Microbial Diversity, Taxonomy, and Phylogeny

Today's Lecture

I. **Microbial Diversity: range of habitats that support microbial life**

II. **Taxonomy/Phylogeny**

- **A. Three domains of life**
- **B. Endosymbiosis**
- **C. Phenetic system for classifying microorganisms**
  1. **%GC (G + C ratios)**
  2. **DNA:DNA hybridization**
  3. **Fatty acid methyl ester analysis (FAME)**
- **D. Phylogenetic system for classifying microorganisms: rRNA as a molecular chronometer**
  1. **rRNA sequence analysis**
  2. **ribotyping**
  3. **MLST**
  4. **metagenomics**
  5. **phylogenetic analysis**

The Common Denominator of cells:
All cells need carbon and energy sources.

Energy:

Chemoorganotrophs obtain their energy from the oxidation of organic compounds.

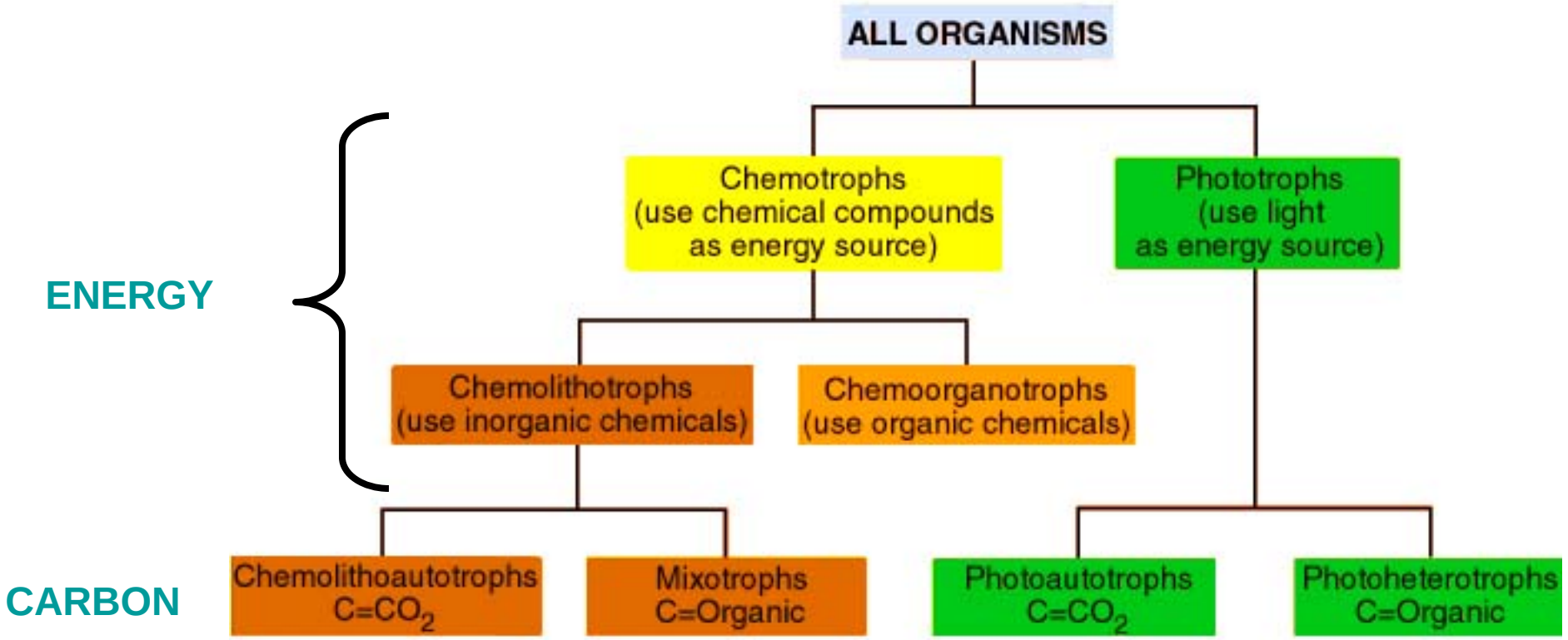
Chemolithotrophs obtain their energy from the oxidation of inorganic compounds.

Phototrophs contain pigments that allow them to use light as an energy source.

Carbon:

Autotrophs use CO_2 as a sole source of carbon.

Heterotrophs use organic compounds as carbon sources.


ALL ORGANISMS
Chemotrophs (use chemical compounds as energy source)
Phototrophs (use light as energy source)
ENERGY
Chemolithotrophs (use inorganic chemicals)
Chemoorganotrophs (use organic chemicals)
Chemolithoautotrophs C=CO2
Mixotrophs C=Organic
Photoautotrophs C=CO2
Photoheterotrophs C=Organic
CARBON

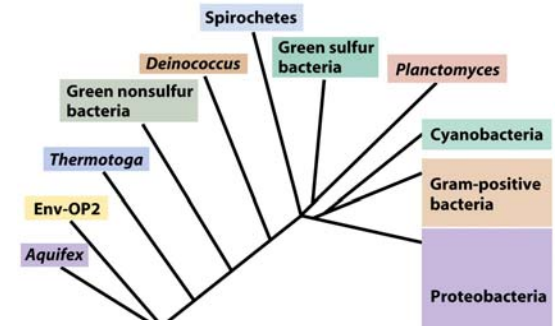
Extremophiles thrive under environmental conditions in which eukaryotic organisms cannot survive.

Table 2.1 Classes and examples of extremophiles[a]

Extreme	Descriptive term	Genus/species	Domain	Habitat	Minimum	Optimum	Maximum
Temperature							
High	Hyperthermophile	*Pyrolobus fumarii*	*Archaea*	Hot, undersea hydrothermal vents	90°C	**106°C**	113°C[b]
Low	Psychrophile	*Polaromonas vacuolata*	*Bacteria*	Sea ice	0°C	**4°C**	12°C
pH							
Low	Acidophile	*Picrophilus oshimae*	*Archaea*	Acidic hot springs	−0.06	**0.7**[c]	4
High	Alkaliphile	*Natronobacterium gregoryi*	*Archaea*	Soda lakes	8.5	**10**[d]	12
Pressure	Barophile	*Moritella yayanosii*[e]	*Bacteria*	Deep ocean sediments	500 atm	**700 atm**	>1000 atm
Salt (NaCl)	Halophile	*Halobacterium salinarum*	*Archaea*	Salterns	15%	**25%**	32% (saturation)

[a] In each category the organism listed is the current "record holder" for requiring a particular extreme condition for growth.
[b] A newly isolated archaeon can apparently grow up to 121°C.
[c] *P. oshimae* is also a thermophile, growing optimally at 60°C.
[d] *N. gregoryi* is also an extreme halophile, growing optimally at 20% NaCl.
[e] *Moritella yayanosii* is also a psychrophile, growing optimally at about 4°C.

Table 2-1 Brock Biology of Microorganisms 11/e



Bacteria

Microbial speleogenesis, or sulfuric acid speleogenesis

Many underground caverns were formed by the dissolution of calcium carbonate deposits by sulfuric acid, which is formed by **sulfide-oxidizing bacteria**.

Example:

1. **Lechuguilla Cave**, NM: deepest cave in the continental USA with 184 Km-worth of passages
2. **Carlsbad Caverns**

How?

1. $H_2S+2O_2 \rightarrow H_2SO_4$
2. $CaCO_3+H_2SO_4 \rightarrow CaSO_4+H_2CO_3$.

Erosion rate: 5 cm/1000 years

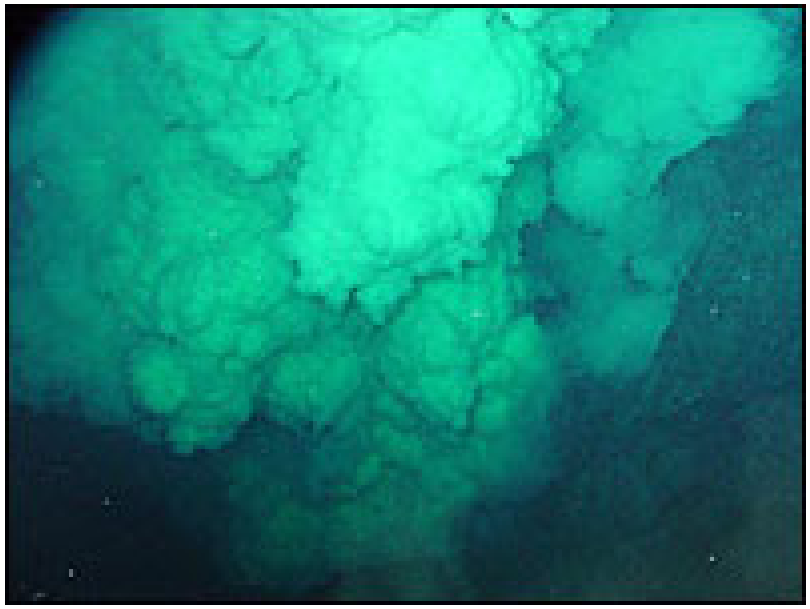
Eubacteria (Proteobacteria):
Beggiatoa, Thiothrix

ACID STREAM. Bacteria accelerate the formation of acidic mine drainage, an environmental problem that taints more than 19,000 kilometers of streams and rivers nationwide. Microbes add to the concentration of H_2SO_4, speeding the leaching of toxic metals (iron, zinc, copper, arsenic) from the rock by a factor of 10.

Pennsylvania Department of Environmental Protection

Glowing, red lava shoots out of the "Brimstone Pit" crater near the summit of NW Rota-1 volcano, located on the floor of the northern Pacific Ocean.

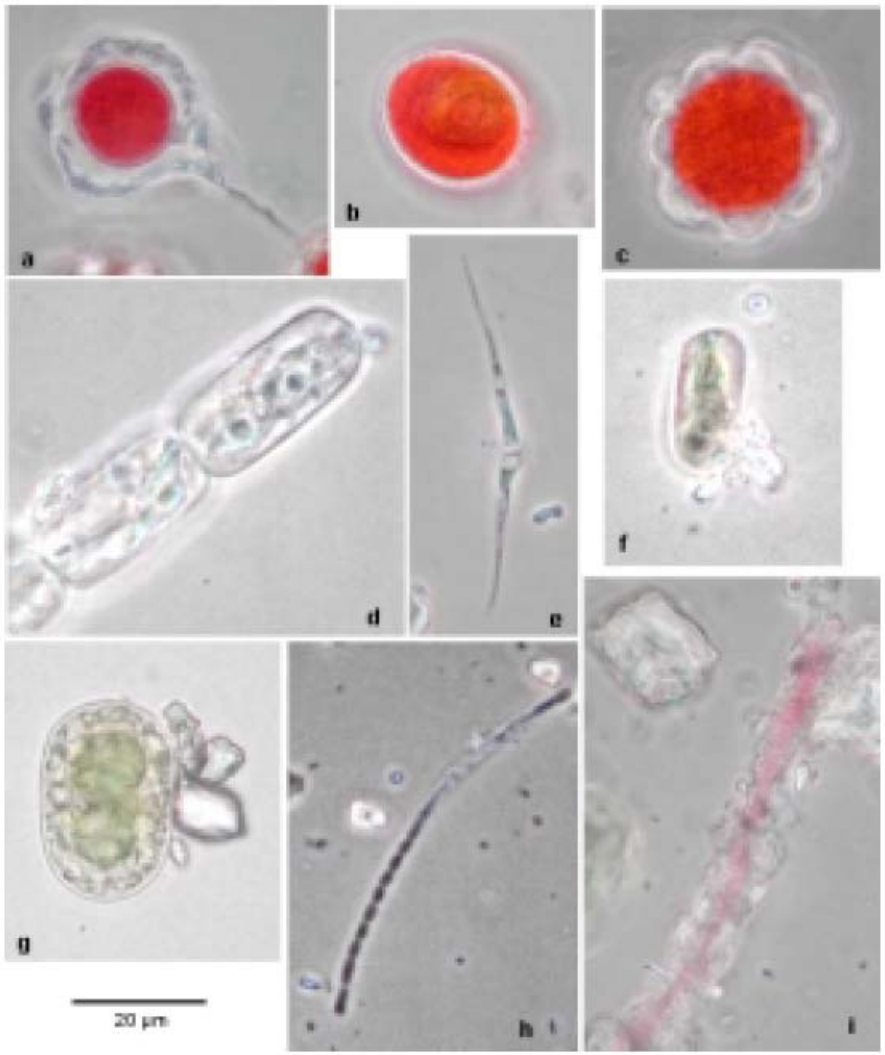
Submarine Ring of Fire 2006 Exploration, NOAA Vents Program



A giant, eruptive plume pours from the "Brimstone Pit" during a 2004 expedition to NW Rota-1.

Submarine Ring of Fire 2004 Exploration, NOAA-OE

http://www.pmel.noaa.gov/vents/

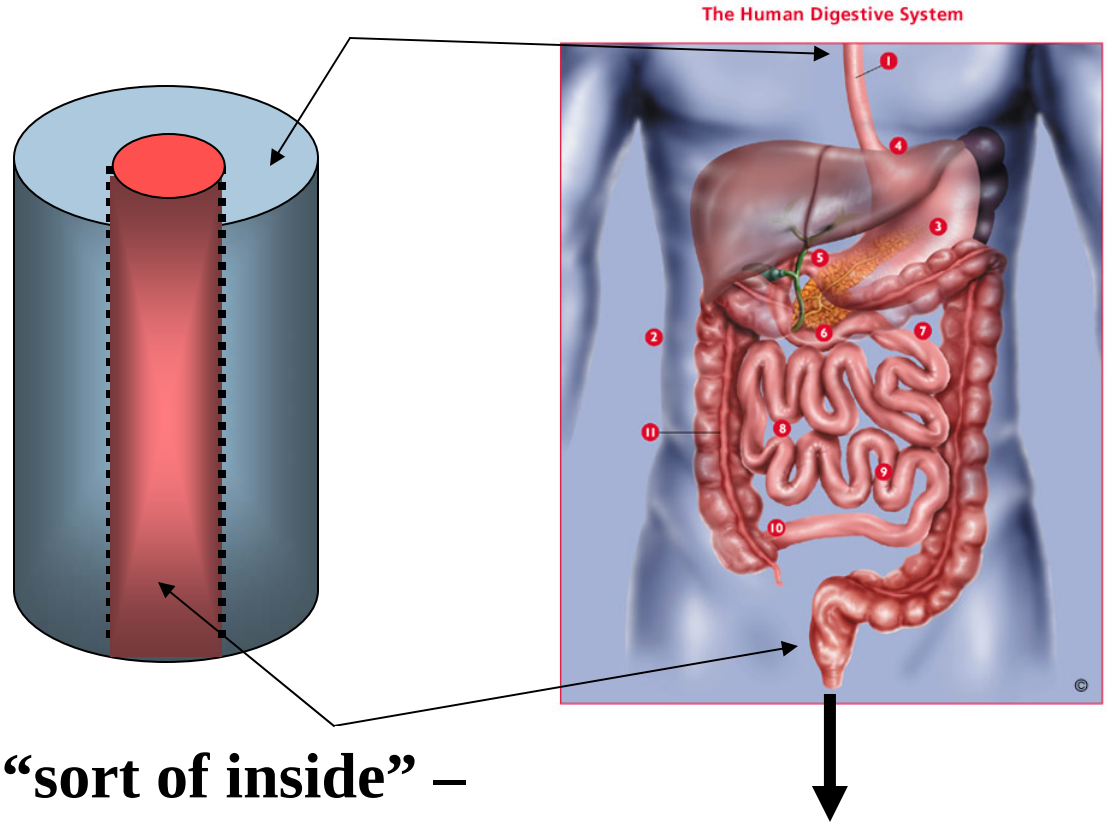


Research and photos:
Dr. Nozomu Takeuchi

Snow algae (Chlorophyta) observed on Gulkana Glacier in the Alaska Range.
a-c: *Chlamydomonas nivalis*
d: *Ancylonema Nordenskioldii*
e: *Koliella* sp.
f: *Mesotaenium bregrenii*
g: *Cylindrocystis brébissonii*
h: Oscilllatriaceae cyanobacteria 1
i: Oscilllatriaceae cyanobacteria 2

You, yourself are a habitat for billions of bacteria.

“really inside” - sterile

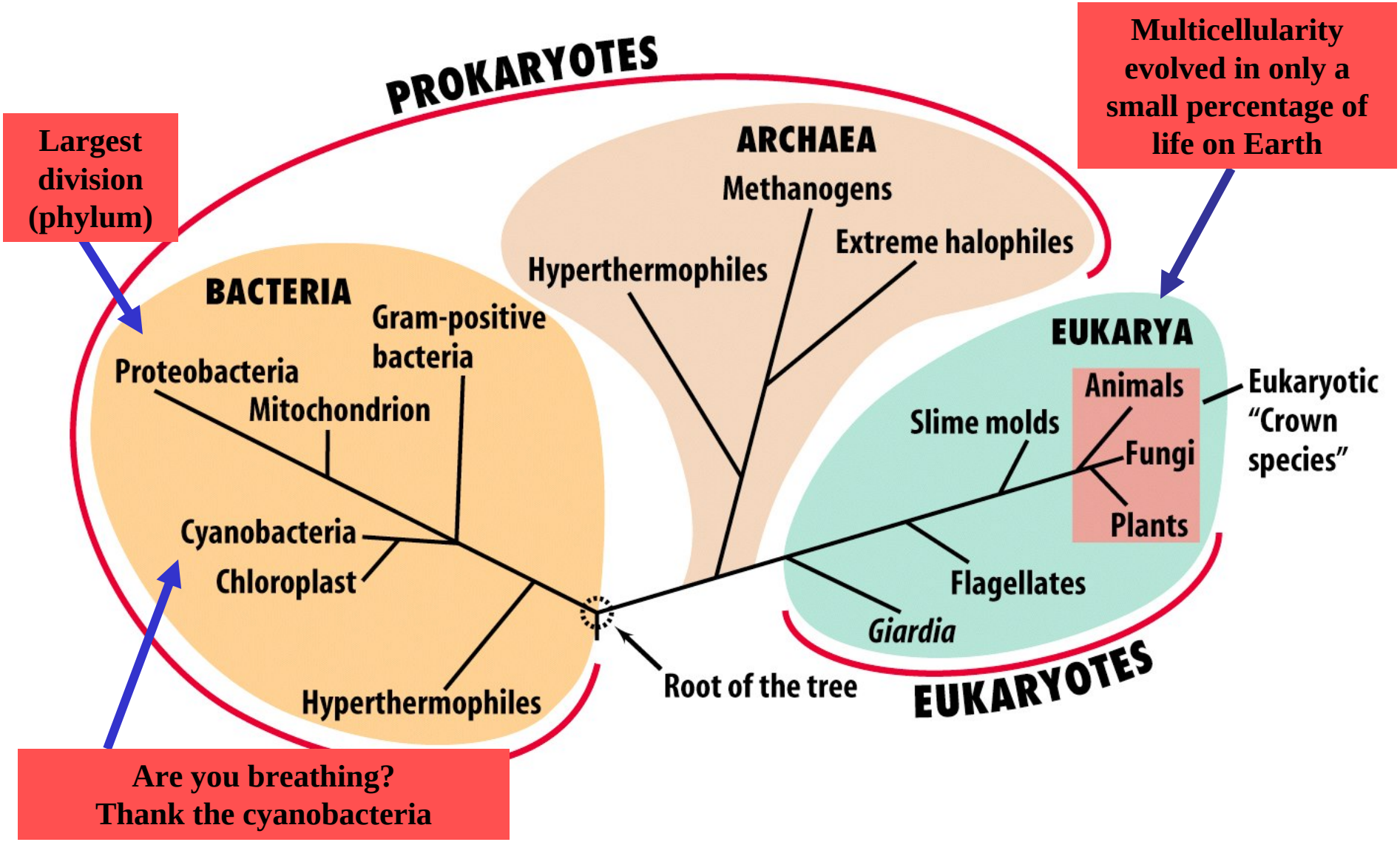


Inside the “you tube”, various microbes find a niche that fits.

“sort of inside” –
a big culture tube

Up to ½ the weight of mammalian fecal matter is bacterial

Prokaryotes represent a huge metabolic and ecological diversity (reservoir of possibilities for life on Earth?)



