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!)

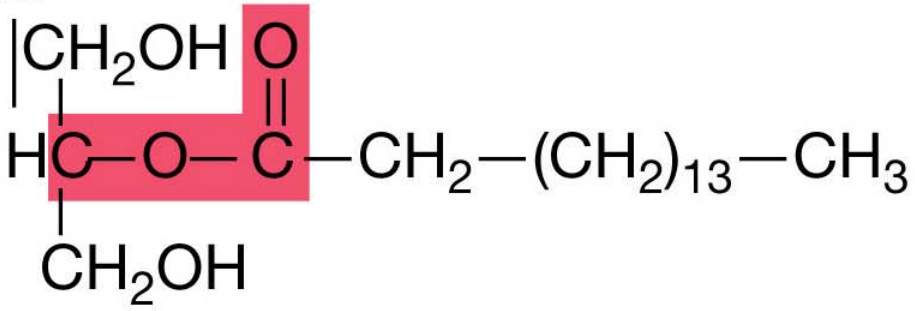
Table 17.1 **Hierarchical classification of the bacterium *Spirochaeta plicatilis***

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

TABLE 11.4 Some phenotypic characteristics of taxonomic value

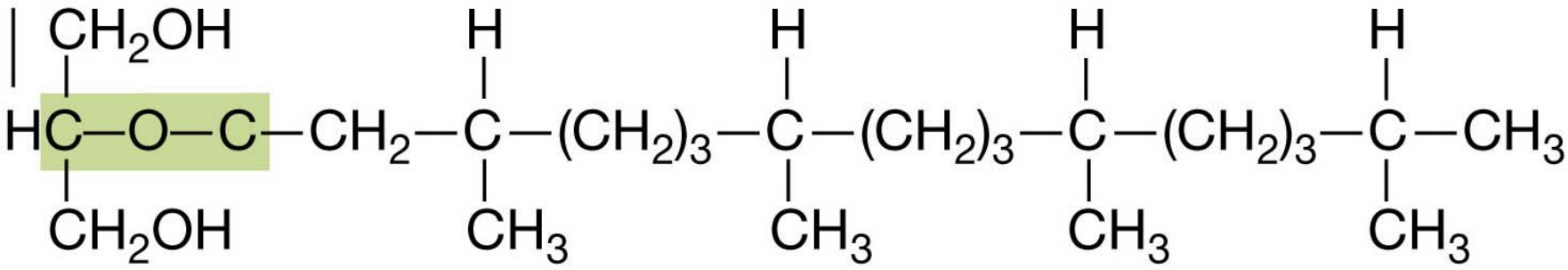
Major category	Components
I. Morphology	Shape; size; Gram reaction
II. Motility	Motile by flagella; motile by gliding; motile by gas vessels; nonmotile
III. Nutrition and physiology	Mechanism of energy conservation (phototroph, chemoorganotroph, chemolithotroph); relationship to oxygen; temperature,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

