UNIT 6

DIVERGENCE OF BACTERIAL AND ARCHAEAL GENOMES

Structure

6.1 INTRODUCTION

There is general agreement among scientists that the Last Universal Common Ancestor (LUCA) probably lived some 3.8-3.5 billion years ago (bya) and represents the most primitive organism from which all organisms now living might have evolved. Prokaryotes appeared around 3.5 bya while, eukaryotes appeared much later. Though all prokaryotes fall in the size range of 0.5-5.0 micrometres and share many characteristics, analysis of ribosomal RNA from several species of prokaryotes led Carl Woese to split prokaryotes into two separate lineages, eubacteria and archaebacteria, which were later renamed as bacteria and archaea, respectively. Biologists now divide all living organisms into three domains: Bacteria, Archaea, and Eukarya.

Both bacteria and archaea are cosmopolitan in distribution. Though majority of the species are mesophiles, there are several others which can be grouped under the category of extremophiles, occupying environments which present extremes of conditions be it temperature, salinity, hydrostatic pressure, or anoxic conditions. Prokaryotic diversity has been described using morphological and biochemical features as well as molecular characteristics. Studies involving gene- and whole genome-sequencing have been very useful in elucidating not only evolutionary relationships but also divergence times of different lineages.

Prokaryotes have played key roles in shaping the evolutionary history of life on earth, three events being the most significant. One such event involved cyanobacteria, which by their photosynthetic activity, converted the primitive, reducing atmosphere of the earth to an oxidising one, which further led to the evolution of aerobic type of metabolism in organisms.

Another crucial event occurred with the engulfment of a facultative anaerobe by a eukaryotic ancestor, which ultimately gave rise to mitochondria responsible for energy production by oxidative phosphorylation. Third event, equally significant, was the uptake of cyanobacteria by eukaryotic ancestors, giving rise to chloroplast found in all plants and algae.

Prokaryotes, once thought to be small, simple organisms, exhibit extraordinary biochemical capabilities (presence of specialised enzymes). They are able to produce a variety of specialised, complex metabolites, using the ordinary primary metabolites, which are finding various applications.

Objectives

After studying this Unit, you would be able to:

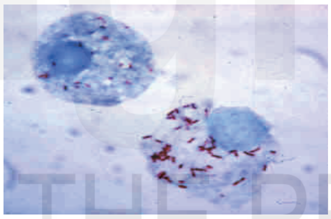
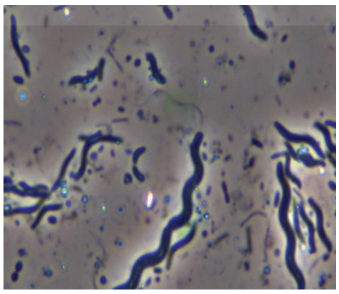
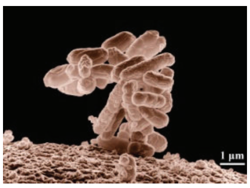
- ❖ comprehend the division of prokaryotes into bacteria and archaea;
- ❖ understand the important differences between the bacteria and archaea as revealed by molecular techniques;
- ❖ explain the extensive diversity in bacteria and archaea in terms of morphological features, metabolic capabilities, and occupation of a variety of habitats;
- ❖ comprehend the divergence in information processing systems that led to a split in the prokaryotes, separating bacteria from archaea;
- ❖ explain the specialisation of ribosomal proteins that occurred early in evolution leading to bacterial and the archaeal lineage;
- ❖ describe the key role played by the prokaryotes in the evolution of aerobic metabolism in organisms, photosynthesis and biogeochemical cycles of the earth; and
- ❖ explain the role of prokaryotes as the source of a number of important enzymes and secondary metabolites which find applications in medicine, biotechnology, and industry.

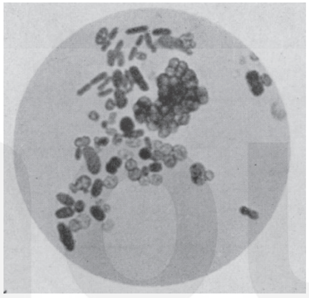
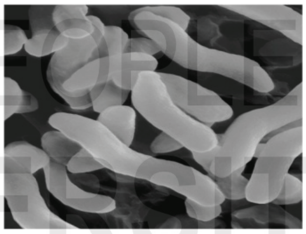
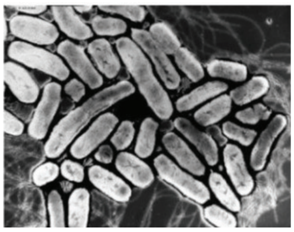
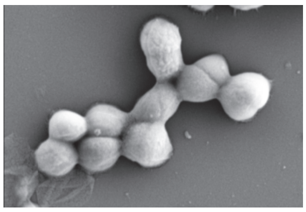
6.2 ORIGIN AND DIVERSIFICATION OF BACTERIA AND ARCHAEA

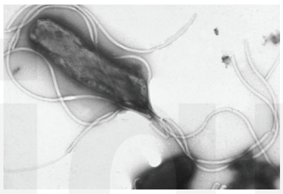
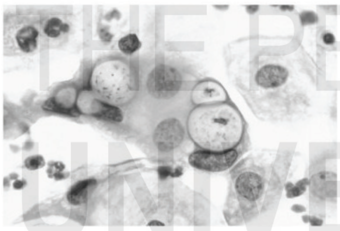
6.2.1 Bacterial and Archaeal Diversity Based on Morphological Attributes

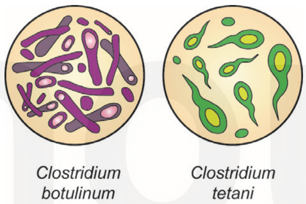
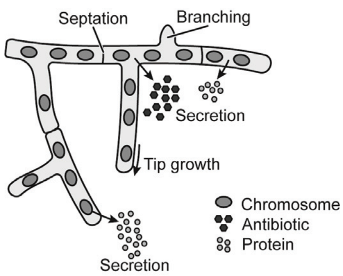
Morphological features of prokaryotes enable us to identify species which can not only reveal microbial diversity but also help in diagnosis of disease and prevention of infection. Characteristics studied mostly include cell shape, cell wall, movement, flagella, Gram staining etc. Four basic shapes can be seen in bacteria: spherical (cocci), rod-shaped (bacilli), arc-shaped (vibrio), and spiral (spirochaete) as first described by Antonie van Leeuwenhoek in 1675 using light microscopy. Observations of cell morphology at a higher resolution (of the order of 2 Angstroms) became available with the development of electron microscopy some 300 years later. Bacteria exhibit enormous diversity in terms of their physical features as can be appreciated from Table 6.1 below, which lists different phyla and classes with their representative examples.

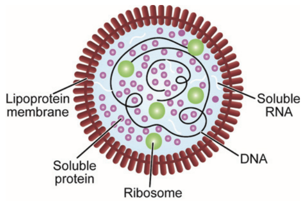
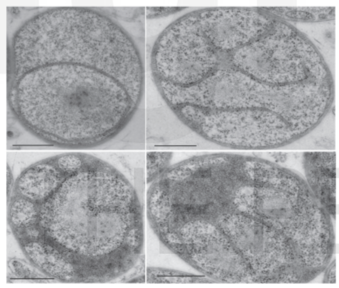
Table 6.1: List of different phyla and classes of bacteria with their representative examples and their mode of life.

Bacteria	Example	Micrograph
Phylum Proteobacteria Class Alpha Proteobacteria May be photoautotrophic, symbionts or parasitic. It has been postulated that aerobic rickettsial ancestor parasitized eukaryotic cells and gave rise to mitochondria.	*Rhizobium:* Endosymbiont in roots of leguminous plants and responsible for fixation of atmospheric nitrogen. *Rickettsia prowazekii:* Obligate, intracellular parasite, responsible for epidemic typhus.	 **(i) *Rickettsia***
Beta proteobacteria Some species of bacteria of this group play a role in the nitrogen cycle.	*Nitrosomonas*: Some species oxidise ammonia into nitrite. *Spirillium minus* is responsible for rat-bite fever.	 **(ii) *Spirillum***
Gamma proteobacteria Some species are pathogenic, while others play a beneficial symbiotic role in the human gut.	*E. coli*: normally has a beneficial role in the human gut though some strains are pathogenic.	 **(iii) *Escherichia coli***

	Vibrio cholerae is the causative agent of cholera	 (iv) ***Vibrio cholerae***
	Salmonella: Some strains cause typhoid fever, while some other strains can cause food poisoning.	 (v) ***Salmonella***
	Azotobacter spp. are soil/water/sediment inhabiting bacteria with the ability to fix nitrogen non-symbiotically and have been in use as a biofertilizer for over a century.	 (vi) ***Azotobacter***
Delta proteobacteria Some species can produce multicellular fruiting bodies for spore formation to overcome unfavourable conditions. Also includes sulphate-reducing species.	*Desulphovibrio vulgaris*: Anaerobic, capable of reducing sulphate.	 (vii) ***Desulfovibrio vulgaris***
Myxobacteria include motile bacteria capable of gliding movements. This grouping includes bacteria which can digest cellulose and have been found to reside in a number of niches such as soil, decaying vegetation, hot springs, organic matter and faeces of ruminants.	*Sorangium cellulosum* is a soil-dwelling bacterium.	 (viii) ***Sorangium cellulosum***
	Micrococcus roseus resides within the gut of termite and is capable of producing several hydrolytic enzymes which can digest cellulose. *Micrococcus luteus* is Gram-positive and	 (ix) ***Micrococcus luteus***

	non-motile coccus, can be seen arranged in tetrads. Shows various modes of life such as saprophytic, commensal or opportunistic pathogens	
Epsilon proteobacteria Includes members occupying deep sea hydrothermal vents (where they can oxidise sulphur and reduce nitrogen oxide) as well as species which are opportunistic pathogens.	*Campylobacter*: Some species are responsible for causing blood poisoning and intestinal inflammation.	 **(x) *Campylobacter***
	Helicobacter pylori: Colonises the lower part of the stomach and duodenum, and plays a role in the development of gastritis, peptic/duodenal ulcers.	 **(xi) *Helicobacter pylori***
Phylum Chlamydias All members of this group lack peptidoglycan in their cell walls and lead obligate intracellular parasitic modes of life.	*Chlamydia psittaci* causes psittacosis, *C. pneumoniae* causes respiratory tract infections, and *C. trachomatis* strains are responsible for chlamydia, conjunctivitis, trachoma, lymphogranuloma venereum.	 **(xii) *Chlamydia psittaci***
Phylum Spirochetes Majority of the members are free living (some are pathogenic), having anaerobic type of metabolism. Cells are spiral shaped and flagella run lengthwise in the periplasmic space between the inner and outer membrane.	*Treponema pallidum syphilis* is responsible for causing syphilis. *Borrellia burgdorferi* is the causative agent of Lyme disease.	 **(xiii) *Treponema pallidum***

