic methods can be used for identification but most of these techniques are time consuming and are affected by changes in several environmental factors. The molecular techniques such as sequence based, gel based, and protein based systems are fast and have high specificity and less chance of error.

2.3.2 Molecular Characterization

Molecular characterization is usually referred to the description or analysis of an organism based on nucleic acid (DNA or RNA) or protein information. Due to lack of fossil record, phylogenetic analysis was not possible until the methods for nucleic acid sequence analysis developed. The recent advances have enabled us to sequence bacterial DNA of even non- culturable organisms, i.e. without culturing them in the laboratory which has divulged the great diversity of microorganisms. There are several methods which are used for molecular characterization or typing or strains. The term **'typing'** is used for designating a strain to the certain specific pattern of DNA profile. However, the resolution of each method varies and range from domain to species or even strain level identification. These methods include: (i) determination of nucleic acid base composition, (ii) nucleic acid hybridization, (iii) nucleic acid sequencing and DNA fingerprinting. The resolution of various genotypic methods shown in Fig. 2.4.

Nucleic acid base composition

One of the many ways of taxonomic determination involves comparison of bases in genome by determining **GC (Guanine and Cytosine) content**. The **mol%GC** can be calculated using following formula:

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

This can be done directly from the sequence data of genome. For bacteria whose genome sequence is not available, the base composition can be estimated from DNA **melting temperature (T_m)** curve. This is based on the fact that the genome having high G+C content will melt at higher temperature as three hydrogen bonds are involved in making pair of G with C whereas only two bonds are there between A (Adenine) and T (Thymine). **Melting temperature** (T_m) is the temperature at which 50% of DNA is denatured. For instance, Tm of *Mycoplasma hominis,* having 29% G+C, is 65 ˚C whereas it is 85 ˚C for *Micrococcus luteus* which has 79% G+C.

For DNA melting analysis, the DNA is slowly heated at increasing temperature. As the temperature increases, the bonds between the base pair start breaking until the complete DNA denatures (strands are separated from one another). While doing this, the absorption of DNA is measured spectrophotometrically at 260 nm (A_{260}). The absorbance increases with the separation of strands and reaches to plateau when the denaturation completes yielding single-stranded DNA (ssDNA). The resultant melting curve is used for estimation of T_m. The mid-point of rising curves gives T_m (Fig. 2.8).

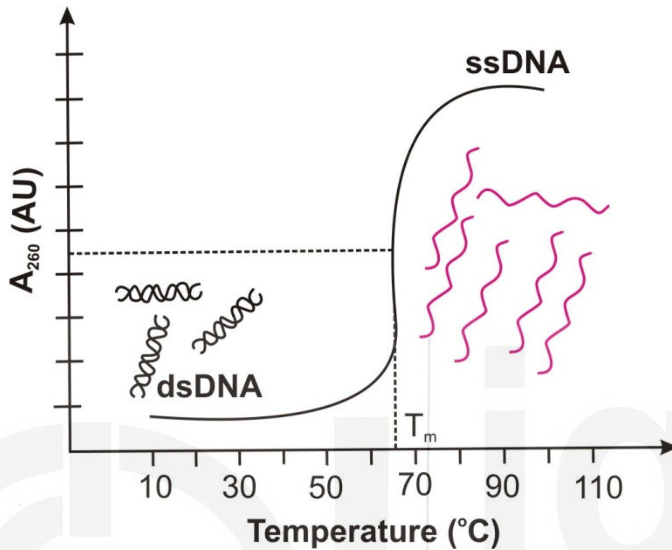


Fig. 2.8: A melting curve of DNA. A Tm is indicated at mid-point.

The range of GC % in prokaryotes (25 and 80%) is very wide as compared to the higher eukaryotic organisms (30 to 50). This %G+C content within a species remain constant and varies very little within the genus. It is estimated that the two organisms having more than 10% difference in G+C content are not closely related. However, it may not be applied that if the organisms have similar G+C content will be closely related to the sequence may vary even if the composition is similar.

Table 2.2: Range of G + C content in various microbial groups

Microbial Groups	% G + C
Eubacteria (Bacteria)	25 to 80
Archebacteria	27 to 62
Fungi	22 to 62
Protozoa	21 to 65
Algae	37 to 68

Nucleic acid hybridization

One of the oldest and direct molecular methods for comparing a pair of microbes is **DNA-DNA hybridization** (**DDH**). In this method, the DNAs are heated to get single-stranded DNA (ssDNA). Then the ssDNAs of two microbes are allowed to cool and hold at temperature 25 ˚C below the T_m. The extent of annealing two strands depends on the similarity between the

sequences of two organisms being compared. By convention, the organisms showing 70% similarity based on the extent of hybridization are considered a member of same species. A schematic of DDH is shown in Fig. 2.9.

