

UNIT 3

CHARACTERIZATION IN MICROBIAL TAXONOMY

Structure

3.1	Introduction	Phenotypic or Classical Characterization
	Expected Learning Outcomes	Molecular Characterization
3.2	Introduction to Phylogeny	3.4 Summary
	Phylogenetic Trees	3.5 Terminal Questions
3.3	Characterization	3.6 Answers

3.1 INTRODUCTION

In the previous unit, you learnt about different approaches used for the classification of microorganisms. As more and more information about these tiny living beings started pouring in with the technological advances, changes were done in the classification systems and rules were decided for their systematic naming. Evolutionary relationships between different organisms are being explored and are being used for taxonomy.

In this unit 3, we will discuss about the phylogeny of microorganisms and the challenges in this field. We shall also describe various techniques which have been extensively used for their identification, taxonomic classification and determination of phylogeny.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ Understand the meaning and significance of phylogeny;
- ❖ Know the challenges in determining phylogeny in microorganisms; and
- ❖ explain the methods available for identification, characterization and taxonomic affiliation of the microorganisms.

3.2 INTRODUCTION TO PHYLOGENY

Phylogeny is the study of evolutionary relationship between different groups of organisms. Probably we would never be able to know the true origin of life, but methods of classification which you will study in this unit are important tools to determine common origin of a group of organisms. Traditional way of determining the relatedness at the ancestral level using study of morphological features is not always possible for microbes because of limitation of getting pure microbial culture. With advancement of technology, availability of molecular sequences (DNA, RNA and proteins) revolutionized these studies. It was understood that these sequences are like 'molecular clocks' which have evolved with time conserving the crucial information of survival. Zuckerkandl and Pauling in 1965 for the first time gave the concept that nucleic acid and protein sequences change with time and can be considered as molecular chronometer. It is assumed that these changes increase with time but are neutral in nature so that the protein is still functional. If sequences of two similar molecules are quite different in two groups of organisms, it implies that the groups have diverged from one another long time ago. Therefore, these molecular sequences could be rich and reliable source of information to understand the phylogenetic relationship between different organisms.

There are some limitations to this method:

1. Changes in sequences occur at different rates. There are periods when these changes are fast paced. Moreover, different molecules and different parts of the same molecule may change at different rates.
2. Care needs to be taken which sequence is chosen to study the phylogeny. It is because horizontal gene transfer greatly influences bacterial genome. It is a mechanism which permits a bacterium to acquire one or more genes from a completely unrelated organism.

3.2.1 Phylogenetic Trees

Phylogenetic relationships are represented in the form of phylogenetic trees. Phylogenetic tree is a graphical representation comprising branches which connect nodes. Nodes represent taxonomic units at any level of hierarchy. It could occur at the Kingdom level and all the way to the species level (Fig. 3.1). A basal taxon is a taxon that diverges early in the history of that group. A basal taxon typically originates near the common ancestor of the group; for example, in Fig 3.1 taxon G represents a basal taxon because 1) it diverged early in time and 2) it diverge near the time of the ancestral taxon. A polytomy is a branch from which more than two groups or taxons emerged, forming a pitchfork shape. Polytomy represents an unknown cause of the divergence. (See Fig. 3.1). For example, we do not have a clear knowledge of why Taxa D, E, and F diverged from their ancestor.

It is important to note that phylogenetic trees show patterns of descent, not phenotypic similarity. These do not indicate when species evolved or how much change occurred in a lineage. It should not be assumed that a taxon evolved from the taxon next to it.

