Chapter V

Bacterial classification

1- Introduction

To facilitate communication among different actors in epidemiological surveillance, it is essential to have precise systems for classifying and naming circulating bacterial strains, as well as their potentially pathogenic variants.

2. Definition of Taxonomy

Taxonomy is the science of **classification rules**. The term comes from the Greek words "***taxis***" which means (order) and ***nomos*** which means (laws). Taxonomy allows the naming of living organisms (**nomenclature**) and their classification into units (*taxa*), within which they share a large number of common characteristics.

"Bacterial taxonomy aims to establish groups of bacteria with common characteristics, known as **taxa**, to which it assigns a name."

Taxonomy can be divided into three complementary and interdependent operations: **classification**, **identification**, and **nomenclature**.

2.1. Classification

It involves grouping organisms into taxonomic units (*taxa*) depending on their phenotypic and/or phylogenetic relationships. Currently, two types of phenotypes are distinguished:

- **Expressed (observable) phenotype**: This refers to observable characteristics such as morphological traits, staining properties, growth conditions, biochemical, physiological, antigenic traits, sensitivity to bacteriophages, or antibacterial substances.
- **Cryptic (non-observable) phenotype**: This can be analyzed through chemotaxonomy by studying the structure of informational molecules (eg. DNA, RNA, proteins).

2.2. Identification

Identification assigns unknown strains to one of the described taxa, allowing precise classification.

2.3. Nomenclature

It is the discipline that establishes the rules of an international language, ensuring that identical bacteria are named consistently worldwide. It encompasses all the rules governing the assignment of a **name** to each distinct taxon.

A correct nomenclature allows a taxon to be designated without describing all its characteristics, enabling specialists (epidemiologists, researchers, clinicians, vaccine manufacturers, etc.) to communicate unambiguously. For example, referring to a bacterial strain as part of the *Enterobacteriaceae* family immediately informs any bacteriologist that the strain consists of Gram-negative bacteria, mostly non-spore-forming, facultative anaerobes, etc.

2.3.1. Types of Names

Two categories of names are distinguished:

- **Informal (vernacular) names**: These are used in national languages, such as *Koch's bacillus* for *Mycobacterium tuberculosis*

- **Scientific names (Latin words)**: Scientific nomenclature follows the rules established by Swedish botanist Carl von Linné in 1753. Like other life sciences (botany, zoology, etc.), bacterial nomenclature follows a **binomial system**:
    - The first part of the name is the **genus**.
    - The second part is the **species**.

Example: *Escherichia coli, Staphylococcus aureus, Salmonella enterica*.

This nomenclature is **universal** and defined by the **International Code of Bacterial Nomenclature**.

2.3.2. Rules for forming scientific names (Taxonomic Hierarchy)

In bacteriology, a **common nomenclature** must be used by all specialists to enable effective collaboration.

The most commonly used taxonomic ranks in medical bacteriology are: **Order, Family, Genus, and Species**.

These taxa are expressed in **Latin words**, which are italicized (or underlined) and follow these suffix conventions:

- ✓ **"-ales"** for orders
- ✓ **"-aceae"** for families
- ✓ **"-us"**, **"-er"** (masculine) or **"-a"** (feminine) for genus and species (grammatically adjusted).

In practice, genus and species are sufficient to characterize a strain (**binomial system**).

- **Genus**: Italicized, with the first letter capitalized.
- **Species**: An "epithet" in italics (or underlined), with the first letter in lowercase.

Example: *Salmonella enterica* (Genus: *Salmonella*, Species: *enterica*).

After the first mention, the **abbreviated form** can be used: the first letter of the genus followed by a period (point) and the species name (e.g., *S. enterica*, *E. coli*).

Subspecies names: A name in which the species name is followed by *subsp.* and a third italicized term unique to the subspecies. (Example: *Salmonella enterica subsp. Arizonae*).