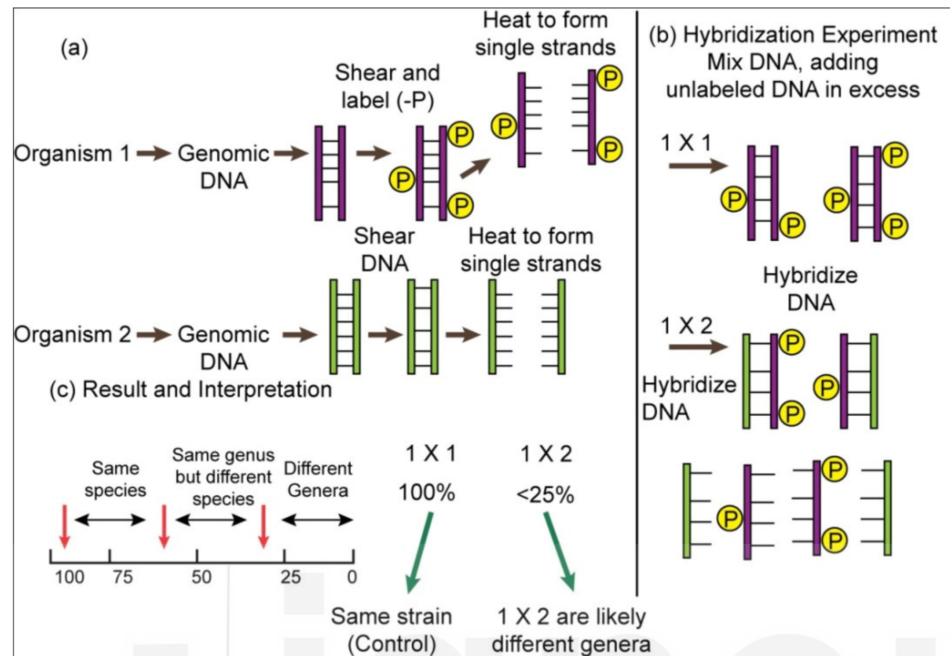


Fig. 2.9: Methodology for DNA-DNA hybridization to identify an organism. (A) Genomic DNA of test isolate is labeled (shown here as radioactive phosphate in the DNA of Organism 1). (B) Excess unlabeled DNA is added to prevent labeled DNA from reannealing with itself. Following hybridization, hybridized DNA is separated from unhybridized DNA. Radioactivity in the hybridized DNA is measured. (C) Radioactivity in the control (Organism 1 DNA hybridizing to itself) is taken as the 100% hybridization value. The Figure and legend are adapted form Brock Biology of Microorganisms, 13th edition with slight modification.

Though the DDH is still used for confirming the taxonomic affiliation of a microbe, it is cumbersome to use and sometimes crude when compared to genomic data. The best alternative method of DDH is the estimation of **average nucleotide identity (ANI).** In ANI, pair-wise alignments of genome sequences of the two organisms are done. In this method information of whole genome sequence is not required. 20% coverage of genome sequence is sufficient for ANI. In general, an ANI value of 95 to 96% indicates identification at species level.

Nucleic acid sequencing-based characterization

A certain stretch of the genome or specific genes are the potential candidate for identification. These sequences can also be used for analysis of phylogeny. The desirable properties of a gene or sequence required for being used as identification marker are as follows:

i) It should be universal, i.e., present in all the organisms.

ii) The sequence should be long enough for identification with high resolution.

iii) The sequence should be conserved among the organisms so that there will be measurable relationship even between distantly related organisms.

iv) It should have some variable region which is enough for distinguishing the organisms even if they are closely related, and

v) which are not transferred to another organism through horizontal gene transfer.

SSU rRNA sequence analysis

Considering properties mentioned above, the rRNA genes appeared to be the most appropriate candidate for microbial identification. Since **rRNA gene** is one of the most conserved genes, present in all organisms, and long enough, it is routinely used for microbial identification. The rRNA is constituent of ribosome which has two units namely **small subunit (SSU)** and **large subunit (LSU).** Three types of rRNA namely 5S (120bp), 16S (1542 bp), and 23S (3200 bp) are present in prokaryotic ribosomes out of which **16S rRNA**, a component of SSU, is the most frequently used for identification due to its most conserved nature and moderate size for analysis. The 16S rRNA gene (also called 16S rDNA) consists of conserved sequences flanking several variable (referred as **V1 to V9**) region (Fig. 2.10). These **variable regions** enable comparison between closely related microbes while the stable (**conserved**) regions allow the comparison of distantly related microorganisms. For identification, usually full length 16S rRNA gene is amplified by **polymerase chain reaction (PCR)** using a pair of primers designed based on conserved region, sequenced in DNA sequencer, and analyzed using the appropriate computational tool. However, sometimes the partial sequence of 16S rRNA gene can also be analyzed for tentative identification. For identification of eukaryotic microorganisms, **18S rRNA** (component of SSU) gene is used for identification. For identification of fungi at the species level, **ITS (inter-transcribed spacer sequence),** which is located between rRNA genes, are used. The arrangement of fungal rRNA genes is shown in Fig. 2.11. Once the sequence data is obtained, the sequence is searched for the best match on various databases such as NCBI (**National Center for Biotechnology Information**) or **Ribosomal Database project** (RDP, https://rdp.cme.msu.edu/) using BLAST (Basic Local Alignment Sequence Tool).