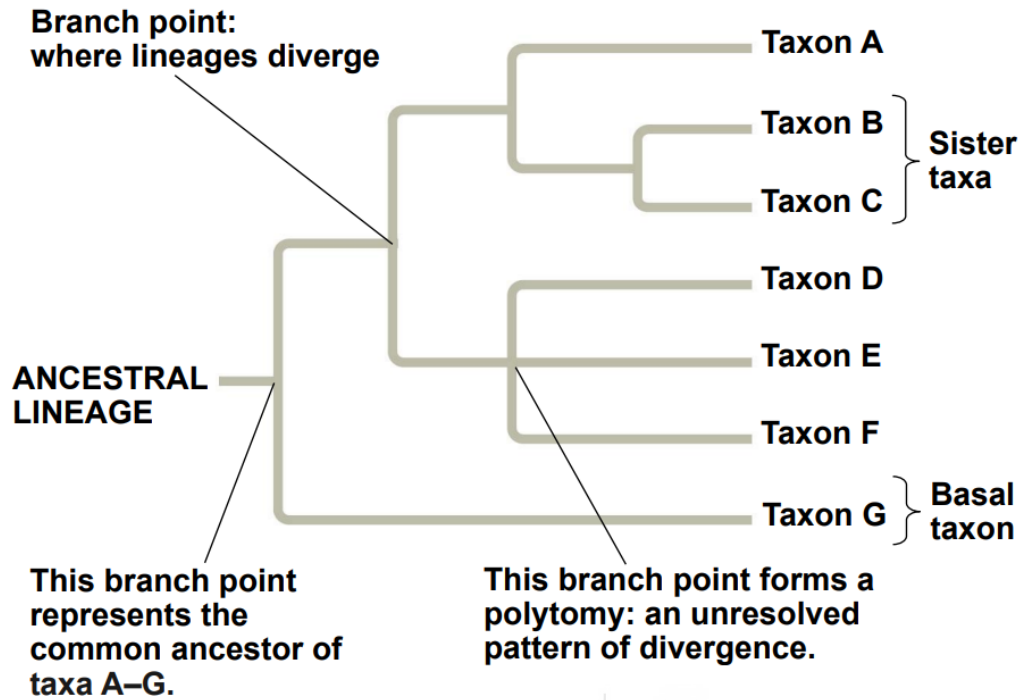


Fig. 3.1: Components of a typical phylogenetic tree.

Depending on the fact if ancestor is known or unknown, phylogenetic tree is rooted or unrooted (Fig. 3.2). In Fig 3.2, Z represents the most recent ancestor from which all other members of the tree have evolved. While unrooted tree indicates the relation among A, B, C and D but does not provide any information about their ancestor.

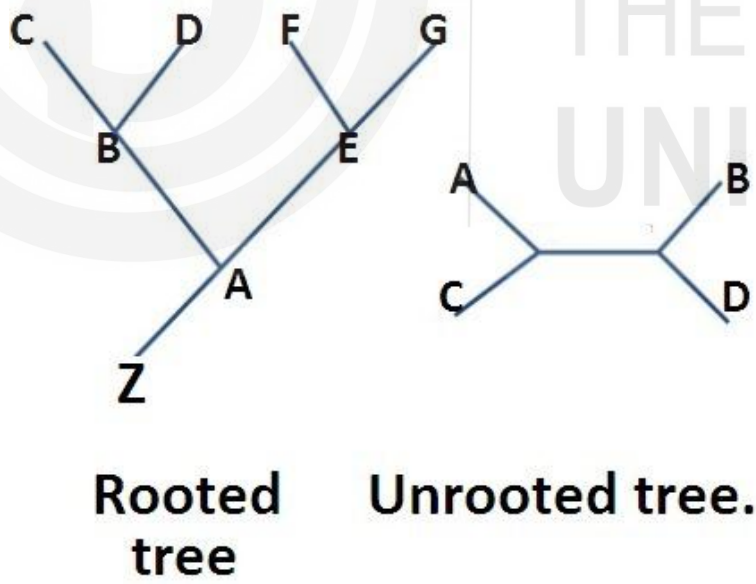


Fig. 3.2: Rooted and unrooted phylogenetic trees.