Diacritical marks (á, à, â, ä, ã, é, è, ê, ë, í, î, ï, ñ, ó, ò, ô, ö, õ, ú, ù, û, ü, ø, œ…) are not allowed, and words must not contain hyphens. (Example: *Bacteroides*, not *Bacteroïdes*).

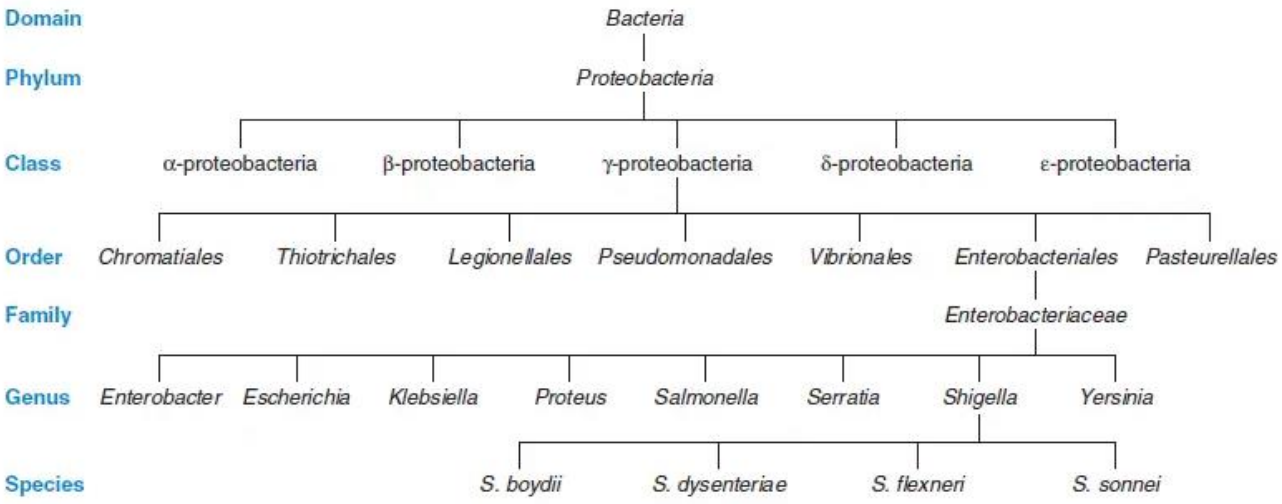

Figure 1. Hierarchical arrangement in taxonomy

3. Principles and methods of bacterial classification

3.1. Phenetic (Phenotypic) classification

This is an approach used to classify bacteria based on their observable and measurable characteristics. It includes all techniques that do not involve nucleic acids and relies on phenotypic similarities (morphological, biochemical, and physiological traits).

3.1.1. Basis of phenetic classification

Numerous laboratory procedures reveal phenotypic differences between bacteria.

- **Macroscopic, microscopic, and staining characteristics**
  - Description of colonies (shape, size, color, odor)
  - Cell morphology (bacilli, cocci) and their arrangements
  - Staining techniques (Gram, methylene blue, acid-alcohol resistant staining)
  - Observation of motility in a fresh state
  - Presence or absence of endospores
  - Aerobic or anaerobic growth

Morphological characteristics are useful for identification but cannot, on their own, demonstrate phylogenetic relationships.

- **Metabolic tests**

These are very important and can distinguish closely related bacteria. They test for the presence of enzymes (oxidase, catalase), the degradation of urea, esculin, lactose fermentation and gas production, utilization of different sugars as carbon sources, citrate utilization, acetoin production, etc. These techniques have been miniaturized into specialized kits (e.g., API strips).

