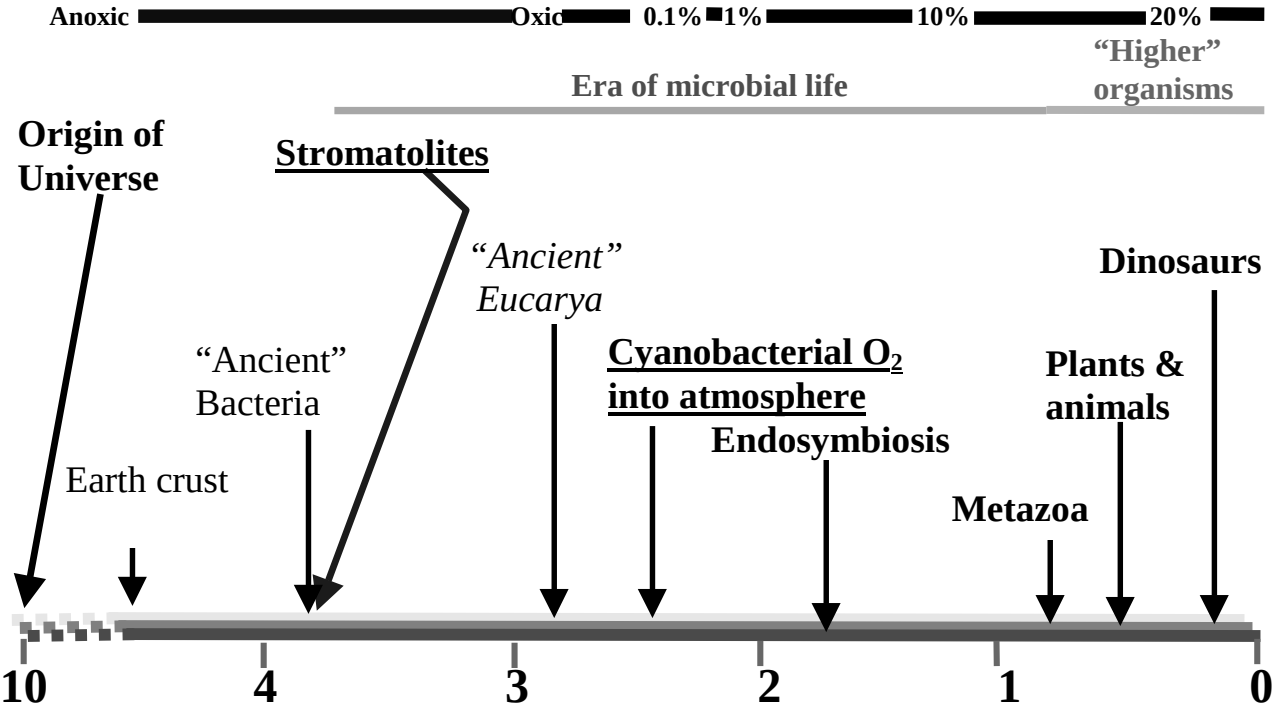


Approximate timing of major events (billion yrs from present day) in the history of life on earth

Molecular Science & Bioinformatics – connecting past with present

Questions:

1. **The early cells: Looks? Genetics? Genomes (plastic vs specialised) Physiology? Metabolism?**
2. **The early cells were microbes – Did they impact evolution?**
3. **The early cells - Can they still exist today? Why not?**
4. **The early cells – Can we recreate them today?**
5. **Microbes are small - Are they impacting evolution today? How?**

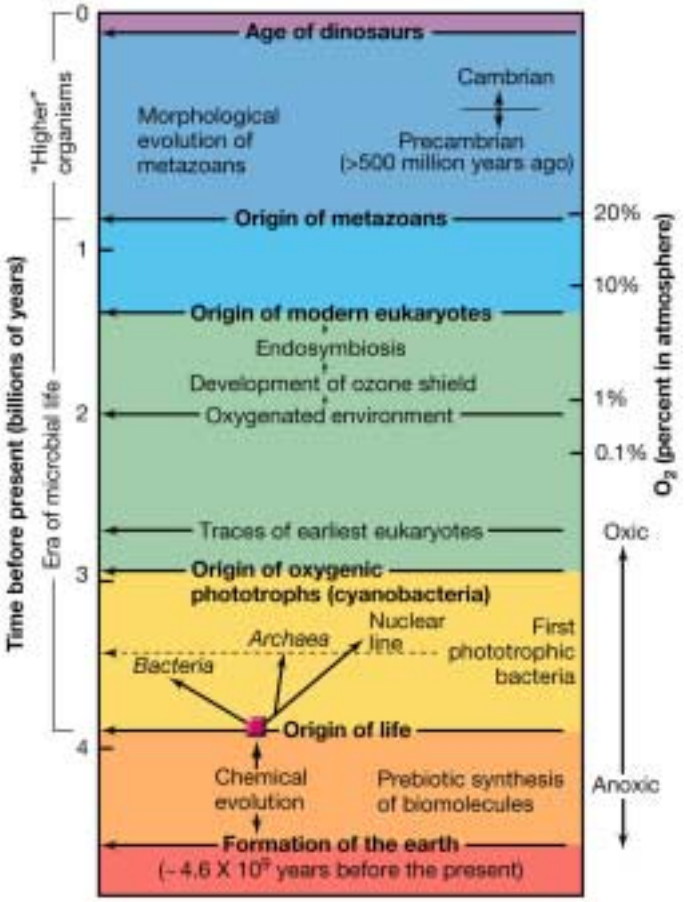
Our understanding of life is based on the technology available to us at the time. Technology advances and so does our understanding (facts / hypothesis) – keep an open mind (debate knowledge not emotion).

In MAM, we will try and connect the past with the present with a view to predicting the future using evolutionary / comparative biology tools and findings. We will also use examples of bioinformatics tools useful in improving outcomes in Biotechnology.

TABLE OF CONTENTS

1. The beginnings - Evolution of earth and earliest life forms

The early earth's atmosphere contributed significantly to evolution of life. Fossil and other evidence suggests that the first cells evolved on earth some 3.8 billion years ago and these first cells then proliferated and evolved as the earth's environment changed. Figure 1 summarises the time scale and the rapid proliferation of eucaryotes, which began some 1.5 billion years ago.



1.1 Earth's origin

- Big Bang – a hot exploding star created our galaxy
- Earth is ~4.6 billion yrs
- Rocks dated to ~ 4 billion yrs have been found:
  - 3.86 billion year old rocks of Greenland
  - 3.5 billion year old rocks of Western Australia (Warrawoona & Pilbara) & Swaziland
- ♦ The rocks are of 3 types – indicates the type of environment existing at that time:
  - Carbonate (rich in carbon dioxide)
  - volcanic (hot & reducing due to volcanic gasses – CO, H_2S, H_2)
  - sedimentary (water present – cellular processes possible)

1.2 Evidence of microbial life on earth

- Fossilised record for microbes dated to ~3.5 billion yrs – rods /cocci in carbonaceous material and in stromatolites:
- Ancient stromatolites = fossilised filamentous, phototrophic, non-oxygen evolving microbes + sediment (similar to present day hot spring microbial mats of *Chloroflexus* species)
  - 60 cm 3.5 billion yrs old mushroom shaped Hamlin stromatolites (WA)

- 1.6 billion yr old cone shaped dolomite stromatolites of MacArthur Basin (NT) – Fig 2A
- Modern Stromatolites = Cyanobacteria layers + $CaCo_3$ + sediments - Fig 2B
  - The Sharks Bay stromatolites (WA) are 3000 yrs old & still growing – Fig 2C
  - Surface hot springs are colonised by microbial mats
  - Hot water emanating from the subsurface Great Artesian Basin of Australia aquifer is colonised by microbial mats

2A

2B

2C

1.3 Condition of early earth

- Reducing environment - no significant oxygen
- Water present
- Significant presence of gases – CH_4, CO_2, N_2 & NH_3 & of sulfides as H_2S & FeS present
  - CH_4 and NH_3 reacted to form HCN
- Traces of CO & H_2 also present
- Early earth was hot –
  - Much hotter (> 100^oC) in the first half billion years of its existence & free water did not exist
  - Free water accumulated only after earth cooled.
  - The rate of earth's cooling is unknown.
  - Self-replicating entities appeared around this period leading to cellular evolution – the first cells were heat resistant and perhaps similar to the hyperthermophiles (see rRNA phylogenetic evidence below).