Archaea in particular represent a huge diversity

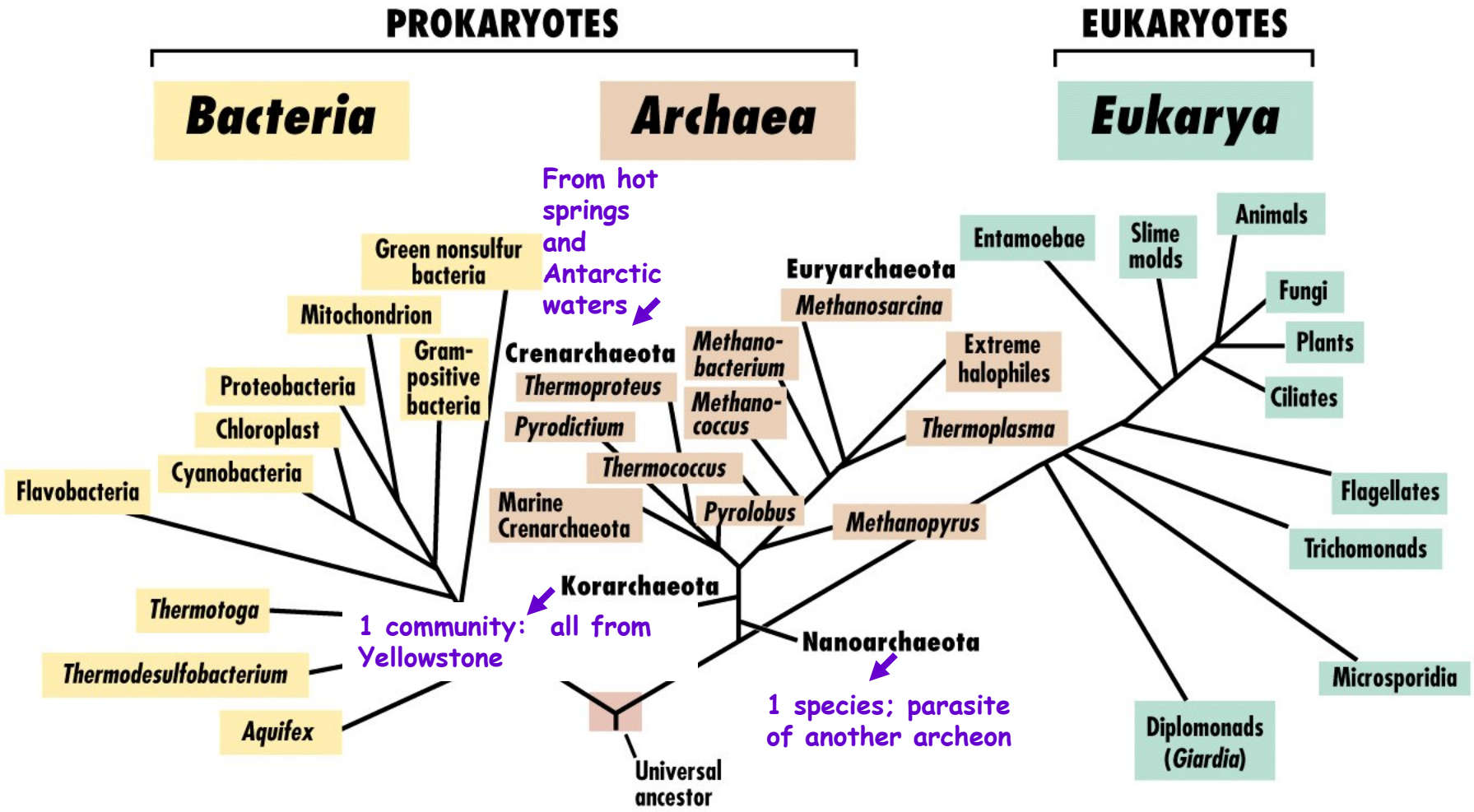


Figure 11-16 Brock Biology of Microorganisms 11/e

Although species of *Bacteria* and *Archaea* share a superficially similar 'prokaryotic' cell structure, they differ dramatically in their evolutionary history and molecular makeup.

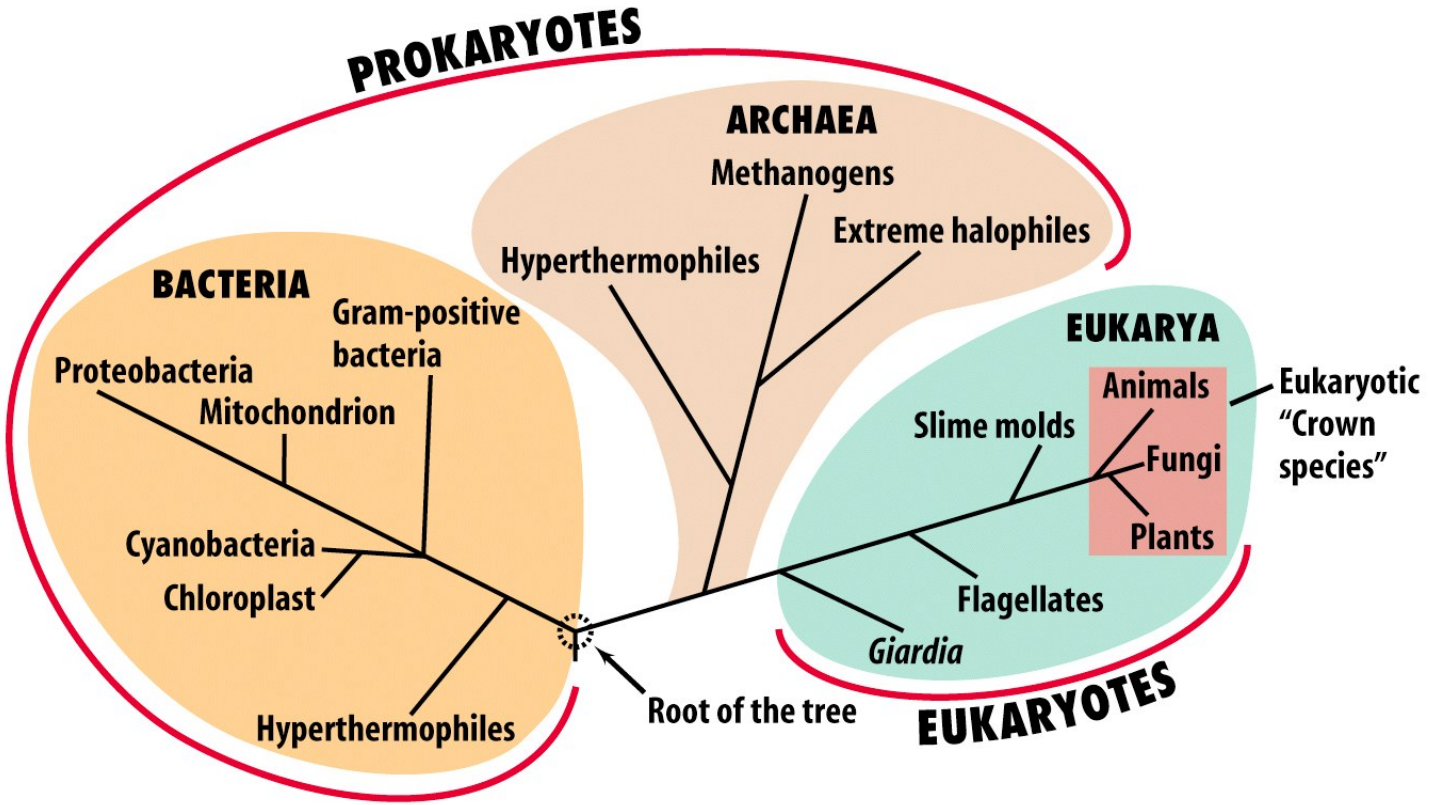


Figure 2-7 Brock Biology of Microorganisms 11/e

Are Archaea more like Bacteria or Eukarya?

	Bacteria	Archaea	Eukaryotes
	Archaea resemble Bacteria		
Cell volume	1 to 100 μm^3 (usually)		1 to 10^6 μm^3
DNA chromosome	Circular (usually)		Linear
Gene organization	operons; few introns		Many introns
Metabolism	Denitrification, N_2 fixation, lithotrophy, respiration and fermentation		Respiration & fermentation
Nuclear membrane	None		Nuclear membrane
Multicellularity	Simple		Complex
Ribosome size	70s		80s

See table 11.3

Are Archaea more like Bacteria or Eukarya?

Bind ribosome →

	Bacteria	Archaea Eukaryotes
		Archaea resemble Eukaryotes
Cell wall	Peptidoglycan (nearly always)	Absent in most species (Methanogens have pseudopeptidoglycan)
Ribosome sensitivity to Cm, Kn, and Sr	Sensitive	Resistant
Ribosomes sensitive to Diptheria toxin	Resistant	Sensitive
Translation initiator	Formyl-Met	Methionine (except mitochondrial F-Met)
RNA polymerase	Bacterial 4 subunit	Eukaryotic 8+ subunits
Transcription factors	Bacterial	Eukaryotic

See table 11.3

Are Archaea more like Bacteria or Eukarya?

	Bacteria	Archaea	Eukaryotes
	Archaea Differ from Bacteria and Eukaryotes		
Methanogenesis	No	Yes	No
Thermophilic growth, Max temp	90°C	113°C	60°C
Photosynthesis	Many species. bacterio-chlorophyll	Halobacteria only; bacterio-rhodopsin	Many species; bacterial chlorophyll
Chlorophyll light absorption	Red and blue	Green	Red and blue
Membrane lipids (major)	Ester-linked fatty acids	Ether-linked isoprenoids	Ester-linked fatty acids
Pathogens that infect animals or plants	Many pathogens	No pathogens	Many pathogens

See table 11.3

Molecular sequencing has also shown that the major organelles of *Eukarya* have evolutionary roots in the *Bacteria*

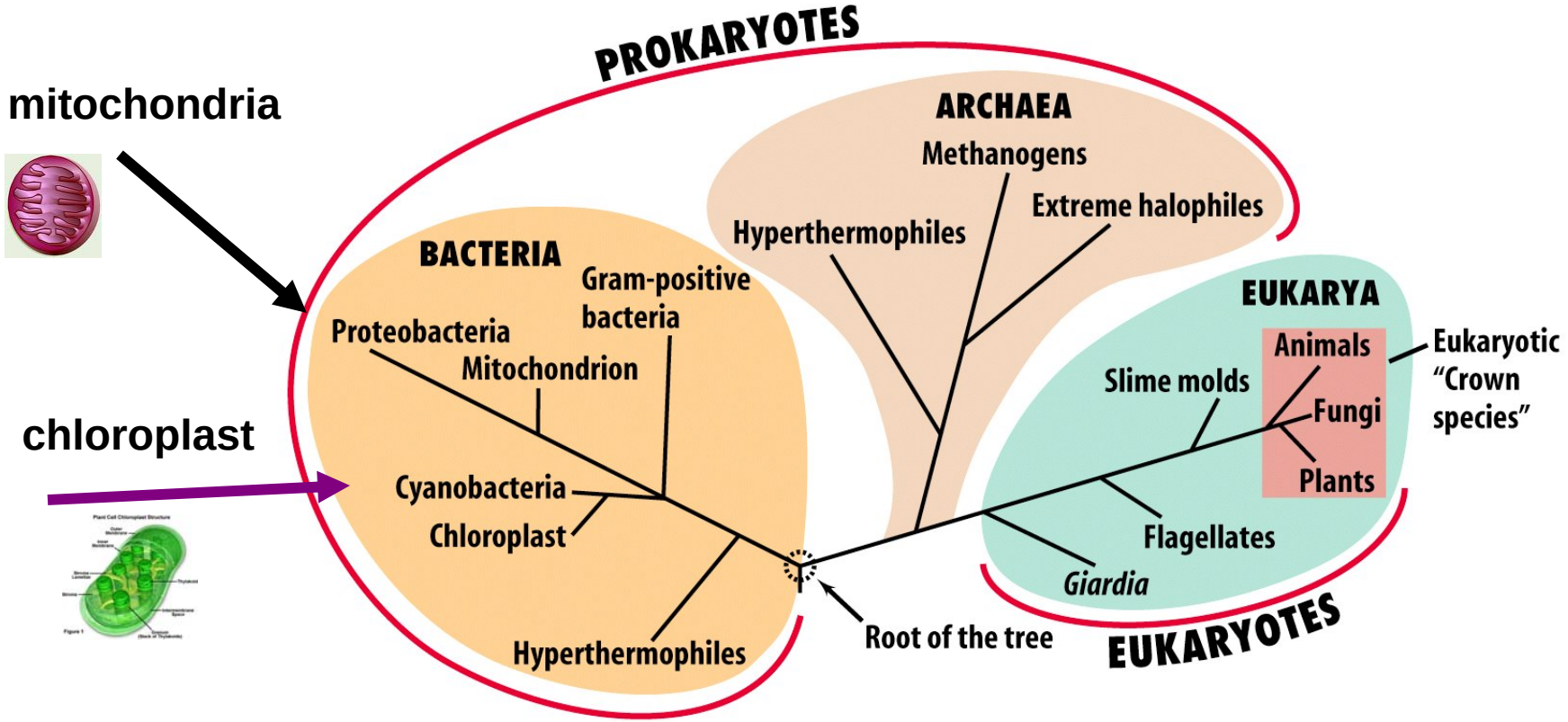


Figure 2-7 Brock Biology of Microorganisms 11/e

Endosymbiotic theory:

Mitochondria and chloroplasts evolved from free-living cells that established stable residency in cells of Eukarya eons ago.

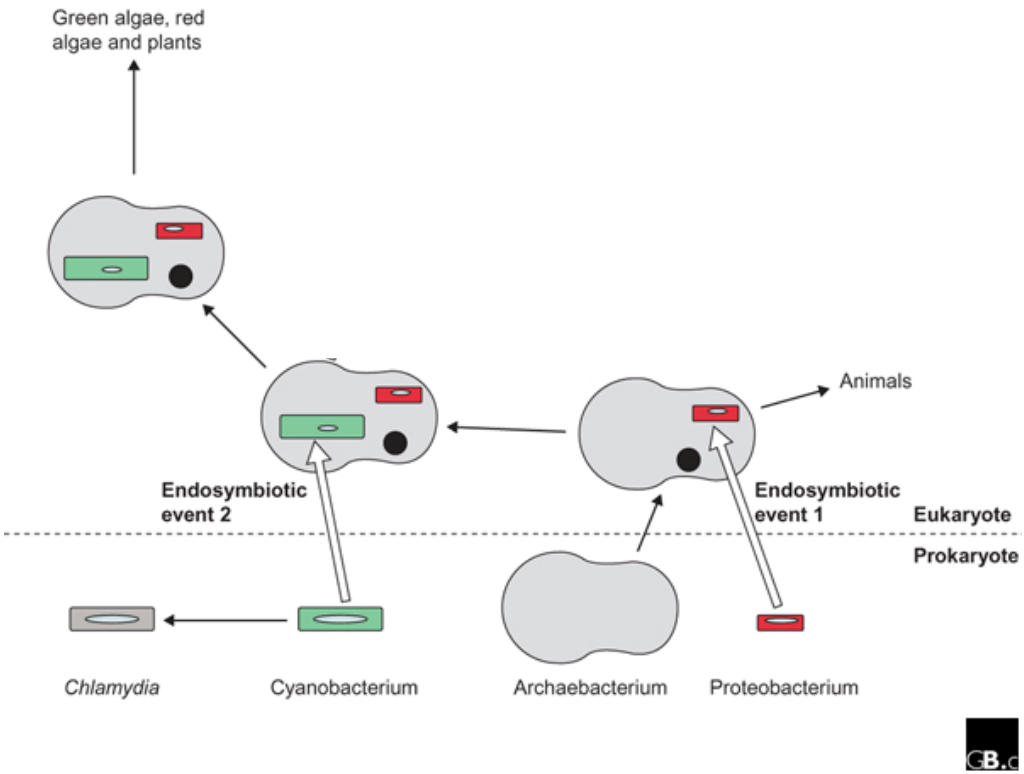
Evidence for endosymbiosis:

Mitochondria and chloroplasts of eukaryotes contain:

-circular genomes

-bacteria-like ribosomes (same antibiotic sensitivities)

Acquisition of genomes and compartments during evolution



Raven *et al. Genome Biology* 2003 **4**:209

Taxonomic classification

Classification - process of arranging organisms into similar or related groups, primarily to provide easy identification and study

Nomenclature - system of assigning names to organisms; binomial system

- Domain - a collection of similar kingdoms
- Kingdom - a collection of similar phyla or classes
- Phylum/division - a collection of similar classes
- Class - a collection of similar orders
- Order - a collection of similar families
- Family - a collection of similar genera
- Genus - a collection of related species
- Species - a group of related isolates or **strains**

Bacillus anthracis

Bacillus cereus

Bacillus anthracis

Escherichia coli (*E. coli*)

E. coli K12 - a specific strain often used in laboratory research
E. coli O157:H7 - a group of strains able to cause a severe diarrheal disease

Table 17.1 Hierarchical classification of the bacterium *Spirochaeta plicatilis*

Taxon	Name
Domain	*Bacteria*
Phylum	*Spirochaetes* (vernacular name: spirochetes)
Class	*Spirochaetes*
Order	*Spirochaetales*
Family	*Spirochaetaceae*
Genus	*Spirochaeta*
Species	*plicatilis*

Microbial Taxonomy

Traditional taxonomy or the classification through **identification** and **nomenclature** of microbes, both "prokaryote" and eukaryote, has been in a mess – we were stuck with it for traditional reasons.

A "natural" taxonomy would be based on evolutionary relatedness: Thus, organisms in same "genus" (a collection of "species") would have similar properties in a fundamental sense.

A natural taxonomy of macrobes has long been possible: Large organisms have many easily distinguished features (e.g., body-plans and developmental processes, that can be used to describe hierarchies of relatedness).

Microbes usually have few distinguishing properties that relate them, so a hierarchical taxonomy mainly has not been possible.

Recent advances in **molecular phylogeny** have changed this picture. We now have a relatively quantitative way to view **biodiversity**, in the context of phylogenetic maps or evolutionary trees.

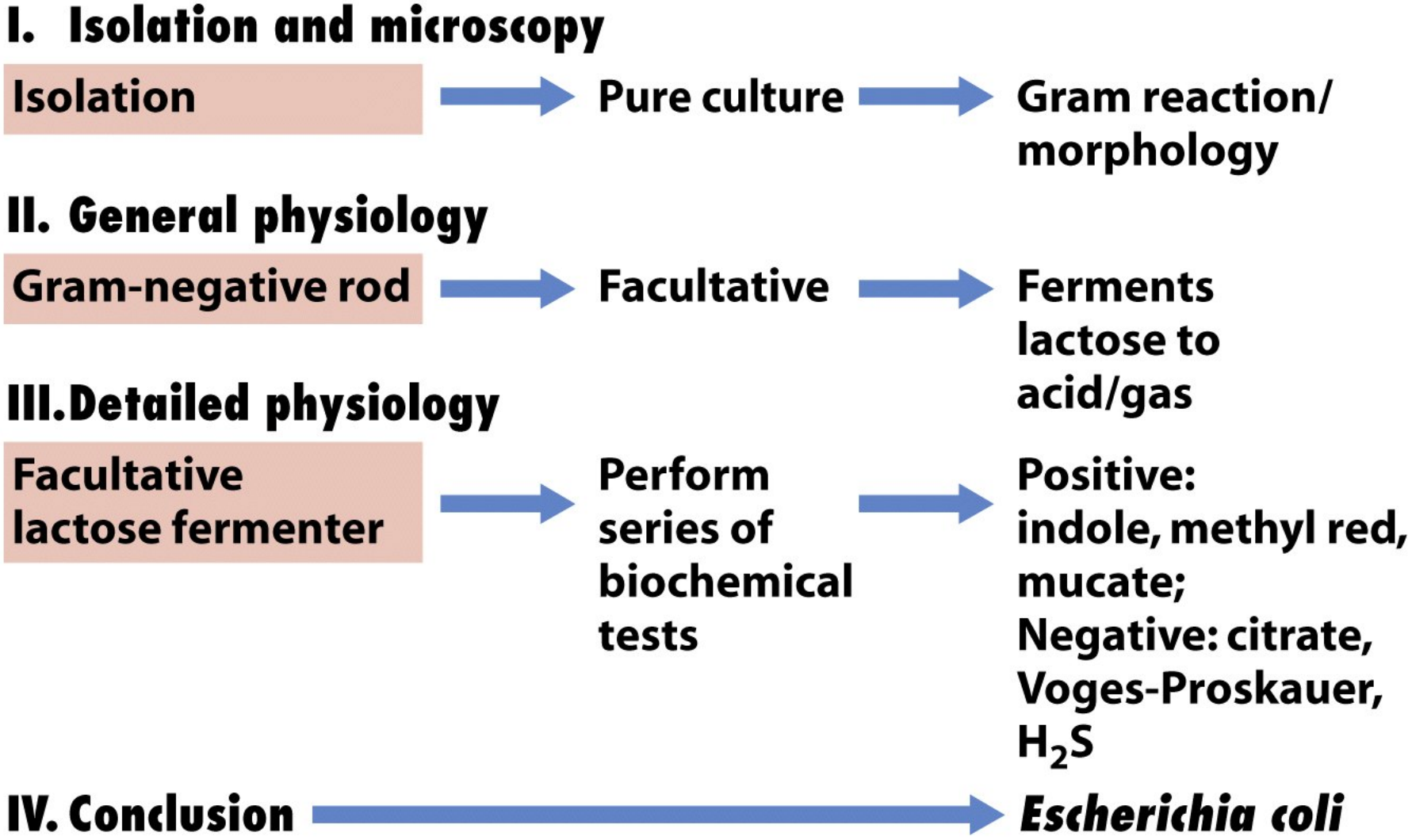
Slowly evolving molecules (e.g. rRNA) used for large-scale structure; "fast- clock" molecules for fine-structure.

The literature language (e.g. "species") and formal nomenclature, however, remain solidly rooted in the tradition of Linnaeus at this time. (You have to call them something!)