Phylum Cyanobacteria Also known as blue-green algae, occur abundantly in all kinds of environments such as terrestrial, marine, and fresh water and are photosynthetic.	*Prochlorococcus* being photosynthetic is responsible for producing enormous amounts of oxygen. *Nostoc* includes a number of species which have different morphology, habitat distribution and ecological roles and can be found in both terrestrial and aquatic environments as well as existing symbiotically inside plant tissues and serve to provide nitrogen to its host plant.	 **(xv) *Scytonema***
Phylum Firmicutes contains **Gram-positive bacteria** Members of this group have a thick cell wall without an outer membrane. **Class Clostridia** They are anaerobic (some members are capable of forming highly resistant spores). Species may be found inhabiting soil, aquatic bodies or raw meat/poultry.	*Clostridium botulinum* is rod shaped, anaerobic, spore-forming, found in soil, dust and sediments of water bodies. It produces a neurotoxin botulinum which causes botulism. *C. tetani*, commonly found in soil, is responsible for causing tetanus. *C. perfringens* can be found on raw meat and poultry and in the intestines of animals and is the causative agent of food poisoning.	 **(a)** **(b)** **(xvi)** **(a) *Clostridium Botulinum*** **(b) *Clostridium Tetani***
	Streptomyces has over 500 species of filamentous bacteria occurring in soil. They are the source of a large number of bioactive compounds such as antibacterial, antifungal, antiparasitic, immunosuppressants and extracellular enzymes. Some species are pathogenic in plants.	 **(xvii) *Streptomyces***

Phylum Firmicutes **Class Mollicutes** Mycoplasma represent the smallest bacteria which lack a cell wall. It includes free-living species as well as commensal and pathogenic species. **Class Bacilli** Includes members which are facultative anaerobes and some species can form spores.	*Mycoplasma pneumoniae,* causes pneumonia, and *M. hominis* and *Ureaplasma urealyticum* are responsible for some genitourinary diseases. *Lactobacillus* is a probiotic species, found in fermented dairy products and contributes to gut health and boosts immunity. *Streptococcus, Listeria, Staphylococcus, Bacillus* are some of the pathogenic species capable of causing infection.	 **(xviii) *Mycoplasma***
Planctomycetes These bacteria occupy a variety of habitats such as marine and freshwater ecosystems, soils, even extremely dry soils of Atacama desert and form living stromatolite microbial mats in the marine shallow waters with extreme salinity. All planctomycetes lack peptidoglycan in their cell wall, making them resistant to beta-lactam antibiotics, including penicillins. They reproduce by budding rather than by binary fission.	In *Gemmata obscuriglobus* membranes divide the interior of the cell into separate compartments, condensed nucleoid surrounded by membranes. Anammox group of planctomycetes exhibit the process of generating energy by anaerobic oxidation of ammonia and plays an important role in nitrogen cycling globally.	 **(xix) *Gemmata obscuriglobus***

Though majority of bacterial species show one of the basic four shapes and are unicellular without intracellular compartmentalisation, there are some exceptions too. Multicellularity encountered in Actinobacteria and Myxobacteria is a simple type of cell aggregation. In *Streptomyces* genus, hyphae are formed by germinating spores. Hyphae have multinuclear aerial mycelium, having septa at intervals. This creates a chain of uninucleated spores (Table 6.1 xvii). Multicellularity encountered in Cyanobacteria, Chloroflexi and some proteobacteria such as *Beggiatoa* is in the filamentous form arising due to cell division and adhesion (Table 6.1 xv). In phylum Planctomycetes, membranes divide the interior of the cell into separate compartments, condensed nucleoid surrounded by membranes (Table 6.1 xviii). Like Planctomycetes, phyla Verrucomicrobia and Chlamydia too show the presence of internal membranes which, though superficially similar to the endomembranes possessed by eukaryotes, bear no homology to the latter and represent a case of convergent evolution.

Based on the cell wall structure and the ability to take up the Gram stain, bacteria are divided into two groups, Gram-positive and Gram-negative (Fig. 6.1). Table 6.2 below compares some of the attributes of Gram-positive and Gram-negative bacteria.

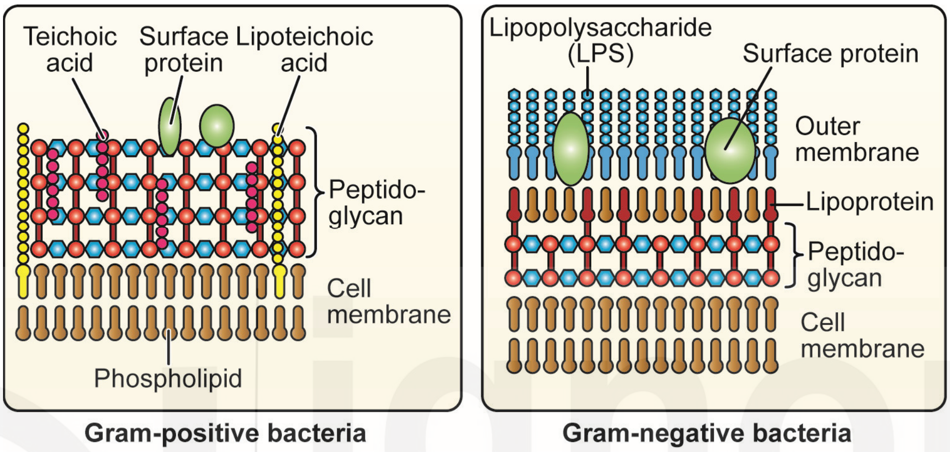


Fig. 6.1: Diagrammatic representation showing cell wall differences between Gram-positive and Gram-negative bacteria.

Table 6.2: Differences between characteristics of cell wall of Gram-positive and Gram-negative Bacteria.

Characteristic of cell wall	**Gram-positive bacteria**	**Gram-negative bacteria**
Gram staining	Retain primary stain after alcohol treatment	Lose primary stain after alcohol treatment
Number of layers	Single; outer membrane absent	Double; outer membrane present to the outside of peptidoglycan layer
Peptidoglycan layer	Thick, about 20-80 nm	Thin, about 5-10 nm
Periplasmic space between cytoplasmic and outer membrane	Small, if present	Large
Teichoic acid	Present	Absent
#Porins	Absent	Present
Lipid content	Very low (2-5%)	Very high (15-20%)
Lipopolysaccharide (LPS)	Absent (except for *Listeria monocytogenes*)	Present

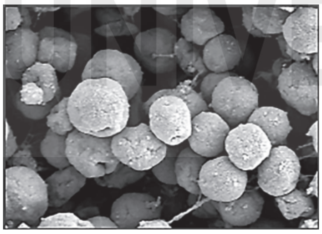
Susceptibility to lysozyme	Sensitive (lysis of cells)	Resistant (cells protected by the outer membrane)
Susceptibility to penicillin and detergents	*Sensitive	*Intrinsically more resistant
Examples of bacteria	*Bacillus anthracis, Staphylococcus aureus, Streptococcus pneumoniae, Enterococcus faecalis, Mycobacterium tuberculosis, Clostridium botulinum*	*Salmonella* species, *Shigella* species, *Haemophilus influenzae, Escherichia coli, Neisseria gonorrhoeae, Neisseria meningitidis, Pseudomonas aeruginosa*

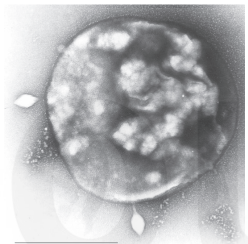
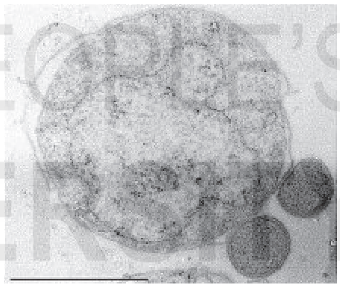
#Porins are surface proteins which form channels to allow influx of nutrient molecules (excluding antibiotics and inhibitors) and efflux of waste products.

* Although antibiotic-resistance genes in members of both the groups can be acquired through HGT, it is the cell wall architecture of Gram-negative bacteria which does not allow antibiotics to enter the cell easily.

Name Archaea has been derived from the Greek word *archaios* which means ancient or primitive and appropriately so as some members belonging to archaea do indeed show primitive characteristics. List of different phyla of archaea with their representative examples and their mode of life is shown in Table 6.3.

Table 6.3: List of different phyla of archaea with their representative examples and their mode of life.

Archaea	Example	Micrograph
Phylum Euryarchaeota includes methanogens and halophiles. Methanogens are the largest microorganisms belonging to archaea and show a cosmopolitan distribution, occurring mostly in anoxic environments such as aquatic sediments, rice paddies, anaerobic digesters, and the gastrointestinal tract of animals. They can grow at a wide range of temperatures (4^oC to 100^oC), salinities (freshwater to brine) and pH (3 to 9). They possess several unique coenzymes. Because of the unique cell wall construction, they are not susceptible to penicillin	*Methanogens:* Produce methane as a by-product of metabolism, which causes flatulence in humans and other animals. *Methanosaeta* is acetotrophic, using acetate fermentation to produce methane.	 **(i) Methanogenic microoganisms**

and other antibiotics. They are chemoautotrophic, H_2 serves as the source of both energy and electrons, and CO_2 works both as an electron sink and the source of cellular carbon.		
Halophiles grow best in environments having salt concentrations between 10% and 35% due to the presence of adaptations like efficient ion pumps, UV absorbing pigments and specific proteins to withstand osmotic stress.	*Halobacterium* is found in waters of extreme salinity and contains bacteriorhodopsin in the membrane which imparts red colour to their blooms.	 **(ii) *Halobacterium***
Phylum Crenarchaeota This includes members which are sulphur-dependent extremophiles and perform an important function of carbon fixation.	*Sulfolobus:* Several species grow in volcanic springs at temperatures of 75 to 80^{o}C and at a pH between 2 and 3.	 **(iii) *Sulfolobus archaea***
Phylum Nanoarchaeota contains only one species, *Nanoarchaeum equitans,* the smallest known organism with a diameter of 400 nm (1/100th the size of *E. coli*). It has the smallest genome and lacks the genes required to synthesise amino acids, nucleotides, lipids and cofactors.	*Nanoarchaeum equitans* is an obligate symbiont on *Ignicoccus,* a genus of hyperthermophilic archaea found in marine hydrothermal vents.	 **(iv) *Nanoarchaeum equitans*** Small dark spheres of *Nanoarchaeum equitans* are in contact with the host cell, *Ignicoccus.*
Phylum Korarchaeota It represents one of the most primitive forms of life and its members have been found to occur only in the Obsidian Pool, a hot spring at Yellowstone National Park.	*Candidatus* Korarchaeum cryptofilum, is an ultrathin, filamentous heterotroph, metabolising peptides and protein into H_2 anaerobically.	 **(v) Korarchaeum cryptofilum**

BOX 6.1: Contributions of cyanobacteria on the evolution of life forms on earth.