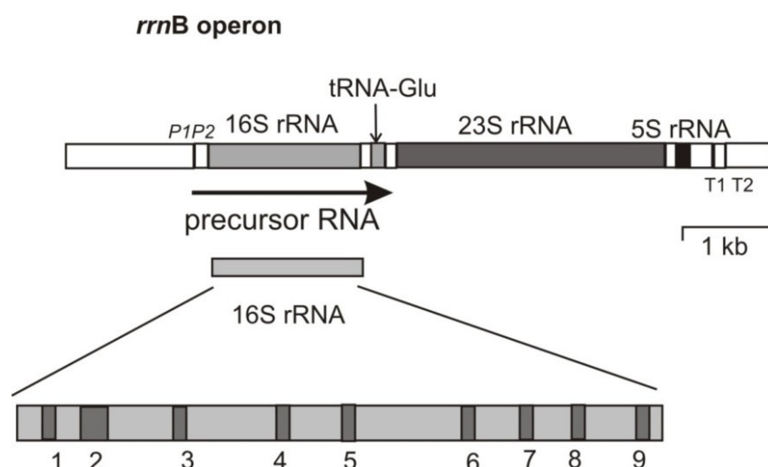


Fig. 2.10: Schematics showing arrangement of rRNA genes of *E. coli*. The lower panel shows various conserved (light grey) and variable regions (dark grey).

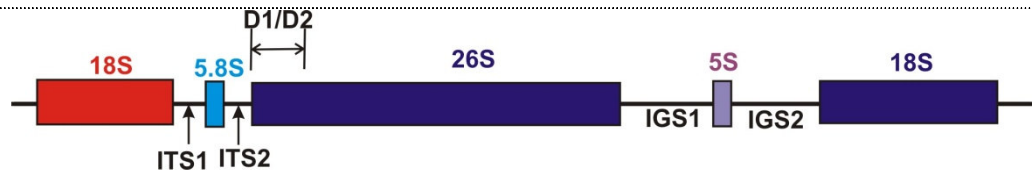


Fig. 2.11: A schematic showing rRNA gene arrangement in fungi. ITS1 and ITS2 are inter-transcribed spacer sequence. D1 and D2 refers to domain 1 and 2 whereas IGS (Inter-genic spacer sequence).

The resolution of 16S rRNA sequence-based analysis is at the genus level. However, it can help identify at species level as well. By convention, but not necessarily, the organism showing more than 97% similarity in rRNA gene sequence can be considered a member of same species. The major advantage of 16S rRNA gene sequence analysis is rapid and accurate identification. Further, it can also allow identification of those bacteria which cannot be cultivated in laboratories. This property has enabled us to explore the diversity of microorganism of given community in a culture-independent manner. Any uncultivated microorganisms which are identified solely on its nucleic acid sequence are called a **phylotype**.

Sequence analysis of other genes than 16S rRNA genes

16S rRNA gene sequence has undoubtedly revolutionized the bacterial taxonomy. However, there are certain other house-keeping genes (whose expression is necessarily required for cell growth) whose sequences are also conserved and can be used for taxonomic affiliation of prokaryotes to supplement the data of 16S rRNA sequence data. The name of some of these genes is enlisted in Table 3. A gene ***rpoB*** encoding β-subunit of RNA polymerase is widely used for bacterial identification. It exists as single copy gene and shows much more discriminatory power than 16S rRNA gene. For identifying fungi (eukaryotic microbes), genes encoding RNA polymerase (*RPB1* and *RPB2*), translation elongation factor 1-ɑ (*tef1*) and **β-tubulin** (*benA, tubC*) are preferred for identification. The ***benA*** encodes beta 1, and beta 2 and ***tubC*** encodes beta 3 protein of β-tubulin. Among the alternative genes, β-tubulin are most commonly used. However, due to lack of an appropriate database, these genes are not solely used for identification.

Table 2.3: Other genes than 16S rRNA genes which can be used for bacterial identification.