Table 11.4 Some phenotypic characteristics of taxonomic value

Major category	Components
I. Morphology	Shape; size; Gram reaction; arrangement of flagella, if present
II. Motility	Motile by flagella; motile by gliding; motile by gas vesicles; nonmotile
III. Nutrition and Physiology	Mechanism of energy conservation (phototroph, chemoorganotroph, chemolithotroph); relationship to oxygen; temperature, pH, and salt requirements/tolerances; ability to use various carbon, nitrogen, and sulfur sources; growth factor requirements
IV. Other factors	Pigments; cell inclusions, or surface layers; pathogenicity; antibiotic sensitivity

ID of an enteric bacterium

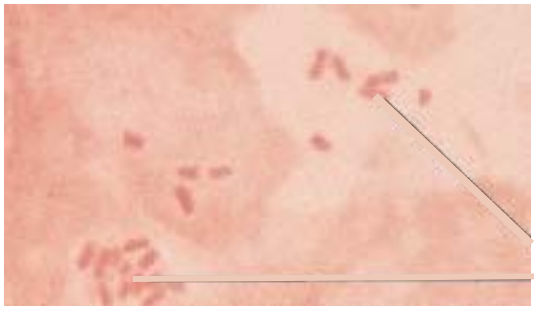


Note: requires isolation in pure culture!

Methods for Microbial Taxonomy and Determination of Evolutionary Relationships

Physiological descriptions of microbes constitute a **taxonomy**, but do not describe evolutionary relationships.

Phenetic system: groups organisms together based on similar phenotypic characteristics.



Appearance (size, shape, staining characteristics)
Metabolic capabilities (ability to break down various compounds)
Other easy-to-observe characteristics (flourescence, pathogenicity)

How do we classify all these diverse life forms?

Table 11.6 **Taxonomic ranks and numbers of known prokaryotic species[a]**

Rank	*Bacteria*	*Archaea*	Total
Domains	1	1	2
Phyla	25	4[a]	29
Classes	34	9	43
Orders	78	13	91
Families	230	23	243
Genera	1227	79	1306
Species	6740	289	7029

[a] Numbers represent validly named genera and species of *Bacteria* and *Archaea* as of 2005. The phyla category for *Archaea* includes the Korarchaeota and the Nanoarchaeota, not yet officially recognized phyla.

Source: Garrity, G.M., Libum, T.G., and Bell, J.A. 2005. *Bergey's Manual of Systematic Bacteriology*, 2d ed., Vol. 2, part A, pp159–220. Springer-Verlag, New York.

We know < 1% of prokaryotes.

Estimates of actual prokaryotic species: 100,000 to 10,000,000

Methods for Microbial Taxonomy and Determination of Evolutionary Relationships

Phenetic system: More in-depth methods may establish relationships, but only if organisms are closely related. Not applicable on broad evolutionary landscapes.

%GC (G + C ratios)

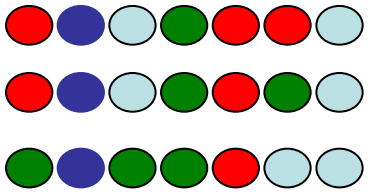
DNA:DNA hybridization

Fatty acid methyl ester analysis (FAME

Ribotyping

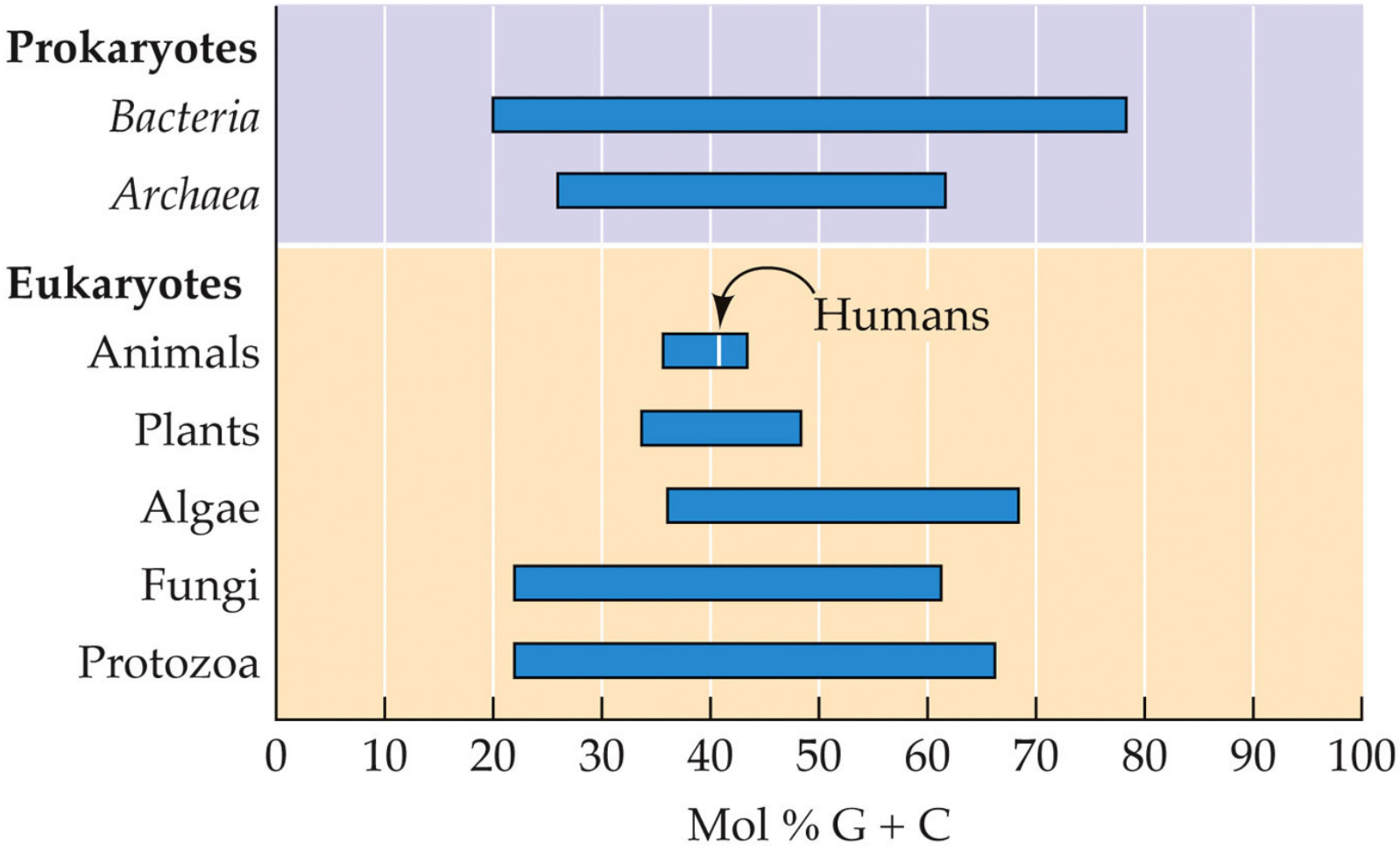
Phyletic system: compares organisms based on evolutionary relationships.

rRNA sequence comparison

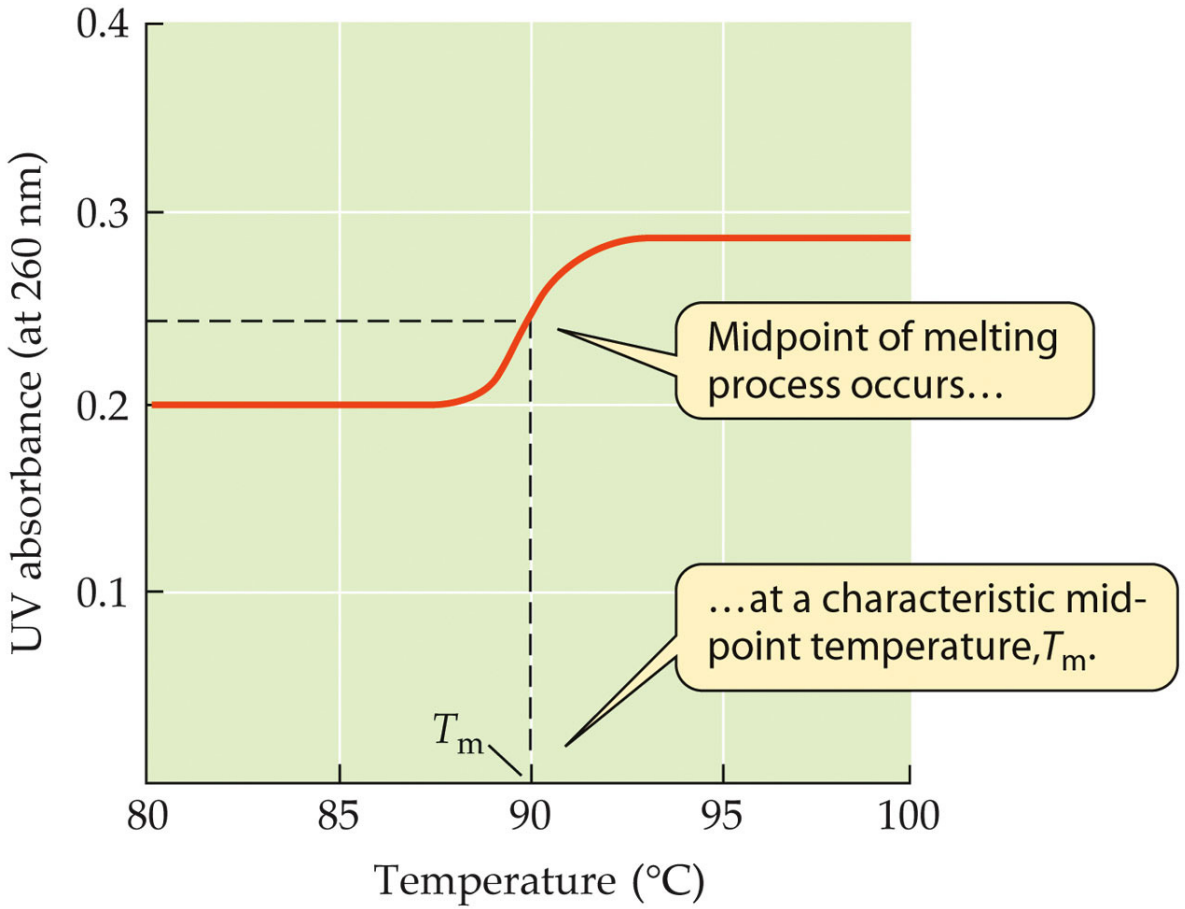


Multilocus sequence typing (MLST)

%GC (G + C ratios): range of DNA base composition

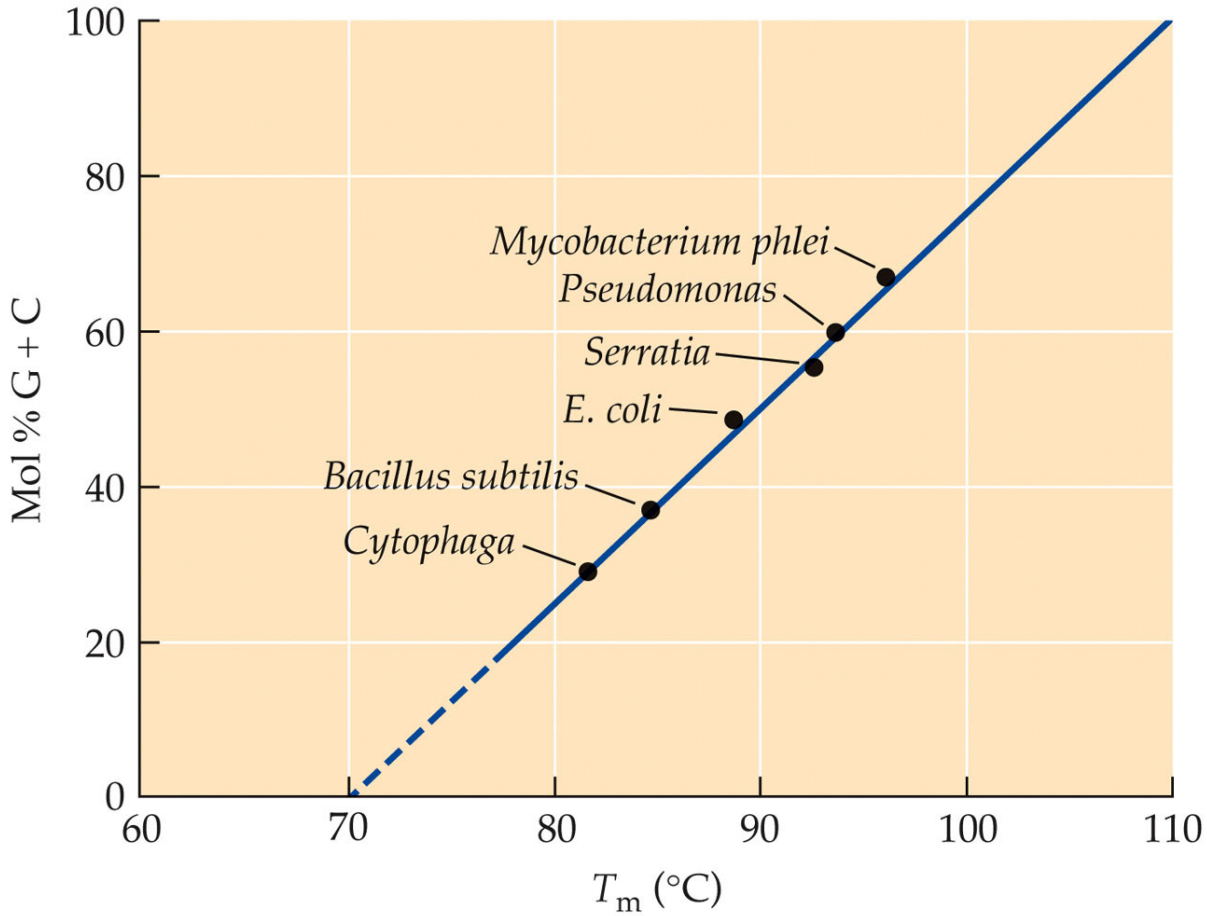


%GC (G + C ratios): Hyperchromic effect



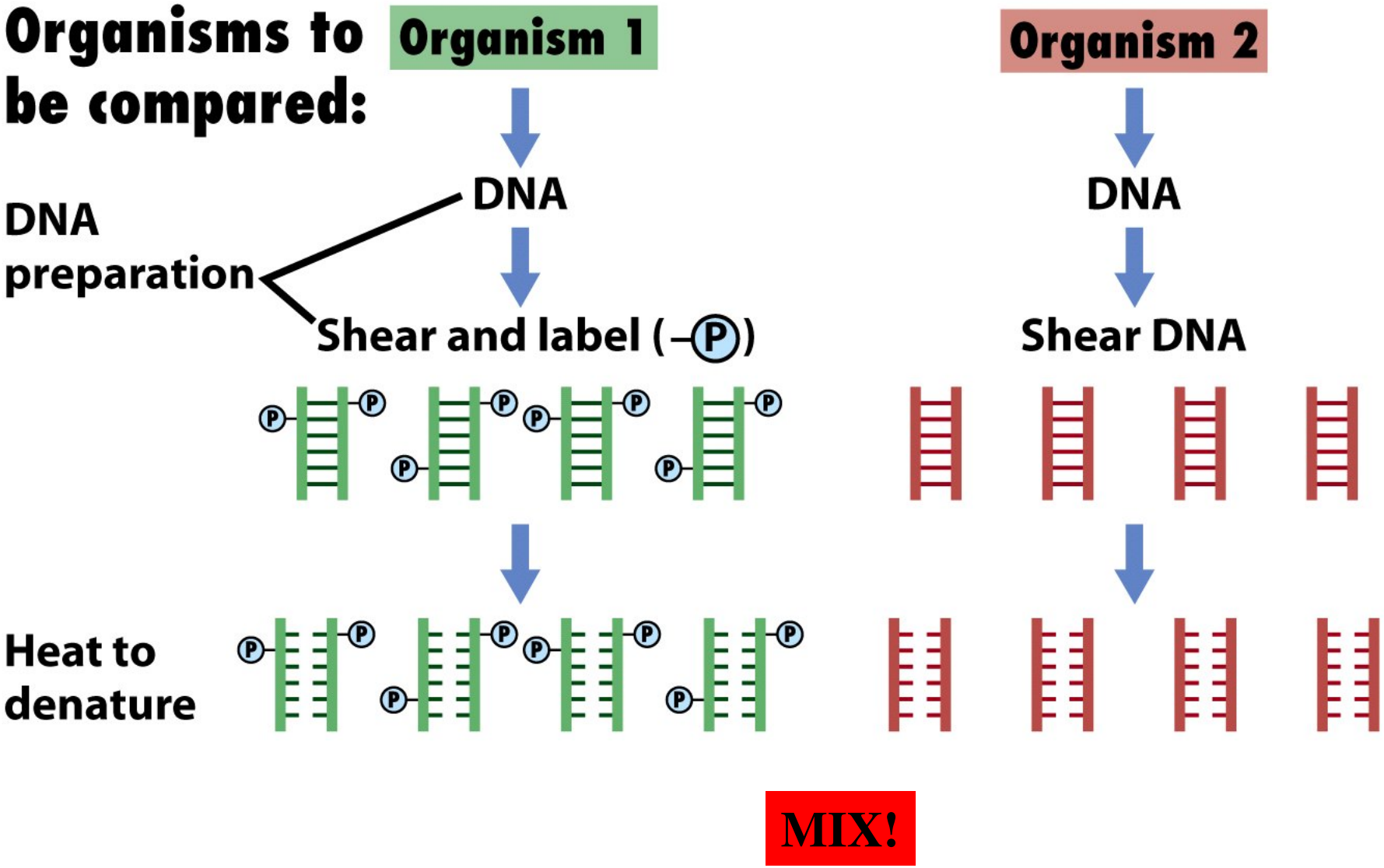
In ds DNA, absorption is less than in ss DNA due to base-stacking interactions. When DNA is denatured, these interactions are disrupted and an increase in absorbance is seen. This change is called the **hyperchromic effect.**

%GC (G + C ratios): inferences



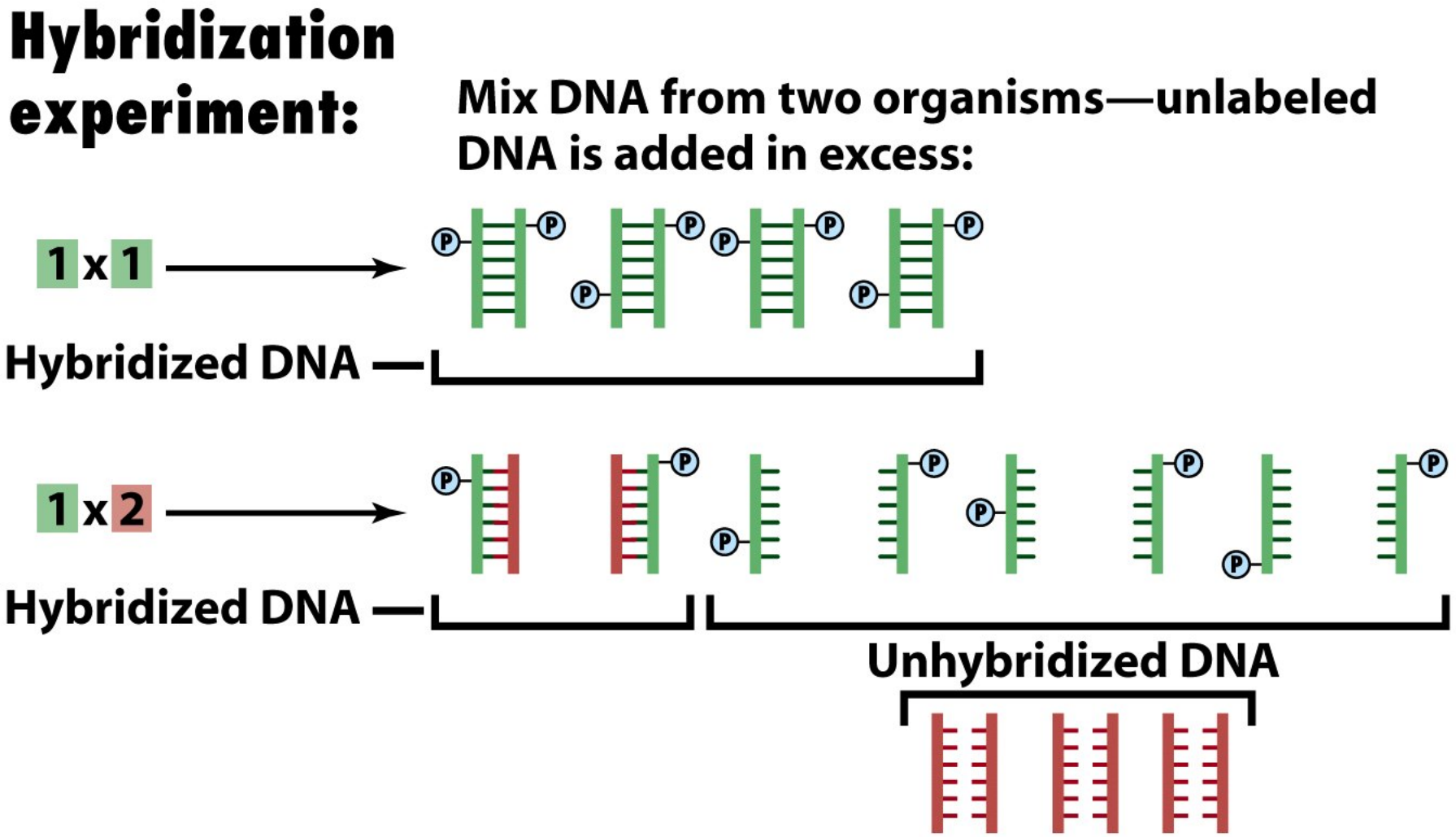
Similarity ≠ sequence identity/ relatedness, BUT
Dissimilarity = sequence differences/unrelatedness.