Cyanobacteria, one of the oldest among prokaryotic phyla (was already in existence around 2.45-2.32 billion years ago [bya]), include a very large and diverse assemblage of 374 genera and 1468 species and have been found to occupy marine, freshwater, or terrestrial environments as well as varied climatic zones. Their evolution had a very profound influence on the evolution of subsequent life forms on the earth due to various reasons such as:

- **Earth's climate:** To begin with, the earth had a reducing atmosphere consisting of methane, carbon dioxide and water vapour. It is hypothesised that photosynthetic activity of cyanobacteria led to the production of oxygen as a by-product which was released into the sea water (Cyanobacteria have no membrane-bound organelles and folds in the outer membrane of the cell are the site of photosynthesis). Gradually as the amount of oxygen produced increased over a span of 200-300 million years, oxygen started escaping into the atmosphere where it could react with methane, gradually displacing it and converting earth's atmosphere into an **oxidising** one. This event occurred probably between 2.4- 2.1 bya and is referred to as the "Great Oxidation Event".
- **Ice age:** Displacement of the greenhouse gas, methane, to a large extent, by oxygen led to cooling down of temperatures globally resulting in the formation of widespread ice sheets (one of the earliest ice ages) on earth.
- **Ozone layer:** UV radiation from the sun by acting on the atmospheric oxygen led to the generation of an **ozone layer** near the upper part of the atmosphere. Protective effect of the ozone layer by blocking harmful UV radiation from reaching the earth allowed colonisation of the surface of the ocean and ultimately the land by living forms.
- **Aerobic metabolism:** It is postulated that anaerobic organisms already inhabiting the earth were adversely affected by the presence of oxygen contributed by the photosynthetic activity of cyanobacteria, around 2.4-2.1 bya, wiping them out in large numbers. This was followed by the evolution of **aerobic metabolism** by organisms, where oxygen served as the final electron acceptor to generate energy after nutrient breakdown as well as evolution of enzymes to detoxify reactive oxygen species arising due to aerobic metabolism. Aerobic respiration paved the way for the evolution of more complex life forms.
- **Biogeochemical flux:** Being primary producers, and present in great abundance and diversity, various species of cyanobacteria play an important role in the biogeochemical cycling of carbon, nitrogen, and phosphorus in a very significant way besides being an important link in food webs.
- **Evolution of chloroplasts:** Endosymbiosis of a cyanobacterium within a unicellular eukaryote is postulated to have given rise to **chloroplasts** found in algae and plants.

On the basis of analysis of dataset of 16S rRNA sequences of several cyanobacterial species, researchers have postulated that evolution of **multicellularity** occurred around the time of "Great Oxidation event". Multicellular level of organisation in the form of filaments might have proved advantageous by improving motility and metabolic fitness as compared to single cells and could have contributed to abundance and diversification of cyanobacteria into newer ecological niches. Filamentous cyanobacteria exhibit directional growth as well as intercellular communication and exchange of resources indicative of evolution of division of labour and terminal cell differentiation.

6.2.2 Bacterial and Archaeal Diversity Based on Preferred Range of Environmental Conditions

According to the preferred habitats, bacteria and archaea can be classified into various groups as shown in Table 6.4 below.

Table 6.4: Classification of bacteria and archaea into various groups based on their preferred habitats.

Category	Preferred range	Examples of Bacteria	Examples of Archaea
Mesophiles	25°C-40°C and pH 5-9	*Escherichia coli, Listeria monocytogenes, Staphylococcus aureus*	*Methanobrevibacter smithii*
Thermophiles	50-55°C	*Thermus aquaticus Thermus thermophilus*	
Hyperthermophiles	80°C	*Thermotoga maritima, Chloroflexus sp., Thiobacillus sp.*	*Methanopyrus kandleri* *Sulfolobus solfataricus Pyrodictium abyssii* [3]*Pyrodictium occultum* [3]*Pyrolobus fumarii*
[1]Psychrophiles	10-15°C	*Moraxella sp., Pseudomonas sp., Vibrio sp., Flavobacterium sp., Bacillus sp.*	[5]*Methanococcoides burtonii, Methanogenium frigidum*
[1]Psychrotrophs	15-25°C	*Pseudomonas sp., Acinetobacter sp., Listeria monocytogenes*	
Piezophiles	[2] High hydrostatic pressure in the depths of ocean, 38 megapascals (MPs)	*Shewanella benthica, Photobacterium profundum, Psychromonas sp.*	[4]*Pyrococcus yayanosii*
Acidophiles and thermoacidophiles	pH 0.5-5.0	*Thiobacillus ferrooxidans*	[8]*Picrophilus oshimae* *Pyrodictium abyssi,* *Ferroplasma acidiphilum*

Halophiles	15-23% salt concentration	*Halorhodospira halophila*, [6]*Salinibacter ruber*	[6]*Halobacterium salinarum, Haloferax mediterranei*
Alkaliphiles and Haloalkaliphiles	High salinity and pH 9-12	*Bacillus halodurans*	[7]*Natrialba magadii*, [7]*Halorubrum vacuolatum*

Hyperthermophiles show optimal growth at 80°C or higher but are also capable of growing at temperatures up to 105°C while thermophiles grow optimally at temperatures ranging from 50° to 70°C.

[1]can grow even at 0°C.

[2] can be accompanied by either high temperature as in hydrothermal vents (thermophiles) or permanently cold conditions as in the deep oceans (psychrophiles).

[3] grow fastest at 105°C and can withstand autoclaving for one hour at 121°C.

As most acid environments have high temperatures, thermophiles and acidophiles may be grouped together.

[6]grows in saltern crystallizer ponds.

[7]grow in Lake Magadi, pH of the brine being 10.

[8]can grow in 1.2M sulphuric acid at 60°C.

[4] can tolerate pressures up to 150MPs (as compared to a pressure of 0.1 MPs at sea level) and is a thermophile, found near hydrothermal vents.

[5] resides in Ace lake in Antarctica at temperature of 1 to 2°C.

Not all archaea are extremophiles, and are found to play important ecological roles in a variety of habitats. For instance, organisms belonging to Crenarchaeota are quite abundant in soil (where they function to oxidise ammonia) as well as in the oceans of the world constituting a good proportion of the planktonic organisms.

Organisms belonging to Euryarchaeota include those dwelling in deep-sea sediments where they function to oxidise anaerobically methane stored in these sediments as well as methanogens living in terrestrial anaerobic environmental conditions which are responsible for a large percentage of methane emissions globally.

Box 6.2: Prokaryotic adaptations under different environmental conditions.

Biomolecules, especially proteins, generally remain stable and functional within a narrow range of temperature and pH conditions. As a number of prokaryotic species are extremophiles, several adaptations have arisen which allow the cellular proteins to remain stable and active under the extreme conditions.

- **Thermophilic adaptations of proteins:** At high temperatures, proteins without appropriate modifications show changes in their folding patterns which expose the hydrophobic cores resulting in aggregation and loss of activity. Proteins in these organisms show higher numbers of hydrophobic amino acids, disulphide bonds and ionic interactions so as to retain protein structure and activity at high temperatures.

- **Acidophilic adaptations of proteins:** Species inhabiting acidic environments maintain cytoplasmic pH of 5.0 to 6.5 by pumping protons out of the cell. As under conditions of low pH many charged polar amino acids become protonated changing their charges, enzyme structure can get disrupted making the enzymes non-functional. Acidophiles possess enzymes and other proteins which can retain optimal structure and catalytic function under highly acidic conditions due to the predominance of acidic amino acids (glutamic and aspartic acid) on the surface of these enzymes and proteins. As most of the acidophiles experience high temperature, their proteins show thermophilic adaptations too.
- **Halophilic adaptations of proteins:** Proteins show a higher content of acidic amino acids which confer increased negative charge on the surface of proteins. Thus, extreme ionic conditions can be compensated and stability as well as activity of proteins is maintained.
- Membrane system of halophiles allows them to pump potassium in and to pump sodium out to cope up with extreme osmotic stress and maintain osmotic balance. In different species of halobacteria, intracellular concentration of potassium may have a range of 1.2 to 4.5 M.
- **Psychrophilic adaptations of proteins:** Reduced hydrophobic core and less charged surface of proteins ensures flexibility and functionality at low temperatures.
- **Piezophilic adaptations of proteins:** Piezophilic adaptations of proteins include possession of a dense, compact hydrophobic core, predominance of smaller amino acids capable of forming hydrogen bonds with a concomitant decrease in large hydrophobic residues and a multi-unit organisation with individual monomers tightly packed helping to protect hydrogen bonds. This pattern of amino acid residues allows the protein to pack very tightly conferring stability to the protein structure under high pressure. Thermophilic adaptation of proteins employs higher content of basic amino acids like arginine.