Genes	Function
*rpoA, **rpoB**, rpoC and rpoD*	Encodes for RNA polymerase a key enzyme for transcription process
gyrA, and gyrB	Encodes gyrase, a topoisomerase
dnaJ	It is a molecular chaperone
ppk1	Codes for polyphosphate Kinase 1

Multi-locus sequence alignment (MLSA)

For robust identification of a bacterium at species or even strain level, it is better to rely on more than one gene. **MLSA** employs an approach of sequencing and analyzing 5 to 7 **house-keeping genes** which are relatively

more conserved and are usually not horizontally transferred. The selection of genes may differ based on group of bacteria to be analyzed. MLSA is derived from the similar technique multi-locus sequence typing (MLST) which was used to type the strain based on analysis of partial sequence of many house-keeping genes. The MLSA is also used for phylogenetic analysis with more confidence.

Oligonucleotide signature sequence

These are short conserved sequence that is identified after comparing thousands of rRNA genes of organisms and are specific to particular phylogenetic group. This is useful for grouping or identifying microorganisms at domain level.

Indels

Indels (**in**sertion/**del**etion) are specific length of sequences which are inserted or deleted in many genes and are specific to certain phylum. Indels are particularly useful in phylogenetic analysis when they are flanked by conserved region. The indels in housekeeping genes are less prone to horizontal transfer and can be used for phylogenetic analysis.

DNA fingerprinting

The bacterial strains can be identified or typed by sequencing-independent approaches that analyze DNA profiling which can either be generated by restriction digestion or by amplification of repetitive DNA present in the genome. Some of the DNA or genomic fingerprinting techniques are described below.

rRNA gene-based fingerprinting

There are two methods by which profiling of genome based on rRNA gene is done. One method called **RFLP** (**R**estriction **f**ragment **l**ength **p**olymorphism) employs digestion of PCR-amplified rRNA gene with restriction enzymes. The digested product is then analyzed by gel electrophoresis. The differential pattern of DNA fragments is generated due to variable sequence of rRNA which can lead to identification or typing of a strain using type strain as a reference (Fig. 2.12a). This method is also called **ARDRA** (**A**mplified **rD**NA **r**estriction **a**nalysis).

Another method is PCR-independent as it is based on hybridization approach. In this method, a genome is digested with one or more restriction enzymes, electrophoretically separated on agarose gel, transferred on nylon membrane, and hybridized with rRNA gene-specific molecular **probe** (a labeled oligonucleotide). After hybridization, the developed image of DNA profile is analyzed. This method is called **ribotyping** (Fig. 2.12b).

Repetitive sequence (Rep)-PCR fingerprinting

Bacterial genomes possess multiple copies of **repetitive DNAs** dispersed across the genome. They are usually located in the intergenic region. Based on the sequences, three families of repetitive DNAs namely **BOX, ERIC** (**E**nterobacter **R**epetitive **I**ntergenic **C**onsensus), and **REP** (**R**epetitive **E**xtra-

ignou THE PEOPLE'S UNIVERSITY

genic **P**alindrome) have been recognized and used for typing and identification. The length of BOX, ERIC, and Rep is 154, 124-127, and 35-40 bp respectively. The corresponding protocol is referred to BOX-PCR, ERIC-PCR, and REP-PCR respectively. The length and sequence of these repetitive DNAs are same across the majority of Gram-positive and negative bacteria. However, they differ in their location and number of repeats which form the basis of variation among the bacteria. In rep-PCR approach, primers specific to the repetitive sequences are used to amplify DNA fragments between repetitive elements. The resulting amplified products are separated by gel or capillary electrophoresis. The given profile is then analyzed using a computer program. The resolution of this method of identification is at the species level.