1.4 Origin of Life

- Primitive earth had intense energy resources generated from UV radiation from sun, lightning discharges, meteorite impacts, radioactivity, & volcanic activity
- Important biologically molecule building blocks can be
  - synthesised under hot and reducing conditions from CH_4, CO_2, N_2 & NH_3 & sulfides (H_2S & FeS) using such energy sources
  - subsequently spontaneously polymerised on relatively anhydrous surfaces such as clays, pyrite and basaltic glass and
  - this polymerised material could accumulate to form organic films from which primitive, self-replicating structures emerged (see "metabolism in primitive cells, later).
- Laboratory experiments has provided evidence -
  - Building blocks synthesised (sugars, amino acids, purines, pyrimidines, nucleotides, thioesters & fatty acids)
  - Building blocks polymerised to polypeptides, polynucleotides & other important macromoleculae and subsequently polymerised

2. Primitive Life: The RNA world, molecular coding and energy generation

- The first systems were simple self replicating entities.

- Ancient cells were self replicating systems which perhaps
  - required energy and hereditary means to make copies of itself.
  - Similar to modern cells but had limited transcription & translation genes.
  - May not have had cellular structures
- Scientist now believe that the first cells lacked had RNA (not DNA) and few proteins if any. RNA was involved in catalysis and genetic coding – the age of RNA life may have predated cellular life.

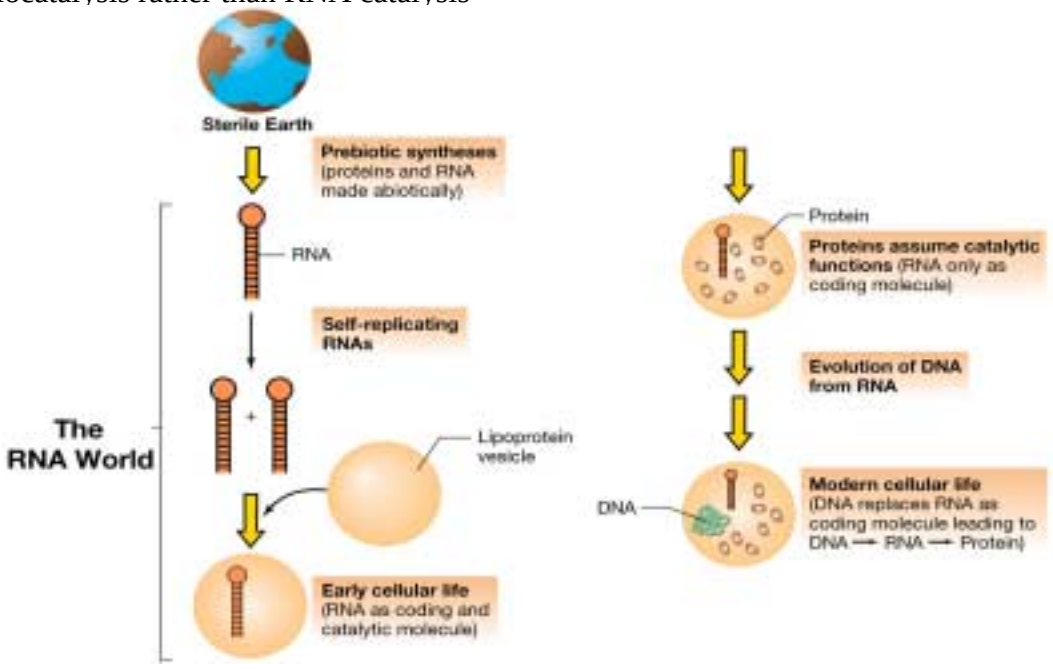
2.1 The RNA life

- Polynucleoides synthesis followed by polymerisation (as described above) to produce RNA.
- The RNA functioned simply to replicate them selves & carried out minimal number of catalytic reactions necessary for this purpose
- RNA mediated catalysis is also found in modern cells today (see box, below)

Background information on RNA processing by ribozymes in modern cells

- 3 types of RNA exist: mRNA = "functional"; tRNA & 16S rRNA = "structural"
- RNAs may initially exist as non-functional "precursor" molecules requiring processing to become "mature" functional RNA – use different RNA processing mechanisms (eg splicing & / or base modifications)
- RNA processing by splicing using spliceosomes (complex of RNA and proteins):
  - It is a common mechanism for removing the intervening introns (non-coding) from exons (coding) in precursor eucaryotic mRNA, and less common in *Bacteria*, *Archaea* and bacteriophages.
  - Specific processes for different genes.
  - Removes introns and joins exons in an ordered manner (not random) to form mature mRNA.
  - Refer to Figure 7.30 Brock Biology of Microorganisms, Madigan, Martinko & Parker, Prentice Hall,10th edition, 2003**.**

- Cellular life arose (see Figure 3)
  - when self-replicating RNA became enclosed within lipoprotein vesicles – spontaneous aggregation of lipids & proteins to form membranous structures
  - Natural selection led to further evolutionary development.
  - Proteins are more catalytic than RNA and hence the evolutionary push drove protein biocatalysis rather than RNA catalysis



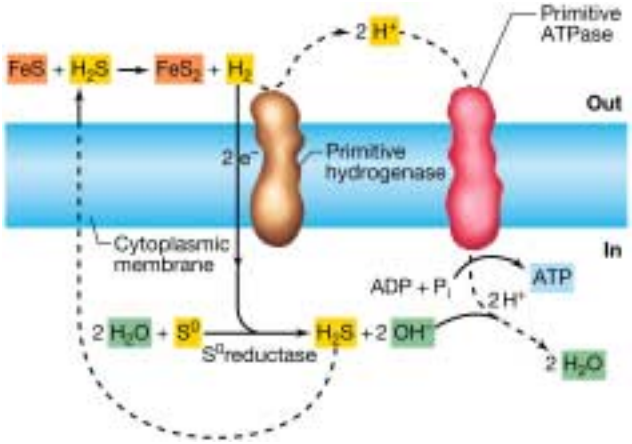
- This efficient and specific protein mediated catalysis replaced RNA mediated catalysis.

2.2 Modern cells and the central dogma

- DNA as the genome was perhaps established as a result of evolutionary pressure for greater efficiency and fidelity (accuracy) – DNA polymerase is more accurate than RNA polymerase
- Storage of genetic resources in one place and processing only what was required under a set of condition (gene expression) conserved energy and therefore improved competitiveness.
- Consequently, during early evolution, the 3 stage process of DNA⟶ RNA⟶ protein got fixed as the best solution for biological information processing & has become a success story.