DNA:DNA hybridization



MIX!

DNA:DNA hybridization

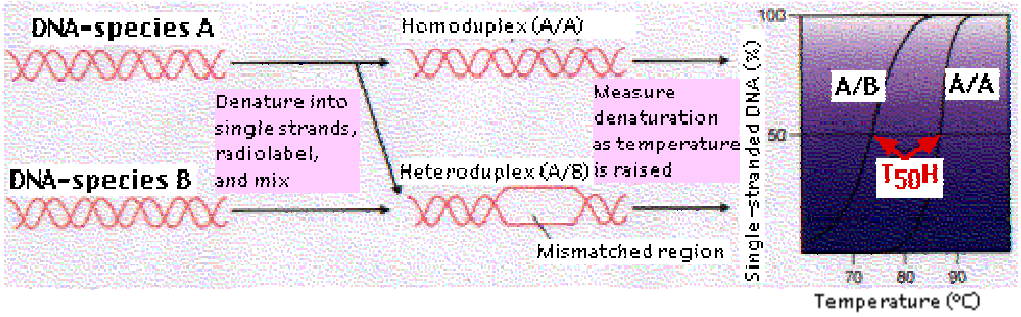


1. **The ssDNA of species A is made radioactive.**

2. **The radioactive ssDNA is then allowed to hybridize with nonradioactive ssDNA of the same species (A) as well as — in a separate tube — the ssDNA of species B.**

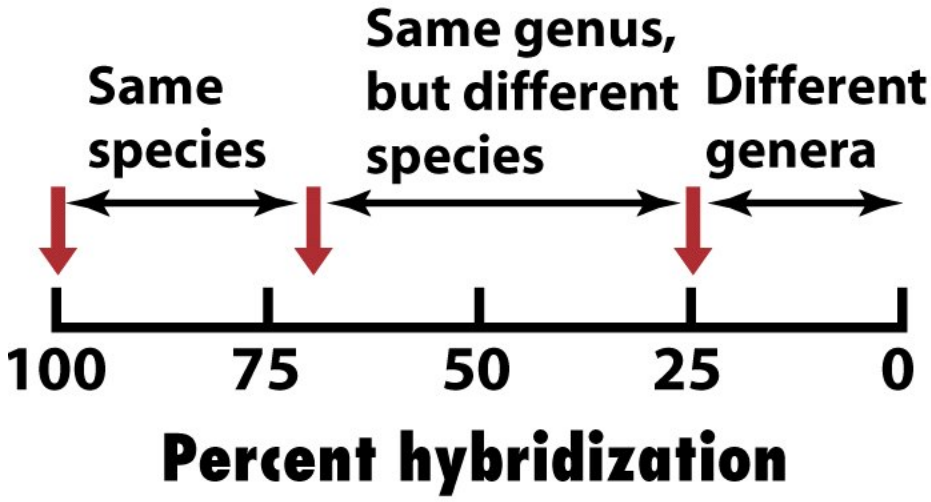
3. **After hybridization is complete, the mixtures (A/A) and (A/B) are individually heated in small (2°–3°C) increments. At each higher temperature, an aliquot is passed over hydroxyapatite. The dsDNA sticks to the hydroxyapatite; ssDNA does not and flows right through. Any radioactive strands (A) that have separated from the DNA duplexes pass through the column, and the amount is measured from their radioactivity.**

4. **A graph showing the percentage of ssDNA at each temperature is drawn. The temperature at which 50% of the DNA duplexes (dsDNA) have been denatured (T50H) is determined.**



DNA:DNA hybridization: inferences

Results and interpretation:



1 x 1

100%

Same strain (control)

1 x 2

25%

1 and 2 are likely different genera

70% or greater; considered same species

Fatty acid methyl ester (FAME) analysis

Classes of Fatty Acids in *Bacteria*

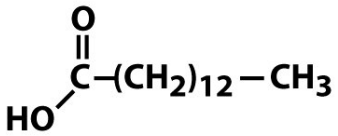
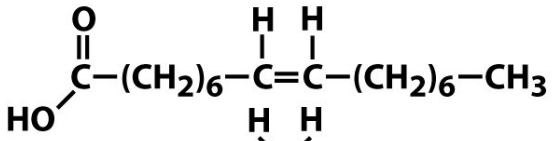
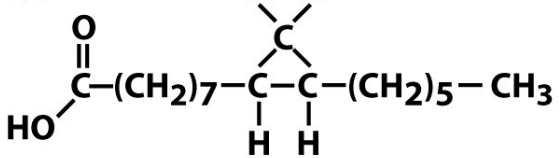
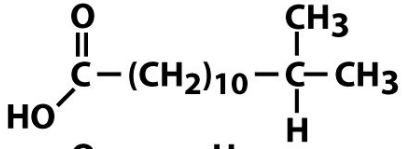
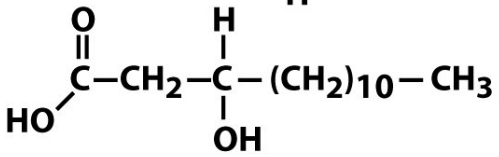
Class / Example	Structure of example
I. ***Saturated:*** tetradecanoic acid	
II. ***Unsaturated:*** *omega-7-cis* hexadecanoic acid	
III. ***Cyclopropane:*** *cis 7, 8* methylene hexadecanoic acid	
IV. ***Branched:*** 13-methyltetradecanoic acid	
V. ***Hydroxy:*** 3-hydroxytetradecanoic acid	

Figure 11-24a Brock Biology of Microorganisms 11/e

Fatty acid methyl ester (FAME) analysis

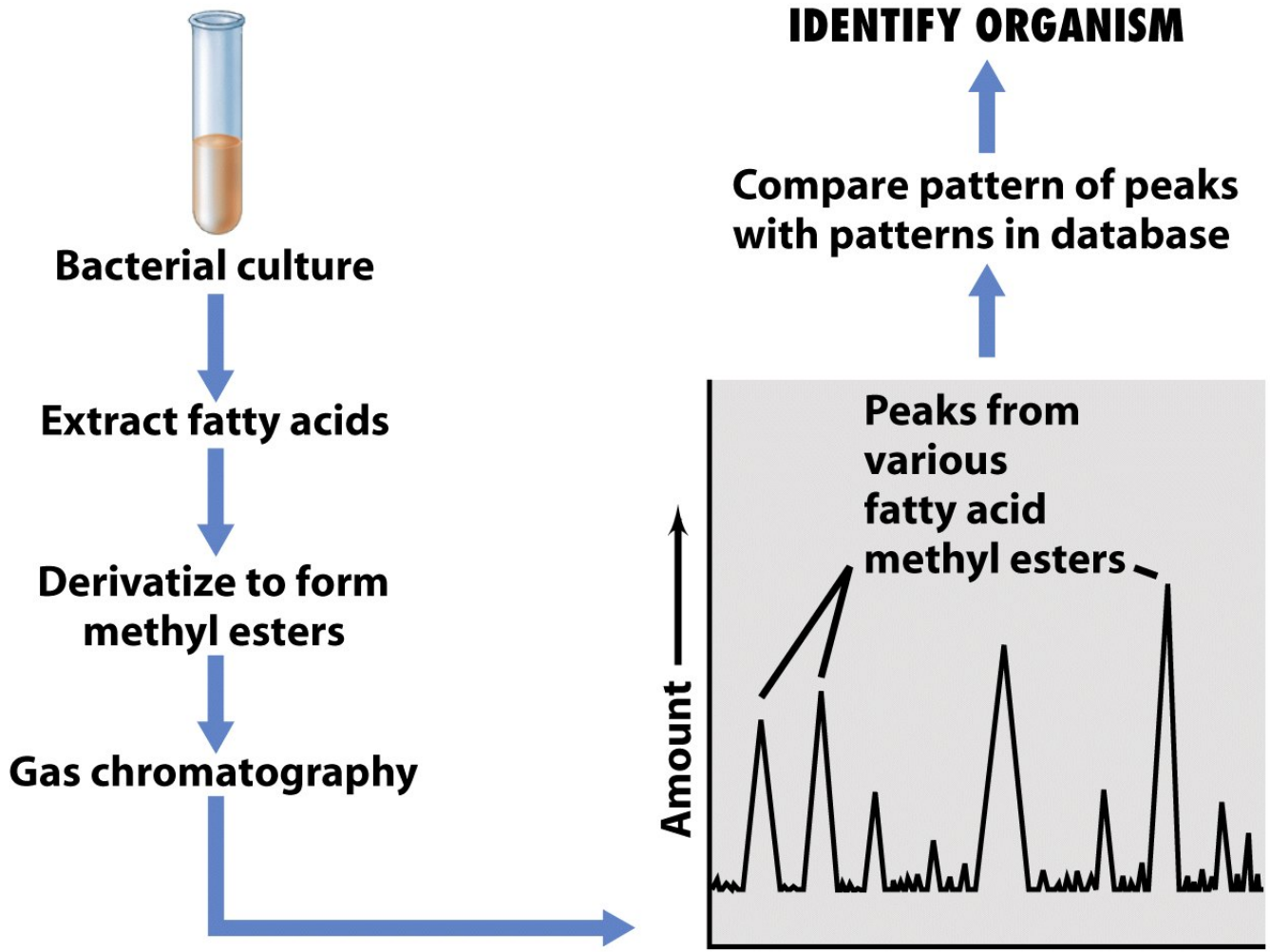
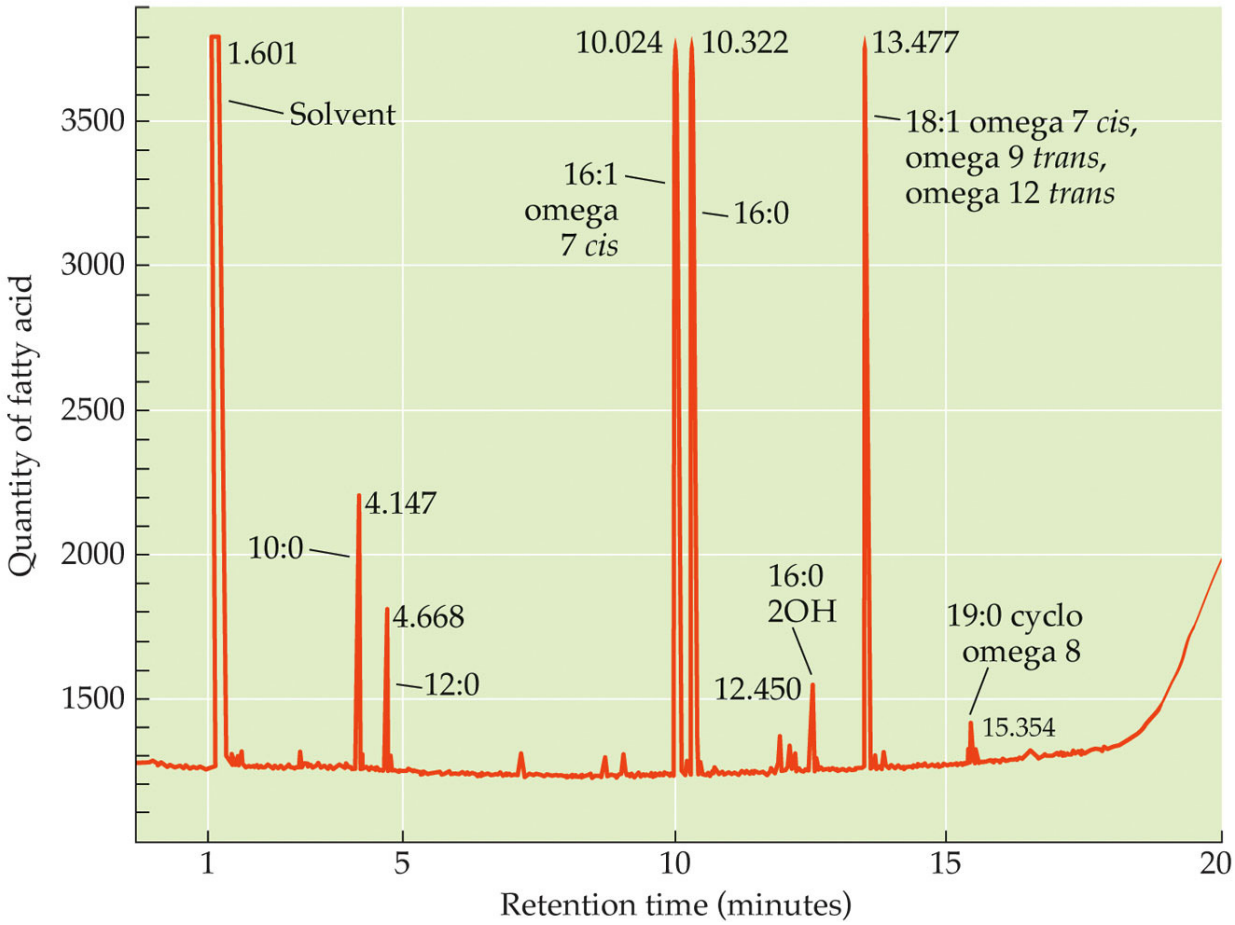
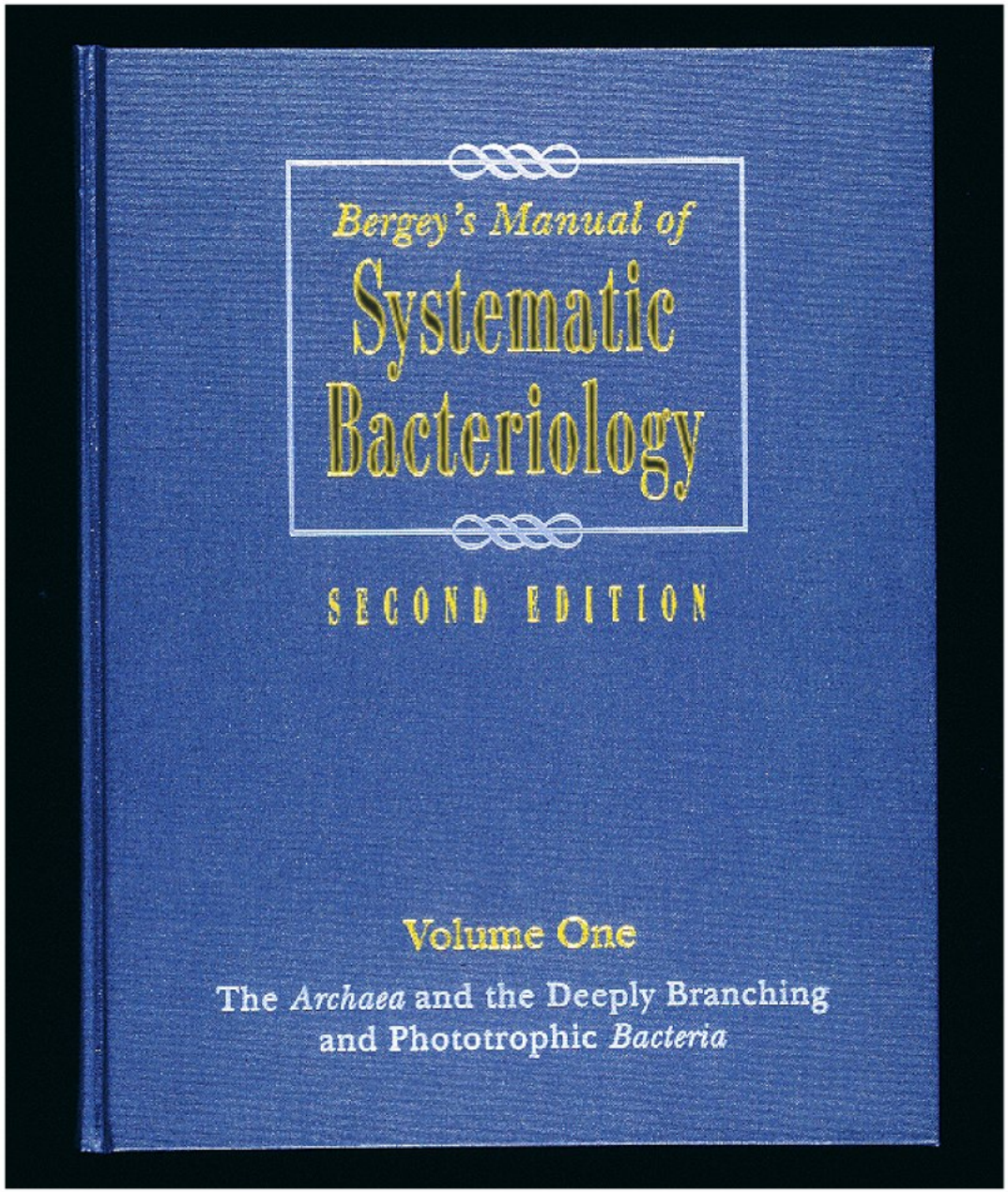


Figure 11-24b Brock Biology of Microorganisms 11/e

Fatty acid methyl ester (FAME) analysis: inferences



FAME analysis can differentiate closely related prokaryotes, but it's not so useful for distantly related organisms


Bergey's Manual of
Systematic Bacteriology
SECOND EDITION
Volume One
The Archaea and the Deeply Branching and Phototrophic Bacteria

Taxonomy Summary

Phenetic system: groups organisms together based on similar phenotypic characteristics

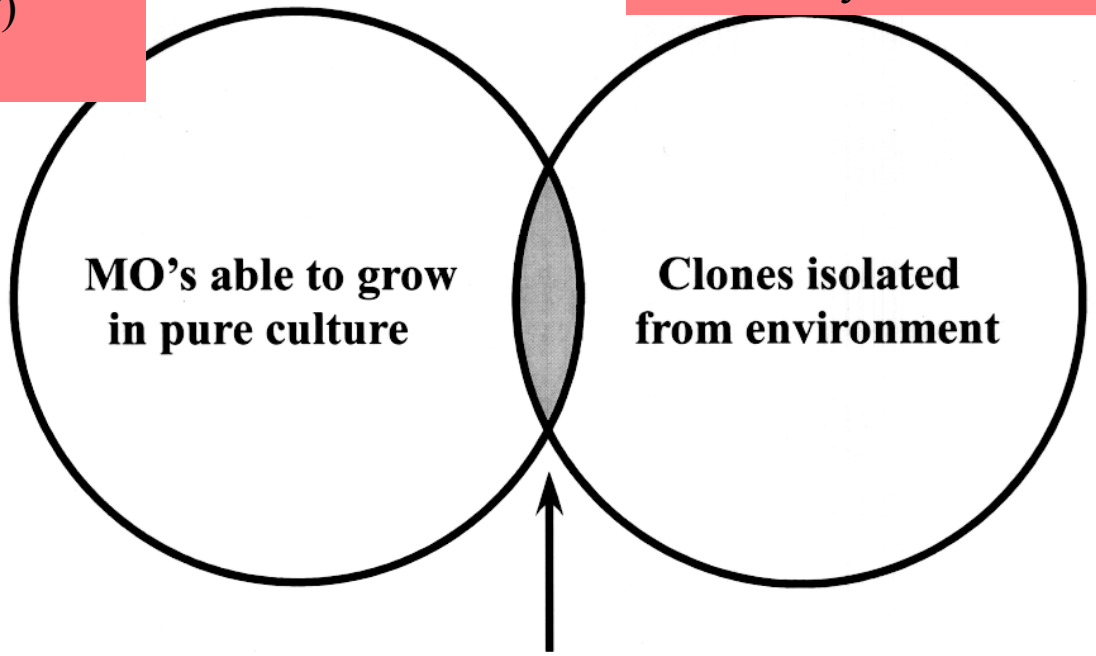
Classical physiological descriptions of microbes constitute a taxonomy, but do not provide relationships (except as might be inferred subjectively).

Methods such as G+C ratios, FAME, DNA-DNA hybridization, or REP PCR establish relationships, but only if close, i.e., they are not sufficiently general to be broadly applicable.

All these methods require pure-cultivation of organisms for characterization, but we can't cultivate much of what is out there.

Traditional culturing techniques isolate ~1% of the total bacteria in marine ecosystems, severely underestimating diversity and community structure

Nutrient-rich **culture media** select for copiotrophic (weedy) bacteria



Less than 1% Crossover between these groups

…how to classify organisms that can neither be seen with the naked eye, nor cultured?

Yet these make up the majority of Earth's biomass!

rRNA sequencing and the Tree of Life

Phyletic system: compares organisms based on evolutionary relationships.

rRNA methods: NOT culture-based.

Ribosomal RNAs (rRNAs) and its respective genes (DNA) are excellent descriptors of microbial taxa based on **phylogeny**.