Temperature, pH, and hydrostatic pressure also affect the function of the cell membrane by impacting the fluidity and permeability of the cell membrane for diffusion of nutrients. Therefore, organisms try to maintain membrane integrity continuously despite variations in external environmental conditions by modifying lipid composition of their membranes. For example, psychrophilic bacteria tend to have higher proportion of unsaturated and shorter-chain fatty acids in their plasma membrane so as to maintain flexibility at low temperature while thermophilic bacteria contain higher proportion of saturated fatty acids in their cell membrane so as to remain stable at high temperature. Thermophilic and hyperthermophilic archaea achieve higher thermostability of their cell membranes by having repeating subunits of the C5 compound, phytane (branched saturated isoprenoid) joined by ether linkage (as against ester linkage of phospholipids found in bacterial cell membranes) (Fig. 6.2) as well as monolayer structure constituted by membrane-spanning tetraether lipids in order to avoid separation of inner and outer layers of a membrane bilayer which can occur at very high temperature. This structure is almost impermeable to ions and protons helping to maintain membrane integrity under conditions of extreme temperature and acidity.

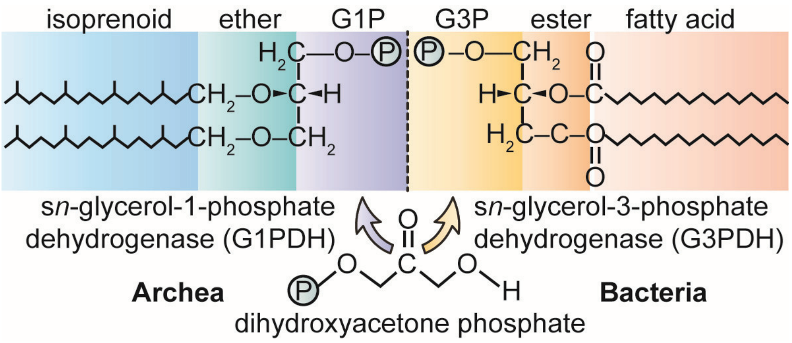


Fig. 6.2: Comparison between membrane lipids of archaea and bacteria.

6.2.3 Bacterial and Archaeal Diversity Based on Metabolic Properties and Production of Secondary Metabolites

Metabolic properties such as the ability to use different sources of energy, specific requirements of growth factors, ranges of various environmental factors conducive to growth and susceptibility to different classes of antibiotics have been used to distinguish different species of the microorganisms. Production of secondary metabolites is employed by several microbial species in order to compete with other microorganisms in their surroundings. As production of most secondary metabolites has been found to be species- and/or strain- specific, it can be used to classify different species (Table 6.5).

Table 6.5: Secondary metabolites produced by bacteria.

Organism	**Substance/metabolite produced**
Streptomyces avermitilis	Antibiotic Avermictin
S. griseus	Antibiotic Streptomycin
S. bingchenggensis	Anthelmintic Milbemicin
S. venezuelae	Antibiotic Chloramphenicol
S. kanamyceticus	Antibiotic Kanamycin
Clostridium botulinum	Neurotoxin botulinum
C. tetani	Exotoxins, tetanolysin and tetanospasmin
C. perfringens	Enterotoxin (CPE)
Sorangium cellulosum So ce56	Epothilone
Angicoccus disciformis	Myxochelin A

6.2.4 Bacterial and Archaeal Diversity Based on Molecular and Genetic Features

Molecular and genetic features have been studied to investigate microbial diversity using several approaches such as nucleic acid hybridisation, DNA cloning and sequencing, RFLP and other PCR-based methods. Molecular approach for studying diversity, being sequence-specific, also allows us to gain an insight into the evolutionary relationships among the species sampled. Phylogenetic analysis using the 16S rRNA gene has been very extensively used in the classification of different species of prokaryotes (Fig. 6.3). Several reasons support the use of the 16S rRNA gene: every prokaryotic genome has at least one copy of this gene, sample identification using PCR is possible due to the presence of conserved regions in this gene and information provided by its sequence has been found to be reliable in microbial classification.

Woese's delineation of Archaea as a third domain of life, bacteria and eukarya being the other domains, is supported by molecular genomics and phylogenetics data. According to these studies, hyperthermophilic and non-methanogenic ancestors could have given rise to present-day archaeal lineages followed by divergence into two major phyla, the Crenarchaeota and the Euryarchaeota (Table 6.3).

Based on analysis of 16S rRNA gene homologies, one study estimates that there may be 1.4-1.9 million extant bacterial lineages. Although single gene comparisons have been useful, new approaches such as whole genome sequencing and generation of metagenomic data (analysis of all the DNA present in a given sample), are being increasingly employed to gain comprehensive knowledge of prokaryotic diversity and genetic relationships. Availability of whole genome sequences not only allows determination of orthologous genes but also presence or absence of any specific genes in any particular genome (giving an insight of metabolic functions of the organism). For instance, information about functional diversity and identification of species can be elucidated by using specific genes coding for enzymes such as nitrogenase, nitrate reductase etc. in DNA microarrays combined with DNA-DNA hybridisation.

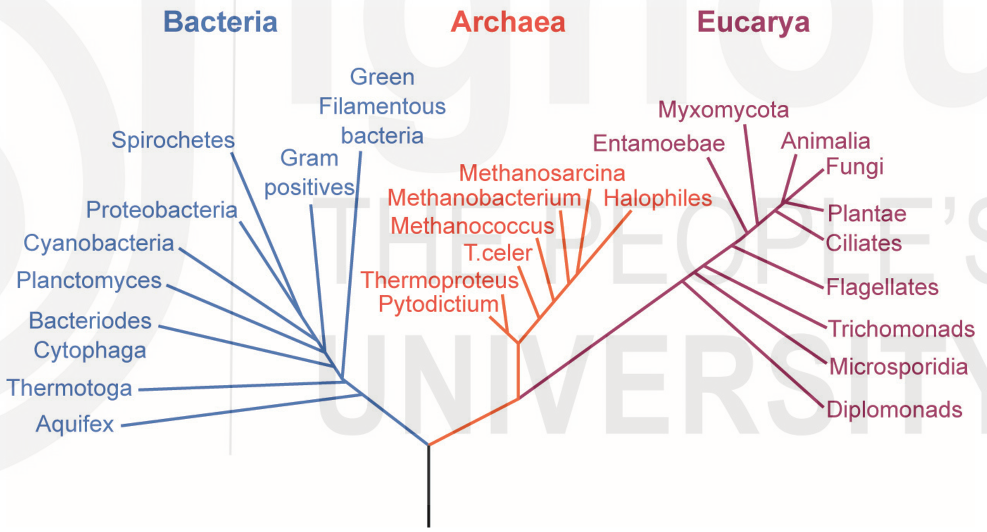


Fig. 6.3: Phylogenetic tree of life based on rRNA sequences.

SAQ 1

a) Write true or false against the following statements:

i) Bacteria have two layers of phospholipids in their cell membranes.

ii) Gram-positive bacteria have a single cell wall surrounded by an outer membrane containing lipopolysaccharides.

iii) The cell wall of Gram-positive bacteria is thick and composed of peptidoglycan.

iv) Porins allow entry of substances into only Gram-positive bacteria.

v) Cell wall of Gram-positive bacteria is anchored to the cell membrane by lipoteichoic acid.

b) State the appropriate reason for the following statements:

i) Psychrophiles seldom cause disease.

ii) Food can get spoiled even under refrigerated and freezing conditions.

iii) *Pyrolobus fumarii* can survive at temperatures as high as 113°C.

iv) Presence of multiple origins of replication in several species of archaea is advantageous.

v) Infections caused by *Mycoplasma* species continue to persist despite prolonged treatment with beta-lactam antibiotics.

6.3 NATURE OF GENOMES

6.3.1 Bacterial Genomes

Complete set of genetic information contained in the DNA of the chromosome of an organism constitutes its genome. Clear understanding of the biology and evolutionary relationships of various bacterial species can be achieved from the sequencing data of bacterial genomes. Additionally, genome sequencing helps in clarifying the basis of antibiotic-resistance and pathogenicity of different species of bacteria as well as in identifying novel targets of antibiotics. This is of immense importance as the majority of pathogenic bacteria have evolved resistance to most of the commonly used antibiotics.

Haemophilus influenzae was the first bacterium whose genome was sequenced by J. Craig Venter in 1995. Subsequently, as sequencing technologies improved, more and more genomes could be tested with the result that at present nearly 30,000 sequenced bacterial genomes from 50 different phyla are publicly available. Some of the medically important species whose genomes have been completely sequenced include *Streptococcus pneumoniae, Mycobacterium tuberculosis, Escherichia coli* O157:H7, *Vibrio cholerae, Clostridium difficile* and *Staphylococcus aureus.* Rapid development of newer technologies for next generation sequencing is also greatly contributing to the bacterial genome data. Some of the main features that have emerged from bacterial genome data are briefly described below.

Majority of bacteria have a single, circular chromosome consisting of a double-stranded DNA molecule which is compacted into a structure called nucleoid, there being no nuclear membrane and hence bacteria are categorised as **prokaryotes**. Proteins facilitate packing of the chromosome into a nucleoid which occupies nearly ⅓ space of the cell's interior. Since the length of the chromosome (approximately 1.5 millimetres) is nearly 500 times the size of the bacterial cell (approximately 1-2 micrometres in length), supercoiling and tight packing of the chromosome into the nucleoid area leave space within the cell for other processes like cell metabolism and protein synthesis to take place.

However, there are a few exceptions too as some bacteria such as *Vibrio, Burkholderia, Leptospira, Brucella and Deinococcus* species have two or more chromosomes and *Borrelia burgdorferi* has a linear chromosome rather than a circular one.

As there is only a single chromosome in the majority of bacterial species, they are **haploid** and there is no possibility of masking the recessive alleles. However, about 10% of bacterial species have multiple copies of individual chromosomes i.e., they are polyploids e.g., *Synechococcus elongatus, Azotobacter vinelandii, Deinococcus radiodurans, Sinorhizobium meliloti.*

Typical genome size of bacteria is 5 MB, though the size varies greatly in different bacterial species and even among different strains within a species; 5.1 Mbp found in *Bacillus megaterium,* 13,033,799 bp in *Sorangium cellulosum* So ce56 and 14,782,125 bp in *Sorangium cellulosum* strain So0157-2; *Haemophilus influenzae* HK1212 has 1.0Mbp as against the strain F3047 with 2.0 Mbp; *Burkholderia pseudomallei* THE has 6.3 Mbp as against the strain MSHR520 with 7.6 Mbp. *Deinococcus radiodurans* R1 has two chromosomes with 2,648,638 and 412,348 bp, occurring in multiple copies and a megaplasmid composed of 177,466 bp and a small plasmid having 45,704 bp, total genome amounting to 3,284,156 bp.