Ester



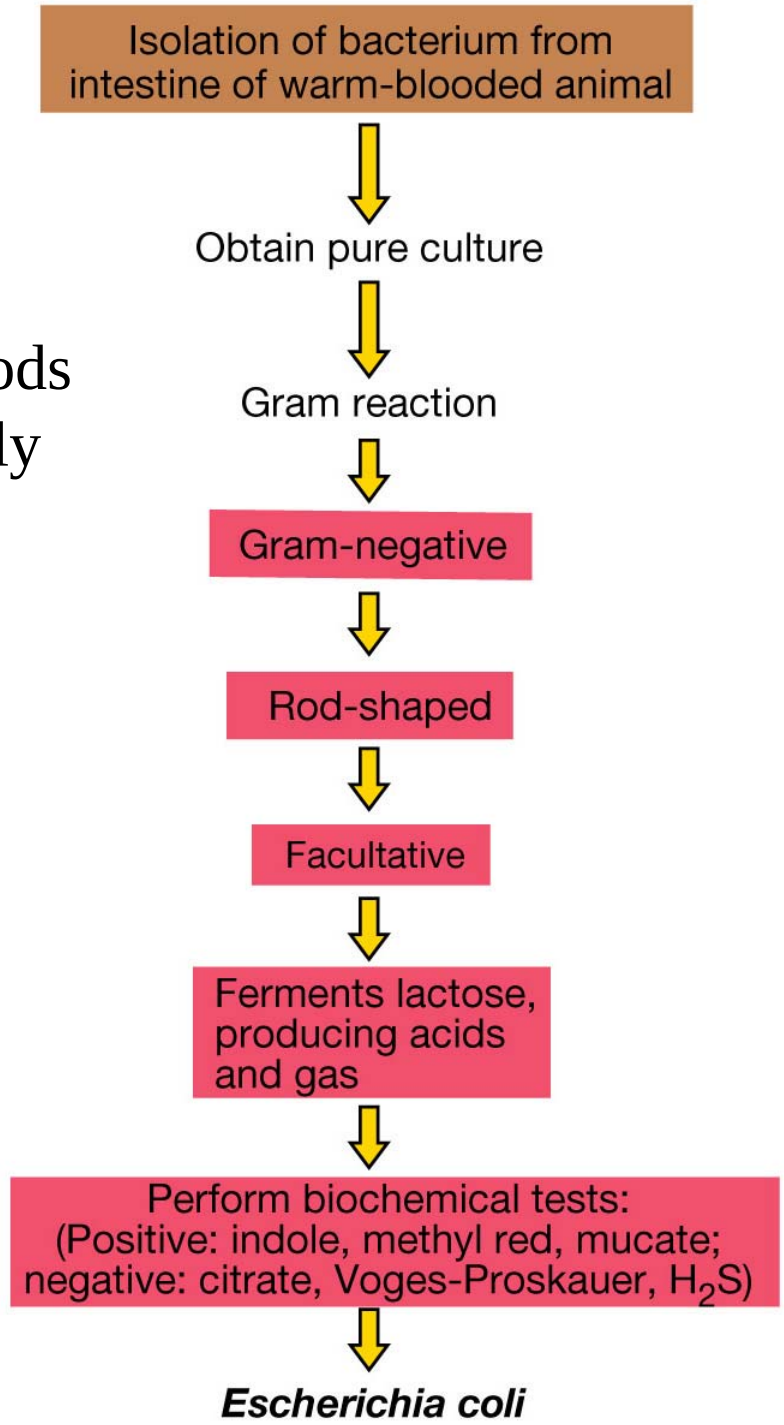
Bacteria, Eukarya

Ether

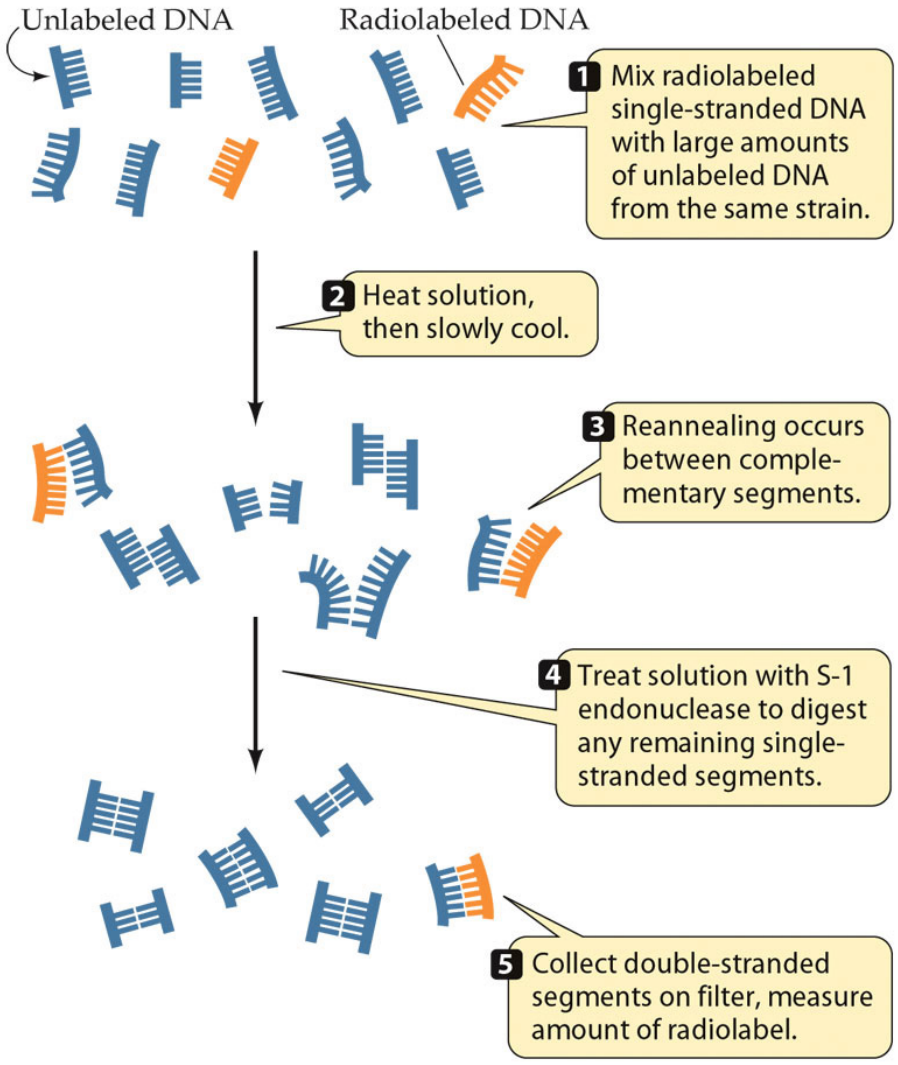


Archaea

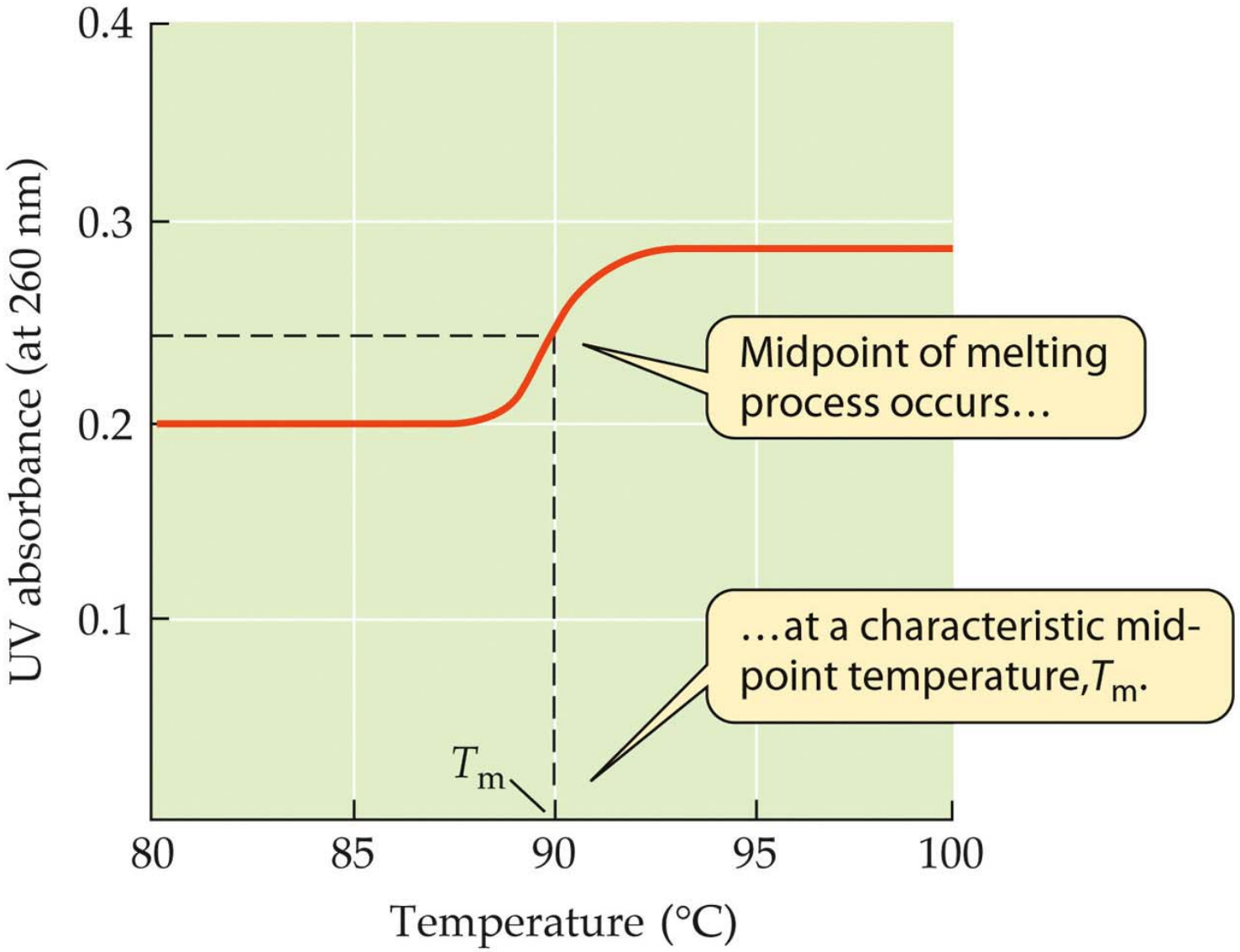