Evolution is the change in a line of descent (*e.g. heritable change*) over time leading to new species or varieties.

The evolutionary relationships between life forms are the subject of the science of **phylogeny**.

Phyletic system: compares organisms based on evolutionary relationships.

Regarding Molecular Phylogeny

The Root of the Problem: Unlike zoology and botany, microbiology developed without the knowledge of phylogenetic relationships among the organisms studied.

Woese (1977): Applied rRNA concept to redefine microbial systematics or the need to understand microbial genealogy.

Pace (1984): Applied rRNA concept to microbial ecology's need to take a census ("see" without culturing).

... the general course of evolution [for bacteria] will probably never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds... For these and other reasons, most modern taxonomists have explicitly abandoned the phylogenetic approach.

(Stanier *et al.*, *1976*)

Why ribosomal RNAs?

Found among all living organisms (for 3.8 of the last 4.5 billion years). Integral part of protein synthesis machinery.

Cell component analyses provide culture-independent means of investigating questions in microbial ecology (lack of morphology).

rRNAs offer a type of sequence information that makes them excellent descriptors of an organism's evolutionary history.

No detectable horizontal gene transfer, especially important for the prokaryotes.

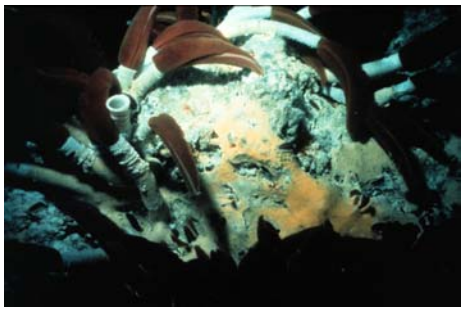
Large and growing database; RDP contains ~100K SSU rRNAs.

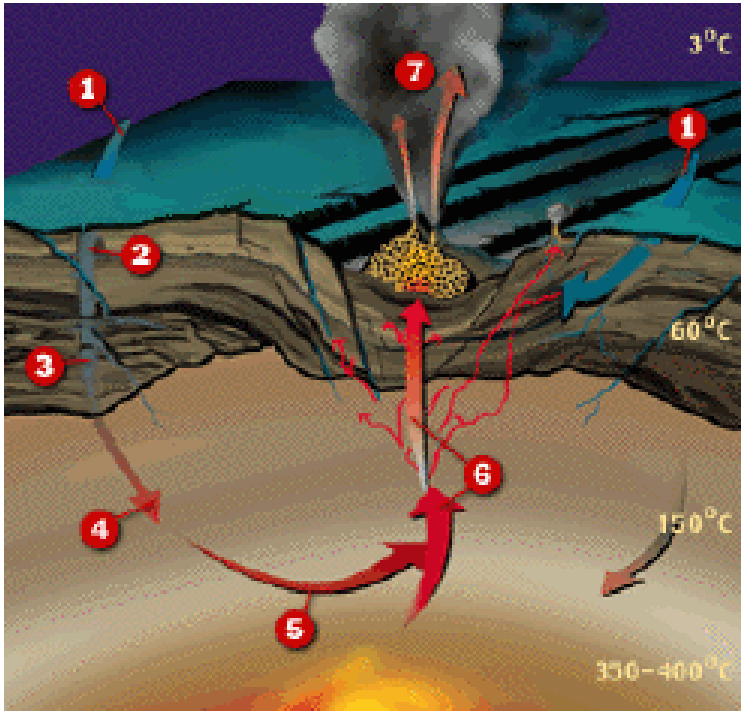
Prokaryotic Cells in the Hydrothermal Vent Tube Worm *Riftia pachyptila* Jones: Possible Chemoautotrophic Symbionts

COLLEEN M. CAVANAUGH, STEPHEN L. GARDINER, MEREDITH L. JONES, HOLGER W. JANNASCH, and JOHN B. WATERBURY

"The existence of a symbiotic association between vestimentiferan tube worms from deep-sea hydrothermal vents and chemoautotrophic sulfur-oxidizing prokaryotes, based on histological and enzymatic evidence… "

Submitted on October 20, 1980



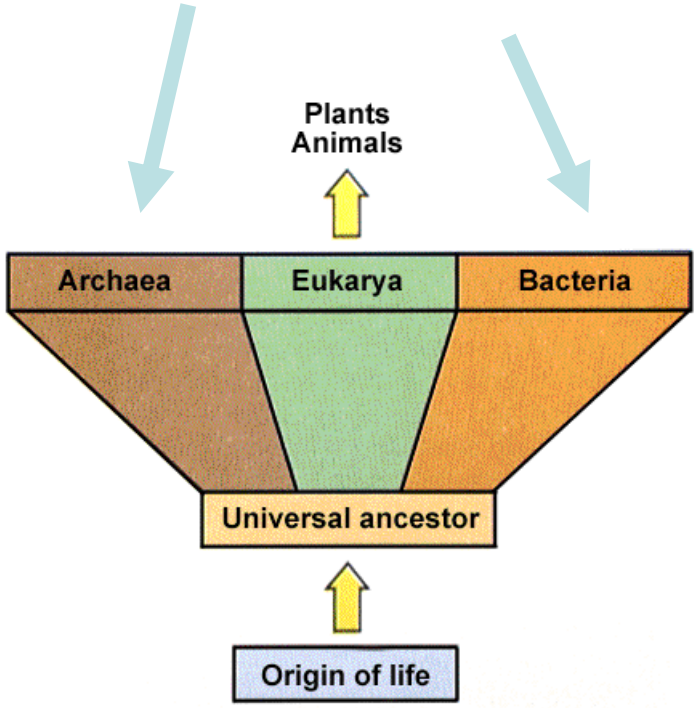


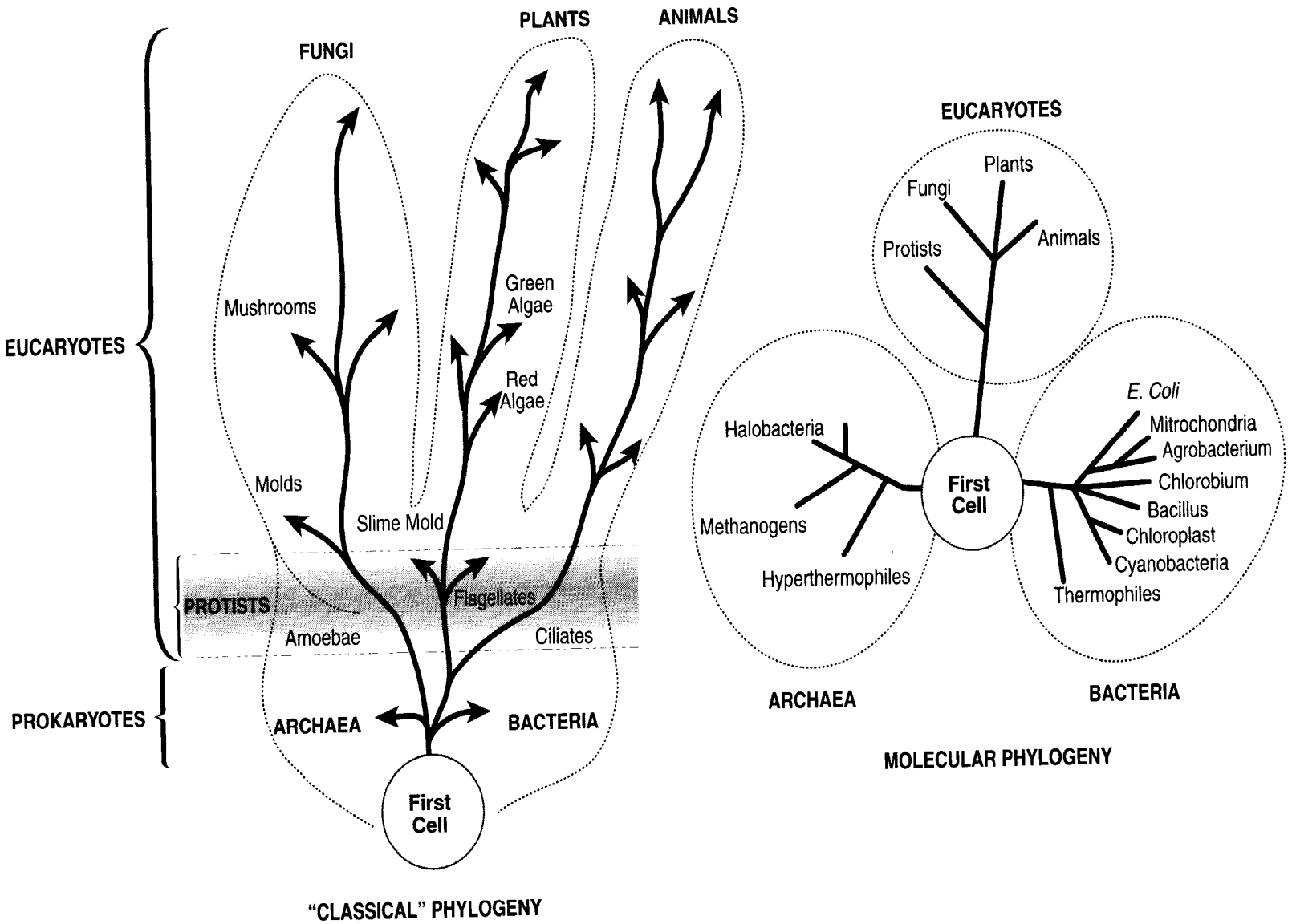
In 1977, Carl Woese and George Fox:

identified Archea (Greek: "ancient ones"), a separate branch of life from eubacteria.

demonstrated that differences in *rRNA sequences* usefully reflect evolutionary relationships.

Two types of prokaryotic cells




PLANTS
ANIMALS
FUNGI
EUCARYOTES
Mushrooms
Green Algae
Red Algae
Molds
Slime Mold
PROTISTS
Flagellates
Amoebae
Ciliates
PROKARYOTES
ARCHAEA
BACTERIA
First Cell
"CLASSICAL" PHYLOGENY
EUCARYOTES
Plants
Fungi
Protists
Animals
Halobacteria
First Cell
E. Coli
Mitrochondria
Agrobacterium
Chlorobium
Bacillus
Chloroplast
Cyanobacteria
Methanogens
Hyperthermophiles
Thermophiles
ARCHAEA
BACTERIA
MOLECULAR PHYLOGENY

Certain molecules are "**molecular chronometers**": differences in nt or aa sequences of homologous molecules are a function of their evolutionary distance.

They are:

-universally distributed among all living organisms (essential for even the most primitive cells)

-functionally homologous

-lack horizontal gene transfer that could confound phylogenetic analysis

Can't accumulate many mutations in such an important macromolecule… so, evolutionary distance between rRNAs reflects evolutionary distance between organisms. Molecular chronometers let us look deep into the evolutionary past.

Useful features of molecular chronometers:

-regions of sequence conservation so DNA can be aligned

-sequence change should reflect evolutionary change in organism as a whole

Examples of molecular chronometers:

rRNA, ATPase, RecA, DNA polymerase, etc.

rRNA is the most widely used.

A huge database of rRNA sequences exists. For example, the **Ribosomal Database Project (RDP)** contains a large collection of such sequences, now numbering over 100,000.

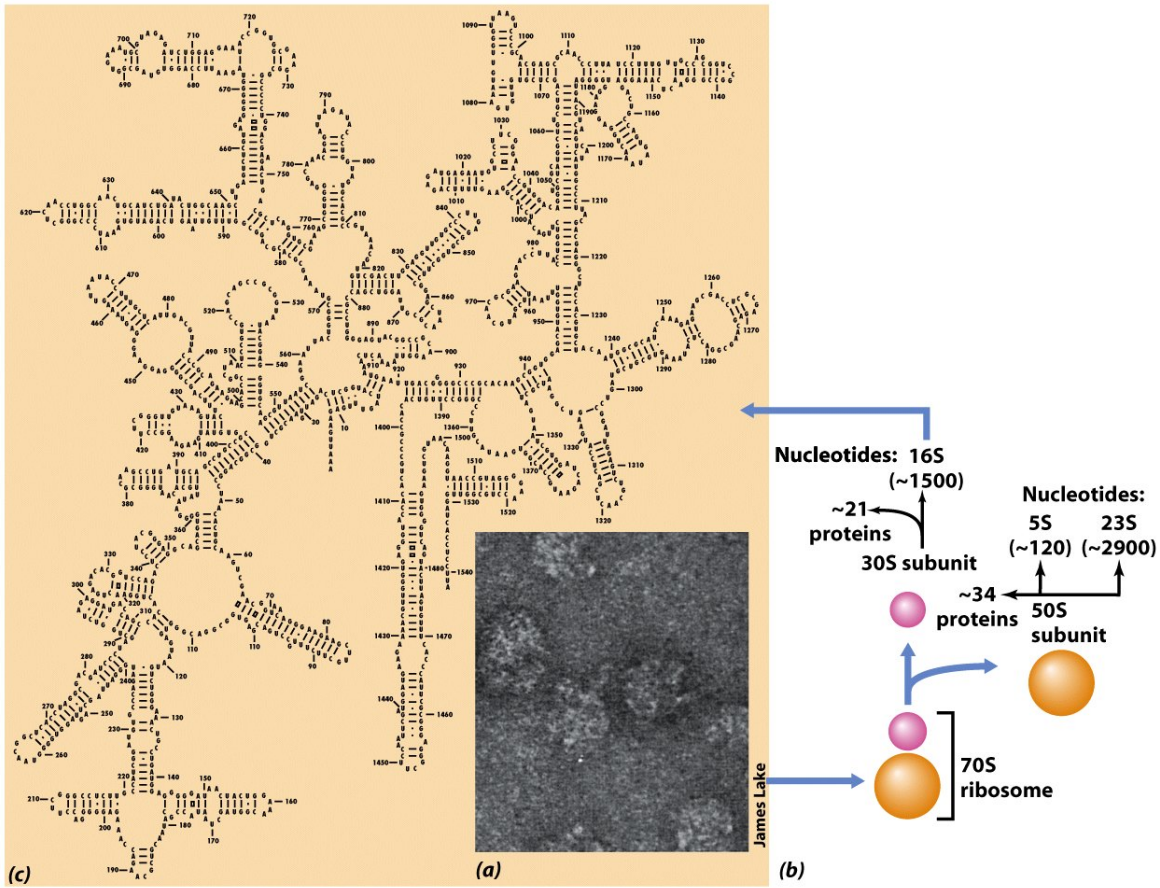
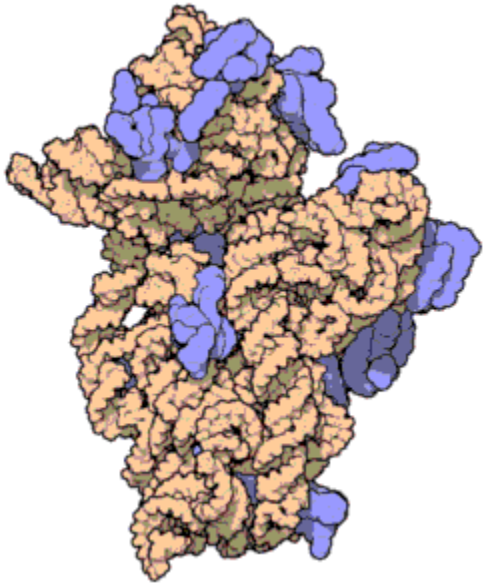
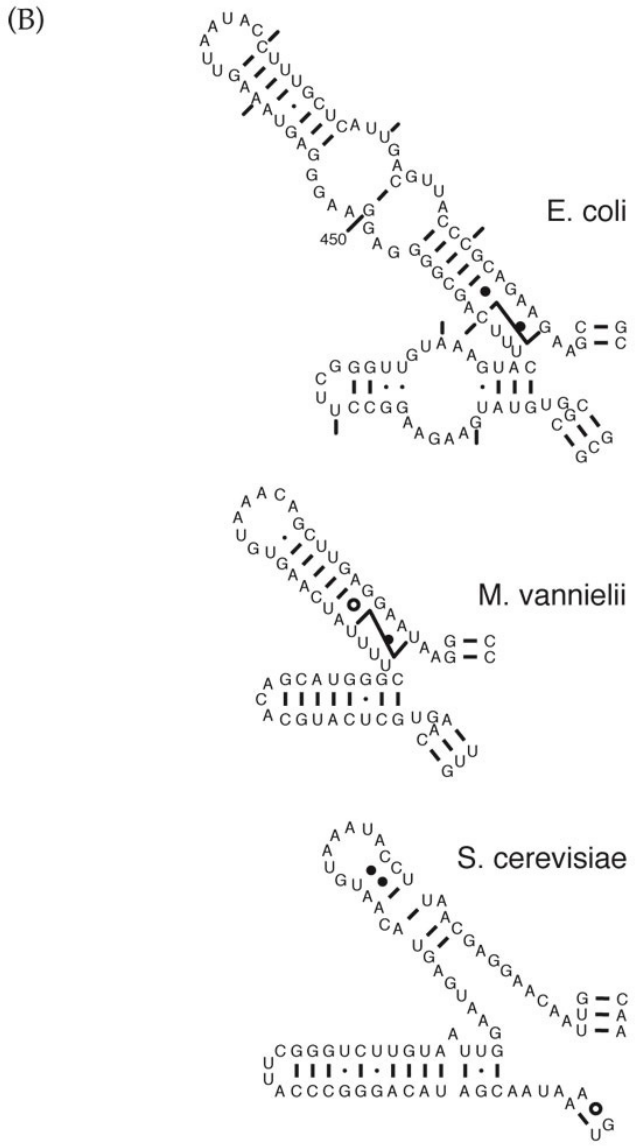


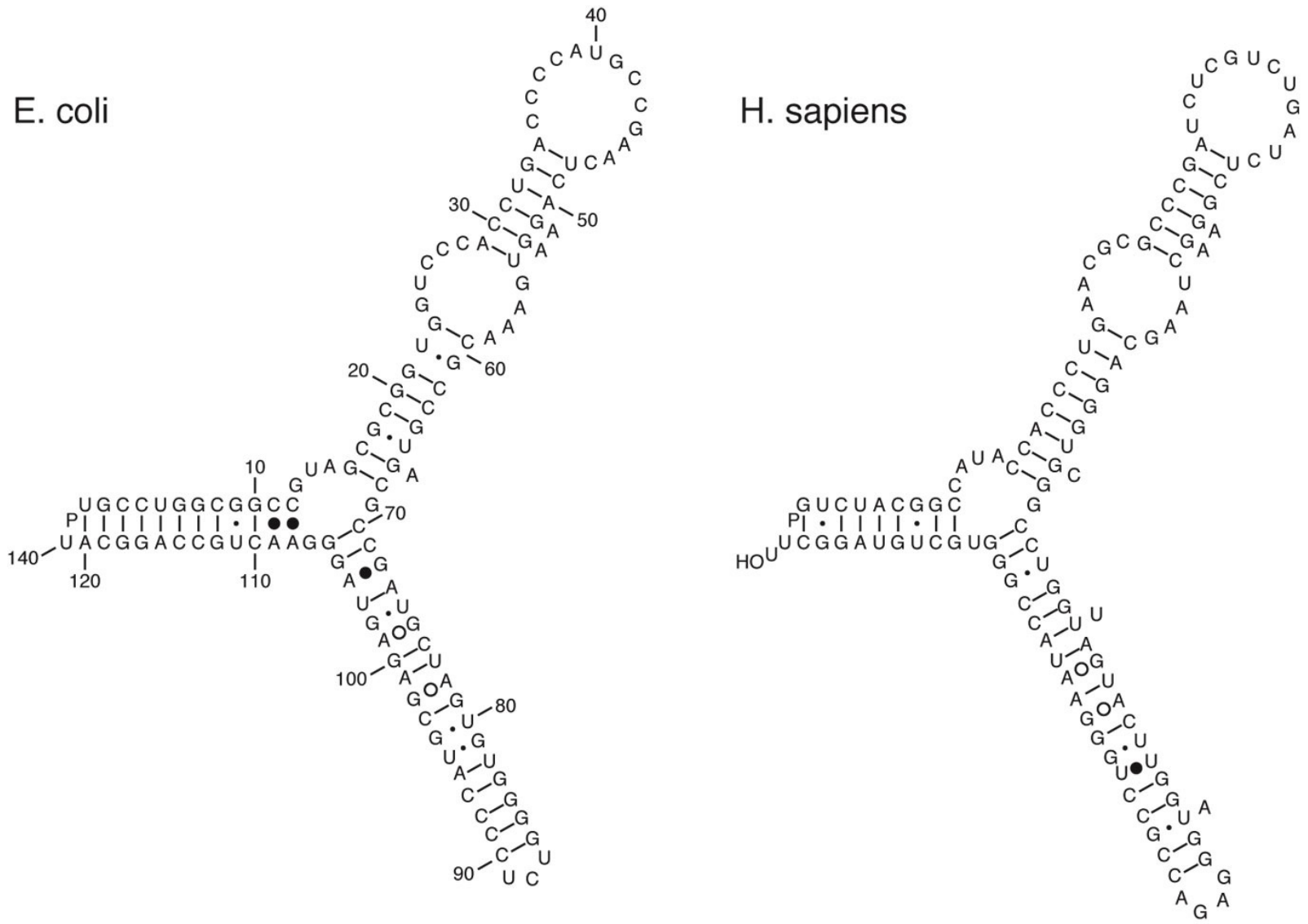
Figure 11-11 Brock Biology of Microorganisms 11/e