An obligate symbiont, *Nasuia deltocephalinicola* strain NAS-ALF has the smallest genome with only 112,091 bp. *Tremblaya princeps* has the second smallest genome consisting of 139,000 bp.

Thus, genome size in bacteria varies due to acquisition of genes by horizontal gene transfer or loss of functional accessory genes due to long term association with the hosts.

BOX 6.3: Plasticity of Bacterial Genomes.

Examination of data on bacterial genomes indicates a major pattern in genome size across different species of bacteria. In general, genome size tends to be larger in free-living species than parasitic species which in turn have larger genomes than obligate pathogens. For instance, studies have shown a progressive reduction in genome size of rickettsial spp. from 1.5 to 1.1 Mb, with the greatest reduction and degradation of genome encountered in the most virulent species when compared with the closely related non-pathogenic species. This trend of convergent evolution shown by a number of bacterial species pathogenic in humans is associated with a selective loss of non-essential genes encoding enzymes required for amino acid synthesis and other pathways involved in ATP, LPS and cell wall component biosynthesis but selective retention of genes coding for recombination and DNA repair proteins and toxin-antitoxin modules needed for evasion of host immune response as well as selective expansion due to plasmids, short palindromic elements, ADP-ATP translocases, type IV secretion system and duplication of gene families, enabling better adaptation to the host-environmental conditions.

However, several studies document the dynamic nature of bacterial genomes as loss or gain of genetic material through various mechanisms (Unit 7) can affect genome size even within related strains. Genomes of pathogenic strains of *Enterococcus faecalis* were reported to be 25% larger than those of commensal strains due to the presence of plasmids, phages and pathogenicity islands (acquired through HGT) (see Unit 7 for HGT mechanisms), responsible for their rapid evolution in an antibiotic-treated patient. Thus, genetic content even within a species varies between non-pathogenic strains and clinical isolates. This has also been confirmed by a study comparing the genome sizes of 3 different *E.coli* strains: *E. coli* K-12, non-pathogenic, 4,639,221 bp.; *E. coli* O157:H7 strain EDL933, enterohemorrhagic, 5,528,445 bp. and *E. coli* strain CFT073, uropathogenic, 5,231,428 bp.

In the large genome (>2 Mb) of *Rickettsia* endosymbiont of *Ixodes scapularis* (REIS), transposons, insertion sequences and other mobile genetic elements constitute nearly 35% of the total genome.

If multiple genomes of the same species are available, calculation of **pan genome** (entire set of all the genes from all strains of a clade) and **core genome** (set of homologous genes present in all the tested genomes) can be achieved to derive information and clearer understanding about species' relatedness and evolution. Comparative analysis of 2000 *E. coli* genomes revealed that the core genome of *E. coli* had nearly 3100 gene families (found in all *E. coli* genomes) and about 89,000 different gene families encountered in various strains and any one *E. coli* bacterium contains less than 10% of the total number of *E. coli* genes in the *E. coli* pan-genome.

Nucleotide composition also shows variation between species: the G+C (guanosine-cytosine) content is relatively uniform within a bacterial genus or species and shows a range of about 25% in *Mycoplasma* species to about 75% in some *Micrococcus* species. In general, analysis of bacterial genomes indicates that bacteria occupying complex environmental niches tend to have larger genomes with a higher GC content as compared to host-associated bacteria.

There are about 2,500 genes in a typical bacterial genome which code for all the metabolic requirements needed for survival; in addition to chromosome, bacteria also have plasmids which confer survival advantage to bacteria against one or more antibiotics, (*Bacillus megaterium* strain QMB1551 harbours seven plasmids, the largest plasmid array in a single bacterial strain). *B. megaterium,* strain QMB1551 while carrying 5300 genes on the chromosome has additional 523 genes located on plasmids. *Sorangium cellulosum* So ce56 with a very large genome contains 9367 protein-coding genes while *Sorangium cellulosum* strain So0157-2 with the largest genome analysed so far contains 11,599 genes. Small genomes are found in mycoplasmas with only about 500 to 1000 genes. It has been postulated that mycoplasmas evolved by degenerative evolution from Gram-positive bacteria and show phylogenetic relationship with some clostridia. *Nasuia deltocephalinicola* strain NAS-ALF with the smallest genome encodes just 137 proteins.

Genes in prokaryotes are arranged in operons which have evolved over time due to the working of natural selection. Coordinate control of all the genes which code for proteins that are related functionally serves to conserve energy for the cell. For example, most of the bacterial genes coding for ribosomal proteins are clustered in a few operons (present near the origin of replication) which allows for coordinated regulation. Placed next to the large RP cluster in the genome are several genes involved in the synthesis of proteins and chaperones.

Gene order and gene content are found to remain quite stable among related species though mobile genetic elements may bring about changes in genome architecture by introducing insertions, inversions, duplications, and translocations.

Bacterial genomes are gene rich, 88% of the genome being represented by protein-coding regions and there are also intergenic spaces (though very limited in extent as compared to eukaryotic genome), repeated elements and inactivated or otherwise functionless genes.

Whole genome sequencing allows to decipher completely the biology of an organism. For instance, the genome of *Streptococcus pneumoniae* TIGR4, a virulent strain (serotype 4, ST 205) has been completely sequenced and has revealed important information about this pathogen. The genome of this strain of *S. pneumoniae* consists of 2,160,837 base pairs and 2,236 putative genes which include several ATP dependent transporters. Some of these transporters facilitate sugar transport indicative of ecological adaptation of the bacteria to sugar rich environments of the oral cavity. Sequencing also revealed a 13-gene cluster responsible for capsular biosynthesis, important for virulence of the bacterium.

BOX 6.4: *Deinococcus radiodurans,* the bacterium with extraordinary capabilities.

Arthur W. Anderson first discovered *Deinococcus radiodurans* in 1956 when he observed red bacteria in a can of ground meat which despite having been sterilised with a megarad range of radiation, had got spoiled. Colonies appear red because of the presence of carotenoid pigments within the bacteria.

D. radiodurans is not only capable of tolerating ionising radiation of very high intensity at a dose of 5,000 Gy (1000 times the lethal dose for a human, and which can kill virtually all other microorganisms), but also desiccation, UV radiation, oxidising agents and electrophilic mutagens.

As there have not been any highly radioactive habitats on the earth over large geologic times, radiation does not seem to be responsible for the evolution of resistance to continuous exposure to radiation in organisms. However, other factors such as UV radiation, oxidising agents or alternating periods of desiccation and hydration or high and low temperatures are quite effective in causing DNA damage and it is postulated that these organisms evolved efficient mechanisms to repair damage to DNA in response to these factors which incidentally also happened to bestow protection against radiation.

Due to the presence of highly expressed genes for several unique proteases and for detoxification and chaperones, *D. radiodurans* is able to maintain integrity of its essential macromolecules like nucleic acids, enzymes and proteins and repair damaged DNA very efficiently. Large chromosome of *D. radiodurans* contains duplicate copies of most of the highly expressed genes of the small chromosome.

16S rRNA analysis reveals that Deinococcus along with Thermus represents an ancient group distinct from other bacterial lineages. *D. radiodurans* lacks the conventional phospholipids found in other bacteria; instead, its membrane has phosphoglycolipids containing alkylamines, not found in other bacterial lineages; a small set of genes found only in *Thermus* and *Deinococcus* (and not in majority of bacteria) have subunits which are found in archaeal genes. Similarities between members of phyla *Thermus-Deinococcus* and the archaeal members in terms of genetics as well as mode of life in extreme conditions indicate that *Thermus-Deinococcus* members may be the closest living relatives of the archaeal species.

By engineering *D. radiodurans* to express the functions of detoxification/degradation of metal and organic compounds, scientists are trying to explore the possible use of this bacterium for cleaning the radioactive waste sites.

In prokaryotic genomes, a gene whose codon usage is very similar to that of the genes encoding ribosomal proteins, translation and transcription processing factors, principal energy metabolism proteins and the chaperone-degradation proteins, and deviates from that of the average genes is referred to as **predicted highly expressed (PHX)**.

Synechocystis genome contains more than 30 PHX genes essential for photosynthesis. *Helicobacter pylori* genome contains the PHX genes encoding urease alpha, urease beta and urease I, which enable the bacterium to convert urea from gastric juices into bicarbonate and ammonia, neutralising highly acidic environment that prevails in the stomach. The *H. pylori* genome is also rich in PHX genes which code for a family of outer membrane proteins (nearly 32 members).

Chlamydia trachomatis and C. pneumoniae are mammalian obligate intracellular parasites and contain PHX gene coding for ATP-ADP translocase which is also present in *Rickettsia* but absent from all other bacteria.

Genome of *Treponema pallidum,* (which causes syphilis in man), has the highest number of PHX flagellar genes whose products ensure its survival by facilitating its movement and spreading to all parts of the body including the brain.

Each region seems to follow its own evolutionary rate (gene-coding regions may be slow to change as against intergenic and repeated sequences).

Gene order and nucleotide sequences of the genome may be conserved among different strains as shown for some bacteria or there may be variation among even related strains due to various factors such as transposon-mediated deletions/insertions, rearrangements and through horizontal gene transfer (you will be studying this in detail in Unit 7).

Genomes of a large number of bacterial species contain genes with defence functions, especially the restriction-modification system which provides innate immunity against viral infections and CRISPR-associated proteins (CRISPR-Cas) (found in nearly 40% of bacterial species) which confer upon bacteria adaptive immunity against viral infections (described in more detail in Unit 7).

6.3.2 Archaeal Genomes

Methanocaldococcus. jannaschii DSM2661 was the first archaeal genome to be sequenced in 1996. Subsequently genomes of other archaeal species from 11 different phyla have also been sequenced revealing important features of the archaeal genomes. Extensive variation of genome size and composition has been documented in archaea.

Genomes of archaeal species consist of a single copy of a circular chromosome with size ranging from 0.5 to 5.8 Mbp. Organisms belonging to Nanoarchaea have the smallest known genome as exemplified by the genome of *Nanoarchaeum equitans* which consists of only 490885 bp., is very compact and although it codes for complete machinery required for information processing, there are no genes for lipid, cofactor, amino acid and nucleotide biosynthesis due to its symbiotic mode of life.