2.3 Metabolism in primitive organisms – chemolithoautotrophic mode of life

- We know that the ordering randomly occurring molecules into highly ordered biological machines (cells) requires energy; How was this met in primitive self-replicating cells?
- We know that (a) anoxic (not a metabolic restrictive environment as modern *Bacteria* and *Archaea* thrive in such an environment), (b) ferrous iron, volcanic gasses (H_2S, CO_2, CO), other chemicals (S^o) and (c) high temperatures were in abundant supply in primitive earth (Figure 4).
- The reaction $FeS + H_2S \longrightarrow FeS_2$ (pyrite)$+ H_2$ produces $\Delta^{0'}$ = -42kJ/reaction.
  - an energy source of primitive cells
  - H_2 used to form a Proton Motive Force (PMF) across a membrane and ATP (chemical energy) generated from a primitive ATPase.
  - S^o used as an electron acceptor and paired with H_2; CO_2 and CO used as carbon sources



- Genome sequencing projects show chemolithoautrophic traits in the most ancient lineages of domain *Bacteria* and *Archaea*.
- Anaerobic processes (e.g.fermentation) require many biological components and must have specialised and evolved subsequently.

2.4 Oxygenation of the atmosphere

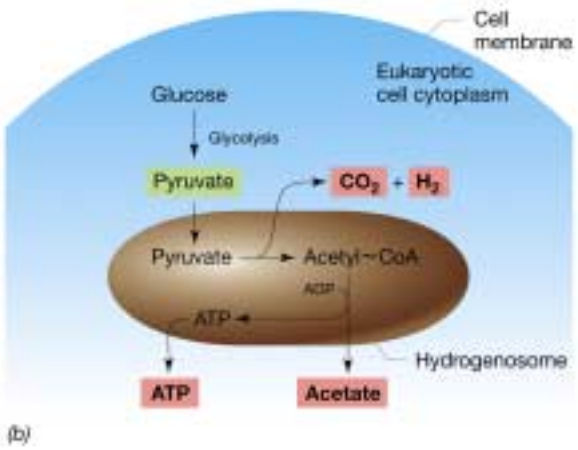
- Anaerobic oxygenic photosynthetic cyanobacteria evolved ~3 billion years ago.
- Due to reactions with reducing substrates (FeS) to produce H_2O, O_2 accumulated gradually
- Consequences of accumulating O_2 in the environment:
  - More energy from oxidation = more microbial densities & increase in diversity.
  - Fossil evidence suggests that the rate of evolution increased after O_2 accumulated - (eucaryotic microbes -> eucaryotic microbes with organelles -> metazoa (multicellular eucarytoes) -> higher eucayotic plants & animals.

- Ozone (adsorbs UV light) formed – provides a protective barrier against sun's UV light; evolution crept out of UV protected areas (eg rocks, deep ocean) to Earth's surface
- A summary of the most likely evolutionary path is depicted in Figure 1.

3. Eucaryotes and organelles

3.1 Origin of the nucleus

- Early primitive eucaryotes were structurally simple – no nuclear-bound membranes, mitochondria – chloroplast less.
- Subsequent cells evolved larger chromosomes and mitosis but the cells could not handle replication as one genome (unlike procaryotes) – partitioned into smaller manageable-sized chromosomes.
- Organelles (mitochondria & chloroplast) were not essential in primitive eucarya but developed later by endosymbiosis (see below)
- Organelle devoid eucaryotes exist today and are found in the deep ancient branches of domain *Eucarya* - *Trichomonas* (a flagellated parasite) *Neocallimastix* (anaerobic rumen fungus) contain hydrogenosomes – membrane bound respiratory organelle:
  - Structure- size similar to mitochondria but lacks cristae, TCA cycle enzymes & electron transport chain (cytochromes)
  - Energy production function- pyruvate oxidised to CO_2, H_2 and acetate; acetate is excreted into the medium (unlike mitochondria in which acetate is used up as it has electron transport chain)
  - Detailed mechanism (Figure 5).

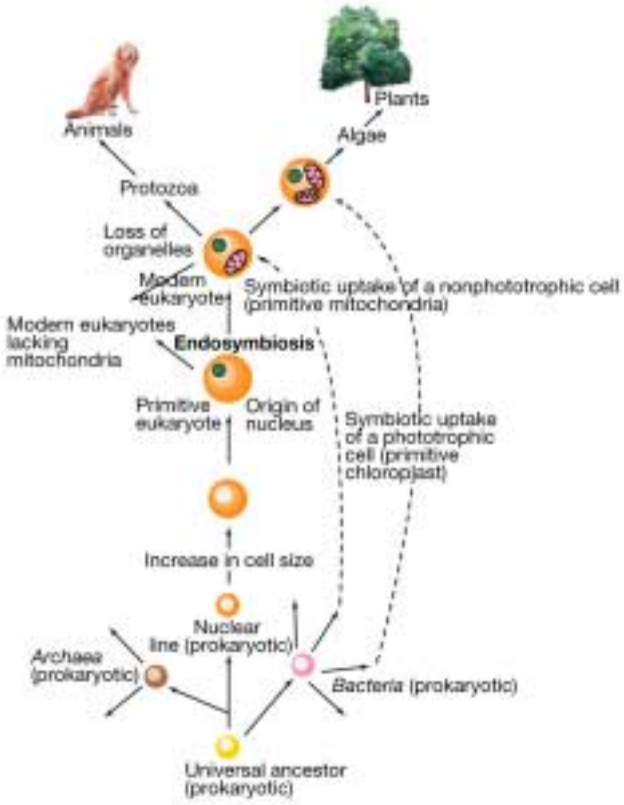


  - In some eucaryotes, symbiotic methanogens utilise the end-products of hydrogenosomes to produce methanogens

3.2 Endosymbiosis

- Evidence suggests that modern organelle containing eucaryotes evolved in steps via stable incorporation of a domain *Bacteria* (not domain *Archaea*) specific chemoorganotrophic and phototrophic symbionts – endosymbiotic theory (Figure 6)
  - Aerobic bacterium (a forerunner of the modern mitochondrion) established itself in the cytoplasm of a eucaryote providing energy and getting protection in exchange.
  - Aerobic phototroph (a forerunner to the modern chloroplast) established itself in the cytoplasm and conferred photosynthetic properties so that the host no longer relied on organic compounds for energy production.

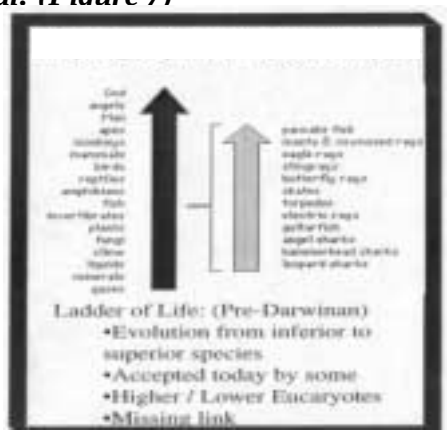
- However, some eucaryotes never incorporated symbionts or disposed them (see 3.1 above).
- Evidence of relationship of mitochondria and chloroplast to Bacteria:
  - Contain covalently closed circular DNA and the genome encodes rRNA, tRNA and genes for the respiratory chain proteins
  - Contain prokaryotic 70S (Svedberg – S- units) specific ribososme.
  - Antibiotics (streptomycin) that interfere with prokaryotic 70S-ribososme structures also inhibit the organelles.
  - Eucaryotic nucleus is "contaminated" by bacterial derived genes. The genes involved in the synthesis of the organelles are encoded by the eucaryotic chromosome – horizantal gene transfer.
  - Molecular phylogeny using comparative rRNA sequence analysis shows that the organelle sequences are closely related to *Bacteria* and thus must have arisen from a common ancestor.