Pink: 16S rRNA. Lots of tertiary structure.
Blue: protein "scaffold"

Secondary Structures of SSU rRNA show homology





Secondary Structures of rRNAs show homology

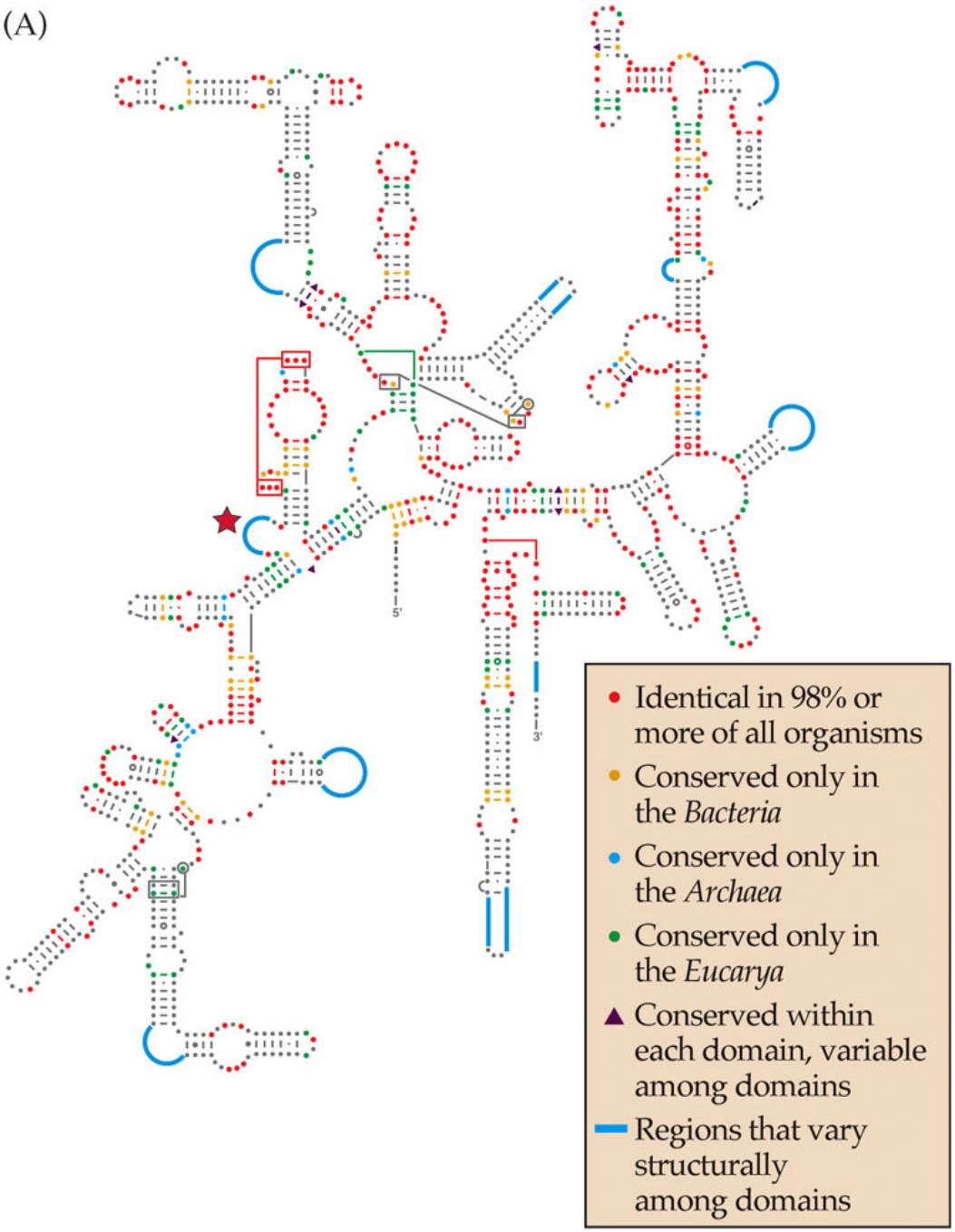
Signature sequences can be obtained at any level of taxonomic hierarchy…

Table 11.1 **Signature sequences from 16S or 18S rRNA defining the three domains of life**

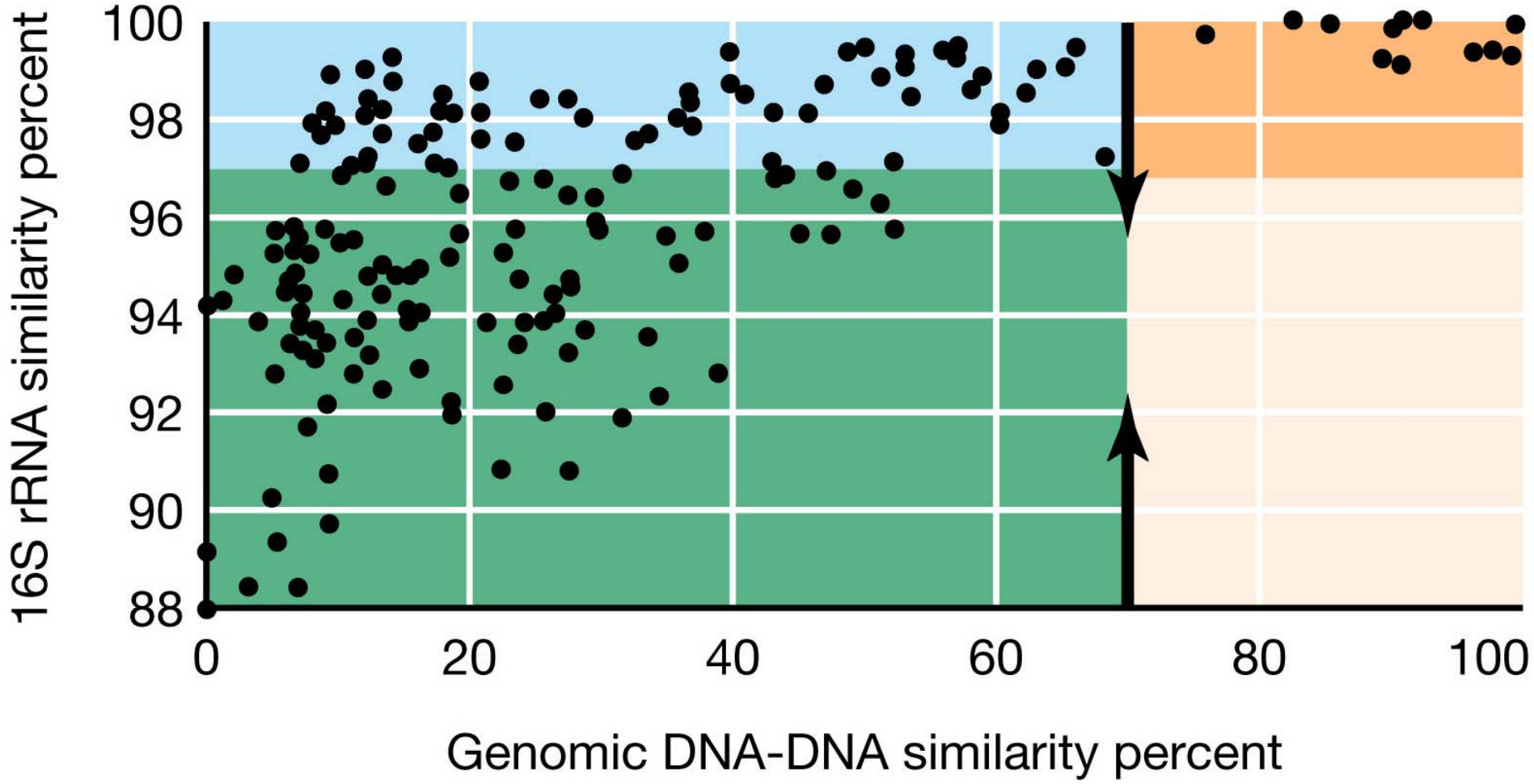
		Occurrence among[c]		
Oligonucleotide signatures[a]	**Approximate position[b]**	***Archaea***	***Bacteria***	***Eukarya***
CACYYG	315	0	>95	0
AAACUCAAA	910	3	100	0
AAACUUAAAG	910	100	0	100
YUYAAUUG	960	100	<1	100
CAACCYYCR	1110	0	>95	0
UCCCUG	1380	>95	0	100
UACACACCG	1400	0	>99	100
CACACACCG	1400	100	0	0

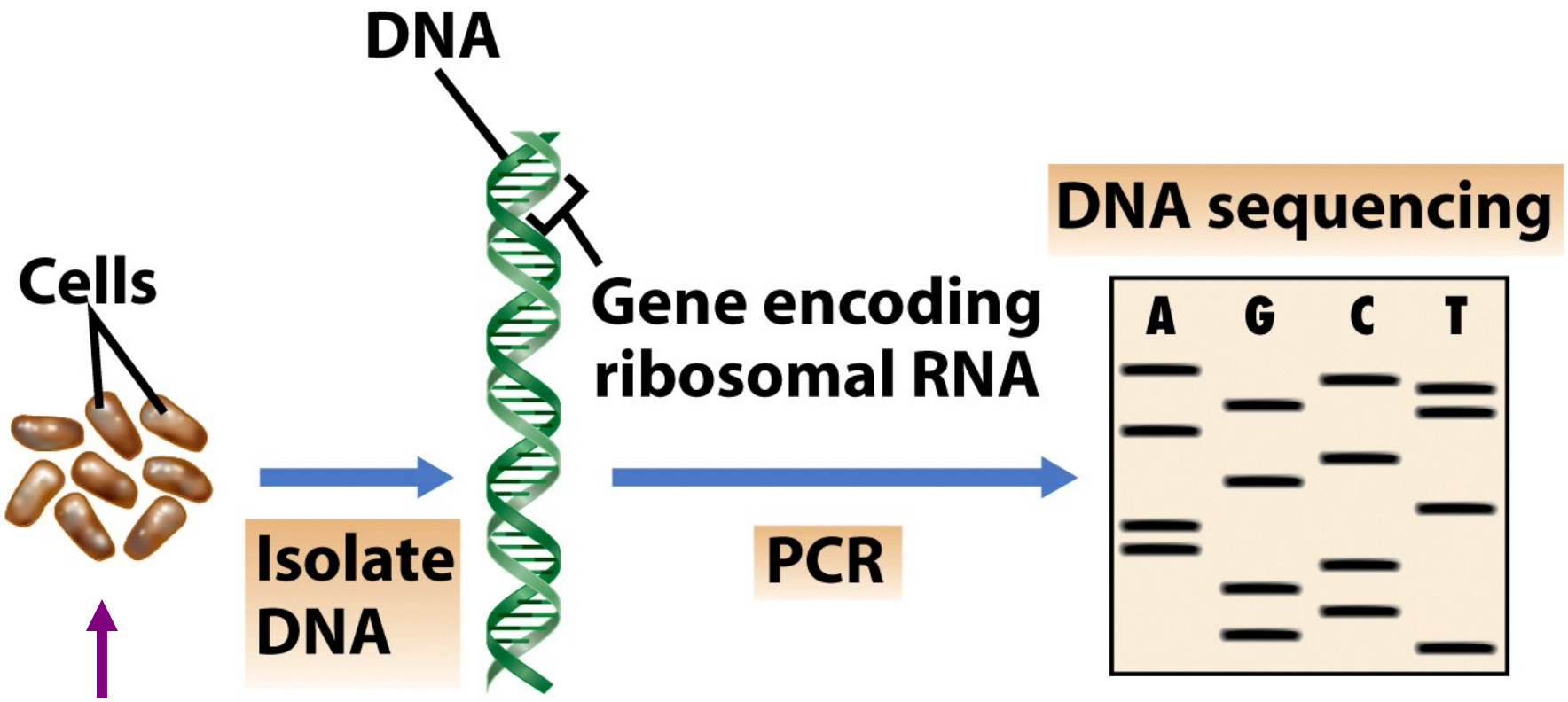
[a] Y, any pyrimidine; R, any purine.
[b] Refer to Figure 11.11*c* for numbering scheme of 16S rRNA.
[c] Occurrence refers to percentage of organisms examined in any domain that contain that sequence.

(A)


5'
3'
Identical in 98% or more of all organisms
Conserved only in the *Bacteria*
Conserved only in the *Archaea*
Conserved only in the *Eucarya*
Conserved within each domain, variable among domains
Regions that vary structurally among domains

Relationship between SSU rDNA and genomic DNA hybridization




DNA
Cells
Gene encoding
ribosomal RNA
DNA sequencing
A
G
C
T
Isolate
DNA
PCR
(cultured or from environmental sample)

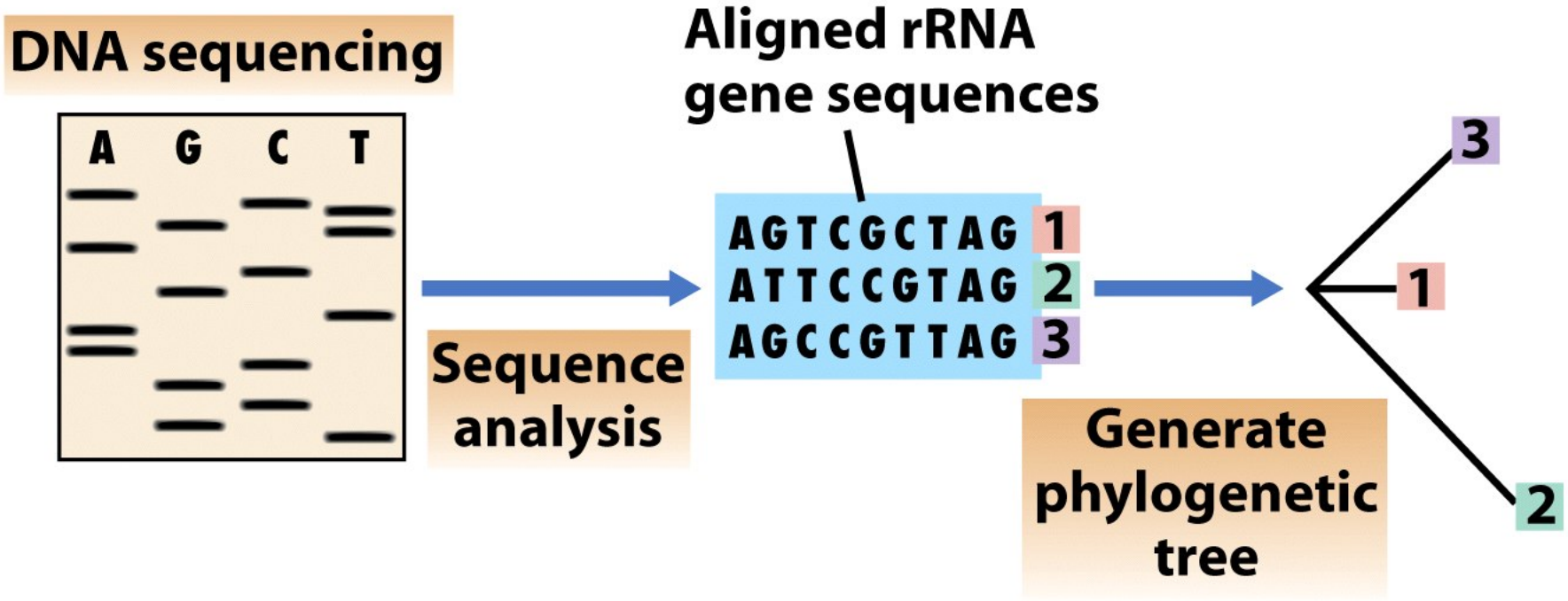


Figure 2-6 part 2 Brock Biology of Microorganisms 11/e

Ribotyping

1. PCR to amplify rRNA
2. Restriction digestion – polymorphisms in sequence = different cut patterns
3. Gel electrophoresis
4. Probe to “light up” sequences of interest
5. Analyze pattern

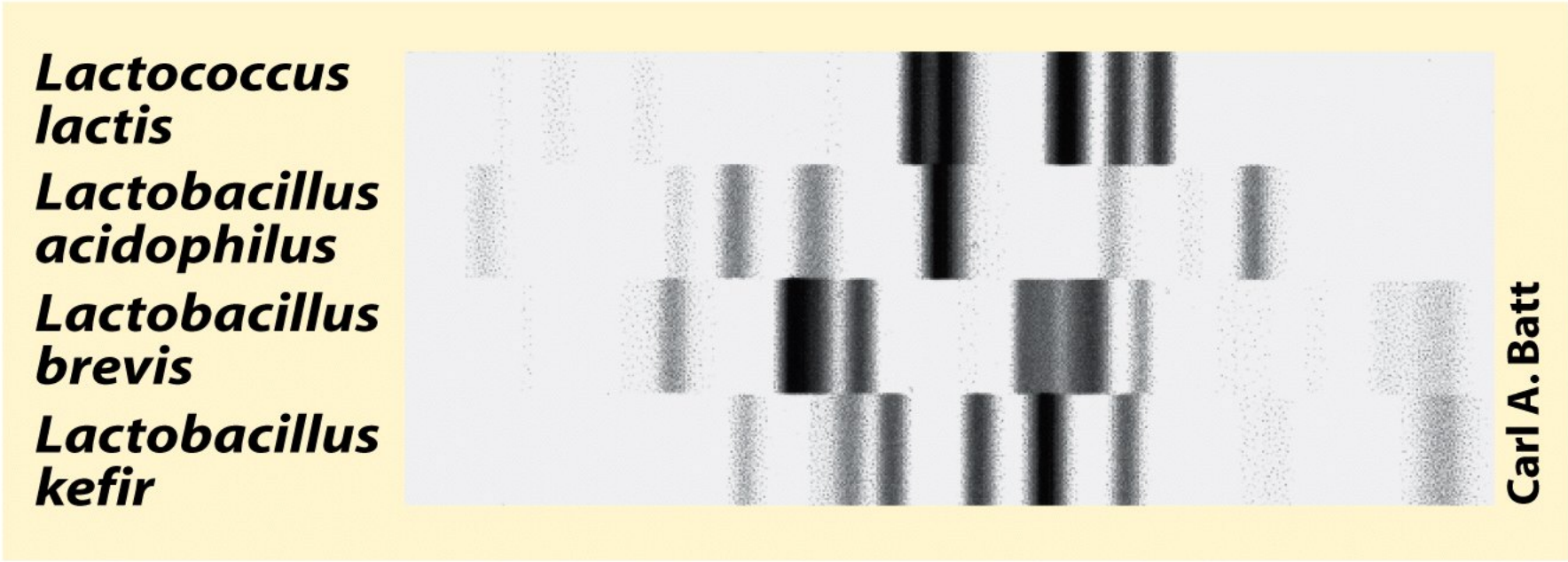


Figure 11-23a Brock Biology of Microorganisms 11/e

Multi-locus sequence typing (MLST)