You will see in the next section of characterization how molecular sequences of proteins, nucleic acids such as DNA, RNA are used for identification and classification of microorganisms. Homologous sequences from different organisms are aligned and mathematical algorithms are used at the backend of computer programmes which determine the degree of relatedness or differences among these sequences. The information is used to calculate

evolutionary distance and generate the phylogenetic trees. You would learn about these computer programmes in another course Bioinformatics and biostatistics (MBC-007).

Tree of life: With increasing knowledge from the field of genomics and information technology also came the realization that all forms of life, from the smallest microorganism to the largest vertebrate, are connected through genetic relatedness on a vast evolutionary tree known as tree of life. This **Tree of Life provides the framework for much of our modern understanding of biology because it reveals the diversity of life as well as the historical basis for similarity and differences among organisms** ([https://www.nsf.gov / bio/pubs/reports/atol.pdf](https://www.nsf.gov/bio/pubs/reports/atol.pdf)). The tree of life is based on three domains; bacteria, archaea and eucarya.

Different designs for tree of life have been proposed (Fig. 3.3 a,b and c). First one is based on early rRNA data which shows origin of three groups from a single ancestor and are equidistant. Second tree indicates that archaea and eucarya have common ancestor and bacteria evolved before other groups of organisms. However, most accepted is the last one (Fig. 3.3) which indicates that eukaryotes evolved by fusion of bacteria and archaea.

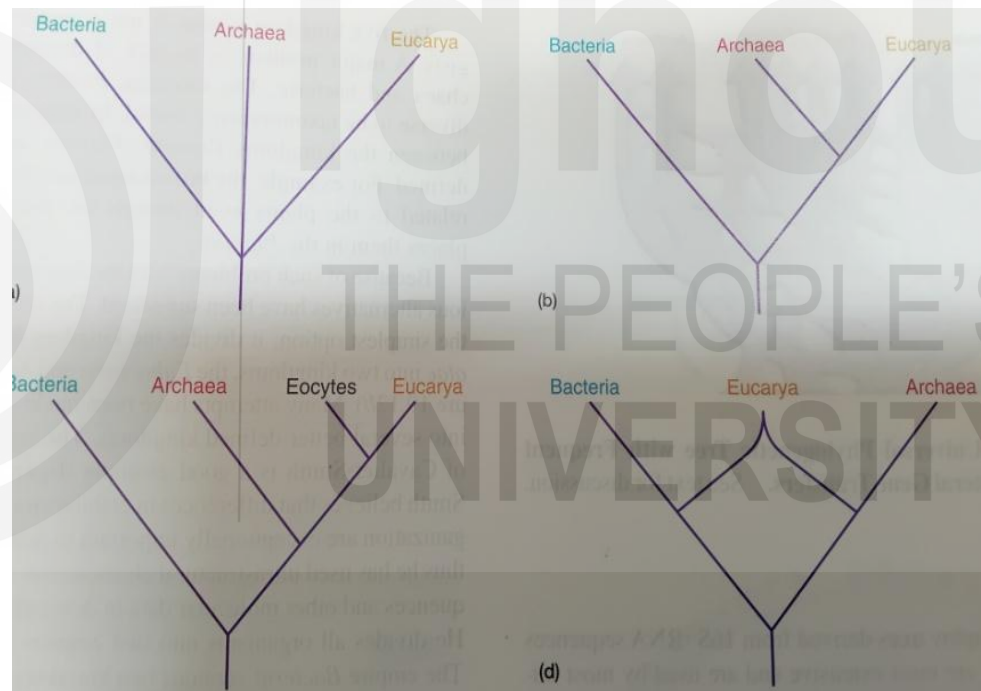
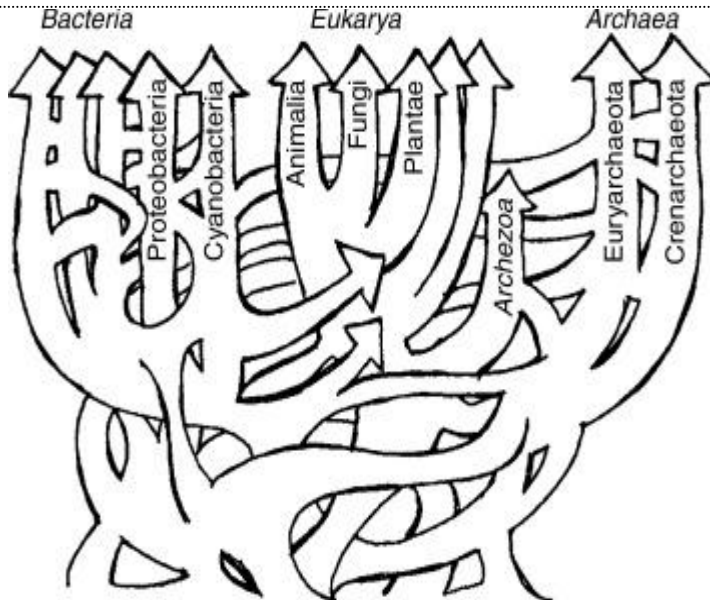