Example of methods used to ID a newly isolated enteric bacterium



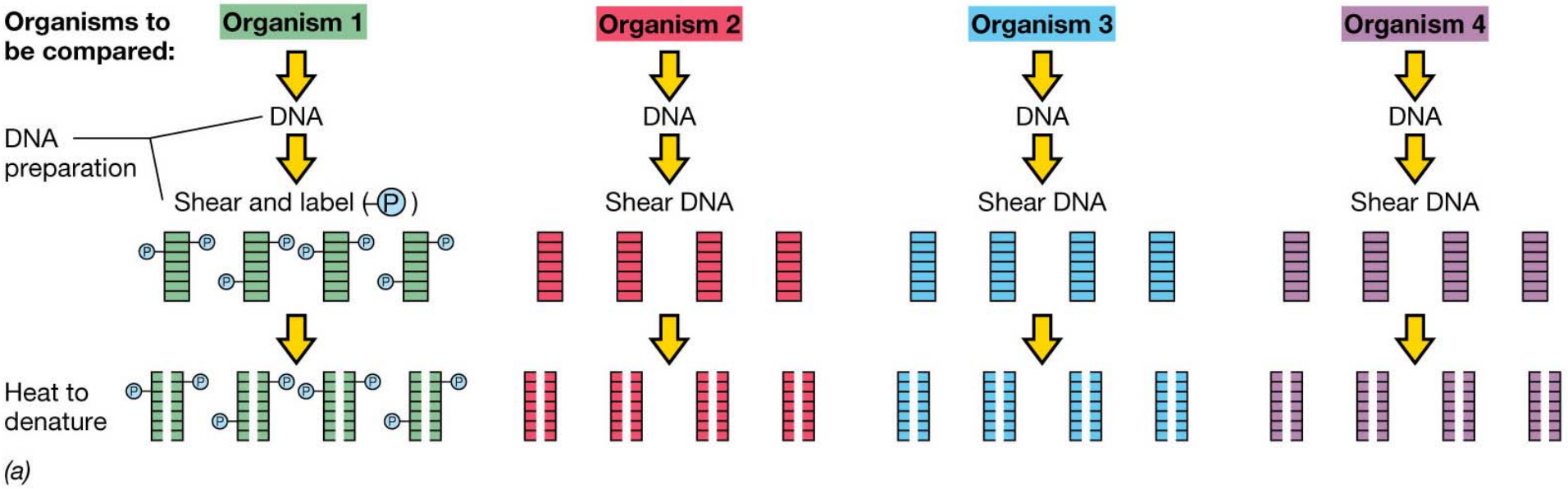
DNA/DNA hydridization or reassociation: A Pair-wise comparison