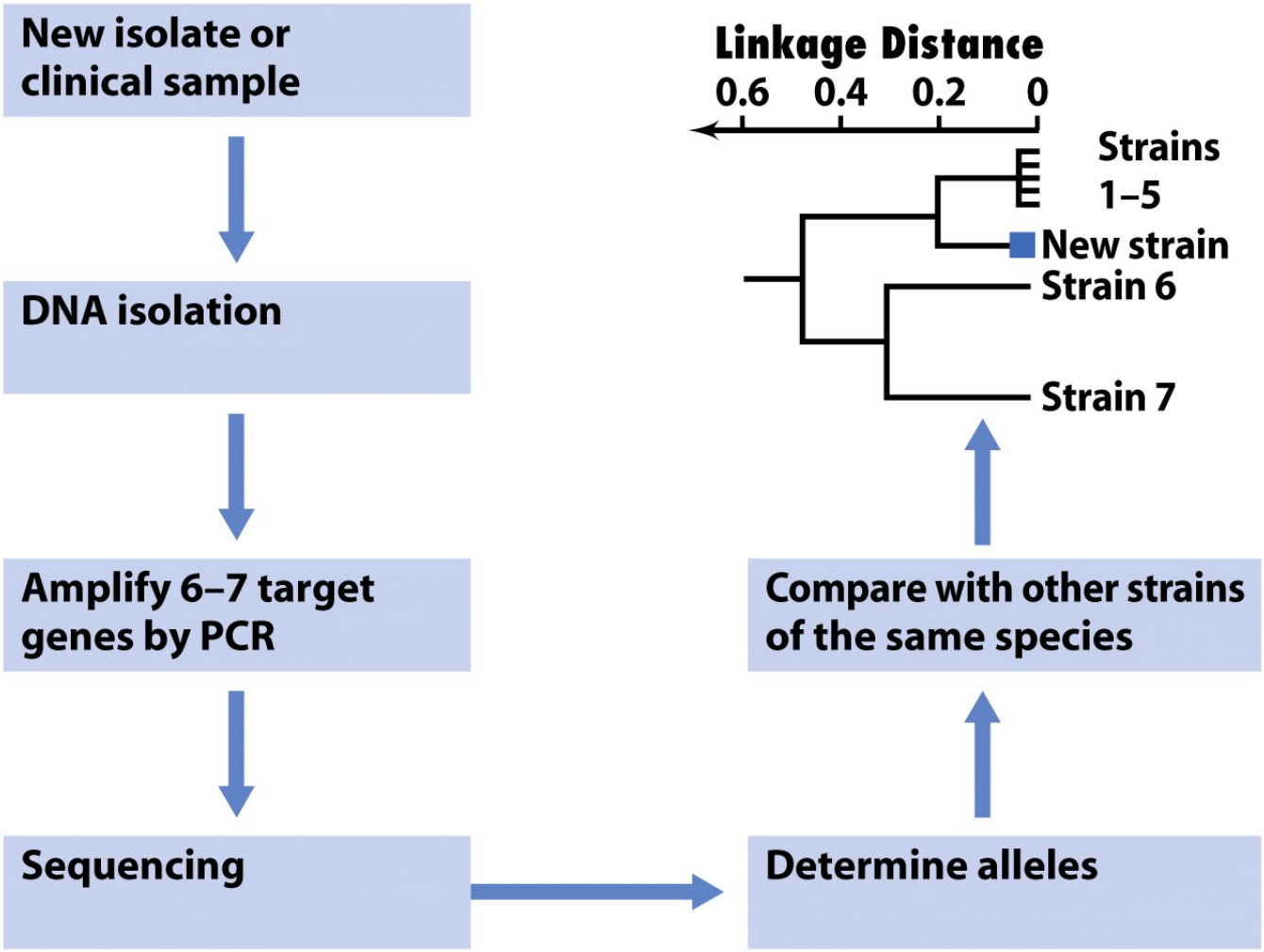
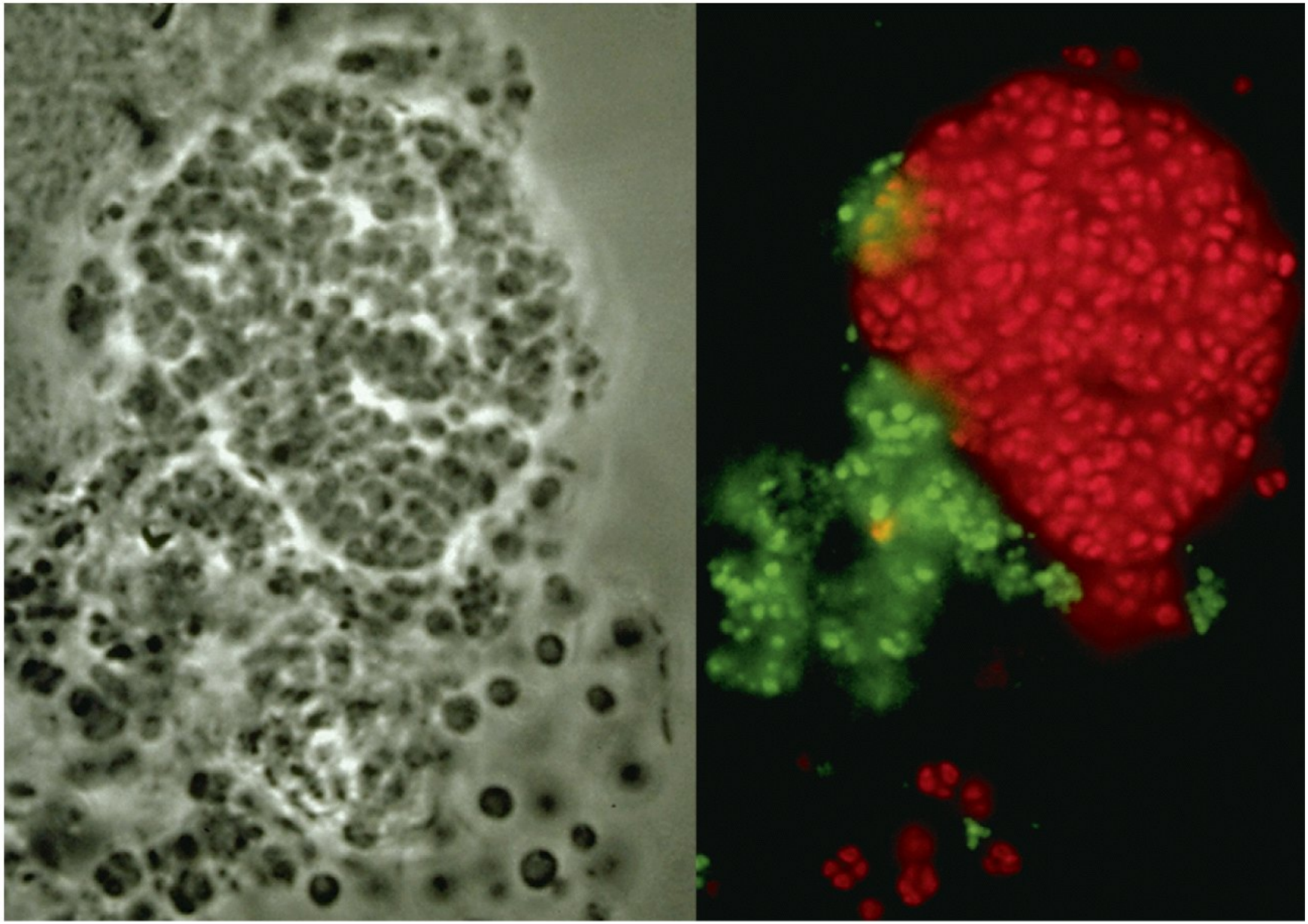


Figure 11-23b Brock Biology of Microorganisms 11/e

Fluorescent *in situ* hybridization (FISH)



1990:

Retrieval and analysis of ***ribosomal RNA*** genes from cells in natural samples have shown that many phylogenetically distinct but as yet uncultured prokaryotes exist in nature.

Filter a lot of seawater

↓

Extract DNA

↓

PCR with rRNA primers

↓

Sequence PCR product

Genetic diversity in Sargasso Sea bacterioplankton

Stephen J. Giovannoni, Theresa B. Britschgi, Craig L. Moyer & Katharine G. Field

Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA

BACTERIOPLANKTON are recognized as important agents of biogeochemical change in marine ecosystems, yet relatively little is known about the species that make up these communities. Uncertainties about the genetic structure and diversity of natural bacterioplankton populations stem from the traditional difficulties associated with microbial cultivation techniques. Discrepancies between direct counts and plate counts are typically several orders of magnitude, raising doubts as to whether cultivated marine bacteria are actually representative of dominant planktonic species[1–3]. We have phylogenetically analysed clone libraries of eubacterial 16S ribosomal RNA genes amplified from natural populations of Sargasso Sea picoplankton by the polymerase chain reaction[4]. The analysis indicates the presence of a novel microbial group, the SAR11 cluster, which appears to be a significant component of this oligotrophic bacterioplankton community. A second cluster of lineages related to the oxygenic phototrophs—cyanobacteria, prochlorophytes and chloroplasts—was also observed. However, none of the genes matched the small subunit rRNA sequences of cultivated marine cyanobacteria from similar habitats. The diversity of 16S rRNA genes observed within the clusters suggests that these bacterioplankton may be consortia of independent lineages sharing surprisingly distant common ancestors.

Today:

Retrieval and analysis of ***genomes*** from cells in natural samples have shown that many phylogenetically distinct but as yet uncultured prokaryotes exist in nature.

J. Craig Venter sails around the world in Sorcerer II, 100 ft sailboat

http://www.sorcerer2expedition.org

In a barrel (~20 L) of seawater in the nutrient-poor Sargasso Sea, Venter found 1800-40,000 new species (depending on how one defines a species)!

1 mL seawater contains 1,000,000 bacteria

1 mL seawater contains 10,000,000 viruses

< 1% have been characterized

(Most don't grow in the lab)

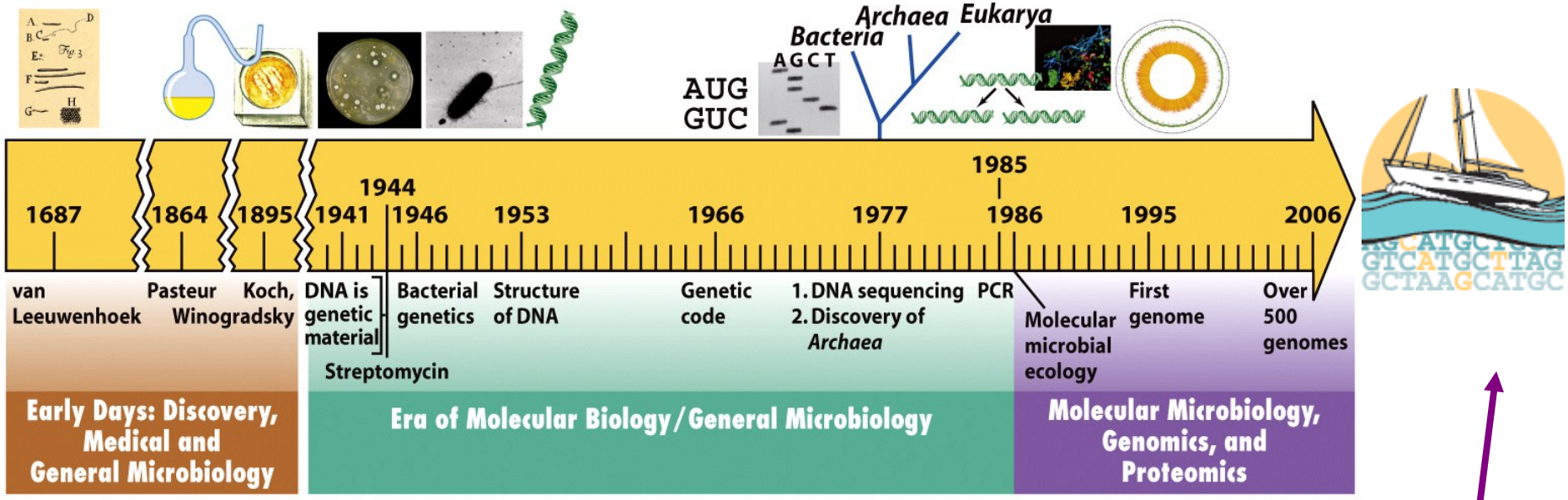


Figure 1-17 Brock Biology of Microorganisms 11/e

Metagenomes

What is a microbial “species”?

Eukaryotic species = interbreeding populations.

Microbes are asexual!

- $\geq$ 70% DNA hybridization
- $\geq$ 97% rRNA sequence identity

(Arbitrary boundary to avoid disrupting existing assigments, rather than on theoretical considerations)

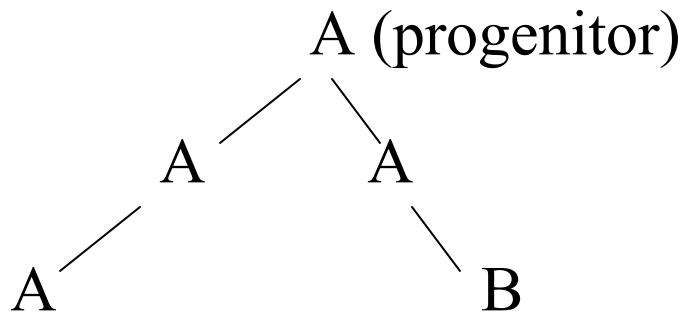
How do new species arise?

Major components of evolution:

1. Vertical inheritance (traits passed from parents to offspring)

2. Descent with modification (traits passed on imperfectly: mutation, recombination)

3. Natural selection (selects among variants)



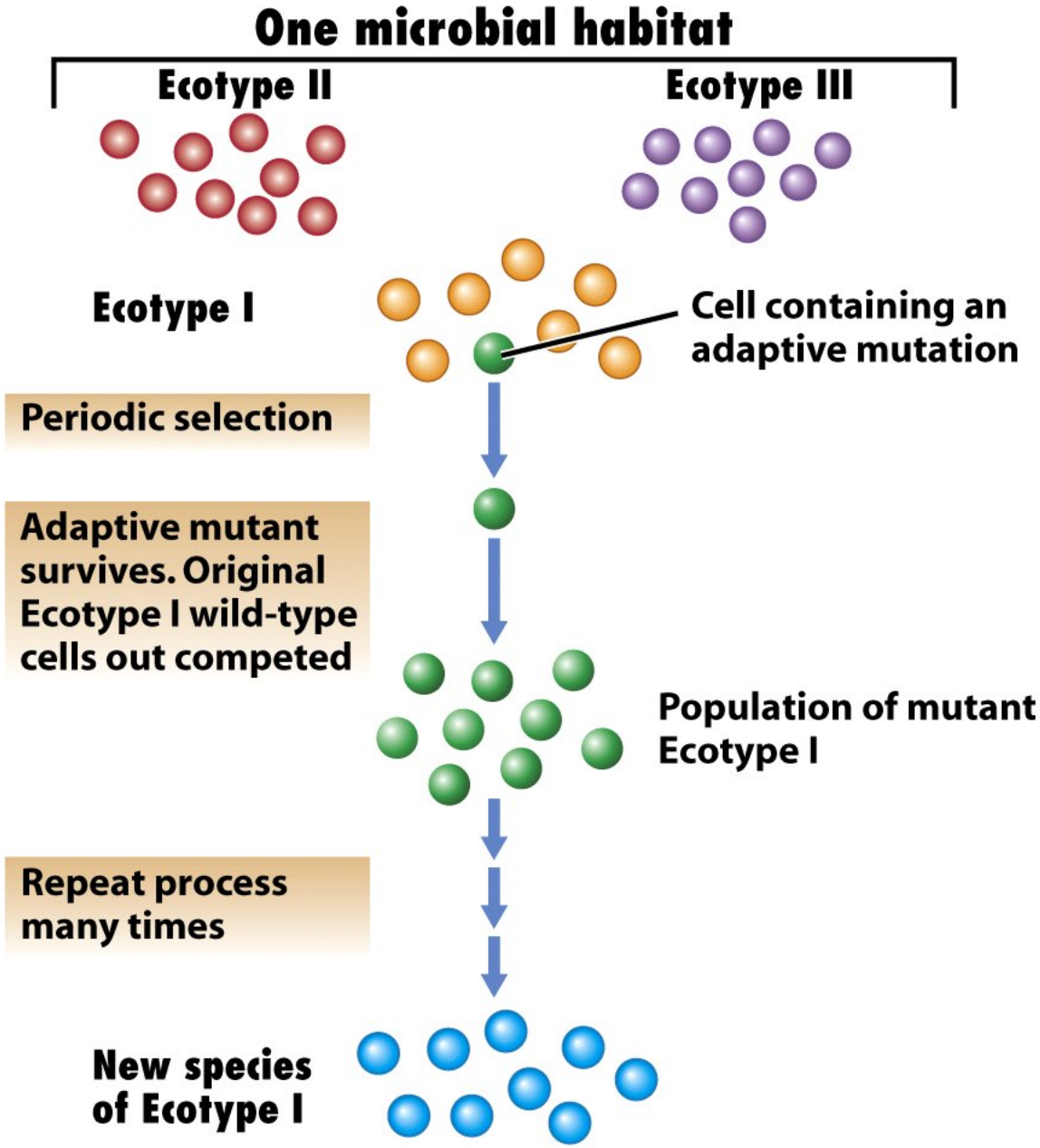
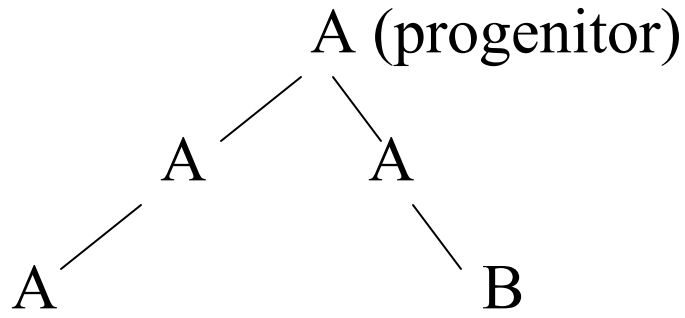


Figure 11-26 Brock Biology of Microorganisms 11/e

How do we reconstruct evolutionary relationships?

1. Observe

2. Biological records
   a. Fossils
   b. Geology (geochemistry)



3. *Infer* from data of current organisms (chemistry, gene sequence, protein sequence)

How do we infer evolutionary relationships?

-Key word is inference (not always correct!)
-none of the organisms in the "Tree of Life" are ancient; they are all modern organisms.
-Some may have characteristics of ancient organisms

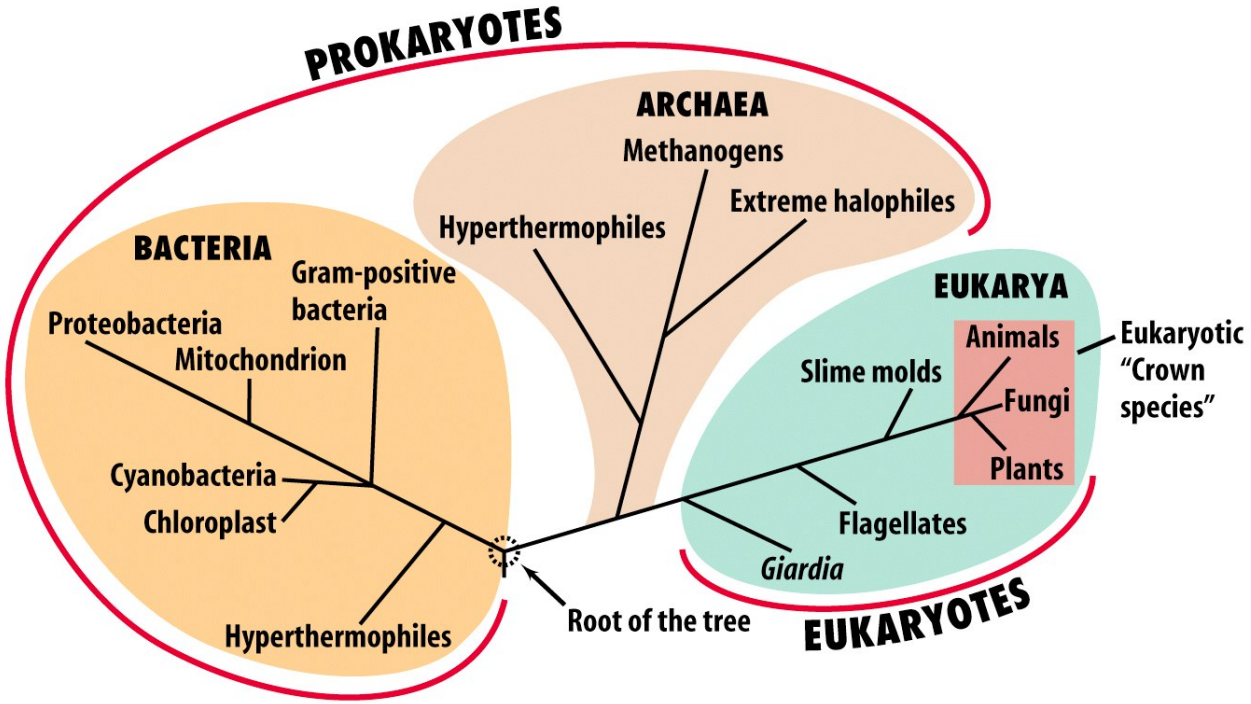


Figure 2-7 Brock Biology of Microorganisms 11/e

What did the first cell look like? We don't know…

Planet Earth is approximately 4.6 billion years old.

The first evidence for microbial life can be found in rocks about 3.86 billion years old… but these shapes lack rRNA to compare with others on the Tree of Life.