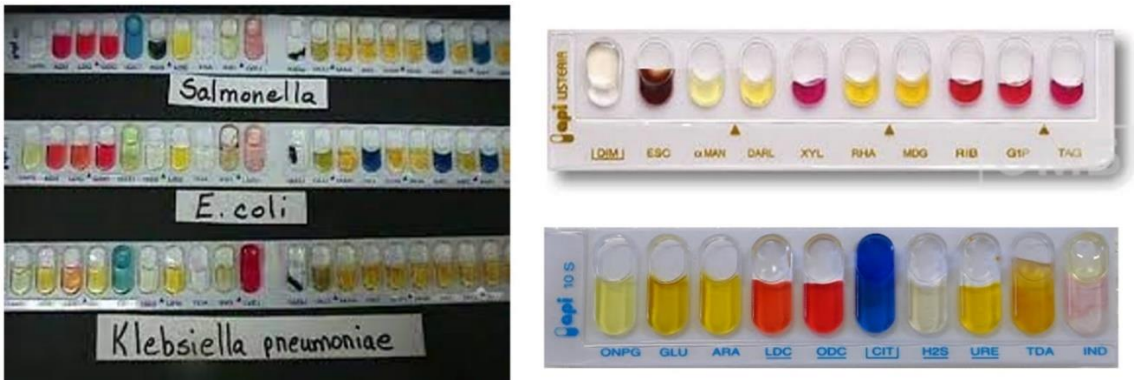

Figure 2. Identification of bacteria strains using API 20E strips

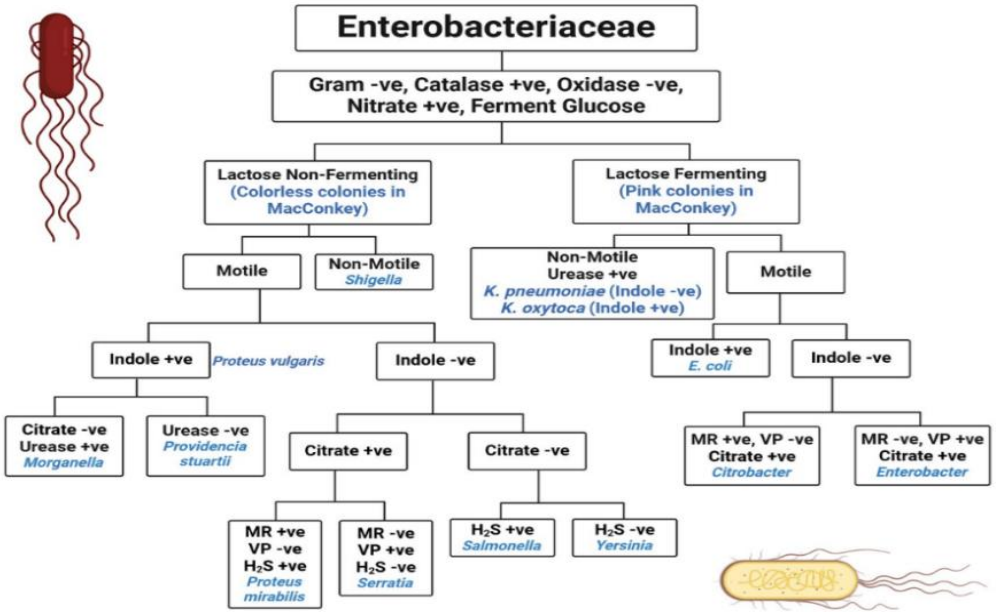

Figure 3. Example of *Enterobacteria* classification based on biochemical characteristics

- **Serological methods:** Serodiagnosis and serotyping are based on the specific antigen–antibody reaction. This method allows differentiation between species and even between strains within the same species. The targeted antigens are the **O antigens** in Gram-negative bacteria**, H (flagellar) antigens**, and **K (capsular) antigens**.