Fig. 3.3 : Various simplified designs of life of tree which depict the evolutionary relation between different domains of life.

Lot of gene transfers takes place among these domains. While extensive horizontal gene transfers are common between bacteria and archaea, surprisingly, eukaryotes have also been found to possess some bacterial and archeal genes. Few bacteria also acquired eukaryotic genes. It is proposed to be mediated by viruses as eucaryotes seldom participate in gene transfer after formation of plants, fungi and animals. In view of these developments, life of tree is actually viewed as a network of various lateral branches connecting different trunks with each branch representing one or more gene transfers. Instead of one, the tree has several trunks which represent groups of primitive cells which contributed to the original gene pool (Fig. 3.4).



If you are interested in reading more about the tree of life, you may read the following article:

<https://www.nsf.gov/bio/pubs/reports/atol.pdf>

Fig. 3.4: Tree of life representing three separate domains linked by one or more gene transfers

SAQ 1

Tick [✓] mark the correct statement:

- Horizontal gene transfer is frequent between eukaryotes and prokaryotes. [True/False]
- Comparison of molecular sequences is the only source to study phylogenetic relations between different organisms. [True/False]
- Tree of life is based on three domains; archea, bacteria and eucarya. [True/False]

We shall now undertake detailed discussion of methods used for characterization and identification of different organisms so that they can be placed in appropriate class.

3.3 CHARACTERIZATION

Characterization which means a description of detailed features of given organism is an important component of taxonomy. 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 is the 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.

3.3.1 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. 3.5

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, CO₂, 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, morphology, 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. 3.5: 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.

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. 3.6).