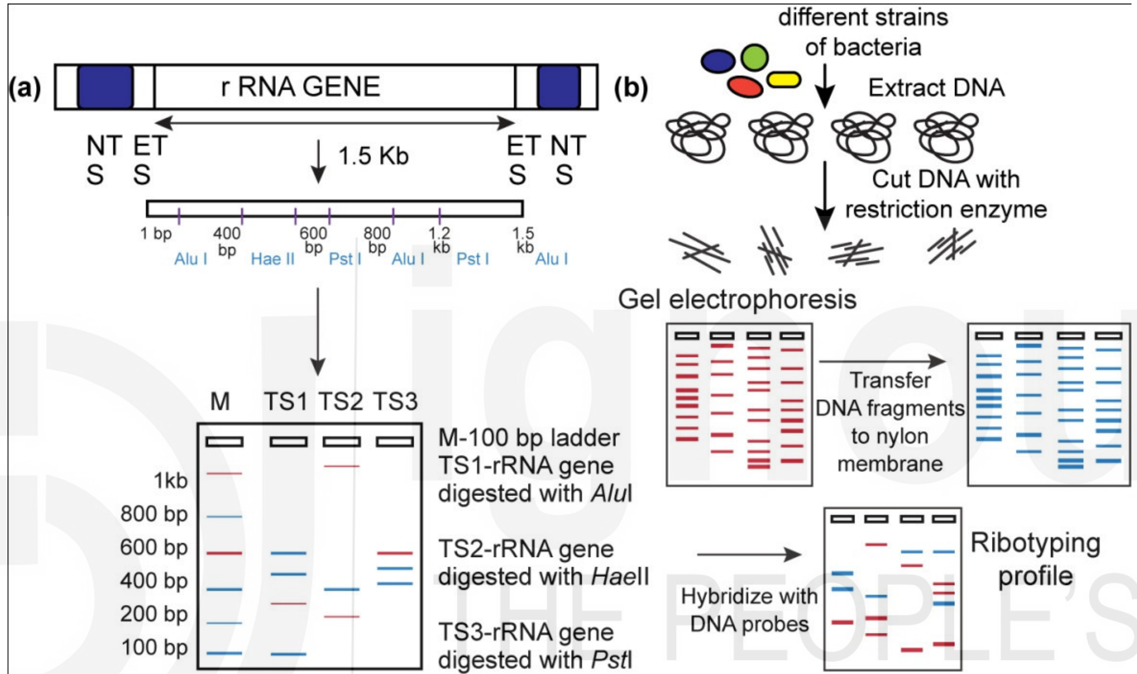


Fig. 2.12: Fingerprinting based on rRNA gene. Methodology for (A) RFLP of 16S rRNA or ARDRA, (B) Ribotyping.

2.4 NOMENCLATURE

Nomenclature is the assigning names to the taxonomic units which have been characterized and classified. Like in higher organisms, the microorganisms are also named following **binomial nomenclature system** proposed by Linnaeus. Accordingly, every organism is assigned two names in which the first indicates *genus* and the second, *species*. The names are written in italics or are underlined as specified by nomenclature rules. The initial letter of genus name is capitalized, and that of the species name is in small letter. For example, in *Escherichia coli, Escherichia* is a genus and *coli* is species. Sometimes the species name is followed by subspecies name. For example, *Salmonella enterica* serovar Typhimurium. Here Typhimurium is a serovar, an epithet at the subspecies level. There may be a different consideration for the naming of the species. It can be based on the location where it was isolated from, the disease it causes, the host it colonizes, etc. For example, *Streptococcus pneumoniae* (Streptococcus: chain of round shaped bacterium that causes **pneumonia**); *Azospirillum brasilense* (Azo-nitrogen, spirillum- spiral in shape, **brasil**iense, i.e., in Brazil), etc.

The naming of an organism follows specific and strict guidelines. These guidelines or rules are published by a specific organization which reviews the nomenclature proposed for new organisms. Rules for assigning names for protozoa and parasitic worms are published in **International Code of Zoological Nomenclature**, whereas it is **International Code of Botanical Nomenclature** for assigning names to algae and fungi. Similarly, **International Committee on Systematics of Prokaryotes** publishes rules for naming prokaryotes in **Bacteriological Code**. Bacteriological Code publishes the name of newly discovered or characterized (both genotypically and phenotypically) bacteria after its review and publication in *International Journal of Systematic and Evolutionary Microbiology* (IJSEM). Finally, the bacterium is listed and described in *Bergey's Manual of Systematic Bacteriology.* Microorganisms that have been sufficiently characterized to assign provisionally at genus and species level but have not been cultivated in pure culture are preceded by the term ***Candidatus*** which means candidate. For instance, a photosynthetic bacterium belonging to phylum '*Acidobacterium*' which can only be grown as coculture with another isolate, has been named as *Candidatus* Chloracidobacterium thermophilum (here genus and species are not italicized). Recently, "**The All-Species Living Tree**" Project (LTP) (https://www.arb-silva.de/projects/living-tree/) has been started with the partnership of five international agencies together with the journal Systematic and Applied Microbiology (SAM). This project provides a valuable resource, particularly for microbial taxonomists. The project aimed to reconstruct separate and curated 16S and 23S rRNA datasets and trees spanning all sequenced type strains of the hitherto classified species of Archaea and Bacteria.

With the development of new and sensitive molecular and metabolic tools, the organisms can be renamed and assigned to new taxonomic ranks. Sometimes a genus can be divided into two new genera, or two genera can be merged to one based on the differences of characteristics. For example, genera *Diplococcus* and *Streptococcus* were merged to one whereas a genus *Enterococcus* was separated from the *Streptococcus.* Therefore, "*Streptococcus faecalis*" and "*Streptococcus faecium*" were renamed as *Enterococcus faecalis* and *Enterococcus faecium* respectively. This example also indicates that the name of species is more stable than the name of genera as the species name did not change even after the formation of new genera.