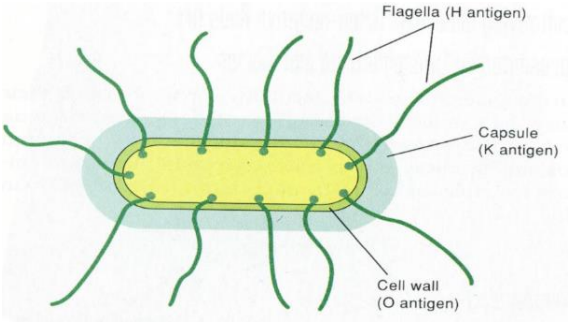

Figure 4. Antigenic structure of Gram-negative bacteria

- **Inhibition tests**: These evaluate the growth of microorganisms on selective media in the presence of antibiotics (antibiogram).
- **Chemotaxonomy**: This involves determining the fatty acid profile of cell walls, and the total protein profile by electrophoresis.

Example

- The composition of peptidoglycan can be used for taxonomic purposes, particularly for Gram-positive bacteria. While the glycan portion is highly conserved between different species, peptidoglycan shows considerable structural variability. The peptide portion, as well as the type and composition of the interpeptide bridge, vary significantly.
- In Gram-negative bacteria, the type of peptidoglycan has limited taxonomic value.
- Teichoic acids can be used as phenotypic markers. Their absence in *Micrococcus* spp. helps distinguish them from *Staphylococcus* spp. Within this genus, the type of teichoic acid can differentiate *Staphylococcus aureus* from other staphylococcal species.

- **Phage typing**: This involves infection by bacteriophages and the formation of lysis plaques. The lysovar or lysotype is then defined.

Limitations of this classification method: This classification method has its limitations, for example:

- Morphology may vary depending on the culture medium and can sometimes be difficult to define.
- A specific enzymatic activity may not be detected when using a synthetic substrate.
- Two phylogenetically distant bacteria may share common antigens.
- The number of phenotypic characteristics used is quite limited (a hundred at most) compared to the number of genes usually present in bacteria (around 5,000).

➢ **Numerical Taxonomy**

Also called **Adansonian classification** (named after the French botanist Michel Adanson). It is a type of **phenetic** classification method that uses a large number of phenotypic characteristics (morphological, biochemical, physiological, etc.) to classify bacteria based on overall similarity. Its aim is to maximize objectivity by considering as many traits as possible.

This method takes into account **all the characteristics** of an organism, with each trait given **equal importance**.

Principle:

The goal is to **group together** in the same similarity class the individuals that are **most alike**. The method involves:

- Studying, for each strain, a number of **50 to 100 characteristics**: morphological, biochemical, cultural, structural, etc.
- Assigning **equal weight** to each characteristic, which is coded as +/1 (presence of the trait) or –/0 (absence of the trait).
- Measuring the **affinity** between bacterial strains using the **Jaccard similarity index**.

Jaccard Index: $S_{(A.B)} = nS^{+} / nS^{+} + nD$

- **$S_{(A,B)}$:** Similarity coefficient between strains A and B
- **nS^{+}**: Number of shared **positive** traits between A and B
- **nD**: Number of **differing** traits

Two strains **A** and **B** are considered to belong to the **same species** if: **$S_{(A.B)} \geq 0,85$ (85%).**

3.2. Phylogenetic classification

From the Greek *Phylon* (race) and *Genesis* (origin or generation), this classification is based on clearly identified, genetically homogeneous species that have evolved differently from a common ancestor. It focuses on lineage and evolutionary divergence among different taxa such as genera, families, and orders. This approach allows the depiction of evolutionary relationships between species in a **phylogenetic tree**.