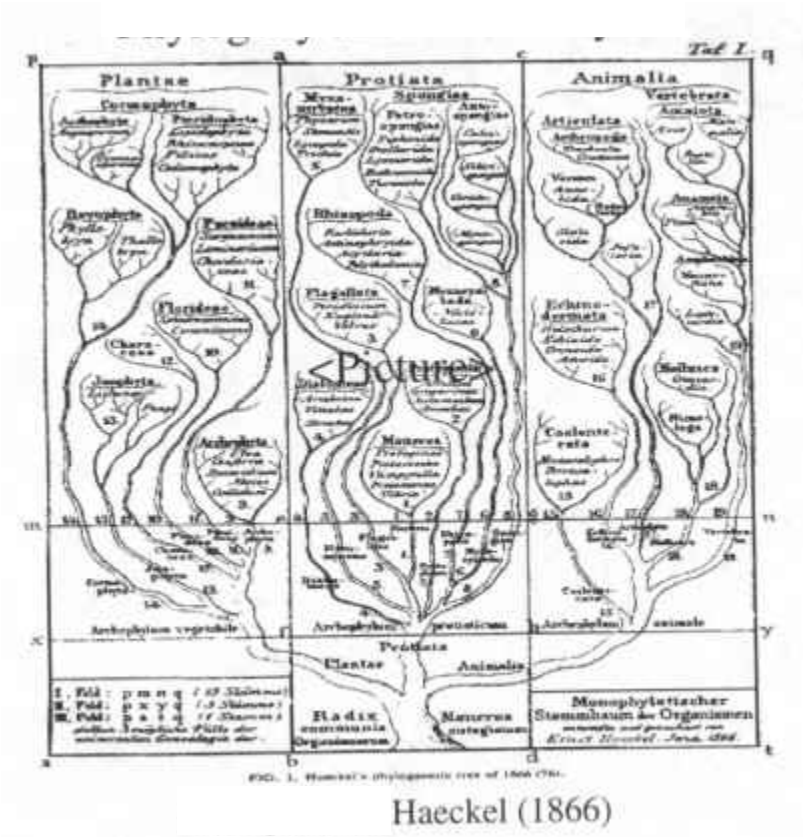
4. The early views on evolution

Mankind has been intrigued with evolution and has proposed a number of theories on the relationships of various life forms and a number of proposals have been put forward. These are discussed below:

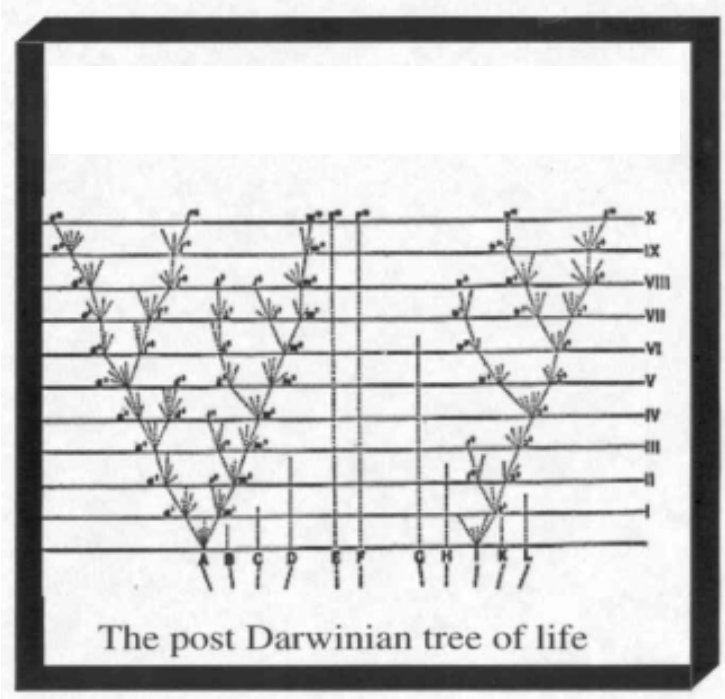
4.1 The ladder of Life proposal: (Fiaure 7)



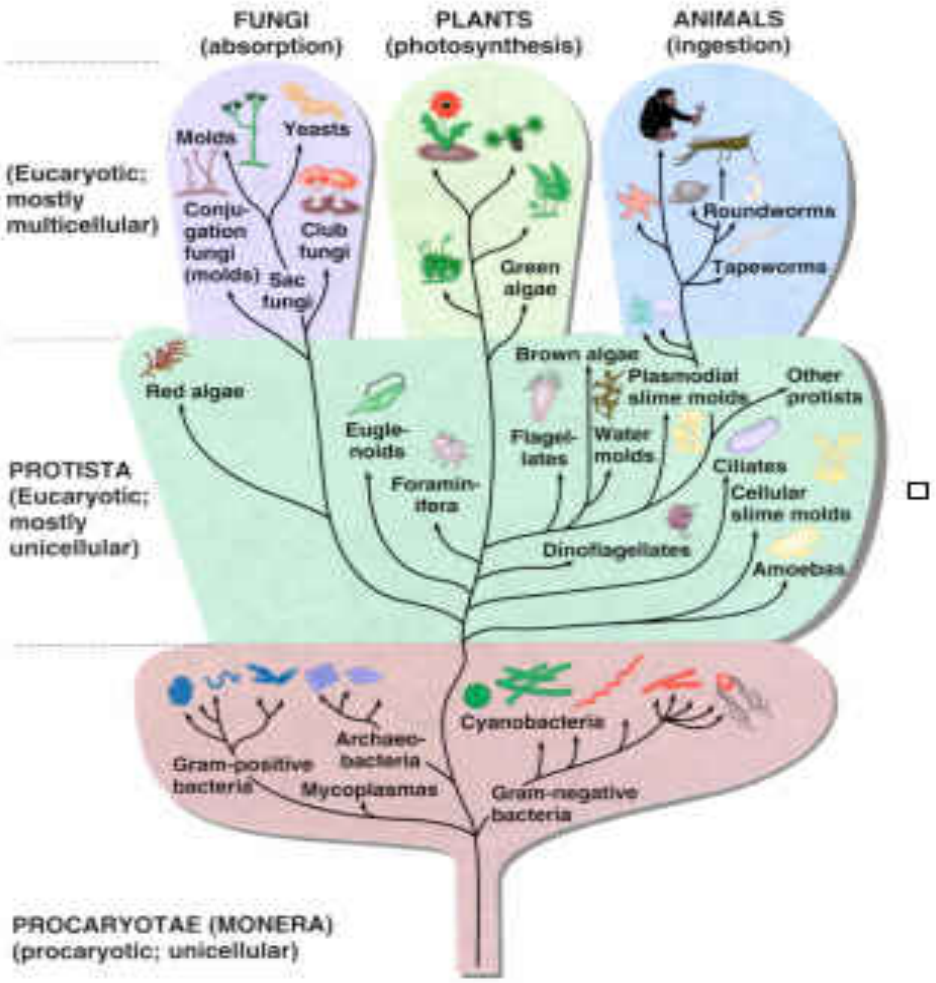
4.2 Hackel's proposal (Figure 8)