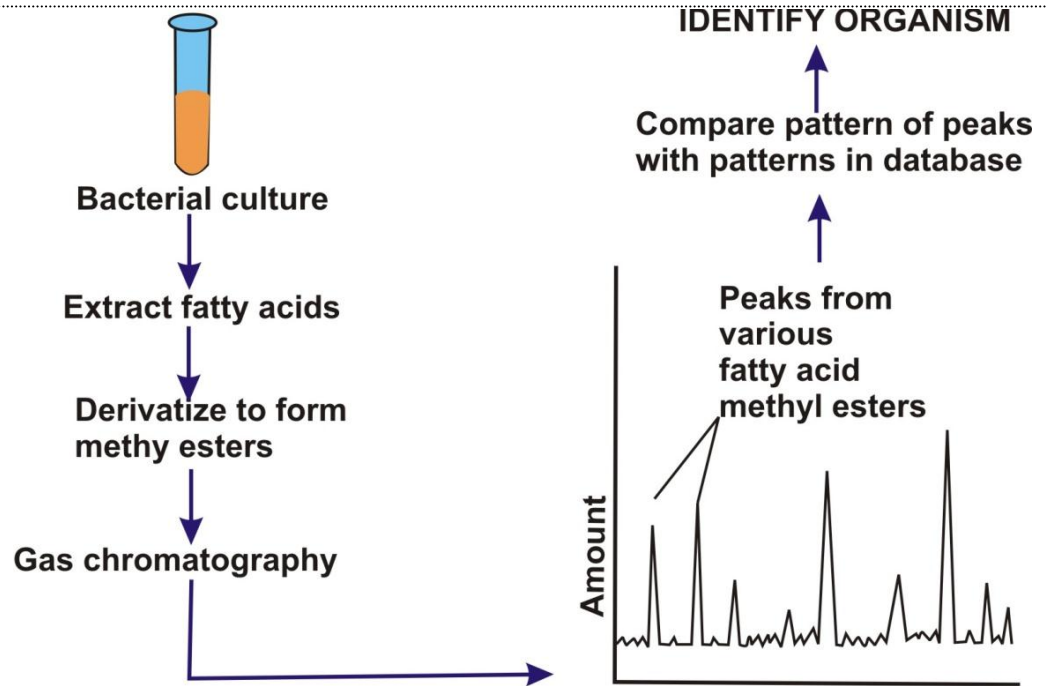


Fig. 3.6: Workflow of microbial identification by FAME analysis. Figure adapted from Brock Biology of Microorganism 13th edition, Pearson.

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. 3.7). 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.

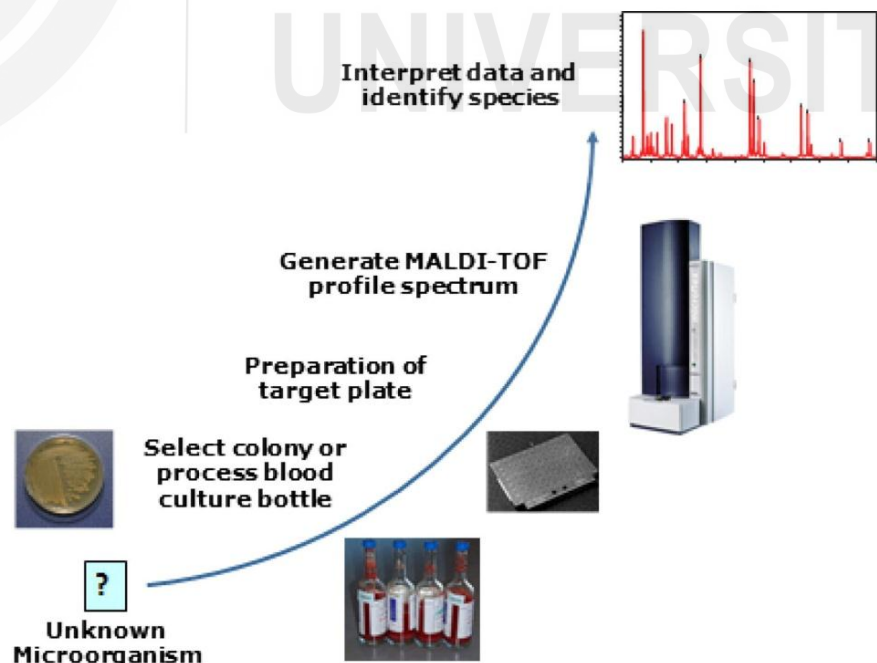
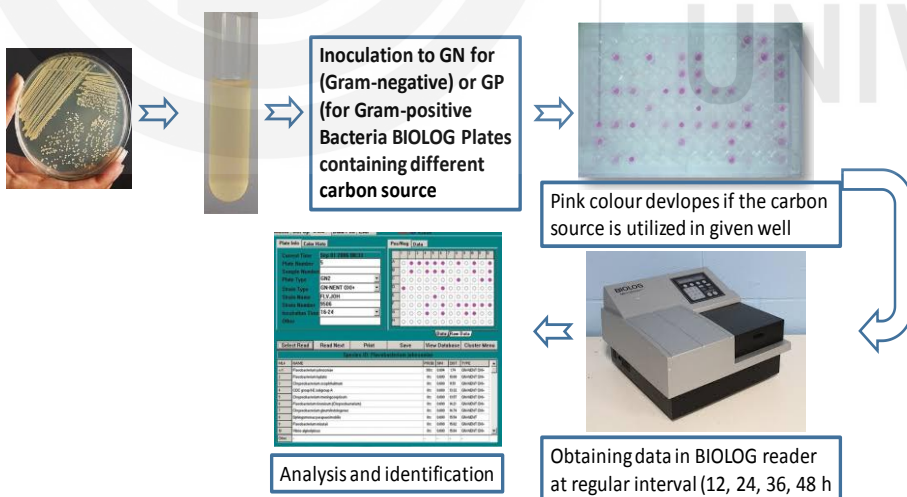