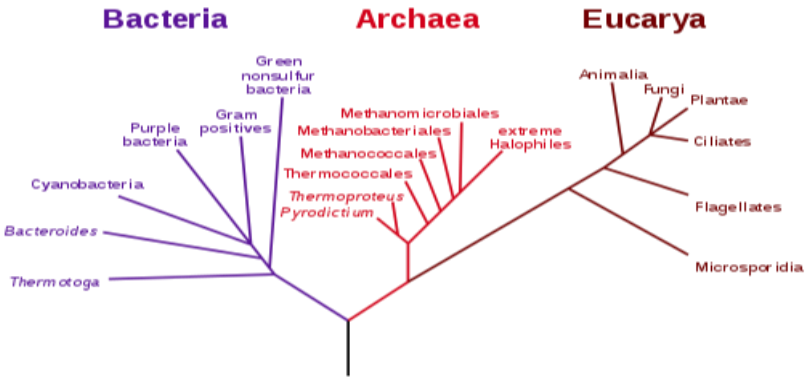

Figure. 5. Phylogenetic trees of life

Bases of phylogenetic classification

These methods are based on the analysis of DNA or RNA molecules, either at the whole genome level or by targeting specific fragments of the bacterial chromosome.

- **GC content (or Chargaff's rules)**

Regardless of the species of origin, DNA always contains equal amounts of purines and pyrimidines:
(A + G) = (C + T) or (A + G) / (C + T) = 1
Moreover, the amount of thymine nucleotides equals that of adenine (A/T = 1), and guanine equals cytosine (G/C = 1).
However, the ratio (A + T) / (C + G) varies significantly and is **characteristic** of the species.
This ratio, called Chargaff's coefficient, can be calculated after sequencing using the following formula:

G + C % = (G + C) × 100 / (A + T + G + C)

GC content varies between species: it is about 50% in *E. coli*, 60% in *Pseudomonas*, and between 25% and 45% in *Clostridium*. It is therefore of taxonomic interest:

- Two bacteria belonging to the same species have identical GC% (with a margin of ±2.5%).
- Two bacteria with different GC% do not belong to the same genetic community.
- However, two bacteria with identical GC% do not necessarily have the same nucleotide sequences and can still be genetically distant.

Thus, GC% can only serve as an **exclusion criterion**: it only helps to state that two organisms are genetically distant. It can, for example, confirm that two strains do not belong to the same genus. **Example:** The GC content of *Staphylococcus aureus* is about 30%, while it ranges from 62 to 70% in the genus *Micrococcus*.

Among living organisms, bacteria show the widest variation in GC content. While the GC content of animals and higher plants varies between 30% and 50%, in prokaryotes it ranges from 23.7% in *Mycoplasma sualvi* to 79% in the genus *Pseudonocardia*.

- **Bacterial genome size**

Genome size refers to the **amount of DNA** contained in a single copy of a genome. It is measured by the **number of nucleotides** (base pairs), with the megabase (Mb)—equal to one million nucleotides—as the unit of measurement.

Differences in genome size can be used to determine the degree of relatedness between bacteria and to construct phylogenetic trees.

Genome size alone is not conclusive, but it can suggest evolutionary relatedness when combined with other data.

- **Determination of DNA-DNA Hybridization rates**

This technique, known as **molecular hybridization**, is based on the idea that similar or related species have similar nucleotide sequences. It relies on the ability of DNA strands to pair with complementary strands after being separated by heat (which breaks the hydrogen bonds between bases).

By heating and then slowly cooling single-stranded DNA from two different organisms, it is possible to reconstitute a double-stranded DNA molecule from previously separated bacterial DNA strands. Homologous (complementary) single strands may come from the same strain—this is called **homologous hybridization** (AxA'), or from different strains (strains A and B)—this is called **heterologous hybridization** (AxB), and the resulting molecule is referred to as **hybrid DNA**.

Hybridization allows for the determination of the genetic relatedness between the organisms being compared (AxB) by assessing their **degree of homology**. The more closely related the organisms are, the stronger the hybridization.

However, **DNA-DNA hybridization** is typically used for the study of **closely related microorganisms**.