4.1 The post Darwian proposal (Figure 9)



4.4 The five kingdom classification (Figure 10)



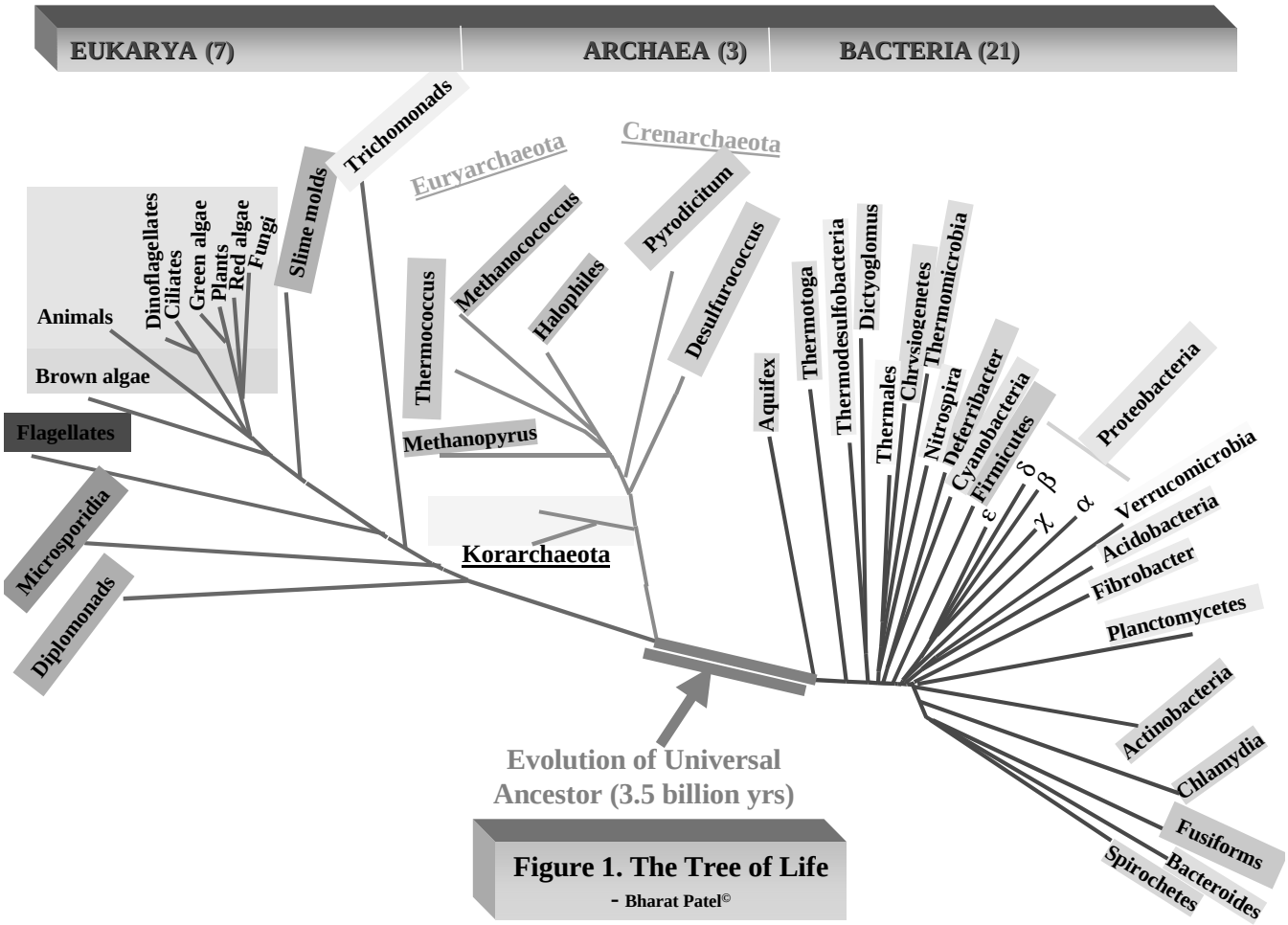
4.5 Procaryotic-Eucaryotic dichotonomy by Electron Microscopy

Electron microscopic examination of cells revealed that cells were of two basic types, namely Procaryotes and Eucaryotes. You have learnt the differences between Procaryotes and Eucaryotes. Can you tabulate the differences and similarities between the 2 cell types

5. Modern views on evolution:

5.1 The 3 Domains concept of Woese based on rRNA

Using rRNA cataloguing, Woese and his colleagues proposed that cells could be place into 3 major Divisions, namely Archaeobacteria, Eubacteria and Eucaryotes. The tree was not rooted, ie did not show which lineage (line of descent) was the most ancient. However, the tree was subsequently rooted using protein sequences giving rise to the concept of a progenote (hypothetical universal ancestor) and 3 domain based distinction of all life on this planet. The 3 domains are given super kingdom status (domain *Archaea*, domain *Bacteria* and domain *Eucarya*). Each domain is then subdivided into phyla, classes, order, families, genera and species (Figure 11)



5.2 Alternate models **(Figure 12)**

A number of alternate models have been proposed in which the branching pattern differs from that of Woese's model (see figure below).

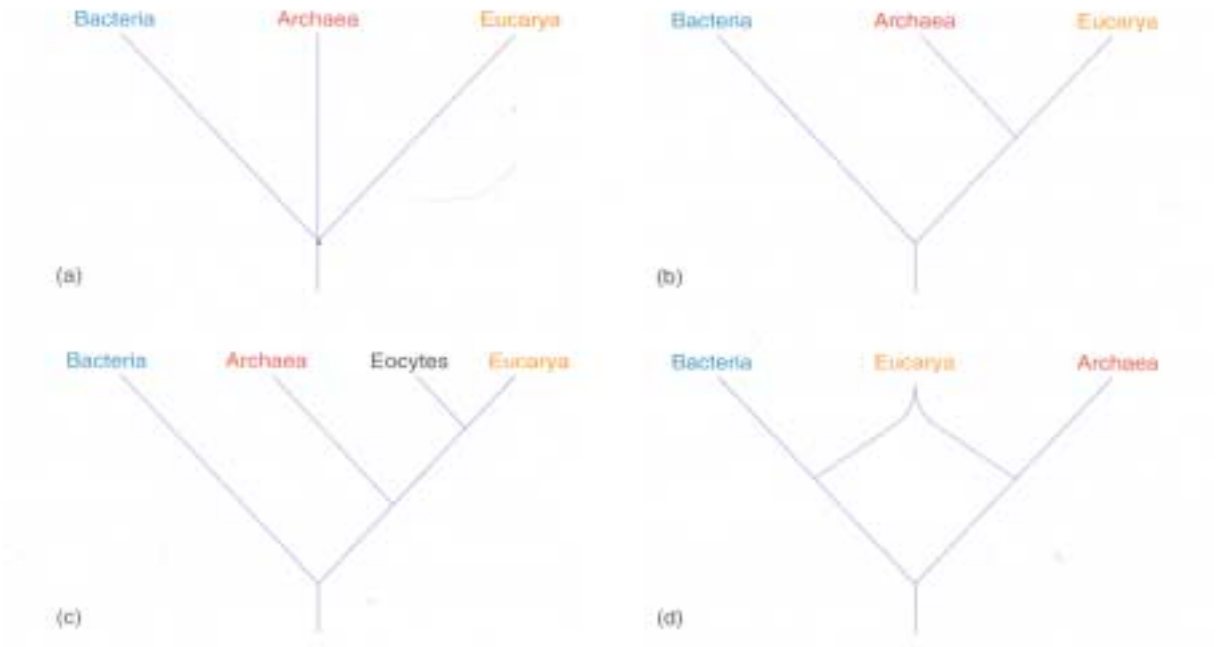
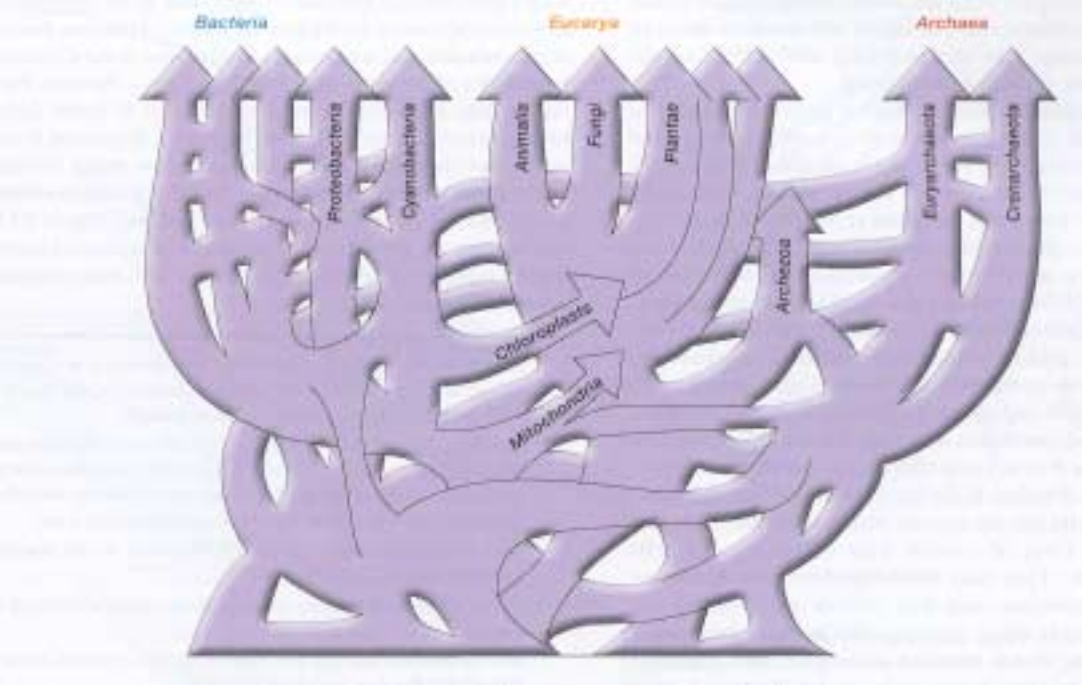