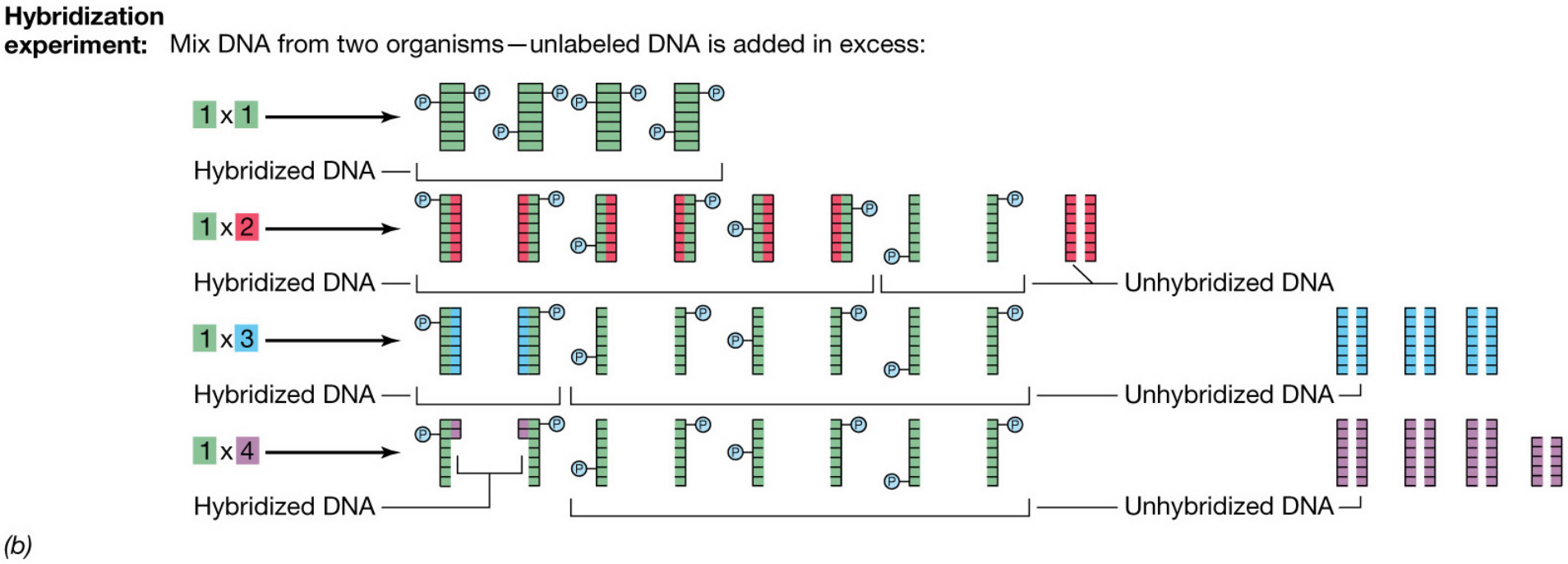
Hyperchromic Effect of DNA



DNA:DNA hybridization Part I

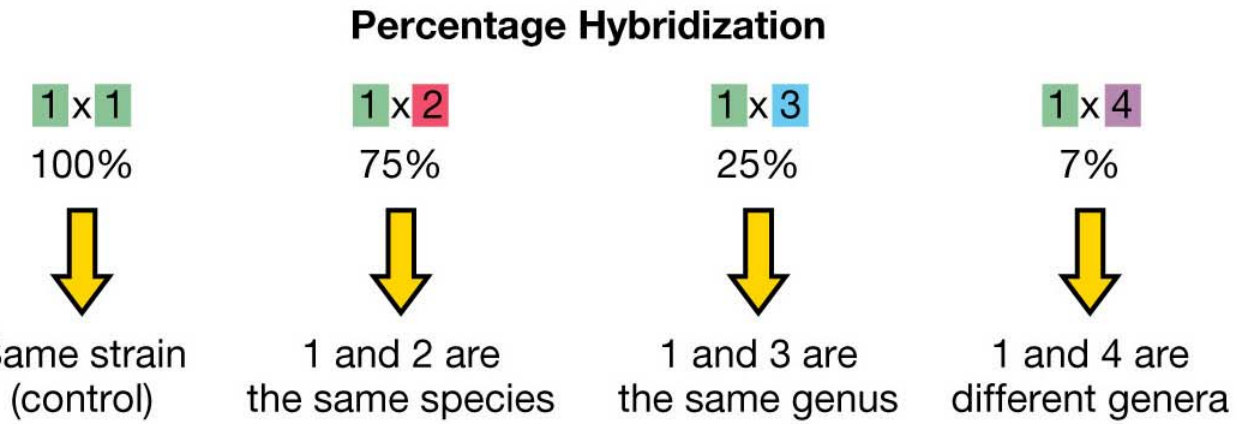