Some species are polyploid and may contain many copies of their chromosome. For example, *Methanocaldococcus jannaschii, Methanococcus maripaludis, Haloferax volcanii* and haloarchaeal isolates from salt deposits. Some of the postulated possible advantages of polyploidy in these species include repair of double-strand breaks, improved survival under prolonged stressful environment, genome equalisation by gene conversion, genomic DNA serving as a phosphate reservoir.

Symbionts which derive their nutrition from a host tend to have smaller genomes due to loss of genes which are not essential for this mode of life.

In the genomes of the majority of archaeal species there is a single origin of replication (regions which are particularly rich in AT nucleotides with conserved sequences termed origin of recognition boxes) which proceeds bidirectionally as in bacteria, but there are exceptions. For example, the genome of *Sulfolobus solfataricus* has three, *Aeropyrum pernix* has two and *Pyrobaculum calidifontis* has four and *Halobacterium* sp. NRC-1, *Haloferax volcanii and Haloarcula hispanica* have multiple origins of replication. Presence of multiple origins of replication in the genome is advantageous as it reduces the replication time.

G+C content in archaeal genomes varies from 28 to 66 mol.% but seems to show no correlation with the temperatures conducive to optimal growth of the organisms. Instead, DNA stability is achieved by counterions and DNA-binding proteins and histones (in most archaeal species) which can compactly pack the genome and overcome the effect of environmental temperature. Some archaeal species show acetylation-dependent compaction and organisation of the genome.

Like bacterial genomes, archaeal genomes are also gene-rich, (nearly 2000 genes) with only minimal noncoding regions in between. Short introns have been found to occur in some protein-coding genes and tRNA genes in Crenarchaeota species but none have been reported from genes in species belonging to Euryarchaeota.

Just like bacterial genomes, insertion sequences, transposable elements and other mobile genetic elements are present in almost all archaeal genomes. For instance, *Sulfolobus solfataricus* shows a lot of metabolic diversity due to the production of a variety of catabolic enzymes attributed to several hundred mobile elements detected in its genome.

Genomes of most of the archaeal species encode histone proteins, which assemble as nucleosome-like structures and are different from the canonical histone octamers found in eukaryotes.

Genomes of a large number of archaeal species contain a number of genes with defence functions, especially the restriction-modification system which provides innate immunity against viral infections and CRISPR-associated proteins (CRISPR-Cas) (found in nearly 80% of archaeal species) which confer upon archaea adaptive immunity against viral infections (described in more detail in Unit 7) and toxin-antitoxin modules which are responsible for induction of apoptosis following phage infection. Such defence genes tend to be organised as genomic islands and horizontal gene transfer seems to play an important role in the maintenance and evolution of these defence islands.

Sequencing of the genome of *M. jannaschii* revealed that while genes with a function in energy production, cell division, and metabolism bore similarity to their counterparts in bacteria, genes involved in replication, transcription and translation were more similar to those found in eukaryotes than to those of bacteria. Its genome houses more than 20 highly expressed genes which are involved in methanogenesis.

SAQ 2

Write true or false against the following statements:

a) *Synechococcus elongatus* is a polyploid archaeal species.

b) In general, bacteria occupying complex environmental niches tend to have larger genomes with a higher GC content as compared to host-associated bacteria.

c) Genome architecture (gene order and gene content) in bacteria remain unaffected by mobile genetic elements.

d) Genomes of most of the archaeal species encode histone proteins.

e) Genomes of 80% of archaeal species contain a number of genes encoding the restriction-modification system and CRISPR-associated proteins which confer immunity against viral infections.

6.4 DIVERGENCE OF BACERIA AND ARCHAEA

6.4.1 Divergence of Replication System

The various steps in the process of DNA replication such as origin of synthesis, unwinding and stabilisation of DNA, initiation and priming, relaxation of supercoiling, repair and ligation rely on a number of products encoded by several genes. Thus, the replication process is quite complex even in seemingly simple prokaryotes.

Though all the three domains of life show a conserved process of DNA replication, modifications of certain proteins involved in the process are evident in each domain of life as shown in Table 6.4. Archaea present a mosaic nature of the information processing system, resembling bacteria in some features while sharing certain other characteristics with eukaryotes.

Table 6.4: Comparison of various components of DNA replication in bacteria, archaea, and eukaryotes.

Attribute	Bacteria	Archaea	Eukarya
Origin of replication	Single	Single/more than one	Multiple
Recognition of origin of replication	DnaA protein	^Orc1/Cdc6 proteins	Origin recognition complex formed of 6 different proteins
Replicative helicase to unwind dsDNA	Homohexamer DnaB with 5'-3' unwinding polarity	*MCM complex, dimer of hexamers, with 3'-5' unwinding polarity and GINS complex (in some)	*MCM complex, a heterohexamer, with 3'-5' unwinding polarity and GINS complex

ssDNA binding proteins	Homodimers or homotetramers of SSB	Diverse across different species. *Pyrococcus furiosus* has heterotrimers of **RPA.	Heterotrimers of **RPA
Primase (DNA-dependent RNA polymerase)	Single subunit protein #DnaG	Homologs of bacterial DnaG primase as well as two-subunit protein (small catalytic *^PriS and large PriL)	Two-subunit protein, (small catalytic PriS and large PriL) complexed with DNA polymerase alpha

^bear homology with the corresponding proteins in eukaryotes

* minichromosome maintenance complex

** replication protein A

DnaG functions in RNA degradation

*^PriS is involved in primer synthesis and DNA repair in archaea.

Though DNA binding proteins, HU and H1, rich in positively charged amino acids, are associated with the DNA in bacterial chromosomes, they do not function to compact the DNA and the bacterial chromosomes are fully functional, capable of being easily replicated and transcribed. Table 6.5 lists the DNA polymerases associated with the process of DNA replication in bacteria.

Table 6.5: DNA polymerases associated with the process of DNA replication in bacteria.

DNA Polymerase	**Properties**	**Function**
I	5'-3' polymerisation, 3'-5' exonuclease activity, 5'-3'exonuclease activity	Removal of the primer and filling the gaps by DNA synthesis and DNA repair
II (Family B)	5'-3' polymerisation, 3'-5' exonuclease activity, no 5'-3' exonuclease activity	DNA repair
III (Family C)	5'-3' polymerisation, 3'-5' exonuclease activity, no 5'-3' exonuclease activity. (Holoenzyme consists of ten polypeptides).	Essential for replication. Also functions in proofreading.
IV		DNA repair
V		DNA repair

Taq DNA polymerase obtained from *Thermus aquaticus* is a thermostable enzyme which shows optimal polymerisation activity at 75-80^{o} C. There are studies to show that *Taq* DNA polymerase bears 50% homology with the *E. coli* DNA Pol I and is placed in family A DNA polymerases. Though it shows 5'-3' exonuclease activity it lacks proofreading activity due to the absence of 3'-5' exonuclease activity. When compared with Klenow Pol (large subunit of *E. coli* DNA polymerase I), Klentaq1 (fragment of *taq* DNA polymerase) has 19 opposite-charge substitutions indicating the role played by redistribution of charges in imparting thermostability to the enzyme.

Table 6.6 shows the DNA polymerases encountered in archaea.

Polymerase D (Pol D), present in most of the archaeal species, consists of two subunits, DP1 with exonuclease activity and DP2 with polymerase catalytic activity.

Catalytic core of archaeal replicative DNA polymerase (Pol D) (responsible for synthesis of both DNA strands) and the large subunits of the universal RNA polymerase (RNAP) required for transcription (in all three domains of life and several large DNA viruses) show homologous sequences. One model suggests that evolution of RNA polymerases and replicative DNA polymerases occurred from a common ancestor which had RNA-dependent RNA polymerase in the RNA-protein world that existed before the start of DNA replication. According to a number of studies, archaeal Pol D might have evolved from the replicative DNA polymerase of the Last Universal cellular Ancestor (LUCA) and subsequently other types of DNA polymerases like Pol A, Pol B and Pol C might have evolved.

Table 6.6: DNA polymerases encountered in archaea.

DNA Polymerase	Property	Function
Family B. PolB3, present in all except Thaumarchaeota. PolB1, present in Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota. PolB2 (inactive). (Encoded by archaeal chromosome as well as several mobile genetic elements including casposons)	Core consists of 3 domains- palm, fingers and thumb, N-terminal 3'-5'exonuclease domain and a uracil-recognition domain	PolB3 involved in the synthesis of the leading strand in Euryarchaea. In Crenarchaeota, two separate enzymes replicate the leading and lagging strands. PolB1 polymerases are replicative
Family D (except Crenarchaeota)	Consist of 2 subunits- small subunit DP1 having 2 domains (one binds ssDNA and second has 3'-5'exonuclease activity, and a multidomain large subunit DP2.	DP2 has polymerase activity, especially for the synthesis of the lagging strand in Euryarchaea. In *Thermococcus kodakarensis,* it can replicate both strands.

The components and the process of replication in archaea show some degree of similarity with those in Eukaryotes.

6.4.2 Divergence of Transcription System

Transcription system makes use of transcription factors and DNA-dependent RNA polymerase (RNAP).

In archaea, general transcription factors for recognition of core promoter include TATA-box binding protein (TBP), transcription factor B (TFB) and transcription factor E (TFE). Presence of homologous regions of helix-turn-helix units in archaeal TFB and bacterial sigma factors has led researchers to postulate that bacterial sigma factors might have evolved from TFB with some modification while TBP and TFE were lost during the course of bacterial evolution. Studies revealing homology between sigma factors and TFB have led to a better understanding of events leading to divergence between archaea and bacteria. Researchers have shown that in the promoters of the archaeal species *Sulfolobus solfataricus*, the promoter-proximal element (PPE) present upstream of the transcription start, that is (-11)-AATATTAA-(-4) bears similarity with regard to its sequence and placement to the bacterial Pribnow box, that is, (-12)-TATAAT-(-7) (Fig.6.4) and proposed that the Pribnow box of bacterial promoters might have evolved from the PPE sequence of archaeal promoters. The initiator element in both archaea and bacteria can be identified by RNA polymerase.