Figure 12 different models showing the branching pattern of the 3 domains

5.3 Models based on whole genomes: the controversy & debate based on genome sequencing and lateral gene transfer data (Figure 13)

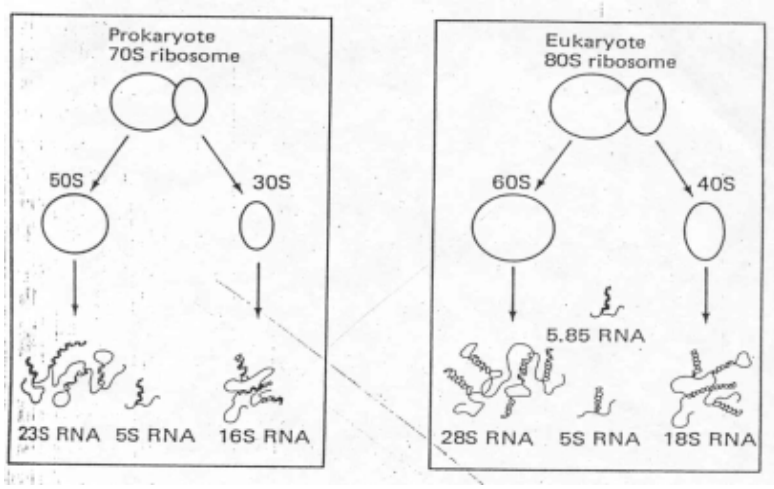
In 1999, Doolittle and his colleagues proposed from the information gleamed from genome sequencing projects that there was extensive gene transfer early in the history of cellular evolution and as a consequence, it would be impossible for cells to have had one universal ancestor. The Doolittle model based on genome comparative analysis is shown



6. rRNA its organisation in a cell and its structure

6.1 Types of rRNA (Figure 14)

Bacteria and Archaea: 16S, 23S, 5S
Eucarya: 18S, 28S, 8S
Mitochondria and Chloroplasts: Bacteria like



100 million ribosomes / bacterial cell and therefore that many rRNA / cell

6.2 Organisation of rRNA genes

- rRNA is coded by rRNA genes. The rRNA genes are organised in a genetic locus called rRNA operons**.**
- A rRNA operon usually consists of 16S –23S-5S genes linked by intergenic intervening spacer regions (ISR) which usual code for different tRNA.(Figure 15)

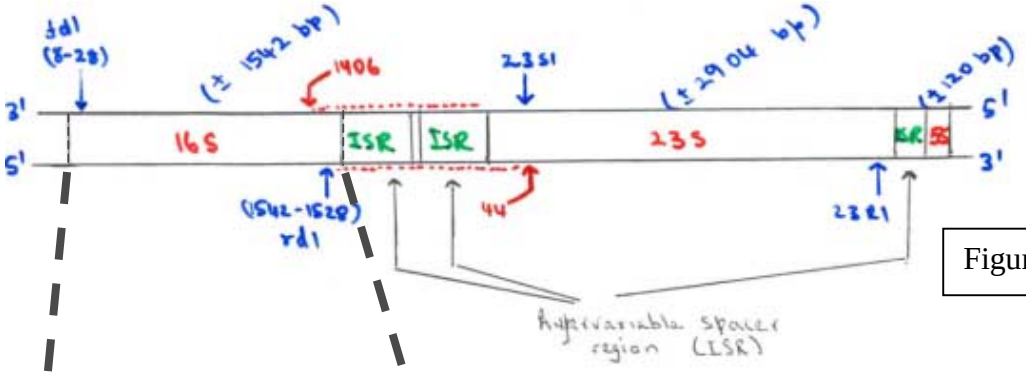


Figure 15a- rRNA operon

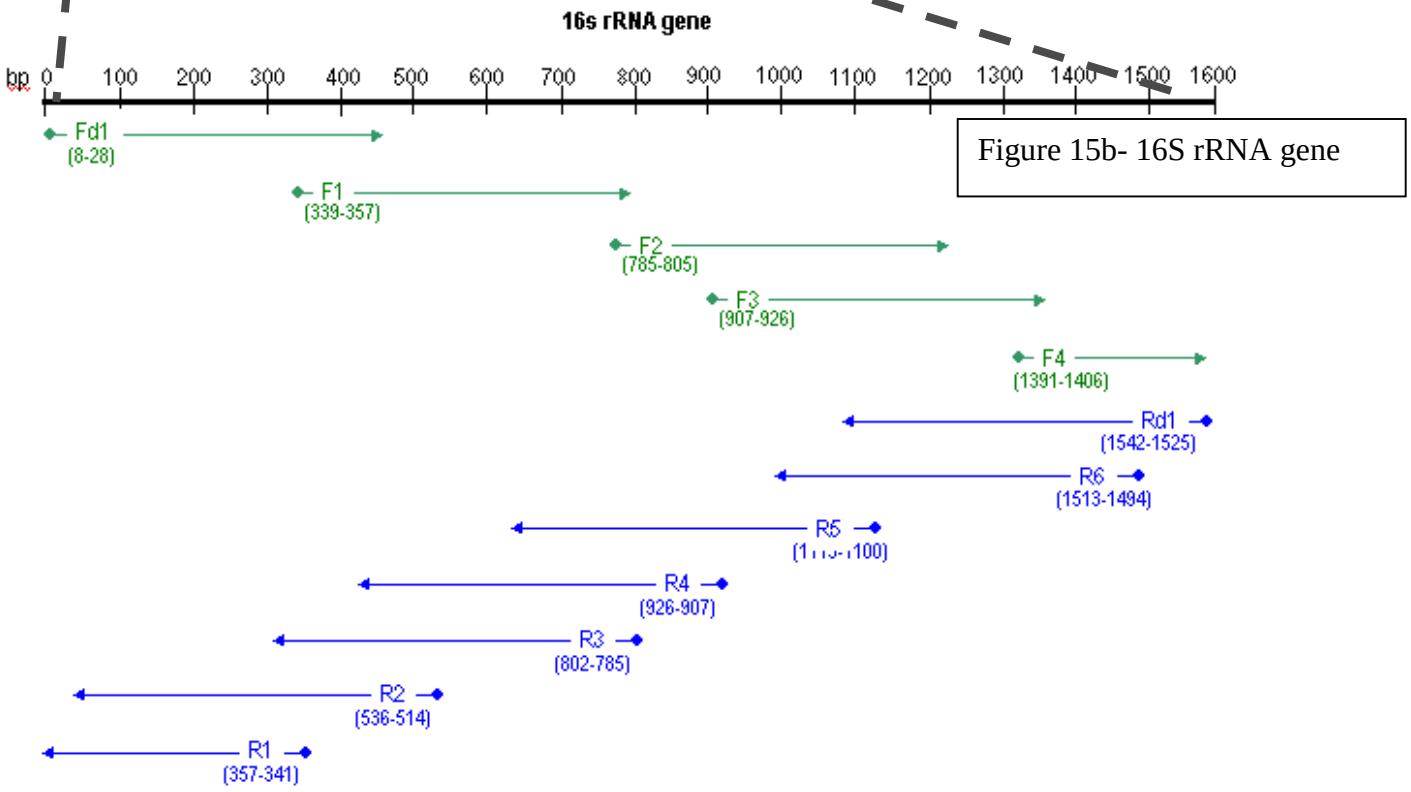


Figure 15b- 16S rRNA gene

- However, not all rRNA genes are organised into operons. In some organisms they are located separately as the following table indicates:
- The number of operons present in the genome differs in different organisms.