Hybridization rate = (concentration of hybrid DNA / total DNA concentration) × 100

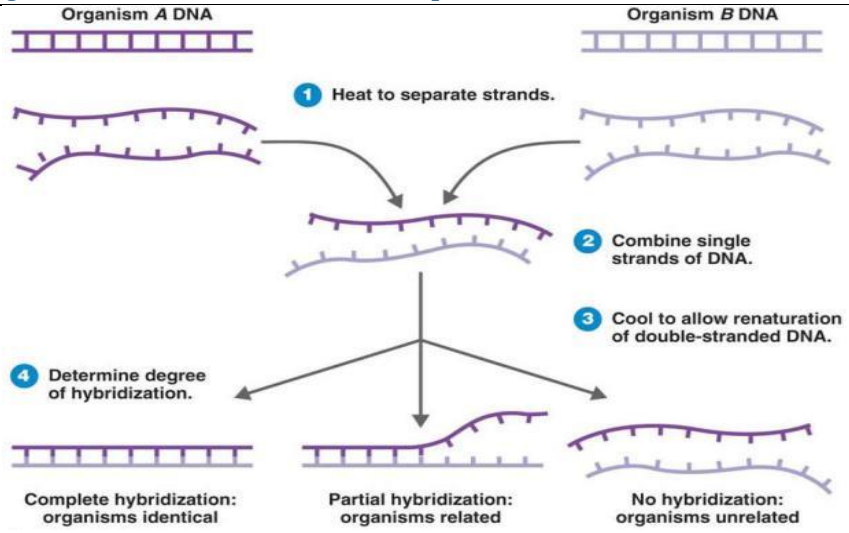

Figure 6. Hybridization principle

The **thermal stability** of hybrids is also taken into consideration, which is assessed by measuring the **melting temperature (Tm)**—the temperature at which DNA strands from different species separate. Similar Tm values indicate a greater genetic similarity, while different Tm values suggest a greater genetic divergence.

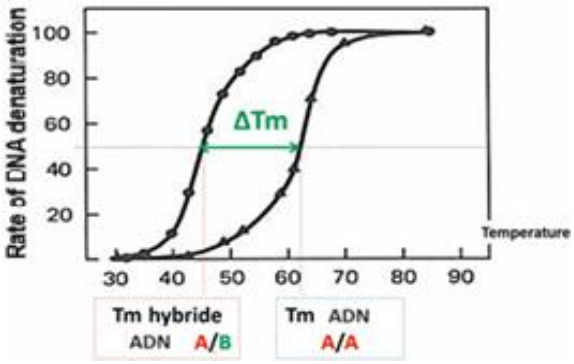

Figure 7. Assessing genetic similarity through DNA melting temperature

Thus, it is considered that:

Two bacteria A and B belong to the same species if: **DNA-DNA hybridization rate ≥ 70%** and **ΔTm ≤ 5°C**

- **Study of ribosomal RNA (rRNA)**

rRNAs are considered "molecular clocks" because:

→ They are present in both prokaryotic and eukaryotic cells,

→ They have a highly conserved structure across all living organisms,

→ They are abundant in the cell, easy to purify, and to sequence.

The most commonly used for taxonomic studies is 16S rRNA, which serves as a molecular marker due to its evolutionary conservation.

Research on 16S rRNA has made it possible to distinguish *Eubacteria* from *Archaebacteria*.

Sequencing of the DNA coding for 16S rRNA is a key step in using 16S rRNA for bacterial taxonomy. It allows the construction of phylogenetic trees and the determination of evolutionary relationships and phylogenetic links between different bacterial species.

- **Study of restriction profiles**

Double-stranded DNA can be cut by restriction enzymes, called **endonucleases**. The bacterial genome is then characterized by a series of fragments (whose sizes can be measured), which are separated by agarose gel electrophoresis. The highly specific mode of action of endonucleases allows the establishment of restriction profiles, making them useful in bacterial taxonomy.