DNA:DNA hybridization Part II



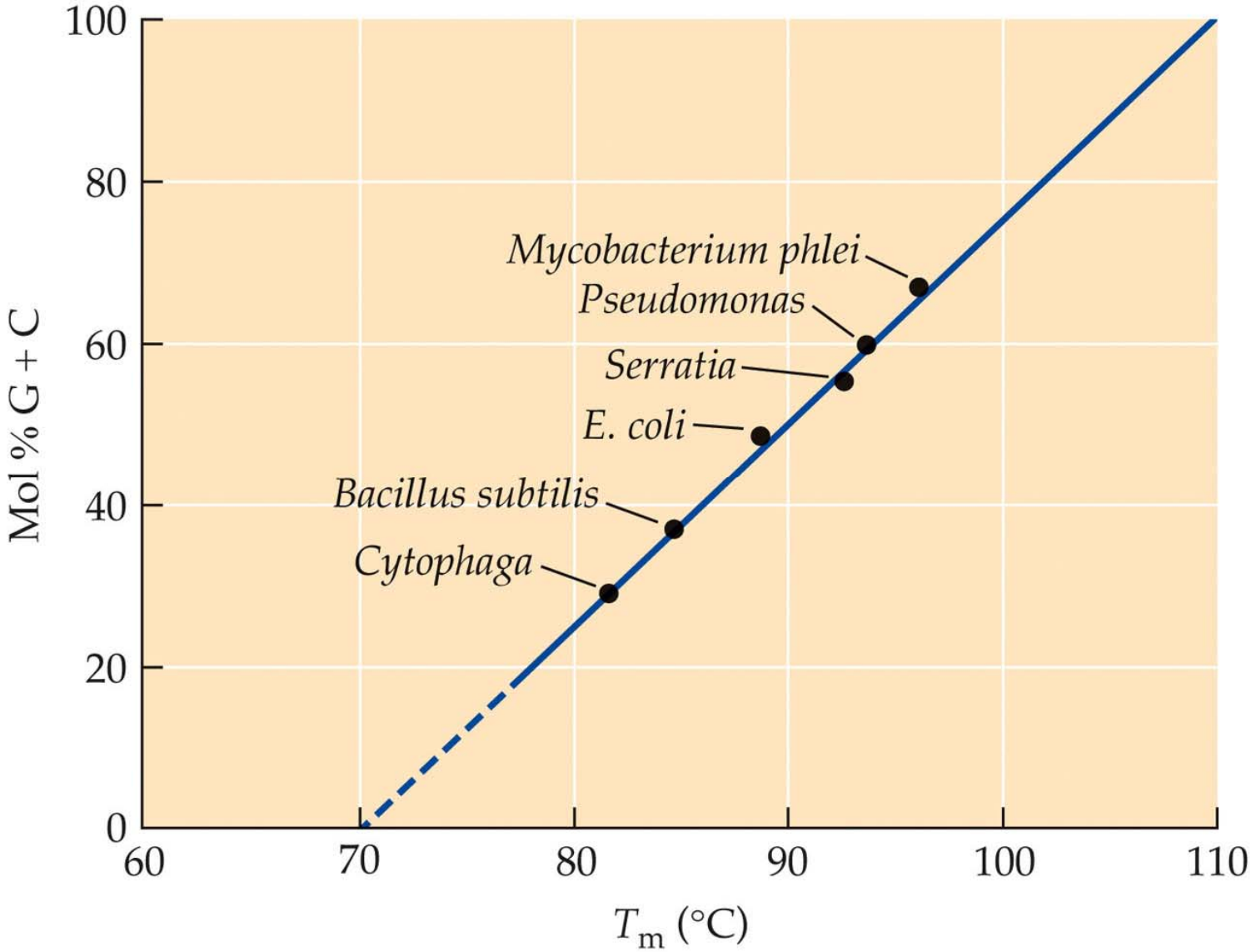
DNA:DNA hybridization Part III

Results and interpretation:

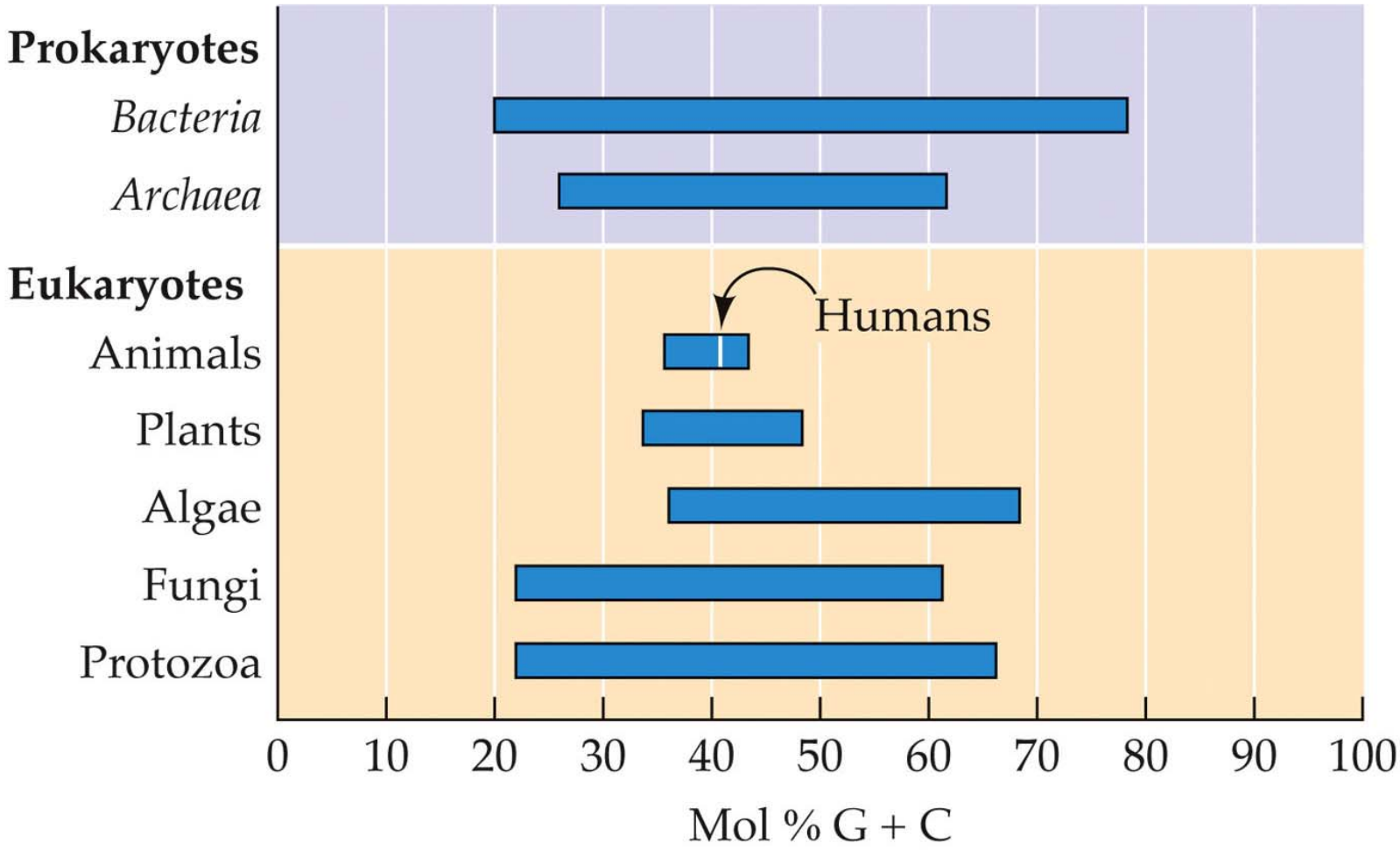


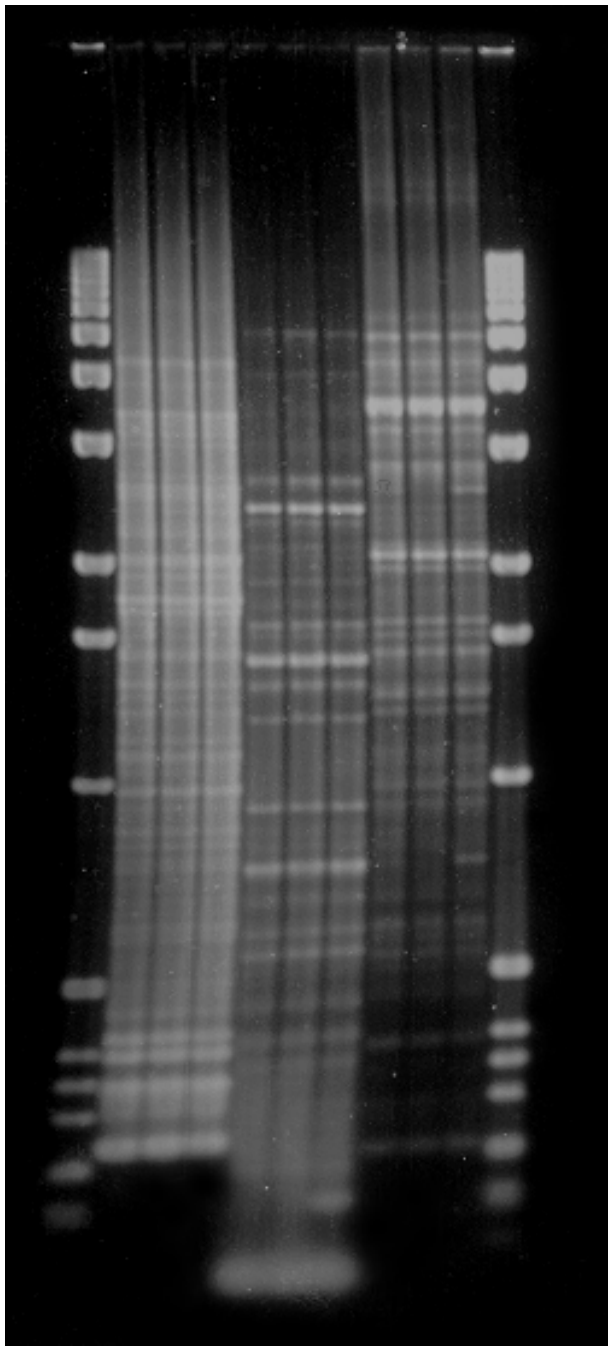
(c)

70% or greater; considered same species


100
80
60
40
20
0
Mol % G + C
60
70
80
90
100
110
T_m (°C)
Mycobacterium phlei
Pseudomonas
Serratia
E. coli
Bacillus subtilis
Cytophaga

Ranges of DNA base composition





REP PCR Fingerprinting

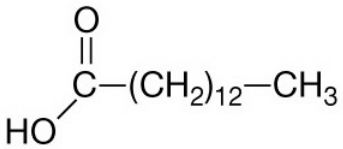
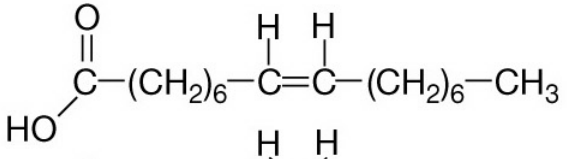
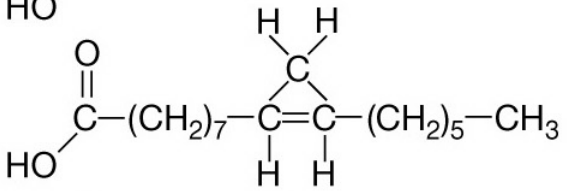
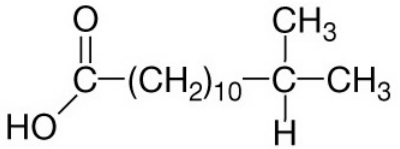
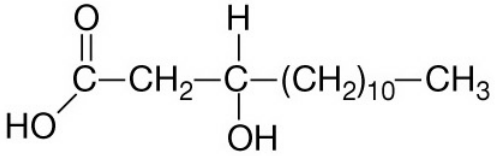
Lanes represent: Strains RL1, ES1, & ES2

- Three different types of PCR based genomic fingerprinting methods. Collectively known as **REP PCR**.