Fig. 3.7: Schematics showing methodology for bacterial identification using MALDI-ToF Mass spectrometry. Figure adapted from web source.

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. 3.8). The profile is then matched with the microbial database, and the bacterium is identified using **numerical taxonomic** approach. 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.



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.

Fig. 3.8: 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.

3.3.2 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 of 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 ranges 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 is shown in Fig. 3.3.

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:

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

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, T_m 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. 3.9).

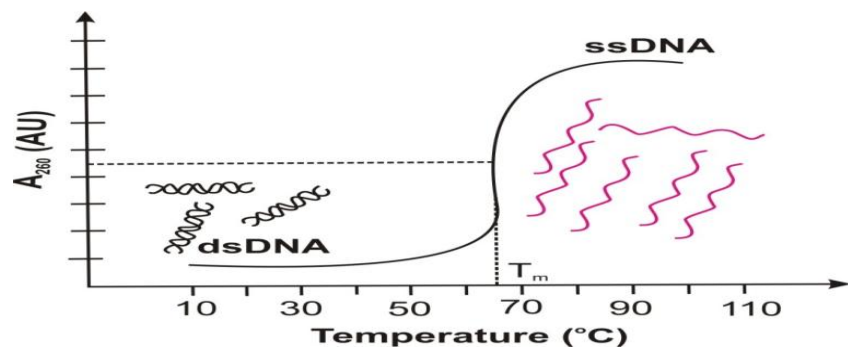


Fig. 3.9: A melting curve of DNA. A T_m is indicated at mid-point.

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 remains 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 3.1: 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. 3.10.