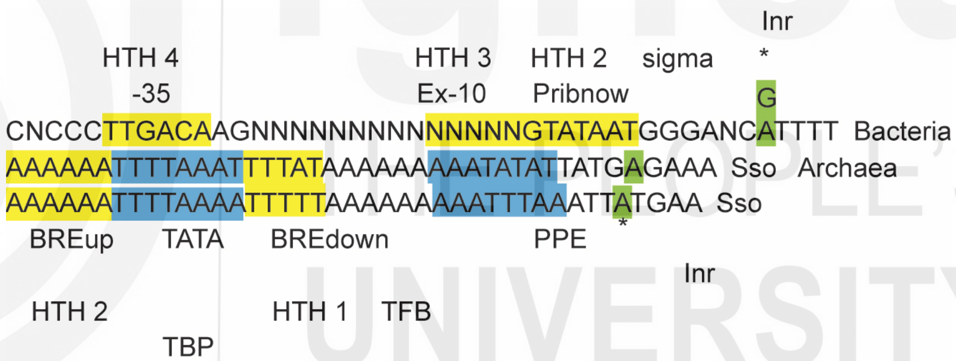


Fig. 6.4: Comparison between bacterial and archaeal promoters. Sso: *Sulfolobus solfataricus*; HTH: helix turn helix; PPE: promoter proximal element; TBP: TATA-box binding protein; TFB: transcription factor B; Inr: initiator element.

Transcription in bacteria is accomplished by a single form of DNA-dependent RNA polymerase (RNAP) consisting of five subunits, two alpha, beta, beta' and omega. The catalytic mechanism and active site for transcription are provided by beta and beta' polypeptides while two alpha subunits function in RNAP assembly and transcriptional regulation. Binding of the core enzyme with a sigma factor constitutes the holoenzyme. Variations in the polymerase holoenzyme are created due to the presence of several different forms of sigma factor. Sigma factor plays the key role of directing the polymerase to specific promoters.

Archaeal DNA-dependent RNA polymerase core enzyme consists of 10 to 12 subunits, designated as A'/A'', B'/B'' (for catalysis), L, N, D and P (for assembly) and F, E, H and K (for auxiliary functions). Several studies have shown that there is homology in sequence and structure between the subunits of bacterial RNAP and the subunits of archaeal RNAP. For instance, beta' subunit of bacterial RNAP shows homology to subunit A'/A'' of archaeal RNAP; subunit beta of bacterial RNAP bears homology with subunit B'/B'' of archaeal RNAP; alpha^{I} and alpha^{II} of bacterial RNAP shows similarity with subunits D and L respectively of archaeal RNAP and omega of bacterial RNAP shows homology to K subunit of archaeal RNAP.

Histone-based chromatin has been shown to play an important role in packaging of the DNA and regulation of transcriptional activity for various cellular processes in eukaryotes. Genomes of several archaeal species encode either histone proteins (in most of the archaea and can form nucleosome-like structures, homologous to those found in eukaryotes) or small basic proteins (analogous to those found in bacteria). However, unlike eukaryotes, linker histones or chromatin-remodelling complexes are not encoded by archaeal genomes nor do archaea show post-translational modification of histones.

6.4.3 Divergence of Translation System

Size and composition of archaeal ribosomes are similar to those of bacterial ribosomes, i.e., they consist of three RNA molecules, 16S, 23S and 5S RNA and 50-70 proteins (varies with species) which join to form a 70S ribosome particle. However, primary structures of ribosomal RNA (rRNA) and ribosomal proteins (RP) bear similarity to those found in eukaryotes rather than to those occurring in bacteria. In general, extremophilic archaea tend to have a more rigid structure of ribosomes as compared to ribosomes found in mesophiles. Halophilic species have acidic ribosomal proteins in order to increase their hydration capacity.

The archaeal genome generally encodes more RP genes as compared to bacterial genomes. The increase in RPs in archaea compared with that in bacteria may reflect a more complex set of interactions in archaea in regulating translation, e.g., differences in structure requiring scaffolding of longer rRNA molecules and expanded interactions with the chaperone machinery. Many genes in the translation category are shared by bacteria and archaea and these genes have been transmitted vertically, and not horizontally. Thus, translation proteins are highly conserved between bacteria and archaea. Bacteria and archaea also share genes for nucleotide transport and metabolism.

A study investigating ribosomal genes in complete genomes from a number of species reported that while thirty-four ribosomal (r) - protein families were represented in all domains of life, thirty-three families of RPs were found only in archaea and eukaryotes and with regard to protein composition, archaeal ribosome appeared to be a miniature version of eukaryotic ribosome. Another important observation of this study was that archaea had plasticity of the translation apparatus as during the course of archaeal evolution, there was a progressive reduction of ribosomal genes.

Researchers engaged in comparative studies of genetic code as well as the various nucleic acids and proteins involved in transcription and translation processes between bacteria and archaea have postulated that archaea represent a lineage that appeared on the earth earlier and from which arose more flexible, successful, and evolutionarily derived organisms constituting the bacterial domain.

6.4.4 Comparison of Characteristics of Bacteria and Archaea

Characteristic	Bacteria	Archaea
Occurrence	Cosmopolitan, found in terrestrial as well as aquatic habitats, some members are extremophiles (living in hot springs, radioactive waste water, organic matter, bodies of plants and animals.	Cosmopolitan, found in terrestrial as well as aquatic habitats, some members are extremophiles (living in extreme conditions such as hot springs, salt lakes, marshlands, oceans, Antarctica); some live within the gut of ruminants and humans.
Cell morphology	rod-shaped/spherical/spirals/coils	rod-shaped/spherical/spiral/coils
Cell wall composition	Peptidoglycan present (except *Chlamydias*, *Planctomycete* spp. and *Mycoplasma* spp.	No peptidoglycan; may have pseudopeptidoglycan, polysaccharides, glycoproteins or proteins
Cell membrane	Fatty acids linked to *sn*-glycerol-3-phosphate by ester linkage and lipid bilayers	Isoprenoid units linked to *sn*-glycerol-1-phosphate by ether linkage, lipid bilayers, though some species have lipid monolayers
Membrane-bound organelles	Absent	Absent
Nucleus	Absent	Absent
Chromosome	Circular	Circular
Origins of replication	Single	Multiple in some species
RNA polymerase	Single type, consisting of 5 polypeptides	Single type, consisting of 10-12 polypeptides
Initiator tRNA	Formyl-methionine	Methionine
Response to antimicrobial agents (which interfere with peptidoglycan biosynthesis). Novobiocin (targets DNA gyrase)	Susceptible Susceptible	Resistant Susceptible
Classical photosynthesis using chlorophyll	Some species can do	Absent
Methanogenesis	Absent	Occurs in Euryarchaeota

Nitrogen fixation, denitrification, chemolithotrophy, and hyperthermophilic growth	Present	Present
Pentose phosphate shunt	Present	Mostly absent
Reproduction	Asexual, by means of binary fission, budding and fragmentation. Some species are capable of sporulation and spores can remain dormant for several years.	Asexual, by means of binary fission, budding and fragmentation. Do not show sporulation.

BOX 6.5 : Contributions of prokaryotes in the evolution of eukaryotes.

The contributions made by cyanobacteria in the evolutionary history of various life forms on the earth is discussed in Box 6.1. Genomes record their own history and investigations on comparisons of prokaryotic genomes with eukaryotic genomes have revealed that prokaryotes have contributed to eukaryotic genomes to a very large extent.

Eukaryotic genes for biochemical and metabolic functions are indicative of bacterial origin. Bacteria have also contributed building blocks of the endomembrane system to the eukaryotic lineage.

Archaeal genes involved in information processing, i.e., the RNA polymerase, the ribosome and aminoacyl tRNA synthetases, and the genes for histones (which enable formation of nucleosomes and exert epigenetic control of gene expression) are conserved in eukaryotic lineage. Chromatin organisation based on histones was instrumental in generating the genomic complexity so characteristic of eukaryotes.

According to one study, archaeal contribution to the eukaryotic genome is about 44% while bacterial genes are nearly 56%. Land plants have a very high proportion of genes derived from bacteria (nearly 67%) and bacterial homologs in the yeast genome comprise nearly 75%.

By their association with ruminants, bacteria have influenced their evolutionary success.

SAQ 3

Fill in the blanks:

a) Sporulation is seen in …….. but not in………..

b) Both bacteria and archaea are susceptible to novobiocin as it targets ………..

c) Binding of the core enzyme RNAP in bacteria with …………constitutes the holoenzyme.

d) Archaeal ribosomes consist of RNA molecules ……… , ……..…. and ……………

e) In archaea, general transcription factors for recognition of core promoter include ………., ………. and ……………

6.5 APPLICATIONS

6.5.1 Medicine and Industry

- Several species of killed- or attenuated-bacteria are used in vaccines while tetanus vaccine is produced from the toxoid of *Clostridium tetani.*
- *Streptomyces* spp. (over 500) are the source of a large number of bioactive compounds such as antibacterial, antifungal, antiparasitic, immunosuppressants and extracellular enzymes.
- Secondary metabolites produced by *Sorangium cellulosum* So ce56 have been found to be of medicinal value, for example epothilone has anti-cancer properties, carolacton has been shown to be effective for the treatment of COVID-19 and myxochelin A produced by *Angicoccus disciformis* has been reported to exhibit activity against certain human cancer cell lines.
- Archaeal membrane lipids are useful in synthesising liposomes for drug delivery.
- Methanogens are added in sewage treatment plants where they use organic pollutants as sources of energy forming methane gas as a byproduct which finds use in cooking.
- Enzymes of *Sulfolobus* spp. exhibit not only great catalytic diversity, (for instance, starch-hydrolysing, cellulolytic, pectinolytic, chitinolytic, proteolytic, and lipolytic) but also stability at high temperatures and low pH. Hence, these enzymes are being used in the food and feed industry, textile and cleaning industry, pulp, and paper industry.
- *Micrococcus luteus* has applications in bioremediation due to its ability to degrade olefinic compounds and hydrocarbons as well as the property to remediate the regions which contain organic pollutants along with metals.
- Anammox group of planctomycetes, due to their ability to oxidise ammonia anaerobically, may find an application in clean-up of nitrogen in wastewater remediation plants.

6.5.2 Biotechnology and Research

- Taq polymerase from *Thermus aquaticus* is used in PCR technique due to its stable nature even at high temperature (70° to 80°C).
- *Sulfolobus* species are used as a model to study molecular mechanisms of DNA replication in archaea.

6.6 SUMMARY

- Woese's designation of three domains of life is supported by molecular data.
- Bacteria and archaea show similarities in overall morphology and basic biochemical features, occupation of a variety of ecological niches as well

as use of a common genetic code; however, the two groups show marked differences in the composition of their cell membranes and cell walls. Membrane lipids in bacteria consist of fatty acids joined to *sn*-glycerol-3-phosphate backbone by ester linkage and in archaea, isoprenoid units are joined to *sn*-glycerol-1-phosphate backbone by ether linkage.