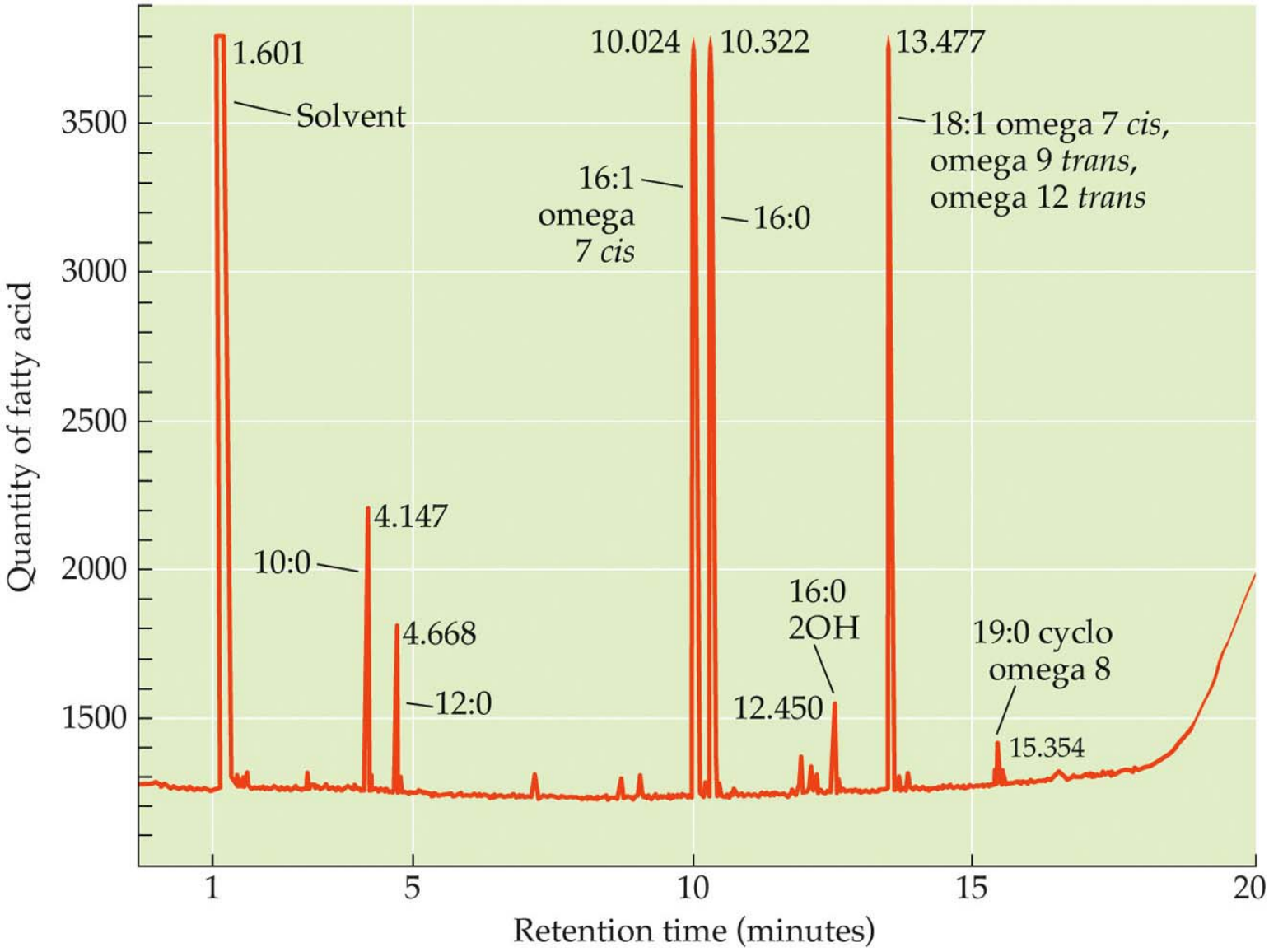
- Minimal genetic variability shown among three strains of bacteria.