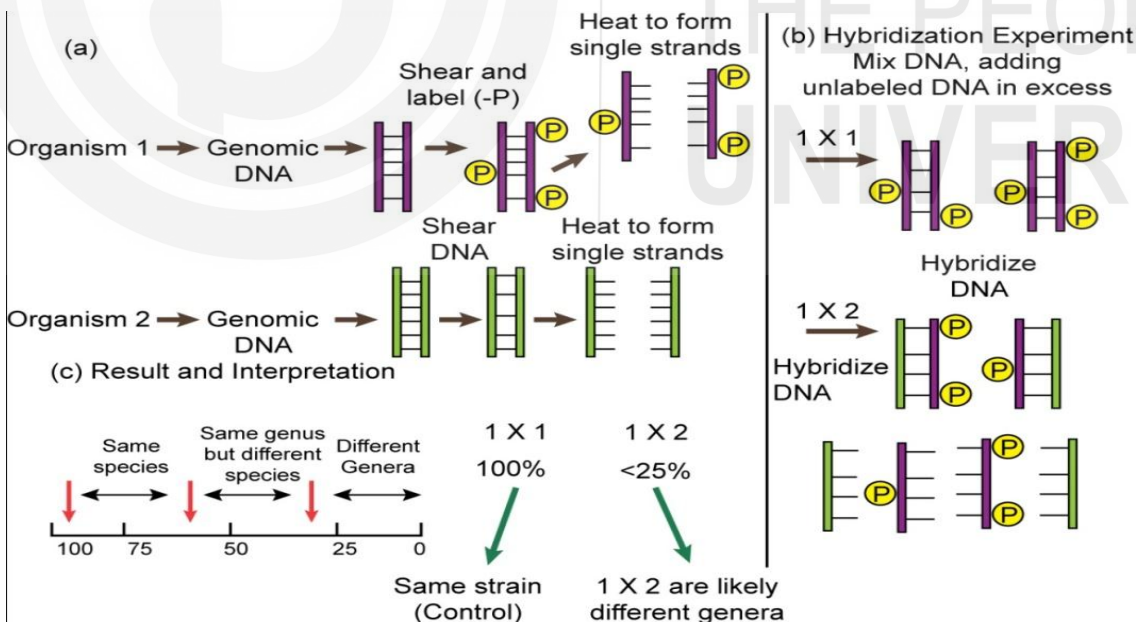


Fig. 3.10: Methodology for DNA-DNA hybridization to identify an organism. (A) Genomic DNA of test isolate is labeled (shown here as radioactive phosphate in the DNA of Organism 1). **(B)** Excess unlabeled DNA is added to prevent labeled DNA from reannealing with itself. Following hybridization, hybridized 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 from 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. 3.11). 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. 3.12. 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).

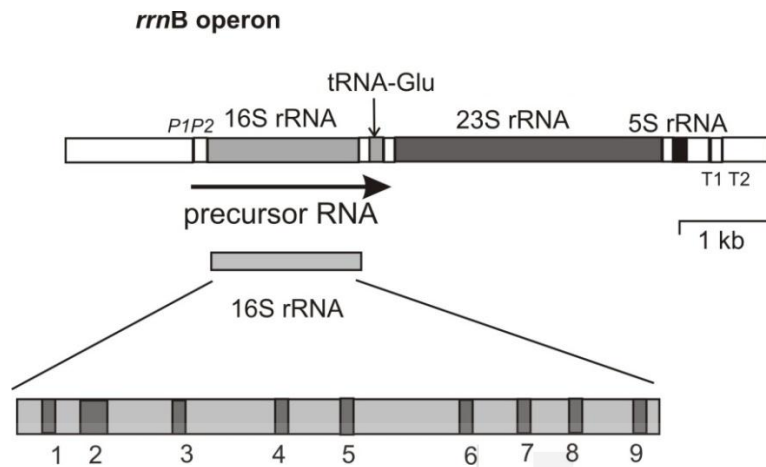


Fig. 3.11: Schematics showing arrangement of rRNA genes of *E. coli*. The lower panel shows various conserved (light grey) and variable regions (dark grey).



Fig. 3.12: 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 3.2. 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- α (*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 3.2: Other genes than 16S rRNA genes which can be used for bacterial identification.

Genes	Function
<i>rpoA</i> , <i>rpoB</i> , <i>rpoC</i> and <i>rpoD</i>	Encodes for RNA polymerase a key enzyme for transcription process
<i>gyrA</i> , and <i>gyrB</i>	Encodes gyrase, a topoisomerase
<i>dnaJ</i>	It is a molecular chaperone
<i>ppk1</i>	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 (**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. 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 (Restriction fragment length polymorphism)** 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. 3.13a). This method is also called **ARDRA (Amplified rDNA restriction analysis)**.

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. 3.13b).

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 (Enterobacter Repetitive Intergenic Consensus)**, and **REP (Repetitive Extragenic Palindrome)** 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.