- In most of the species of bacteria and archaea, the genome consists of a single, circular chromosome and several polycistronic genes organised as operons. However, there are examples of bacterial and archaeal species which exhibit polyploidy. Possible advantages of polyploidy include resistance against double-strand breaks, long-term survival and DNA serving as a phosphate storage polymer.

- Both bacteria and archaea show a remarkable biochemical diversity.

- Prokaryotic evolution is driven by horizontal gene transfer to a great extent, which may result in expansion or contraction of genome size.

- Enzymes involved in the processes of nucleotide biosynthesis, transcription and translation are quite conserved among the three domains.

- Several studies have shown that members belonging to archaea exhibit a high degree of diversity not only in terms of chromosome copy numbers but also in systems controlling DNA duplication and cell division.

- One model suggests that evolution of RNA polymerases and replicative DNA polymerases occurred from a common ancestor which had RNA-dependent RNA polymerase in the RNA-protein world that existed before the start of DNA replication. Archaeal polymerase D might have evolved from the replicative DNA polymerase of the Last Universal cellular Ancestor (LUCA).

- Several studies have shown that there is homology in sequence and structure between the subunits of bacterial RNAP and the subunits of archaeal RNAP.

- The increase in RPs in archaea compared with that in bacteria may reflect a more complex set of interactions in archaea and eukaryotes in regulating translation, e.g., differences in structure requiring scaffolding of longer rRNA molecules, expanded interactions with the chaperone machinery.

- Cyanobacteria have been responsible for converting the reducing atmosphere of the primitive earth into an oxidising one, paving the way for the evolution of aerobic metabolism in organisms around 2.4-2.1bya.

- *Rickettsia* belonging to Proteobacteria have contributed mitochondria to the eukaryotes and cyanobacteria have given rise to chloroplasts of the photosynthetic eukaryotes.

- Some of the archeal species exhibit metabolic features that allow them to cope up with the environmental conditions which seem to have existed on primitive earth in ancient times. Based on this property, scientists postulate that archaea species arose earlier than bacteria.

6.7 TERMINAL QUESTIONS

1. Fill in the blanks:

   a) Members of the bacterial phylum ………. lack peptidoglycan in their cell walls.

   b) ………….. is the causative agent of food poisoning.

   c) …………. are the source of a large number of antibacterial and antifungal compounds.

   d) Engineered bacterium ………….. can be exploited for cleaning up the radioactive waste sites.

   e) Genome of …………. has three origins of replication.

   f) ……………has a linear chromosome rather than a circular one.

   g) ……………. was the first archaeal genome to be sequenced.

   h) Pentose phosphate shunt is present in……… but mostly absent in………

   i) Initiator tRNA carries …………in archaea.

   j) Archaea are better adapted to survive in extreme environmental conditions due to the presence of ………….. in their cell membranes.

2. Write true or false against the following statements:

   a) First photosynthetic organisms belonged to archaea.

   b) Extremophiles are encountered in both archaea and bacteria.

   c) Aerobic rickettsial ancestor gave rise to mitochondria.

   d) *Thermus aquaticus*, being a thermophile, belongs to archaea.

   e) Archaea are not susceptible to beta-lactam antibiotics.

   f) Archaeal genomes do not encode linker histones.

   g) Many genes of both bacteria and archaea are organised within operons.

   h) The origins of replication in archaea are generally GC-rich regions.

   i) Sigma factor in archaea plays the key role of directing the polymerase to specific promoters.

   j) Eukaryotic genes for biochemical and metabolic functions are indicative of bacterial origin.

3. Match items in column I with the items in column II

	Column I		Column II
a)	*Prochlorococcus*	i)	Used as a nitrogen fixing biofertilizer
b)	*Borrelia burgdorferi*	ii)	Highly resistant to ionising radiation
c)	*Azotobacter* spp	iii)	Can withstand temperature of 113°C
d)	*Deinococcus radiodurans*	iv)	Produces enormous amounts of oxygen
e)	*Pyrolobus fumarii*	v)	Can withstand highly acidic soils (even pH 0)
f)	*Picrophilus* sp.	vi)	Causes Lyme disease

4. Distinguish between the following pairs of terms:
   a) Core genome and pan genome
   b) Bacterial cell membrane and archaeal cell membrane
   c) Thermophilic proteins and Psychrophilic proteins
5. Brifely enumerate salient features of bacterial genomes.
6. Write short notes on:
   a) Methanogens;
   b) Applications of thermophiles
7. Describe briefly the role played by cyanobacteria in the evolutionary history of life on earth.

6.8 ANSWERS

Self-Assessment Questions

1. a) i) True; ii) False; iii) True;
   iv) False; v) True

   b) i) Proteins of psychrophilic organisms denature at temperatures higher than their optima and thus are not able to function at body temperatures of warm-blooded animals (37°C) and hence these organisms do not cause disease in warm-blooded animals.

ii) *Listeria monocytogenes* is a psychrotroph. Though 30-37°C is the optimal temperature range for growth, it can grow at a wide temperature range, from 1 to 45°C and survives refrigeration, freezing and drying. This is the most important bacterium responsible for causing food spoilage and poisoning even under refrigerated conditions.

iii) It has salt concentration 10 to 20 times higher than what is found in normal cells leaving very little free water within the cells. This helps in stabilising molecules such as proteins and nucleic acids and ensuring survival of these organisms at such high temperatures.

iv) Presence of multiple origins of replication greatly reduces the time required for replication.

v) Beta-lactam antibiotics affect bacterial growth by inhibiting cell wall synthesis. However, as *Mycoplasma* species lack a cell wall, they remain unaffected by the treatment using beta-lactam antibiotics.

2. a) False; b) True; c) False;
   d) True; e) True

3. a) bacteria; archaea
   b) DNA gyrase
   c) Sigma factor
   d) 16S; 23S; 5S
   e) TBP; TFB; TFE

Terminal Questions

1. a) Chlamydia
   b) *Clostridium perfringens*
   c) *Streptomyces* spp.
   d) *Deinococcus radiodurans*
   e) *Sulfolobus*
   f) *Borrelia burgdorferi*
   g) *Methanocaldococcus jannaschii* DSM2661
   h) bacteria, archaea
   i) methionine
   j) tetraether lipids

2. a) False; b) True; c) True; d) False;
   e) True; f) True; g) True; h) False;
   i) False; j) True.

3. a) Produces enormous amounts of oxygen

   b) Causes Lyme disease

   c) Used as a nitrogen fixing biofertilizer

   d) Highly resistant to ionising radiation

   e) Can withstand temperature of 113°C

   f) Can withstand highly acidic soils (even pH 0)

4. a) Pan genome refers to the entire set of all the genes from all strains of a clade while, core genome represents a set of homologous genes present in all the tested genomes.

   b) Membrane lipids in bacteria consist of fatty acids joined to *sn*-glycerol-3-phosphate backbone by ester linkage while, in archaea, isoprenoid units are joined to *sn*-glycerol-1-phosphate backbone by ether linkage.

   c) Proteins in thermophilic organisms show higher numbers of hydrophobic amino acids, disulphide bonds and ionic interactions so as to retain protein structure and activity at high temperatures. On the other hand, proteins in psychrophilic organisms have reduced hydrophobic core and less charged surface which ensures flexibility and functionality at low temperatures.

5. Refer to Subsection 6.3.1.

6. a) **Methanogens:** Methanogens belong to the phylum Euryarchaeota and are the largest microorganisms belonging to archaea and show a cosmopolitan distribution, occurring mostly in anoxic environments such as aquatic sediments, rice paddies, anaerobic digesters, and the gastrointestinal tract of animals. They can grow at a wide range of temperatures (4°C to 100°C), salinities (freshwater to brine) and pH (3 to 9). They possess several unique coenzymes. Because of the unique cell wall construction, they are not susceptible to penicillin and other antibiotics. As they are chemoautotrophic, H_2 serves as the source of both energy and electrons, and CO_2 works both as an electron sink and the source of cellular carbon. Those residing within the gastrointestinal tract of animals and humans produce methane as a by-product of metabolism, which causes flatulence in humans and other animals. *Methanosaeta* is acetotrophic, using acetate fermentation to produce methane. Methanogens are added in sewage treatment plants where they use organic pollutants as sources of energy forming methane gas as a byproduct which finds use in cooking.

   b) **Applications of thermophiles:** Thermophiles have several adaptations which enable them to survive at high temperatures. Enzymes of *Sulfolobus* spp. exhibit not only great catalytic diversity, (for instance, starch-hydrolysing, cellulolytic, pectinolytic, chitinolytic, proteolytic, and lipolytic) but also stability at high temperatures and low pH. Hence, these enzymes are being used in the food and feed industry, textile and cleaning industry, pulp, and paper industry. Taq polymerase from *Thermus aquaticus* is used in PCR technique due to its stable nature even at high temperature (70° to 80°C).

7. Refer BOX 6.1.

Acknowledgment

Table 6.1 :
- i) Rickettsia - Wikipedia
- ii) Spirillum - Wikipedia
- iii) https://en.wikipedia.org/wiki/Escherichia_coli
- iv) https://en.wikipedia.org/wiki/Vibrio_cholerae
- v) https://en.wikipedia.org/wiki/Salmonella
- vi) Azotobacter - Wikipedia
- ix) https://en.wikipedia.org/wiki/Micrococcus_luteus
- x) https://en.wikipedia.org/wiki/Campylobacter
- xi) https://en.wikipedia.org/wiki/Helicobacter_pylori
- xiii) https://en.wikipedia.org/wiki/Treponema_pallidum
- xv) https://en.wikipedia.org/wiki/Scytonema#
- xviii) https://edurev.in/t/300532/Structure-and-Reproduction-of-Mycoplasma
- xix) https://en.wikipedia.org/wiki/Gemmata_obscuriglobus

Table 6.3 :
- i) Methanogenic microorganisms as workhorses of | EurekAlert!
- ii) Halobacterium - Wikipedia
- iii) https://en.wikipedia.org/wiki/Sulfolobus
- iv) Nanoarchaeum equitans - Wikipedia
- v) Korarchaeum cryptofilum - microbewiki (kenyon.edu)