2.5 BERGEY'S MANUAL OF SYSTEMATIC BACTERIOLOGY

Bergey's manual of Determinative Bacteriology was first published as a guide to identify different species of bacteria by **Professor Bergey** and his collaborators in 1923. It described the classification system for bacteria and provided information which could be used for bacterial identification. The last edition (9th) of this book was published in 1994. In 1984, the first edition of **Bergey's manual of systematic bacteria** with its four volumes published. It had descriptions of all the bacterial and archaeal species identified till then. The traits or charactersmentioned in this book for the classification of bacteria

was phenetic. Further in 2001, the second edition of Bergey's manual of systematic bacteria with its five volumes published. Each volume covered a specific group of organisms. The classification system adopted for a description of the taxons were based on genotypic and phylogenetic approaches. The brief outline of the different volumes of this book is given below.

Volume 1: It deals with all Archaea and deeply branching and phototrophic bacteria. It includes a description of two phyla of Archaea (**Crenarachaeota** and **Euryarchaeota**), nine of deeply branching and two of phototrophic bacteria.

Volume 2: It is divided into parts A and B which describe a single phylum **proteobacteria**. It represents the largest group and contains most of the Gram-negative bacteria which include pathogens (*Escherichia, Salmonella, Vibrio, Helicobacter, Yersinia, Legionellales*, and many other notable genera), many free-living and nitrogen-fixing bacteria.

Volume 3: This volume describes a single phylum **Firmicutes** which includes majorly Gram-positive bacteria having low mol% G + C content. They produce endospores, are resistant to desiccation, and survive extreme conditions.

Volume 4: This volume covers the description of 12 different phyla namely **Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae,** and **Planctomycetes**.

Volume 5: It covers a single phylum **Actinobacteria** which includes Gram-positive bacteria having high mol% G + C content. Many actinobacteria are agriculturally and medically important as actinomycetes are known to decompose organic material and produce antibiotics.

2.6 NUMERICAL TAXONOMY

Numerical taxonomy is a classification system which employs numerical algorithms to quantify and compare various traits of microorganisms. Sneath and Sokal define numerical taxonomy as "the grouping by numerical methods of taxonomic units into taxa by their character states." Each trait is termed as **OTU (operational taxonomic unit)** which is used as an input in numerics. Thus, the information about the properties (OTUs) of organisms is converted into a form suitable for numerical analysis and then compared using a computer. The result of similarity of characters between the microbes is used for grouping. Each character or OTU is given equal weightage. For instance, if the oxidase activity is an OTU, presence, and absence of this activity will be assigned as 1 and 0 respectively. Here oxidase activity is an OTU or character to be compared. Similarly, as many characters as possible (more than 50 or even upto thousands) are considered for comparison and robust classification. It is best to include many different kinds of data based on morphological, biochemical, and physiological characters.

Following data input obtained from different tests, an **association coefficient** is calculated for each pair of organisms. The association coefficient is a function that measures the agreement between characters possessed by two organisms. There are several methods for calculating association coefficient out of which two most common are **simple matching coefficient (S_{SM})**, and **Jaccard coefficient (S_J)**. The S_{SM} is the proportion of characters that match regardless of whether the attribute is present or absent whereas S_J is calculated by ignoring those characters which are absent in both organisms. Both coefficients are represented between 0.0 and 1.0 where 0.0 indicates no match and 1.0 represents 100% match. The S_{SM} and S_J are calculated as given below:

$$S_{SM} = a + d / a + b + c + d$$

$$S_J = a / a + b + c$$

Where **a**= number of characters (OTU) present (shown as 1) in both organisms; **b** and **c**= number of characters differing between the two organisms (0,1 or 1,0, i.e., absent in one organism and present in other or vice versa); **d**= number of characters absent in both (shown as 0)

The values of association coefficient measured for each pair of organisms are then used to construct similarity matrix which forms the basis of clustering the organisms (Fig. 2.13). The clusters are shown in a tree-like representation called as **dendrogram** using suitable computer software. The organisms having great similarity are grouped and separated from dissimilar organisms in the dendrogram. Such groups are called '**phenon**.' Phenons formed at 70or more percent are treated as members of same species.