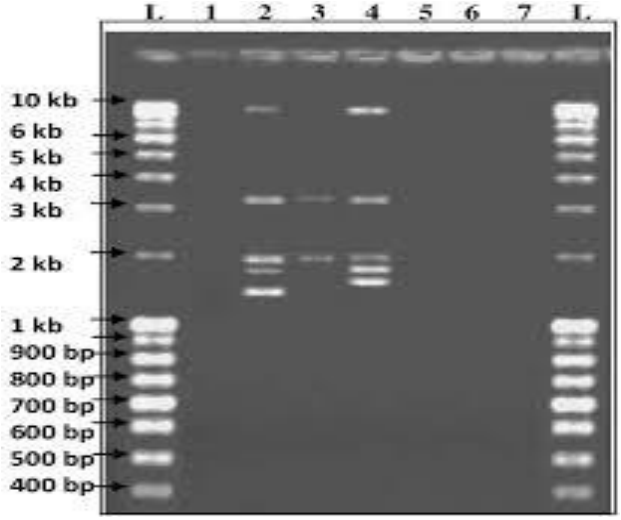

Figure 8. DNA Restriction fragment analysis by gel electrophoresis

- **Whole Genome Sequencing (WGS)**

Whole genome sequencing enables the analysis of the entire bacterial DNA for higher taxonomic resolution than conventional methods. It allows for comprehensive genome comparison to infer phylogenetic relationships and accurately define bacterial species based on genomic data. WGS provides maximal resolution, supporting bacterial taxonomy, epidemiological tracking, and genetic diversity studies.

Principle and main steps of WGS:

- DNA extraction: Isolation of purified genomic DNA.
- Library preparation: DNA fragmentation and adaptor ligation for sequencing.
- Sequencing: High-throughput reading of DNA fragments.
- Assembly: Reconstruction of the genome from sequence reads.
- Bioinformatics analysis: Sequence alignment, phylogenetic inference, and species delineation**.**

3.3. Bergey's classification

The most widely used bacterial classification system among microbiologists today is the *Bergey's Manual of Systematic Bacteriology*. It is a guide for distinguishing bacterial species based on differences between strains, starting from the type strain.

It was first published by David Hendricks Bergey in 1923 and has been continuously expanded and updated since then.

Originally divided into five volumes, its primary aim was to compile **phenotypic** data on recognized bacterial species to aid in the identification of unknown strains. The manual classified bacteria into tribes, families, and orders based on various parameters, particularly structural and functional traits.

However, this approach became highly empirical and has since been replaced by **phylogenetic** data analysis.

The manual provides a classification of the prokaryotic kingdom, dividing it into two domains: *Bacteria* and *Archaea*. Each domain is further divided into phyla, which are subdivided into classes, then orders, families, genera, and finally species.

To keep up with the explosion of knowledge about the microbial world, an **electronic manual** with frequent updates is necessary. *The Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB)* is an essential tool for anyone at the forefront of microbiology research. The digital edition provides up-to-date descriptions of taxonomy, systematics, ecology, physiology, and other biological properties of all named prokaryotic taxa.

4. Basic taxonomic unit: The concept of bacterial Species

The species is the basic taxonomic unit (fundamental unit of classification) in sexually reproducing organisms. A species is defined as a group of organisms that naturally interbreed, capable of genetic exchange through sexual reproduction, resulting in fertile offspring.

This definition is obviously not applicable to prokaryotes. Bacteriologists had to develop a specific definition of species:

"In bacteriology, a species consists of its type strain and all strains considered sufficiently similar to the type strain to be included within the same species."

However, distinct phylogenetic lineages within the same microbial species may display very different ecological or pathogenic properties.

Example: The species *Escherichia coli* includes both commensal strains and highly pathogenic strains, such as enterohemorrhagic E. coli (EHEC).

As a result, different systems are used to distinguish between strains, including:

- Biovars: strains that differ biochemically or physiologically
- Serovars (serotypes): strains with distinct antigenic properties
- Pathovars: strains with different pathogenicity
- Zymovars: strains that differ in enzyme isoforms
- Lysovars (lysotypes or phagovars, phagotypes): strains that differ in sensitivity to bacteriophages
- Antibiotypes: strains that differ in antibiotic sensitivity