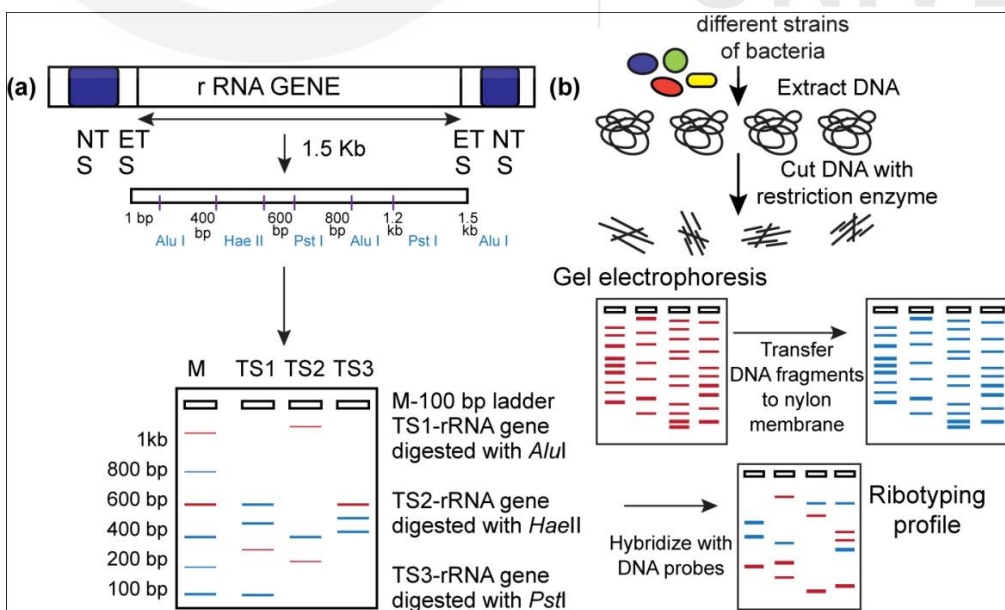


Fig. 3.13: Fingerprinting based on rRNA gene. Methodology for (A) RFLP of 16S rRNA or ARDRA, (B) Ribotyping.

SAQ 2

Fill in the blanks with appropriate words:

- a) The FAME profiling used for identification of microbe analyzes differences in the type ofpresent.
- b) A tree-like presentation which shows the differences among microbial isolates is called.....
- c) 16S and 18S rRNA used for microbial identification is component ofsubunit of the ribosome.
- d)gene sequence analysis can be used as an alternative rRNA gene cluster for identification for fungi.
- e) Families of repetitive DNA namely.....,, andare present in bacteria which can be used for DNA fingerprinting analysis.

3.4 SUMMARY

- Phylogeny is the study of evolutionary relationship between different groups of organisms. This relationship is graphically represented in the form of phylogenetic trees.
- Tree of life represents the evolutionary relation among all forms of life.
- 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.

3.5 TERIMINAL QUESTIONS

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

6. What is molecular characterization? Describe few major methods which can be used under this approach.
7. Describe phenotypic methods which can be used for identification of microbes.
8. Describe sequencing-based approaches for microbial identification.
9. Describe DNA-DNA hybridization method and its best alternative for bacterial identification.
10. What is the significance of indels in microbial identification?

3.6 ANSWERS

Self-Assessment Questions

1.
 - a) False
 - b) False
 - c) True
2.
 - a) Fatty acid
 - b) Dendrogram
 - c) Small
 - d) β -tubulin
 - e) BOX, ERIC, and REP

Terminal Questions

1. 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
2. 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.
3. Please refer to Fig 3.10

4. rRNA gene is one of the most conserved genes, present in all organisms, and long enough, it is routinely used for microbial identification.
5. 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.
6. 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.
7. Serotyping, RAMAN spectroscopy, SDS PAGE, MALDI-Tof, chemotaxonomy are some of the phenotypic methods which can be used for identification of microbes
8. 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 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.
9. 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.
10. 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.