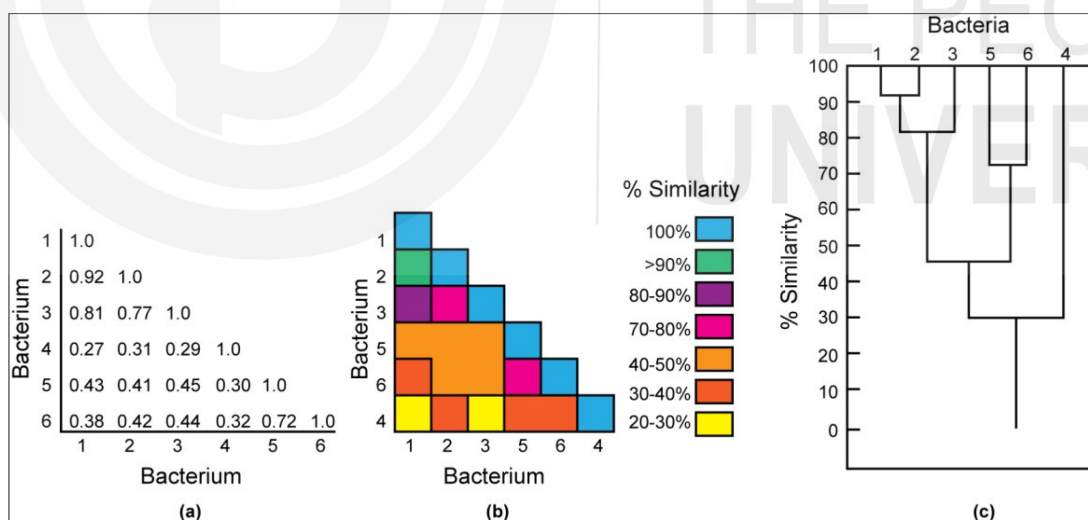


Fig. 2.13: Clustering analysis of 6 bacterial isolates based on the similarity matrix obtained by calculating simple matching coefficient. (a) The similarity matrix is comparing a pair of bacteria. For instance, strain 1 shows the coefficient of 0.92 with strain 2 which indicates 92% similarity. The coefficient value 1 means 100% similarity whereas 0 shows no similarity at all. (b) Based on the % similarity, the bacterial strains are rearranged and grouped with the different colour given based on the range of % similarity. (c) A dendrogram constructed based on the grouping made in 'b,' For instance, strain 1 and 2 shows the highest similarity with 92% (phenon-92) thus shown as the closest from one another. The Figure is adopted from Prescott's Microbiology, 7th Edition, McGraw Hill Publication.

SAQ 1

Tick [√] mark the correct statement:

a) Species concept is applied same to all organism including bacteria [True/False]

b) A bacterium can be identified based on its ability to used different carbon sources. [True/False]

c) Bacteria are placed under kingdom Protista as per Whittaker's five kingdom concept. [True/False]

d) 16S rRNA sequence analysis can be used for identification of protozoan parasites. [True/False]

SAQ 2

Fill in the blanks with appropriate words:

a) Binomial nomenclature was introduced by……………………..

b) The organism causing malaria belongs to domain …………………

c) A descendant of single pure microbial culture is called ………..

d) The FAME profiling used for identification of microbe analyzes differences in the type of ………………present.

e) A tree-like presentation which shows the differences among microbial isolates is called………………

f) 16S and 18S rRNA used for microbial identification is component of ………..subunit of the ribosome.

g) …………..gene sequence analysis can be used as an alternative rRNA gene cluster for identification for fungi.

h) Families of repetitive DNA namely………, …………., and …………..are present in bacteria which can be used for DNA fingerprinting analysis.

2.7 SUMMARY

- Microbial taxonomy or systematics is defined as the study of the diversity of microorganisms with the aim of organizing and prioritizing in an orderly manner.
- It constitutes three separate but integrated components namely classification, characterization/identification, and nomenclature.
- The first classification system which included microorganisms was proposed by Robert H. Whittaker who introduced five kingdom concept and placed microorganisms under kingdom Monera (bacteria, cyanobacteria, and mycoplasma) Protista (algae and protozoa), and Fungi.

- The most recent classification proposed by Carl Woese introduced three domain concepts, i.e., Bacteria, Archaea, and Eucarya based on the differences in rRNA gene sequence.
- The organisms are arranged in various taxonomic ranks which include domain, superphylum, phylum, class, order, genus, and species (top to bottom).
- The microorganisms can be classified as phenetic, genotypic, and phylogenetic. However, the best approach applied for classification is polyphasic taxonomy which employs all the classification approaches mentioned above.
- The microorganisms can be characterized employing observable traits (phenotypic) and/or genotypic traits which lead to the identification of microorganisms.
- The phenotypic methods are based on the morphology of cell or colony, physiological features such as substrate utilization, metabolic characteristics, and growth environment. On the other hand, the genotypic characterization is based on genome sequence, nucleotide base composition (mol%G + C), DNA-DNA hybridization, DNA fingerprinting (ARDRA, Ribotyping, Rep-PCR, etc.), and rRNA or other protein-coding gene sequence analysis.
- For measuring relatedness of microorganisms, the data of different traits are converted into a matrix which gives a value calculated using numerical taxonomic approach.
- The nomenclature of microorganisms follows binomial nomenclature system which indicates the name of genus and species. The guidelines for naming microorganisms are published in International codes for nomenclature after review of the proposed name.
- The description of various cultivated bacteria is provided in *Bergey's manual of systematic bacteriology.*

2.8 TERIMINAL QUESTIONS

1. What is taxonomy and what are its components?
2. How can the cell membrane component used for bacterial identification?
3. What is the principle of BIOLOG microbial identification system?
4. Diagrammatically represent rRNA gene cluster of fungi which can be used for fungal identification.
5. Why is rRNA gene sequence analysis the most preferred approach for microbial identification?
6. What is the principle of bacterial DNA fingerprinting?