- A database at the URL http://rrndb.cme.msu.edu/rrndb/servlet/controller?page=about lists rRNA copy number and gene organisation (operon or separate)

6.3 Primary Sequence of rRNA

- rRNA is made up of the bases A, U, G and C.
- A number of regions of the rRNA are conserved within members of the 3 different domains. Primers derived from universally conserved sequences enables amplification and subsequent sequencing of rRNA genes. For example there are 10 universal regions in the 16S rRNA gene that are conserved. Fd1 (*E. coli* position 8 – 28) and rD1 (*E. coli* position 1542 – 1528) shown above enables amplification of the almost entire 16S rRNA gene corresponding to some 160 bases (see Figure 15b).
- There is substantial variation in the length of the rRNA in inter- and intra- domains (Figure 16)

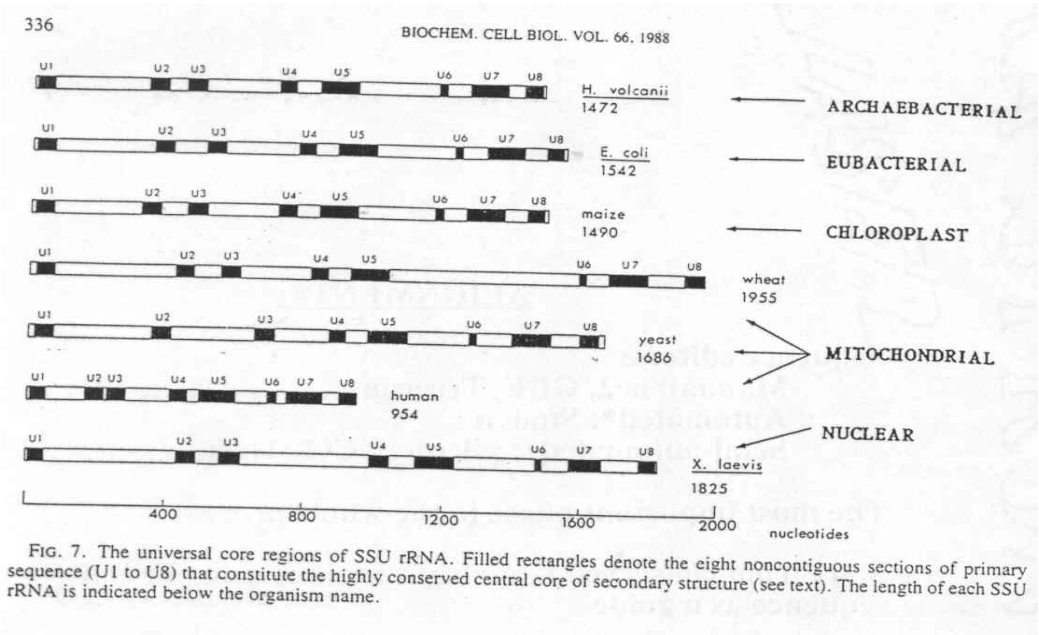


FIG. 7. The universal core regions of SSU rRNA. Filled rectangles denote the eight noncontiguous sections of primary sequence (U1 to U8) that constitute the highly conserved central core of secondary structure (see text). The length of each SSU rRNA is indicated below the organism name.

- The length variability does not involve scattered small insertions or deletions but rather large sequence blocks are added or deleted. This points indirectly to the fact that different regions of the rRNA have different functions. **Phylogenetic comparisons based on rRNA therefore must not be considered as single-marker comparisons but as comparisons using numerous functional domains. As functions are related to secondary structures and tertiary structures alignments are made using "structural constraints" (Figure 17)**

- Therefore, although there is substantial variation in homologous stretches, the distribution of changes is not random. Some segments are conserved while others drift freely, ie there are conserved, variable and hypervariable regions.

6.4 Secondary Structure of rRNA (Figure 17)

- Though there is substantial variation in the primary sequences, there is still a common secondary structure. In fact, the secondary structures from all 3 domains are essentially superimposable over much of the length of the structure. This uniformity of higher order structure in rRNA of all organisms is powerful testimony to its antiquity and functional constancy, i.e it has changed very little over time during evolution.

- There is very little known about the functions of the various structures. It has been suggested that the single stranded loops may be involved in tertiary structural arrangements during protein synthesis.

rRNA are not linear molecules but exist as secondary in a cell. The following structural details can be recogonised

Double stranded helix
Single stranded
Terminal or internal loops
Single buldged bases

rRNA interact with mRNA and tRNA to fold and form tertiary structures; single strands and terminal loops play an important role in the initial recognition processes

Identify and label the structures in the figure below using the terms given above

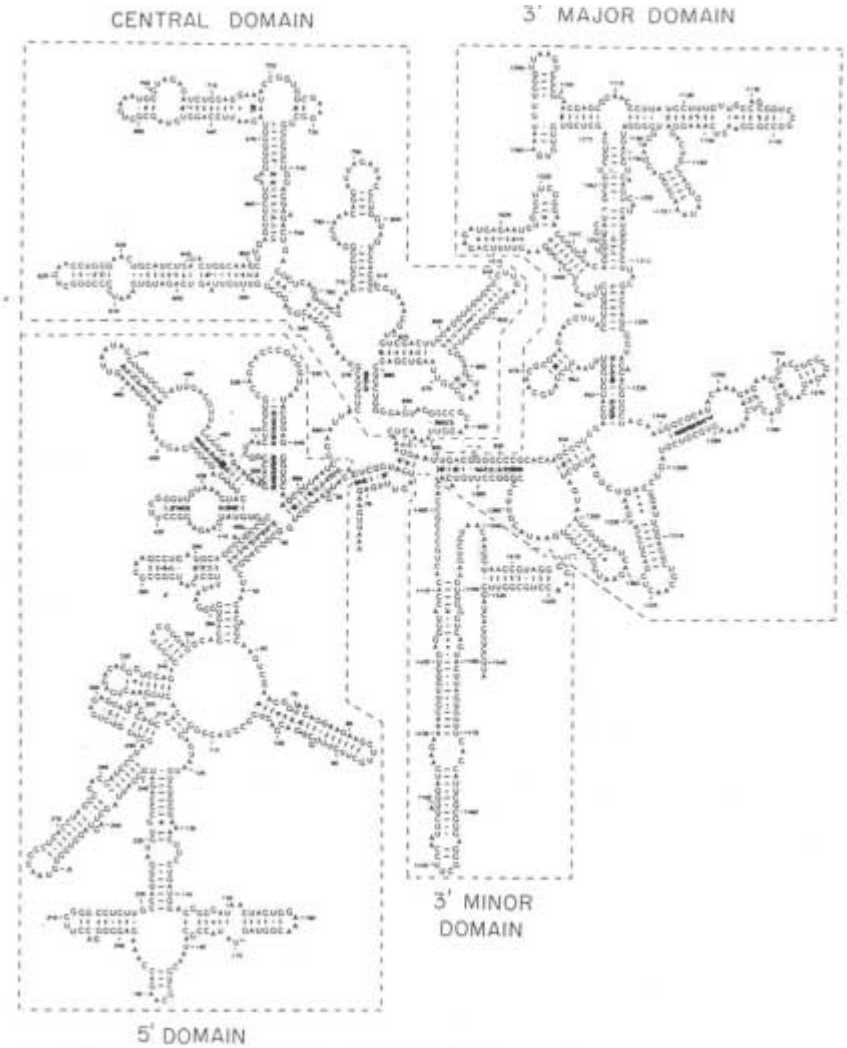


Figure 1 Secondary structure of the 16S rRNA from Escherichia coli. The 5', central, and 3' domains are believed to fold independently of one another.