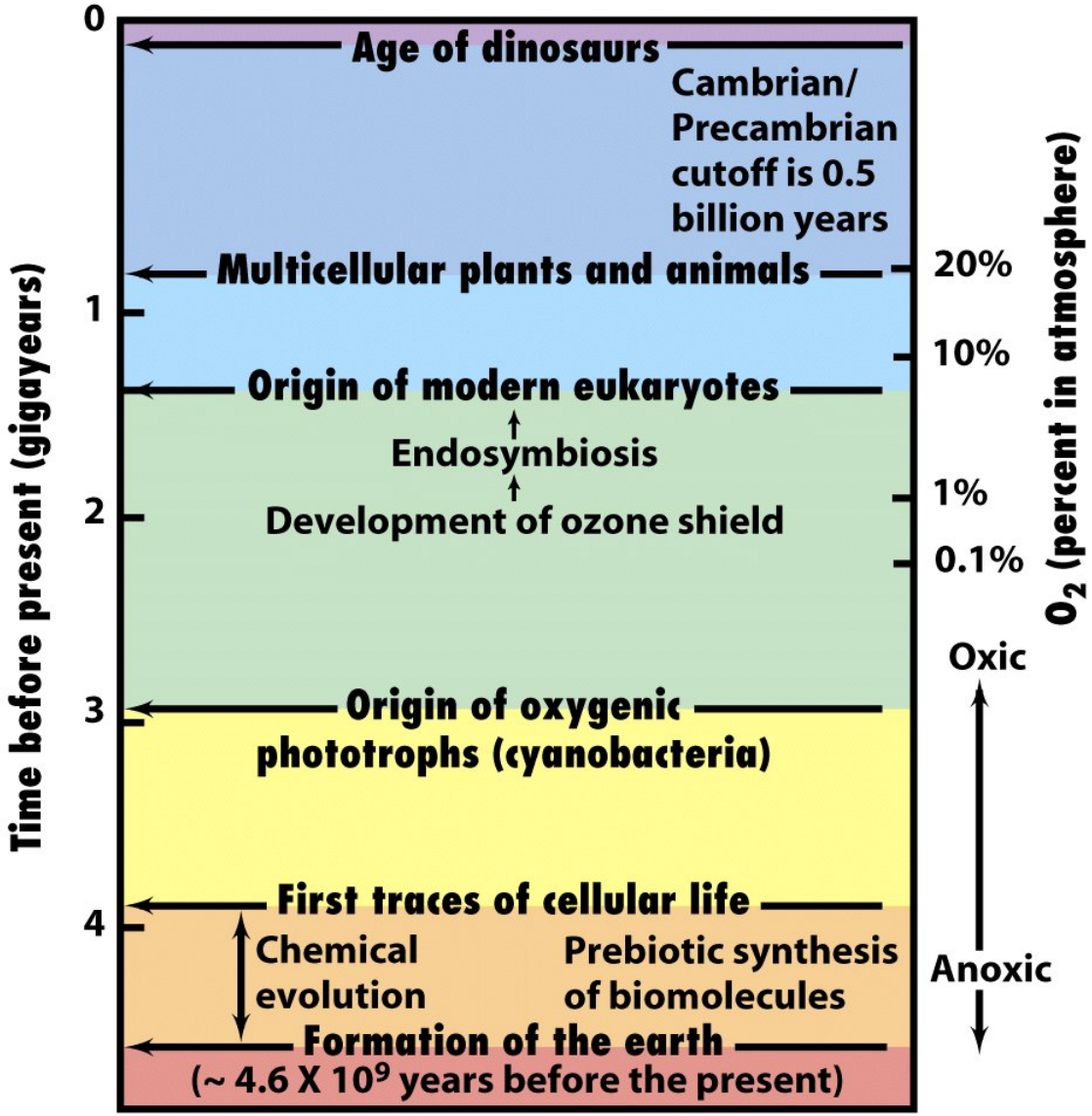
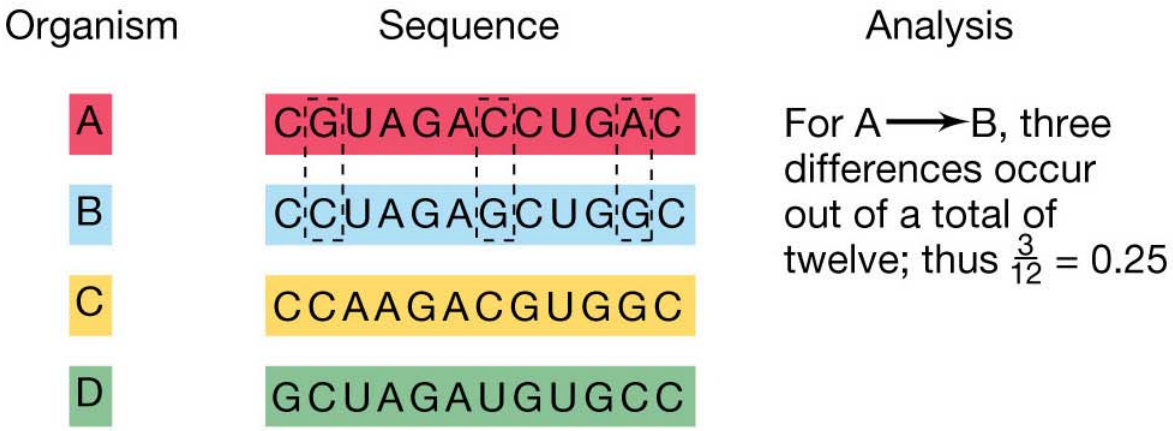


Figure 11-8 Brock Biology of Microorganisms 11/e



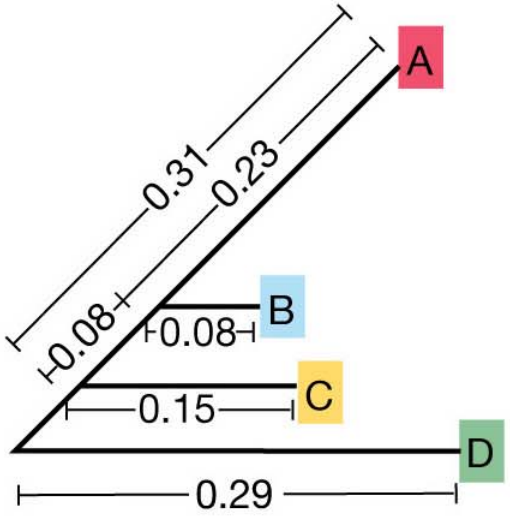
(a) **Sequence alignment and analysis**

Estimating evolutionary distance E_D to map on phylogenetic tree

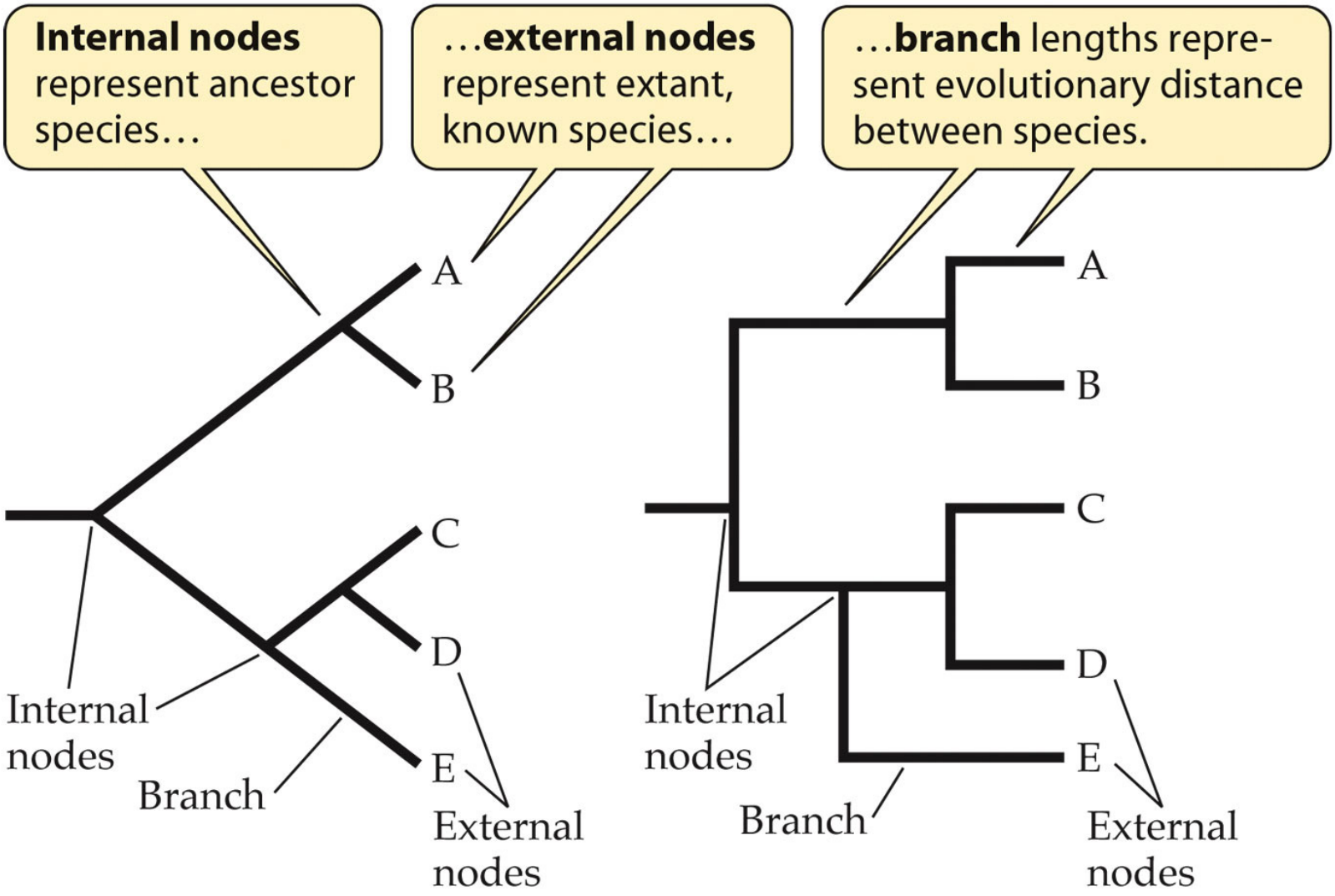
Parsimony: what you see is what you get

Evolutionary distance			Corrected evolutionary distance
E_D	A → B	0.25	0.30
E_D	A → C	0.33	0.44
E_D	A → D	0.42	0.61
E_D	B → C	0.25	0.30
E_D	B → D	0.33	0.44
E_D	C → D	0.33	0.44

(b) **Calculation of evolutionary distance**



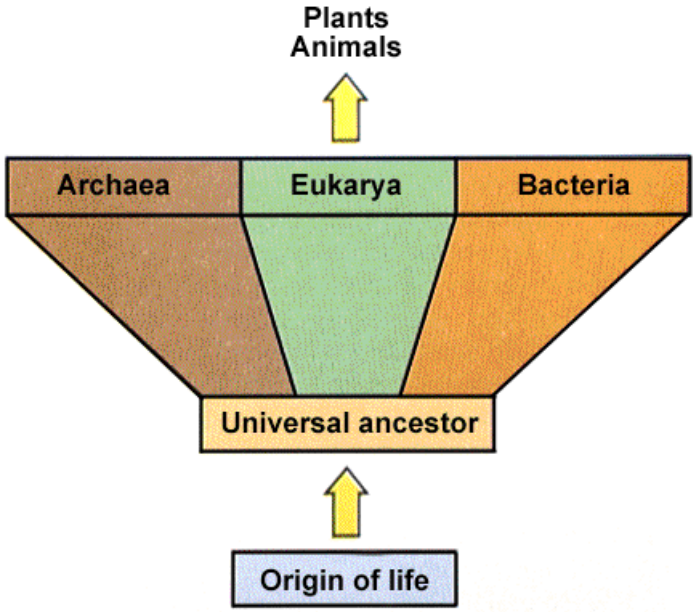
(c) **Phylogenetic tree**



Fan-shaped

Dichotomous

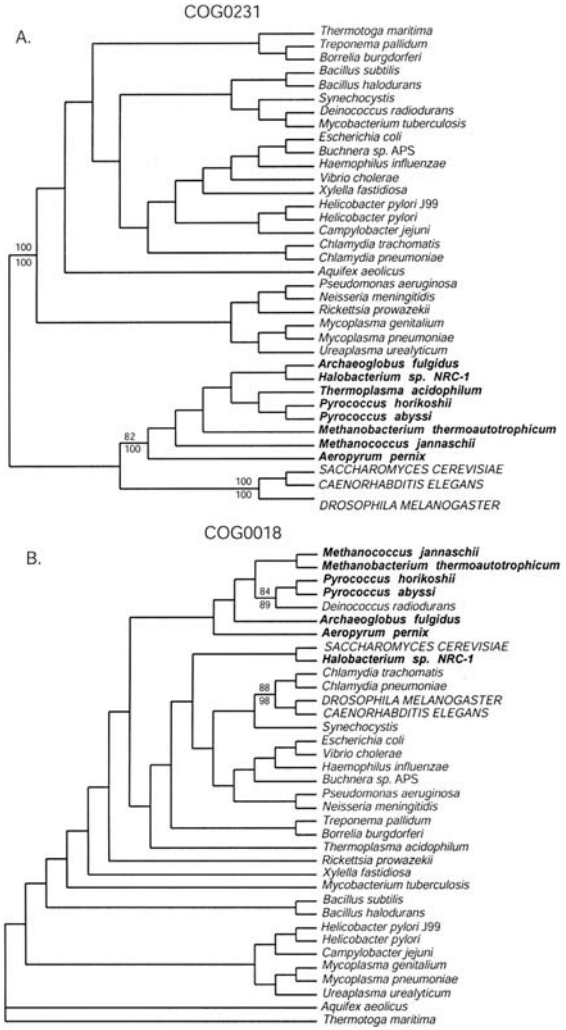
EXAMPLE of a PHYLOGENETIC ANALYSIS:



Question: What is our universal common ancestor like?

In 2003, **Norm Pace's** group did a study using all the known clusters of orthologous groups (COGs), or groups of widely conserved, homologous genes, to find those that are so anciently conserved that they, like rRNA, split neatly along evolutionary lines as far back as the three domains of life. Of 3100 COGs analyzed, only 50 were "three-domain" groups! Most belong to the **nucleic acid-based central information pathway** (ribosomal proteins, DNA/RNA polymerase subunits, elongation factors). However, a few showed little apparent connection to genetic transmission or gene expression (e.g., membrane insertion factors and proteases).

Examples of three-domain and non-three-domain phylogenetic trees from analyses of the COG database protein alignments



J. Kirk Harris et al. Genome Res. 2003; 13: 407-412

Phylogeny allows us to ask **testable questions**, e.g., hypothesis testing.

-**microbial communities** can now be truly examined
(who is out there and how many of them are there??)

-relationships **among microbes** can be studied

-relationships **among microbial genes** can be studied

-can infer **dynamics of sequence change**
(phenetic vs. phyletic = Timex vs. Rolex)

Phylogenies may be right or wrong; we use them to make the best **inferences** we can.

Some Lessons from the BIG TREE: Map of the Biological Record

Single origin for **all life** on Earth...

- Central Dogma intact
- ATP and PMF are universal themes
- Uniformity among chiral carbon compounds (sugars & AAs).
- Hot start origin…

General topology implies:

- Three "primary lines of evolutionary descent."
- The Eukarya "*nuclear*" lineage almost as old as other two.
- Prokaryotes split between *Bacteria* and *Archaea*.
- Tree represents only a limited number of organisms.
- Mitochondria and chloroplasts proven to be of bacterial origin.

Some Lessons from the BIG TREE: Map of the Biological Record

Evolutionary "clock" is NOT constant between different lineages

- Terminal nodes NOT all the same length, so not constant for all organisms either!
- Endosymbionts sped up very fast (semi-autonomous)
- Eucarya – Fast clocks
- Archaea – Slow clocks
- Bacteria – Intermediate

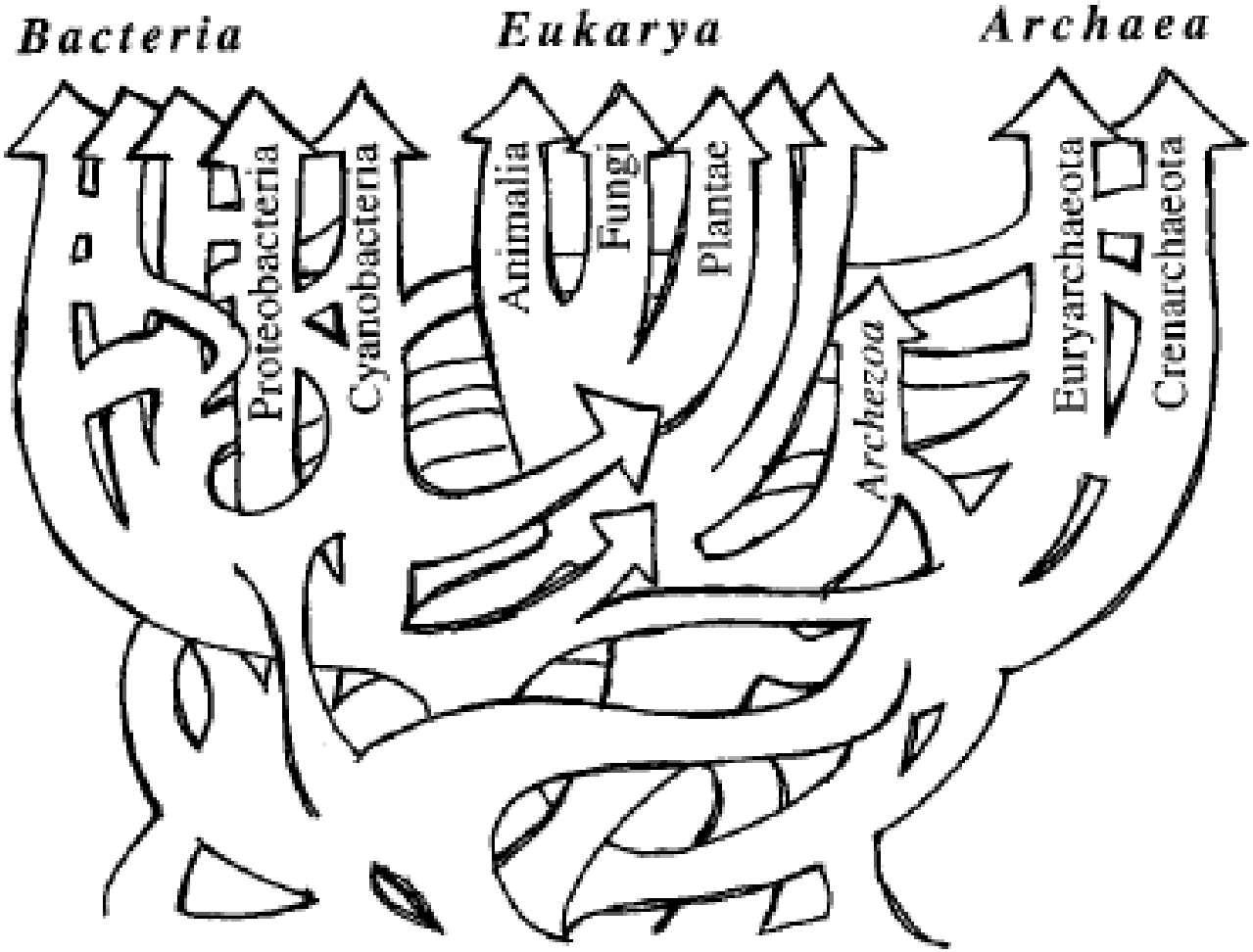
Horizontal gene transfer

This lateral flow of information
across microbial taxa occurs via
the transfer of genes by:

conjugation, transduction, and transformation.

Rem: These are one-way processes!

Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history.



Doolittle's Universal Tree (1999)

A Bit on the Evolution of Evolutionary Thought

A. Prior to the late 19th century, the concept of evolution was on the **evolutionary ladder**. Thus, we still deal in "higher and lower" eucaryotes (I try not to use these terms – they are dumb), "missing links," and "primitive" organisms.

B. In its milieu, *E. coli* is as highly evolved as are we. *E. coli* is **simple** (~5×10^6 bp genome), we are **complex** (~3×10^9 bps); complexity has nothing to do with *evolutionary advancement.*

C. Lineages evolve by diversification, not *progression.* !!!

D. There is no such thing as a *primitive* organism alive today. **Simple**, yes, but still a finely honed product of ~ 4 billion years under the selective hammer of the niches that it and its progenitors have occupied.

C-value paradox: Organism complexity does not correlate to genome size

TABLE 13.3 C values from eukaryotic organisms ranked by size

Species	C value (kb)
Navicola pelliculosa (diatom)	35,000
Drosophila melanogaster (fruitfly)	180,000
Paramecium aurelia (ciliate)	190,000
Gallus domesticus (chicken)	1,200,000
Erysiphe cichoracearum (fungus)	1,500,000
Cyprinus carpio (carp)	1,700,000
Lampreta planeri (lamprey)	1,900,000
Boa constrictor (snake)	2,100,000
Parascaris equorum (roundworm)	2,500,000
Carcarias obscurus (shark)	2,700,000
Rattus norvegicus (rat)	2,900,000
Xenopus laevis (toad)	3,100,000
***Homo sapiens* (human)**	**3,400,000**
Nicotiana tabaccum (tobacco)	3,800,000
Paramecium caudatum (ciliate)	8,600,000
Schistocerca gregaria (locust)	9,300,000
Allium cepa (onion)	18,000,000
Coscinodiscus asteromphalus (diatom)	25,000,000
Lilium formosanum (lily)	36,000,000
Pinus resinosa (pine)	68,000,000
Amphiuma means (newt)	84,000,000
Protopterus aethiopicus (lungfish)	140,000,000
Ophioglossum petiolatum (fern)	160,000,000
Amoeba proteus (amoeba)	290,000,000
Amoeba dubia (amoeba)	670,000,000

Compiled by Li and Graur (1991) from Cavalier-Smith (1985), Sparrow et al. (1972), and other references. The C value for humans is highlighted for reference.

Table 17.2 Comparison of *E. coli* and its primate host species[a]

Property	*E. coli*	*Homo sapiens*	Primates
Mol % G + C	48–52	42	42[b]
16S–18S rRNA variability	>15 bases	?	<16[c]
DNA/DNA reassociation	>70%	98.6%[d]	>70%[e]

[a]Adapted from J. T. Staley, *ASM News*, 1999.
[b]Value for all primates.
[c]Mouse 18S rRNA differs from humans by 16 bases.
[d]Comparison between *Homo sapiens* and chimpanzee.
[e]Comparison between *Homo sapiens* and lemurs.