7. What do you mean by numerical taxonomy and how are simple matching and Jaccard coefficient calculated?

8. What is molecular charaterization? Describe few major methods which can be used under this approach.

9. Describe five kingdom concept of Whittaker.

10. Describe phenotypic methods which can be used for identification of microbes.

11. Describe sequencing-based approaches for microbial identification.

12. Describe DNA-DNA hybridization method and its best alternative for bacterial identification.

13. What is the significance of indels in microbial identification?

2.9 ANSWERS

Self-Assessment Questions

1. a) False

   b) True

   c) False

   d) False

2. a) Carl Linnaeus

   b) Eucarya

   c) Strain

   d) Fatty acid

   e) Dendrogram

   f) Small

   g) β-tubulin

   h) BOX, ERIC, and REP

Terminal Questions

1. Microbial taxonomy may be defined as the study of the diversity of microorganisms with the aim of organizing and prioritizing in an orderly manner. It is also known as systematic. Three components of taxonomy are classification, characterization or identification, and nomenclature.

2. FAME (fatty acid methyl ester) analysis is used to identify bacteria based on type and proportion of fatty acid present in the cell membrane. The membrane lipids of bacteria vary from species to species, and they differ

in chain length, presence or absence of a double bond, rings, branched rings, etc. In this method, the fatty acids are extracted from bacteria grown under standard conditions, chemically treated to generate methyl derivative, and analyzed by gas chromatography. The fatty acid profile is matched with known database and identified

3. BIOLOG has introduced highly automated system with updated microbial database called Omnilog which enables identification of microbes including bacteria, yeast, and filamentous fungi based on ability to metabolize all major classes of biochemicals, in addition to determining other important physiological properties such as pH, salt, and lactic acid tolerance, reducing power, and chemical sensitivity. This system allows profiling of more than 50 different microbial sample at a time and provide data in a minute.

4. Please refer to Fig 2.11

5. rRNA gene is one of the most conserved genes, present in all organisms, and long enough, it is routinely used for microbial identification.

6. The bacterial strains can be identified or typed by sequencing-independent approaches that analyze DNA profiling which can either be generated by restriction digestion or by amplification of repetitive DNA present in the genome. The digested product is then analyzed by gel electrophoresis. The differential pattern of DNA fragments is generated due to variable sequence which can lead to identification or typing of a strain using type strain as a reference.

7. Numerical taxonomy is a classification system which employs numerical algorithms to quantify and compare various traits of microorganisms. For more details, please refer to the section 2.6.

8. Molecular characterization is usually referred to the description or analysis of an organism based on nucleic acid (DNA or RNA) or protein information. Methods include: (i) determination of nucleic acid base composition, (ii) nucleic acid hybridization, (iii) nucleic acid sequencing and DNA fingerprinting.

9. Robert H. Whittaker propose five kingdom concept. The five kingdoms namely Monera, Protista, Plantae, Fungi, and Animalia, were propounded based on (i) cell type, i.e., prokaryotic or eukaryotic, (ii) level of organization, i.e., solitary and colonial unicellular or multicellular organization, and (iii) type of nutrition.

10. Serotyping, RAMAN spectroscopy, SDS PAGE, MALDI-Tof, chemotaxonomy are some of the phenotypic methods which can be used for identification of microbes

11. A certain stretch of the genome or specific genes are the potential candidate for identification. These sequences can also be used for analysis of phylogeny. These include *genome sequence analysis, sequence analysis of 16S rRNA and some other such as rpoB* encoding

ignou THE PEOPLE'S UNIVERSITY

β-subunit of RNA polymerase, genes encoding RNA polymerase (*RPB1* and *RPB2*), translation elongation factor 1-α (*tef1*) and β-tubulin (*benA, tubC*) and determination of GC content.

12. One of the oldest and direct molecular methods for comparing a pair of microbes is DNA-DNA hybridization (DDH). In this method, the DNAs are heated to get single-stranded DNA (ssDNA). Then the ssDNAs of two microbes are allowed to cool and hold at temperature 25 ˚C below the T_m. The extent of annealing two strands depends on the similarity between the sequences of two organisms being compared. By convention, the organisms showing 70% similarity based on the extent of hybridization are considered a member of same species.

13. Indels (insertion/deletion) are specific length of sequences which are inserted or deleted in many genes and are specific to certain phylum. Indels are particularly useful in phylogenetic analysis when they are flanked by conserved region.