<u>NOTE: Understanding rRNA structures will assist you in completing Assignment 1</u>

7. Why is rRNA used as a molecular clock or chronometer

- Key element in protein synthesis
- Ancient and conserved molecule
- Nucleotide sequence defines universal sequence for all 3 domains
- rRNA constitutes a significant component (10,000 to 100,000 copies) as part of the ribosome complex and is easy to isolate, reverse transcribe, clone and sequence (old method)
- rRNA genes (and the intergenic regions) are easy to amplify from DNA using PCR
- 16S rRNA but not 5S rRNA has sufficiently long sequences for statistical significant comparisons. 23S rRNA is much better as it is twice the size of 16S rRNA.
- rRNA genes appear to be free from artifacts of lateral gene transfer between disparate organisms.
- Relationships established by rRNA sequence comparisons represent the evolutionary relationships of organisms.
- Other molecules have been tested but not suitable for determining evolutionary relationships:
  - tRNA: too constrained & too few nucleotides (70 – 90 nucleotides)
  - 5S rRNA: Larger than tRNA (approx 120 nucleotides) but still too small for statistical reasons.
  - Cytochromes: Not universally present in all cells. Comparison not possible
  - Other proteins: Problem of lateral gene transfer, protein domain shuffling etc.

TO ADD THE FOLLOWING

==========

11. Building Phylogeny with rRNA

8.1 DNA extraction, PCR & sequencing
8.2 Sequence Data and its assembly
8.3 Alignments
8.4 DNA Distance calculations
8.5 Neighbor-Joining
8.6 Tree Drawing & editing

12.Phylogeny Software

I have only described a set of PC based phylogeny software that has been used in my laboratory for the past 10 years. There are a number of Unix and Linux phylogeny software which are also been used but have not been described here.

12.1 Alignment editors: One of the most user-friendly alignment editor is Bioedit. A review of the software is available at the URL http://www.molbiol.bbsrc.ac.uk/reviews/bioedit_review.html. It can be downloaded from Tom Hall's homepages at http://www.mbio.ncsu.edu/BioEdit/page2.html.

12.2 Tree construction
Mega
Paup
Phylip

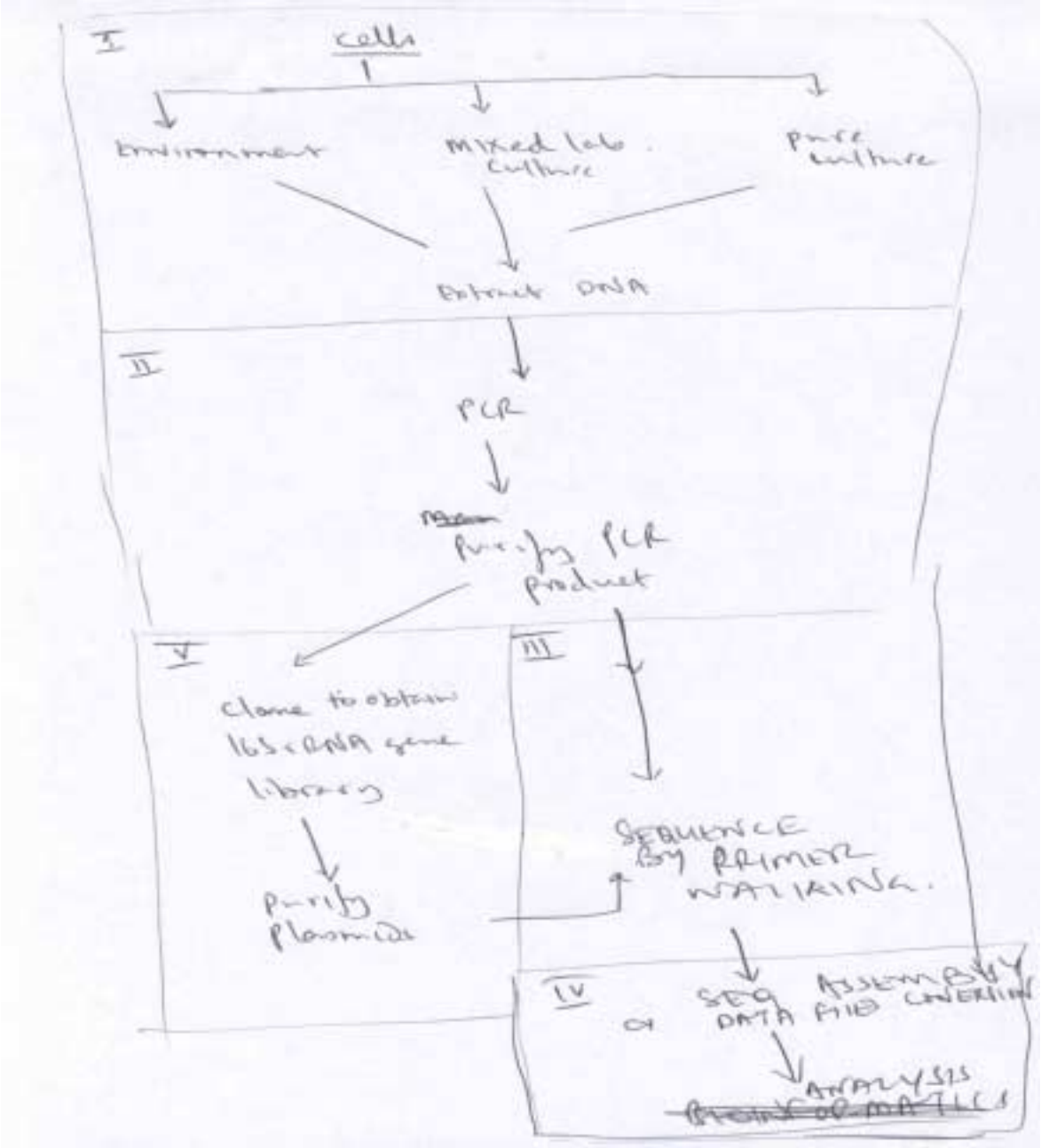
12.3 Tree drawing and graphics
Treecon
Treeview

Sequence manipulators
ReadSeq
Cutoff

12.4
12.5
12.6

13. Major Findings from rRNA studies

10.1 Evolution of life
10.2 Identification and taxonomy
10.3 rRNA database construction
10.4 Microbial Diversity: cultured and uncultured
10.5 Phylogenetic staining and domain specific signatures
10.6 Population dynamics and Real Time PCR



8. Building Phylogeny with rRNA

The objective of most phylogenetic studies is to reconstruct a tree-like pattern that describes the evolutionary relationships between the organisms under investigation. Before examining the methods used in building trees, one needs to understanding the terminology used in phylogenetic analysis.

A typical phylogenetic tree is shown below.

Questions:

1.

1. Explain the arrangement of DNA in microbial cells. Differentiate between nucleoid and nucleus.

2.