FAME analysis Part I

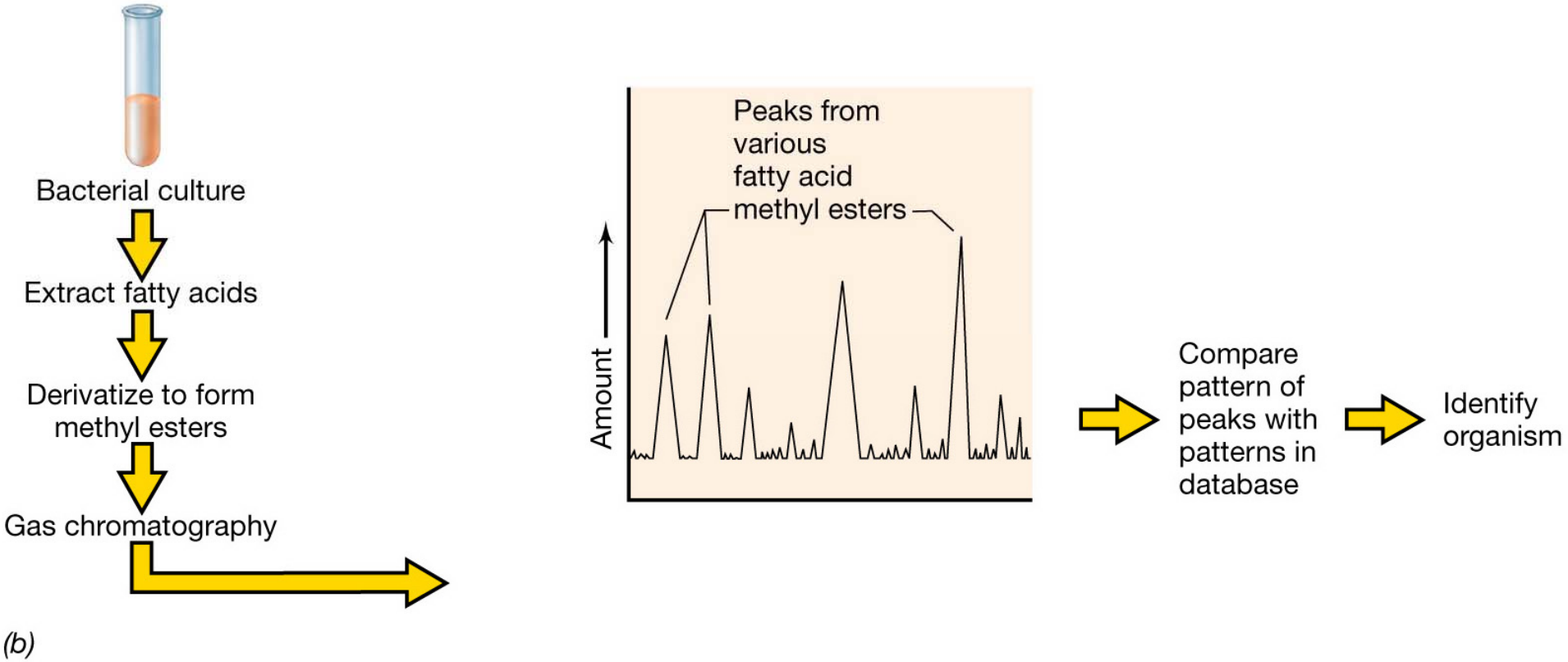
Classes of Fatty Acids in *Bacteria*

Class	Example	Structure of example
Saturated	tetradecanoic acid	
Unsaturated	*omega-7-cis* hexadecanoic acid	
Cyclopropane	*cis 7-8* methylene hexadecanoic acid	
Branched	13-methyltetradecanoic acid	
Hydroxy	3-hydroxytetradecanoic acid	

(a)


1.601
Solvent
10.024
10.322
13.477
18:1 omega 7 *cis*, omega 9 *trans*, omega 12 *trans*
16:1 omega 7 *cis*
16:0
4.147
10:0
4.668
12:0
16:0 2OH
12.450
19:0 cyclo omega 8
15.354
Quantity of fatty acid
3500
3000
2500
2000
1500
1
5
10
15
20
Retention time (minutes)

FAME analysis Part II

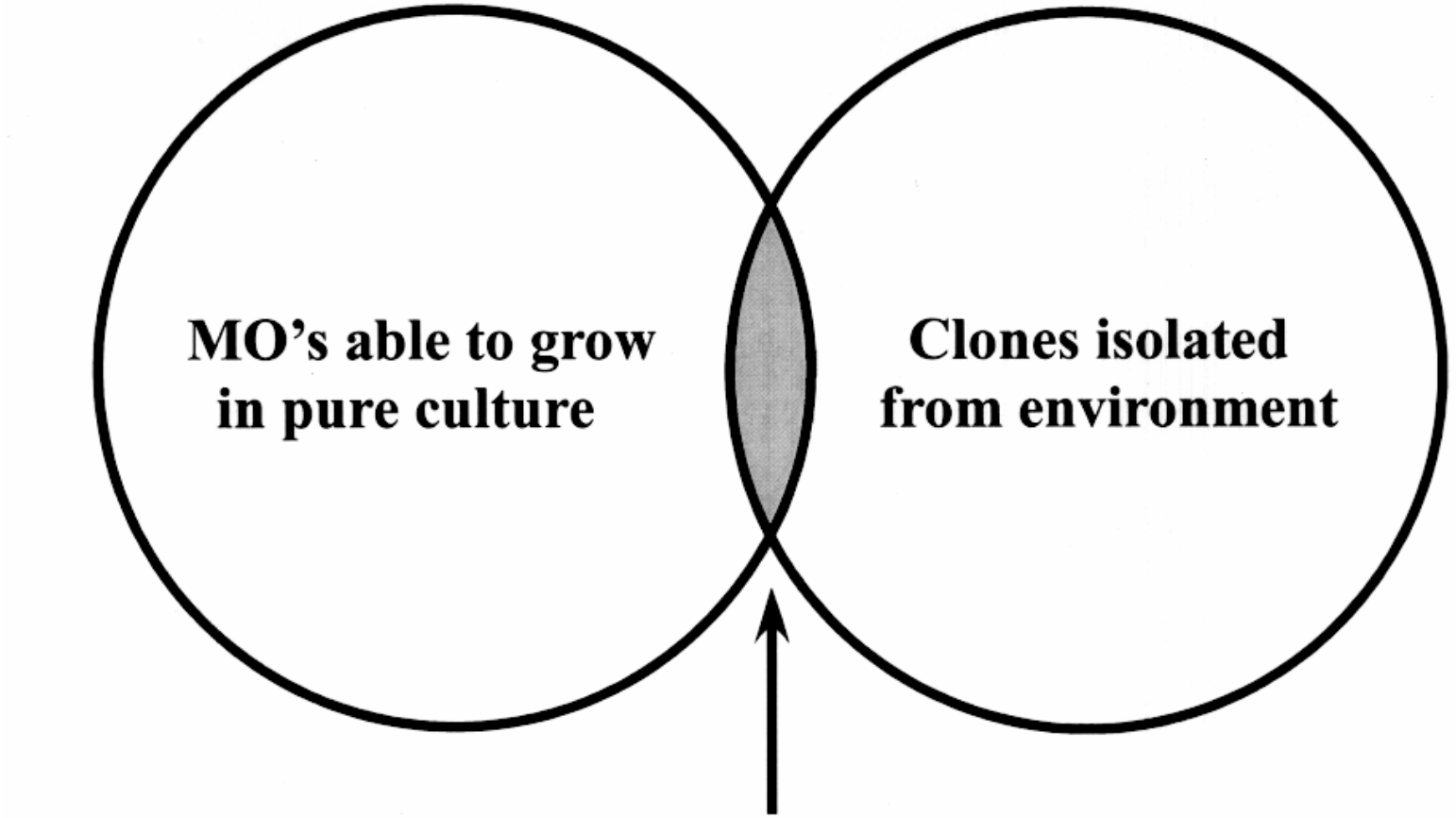


Taxonomy Summary

Classical physiological descriptions of microbes constitute a taxonomy, but do not provide relationships (except as might be inferred subjectively). *Key Words: Classification*, *Identification & Nomenclature*.

Methods such as 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.



>1% Crossover between these groups

Generation of a Clone Library

Total community

DNA extraction

↓

PCR with domain specific primers

↓

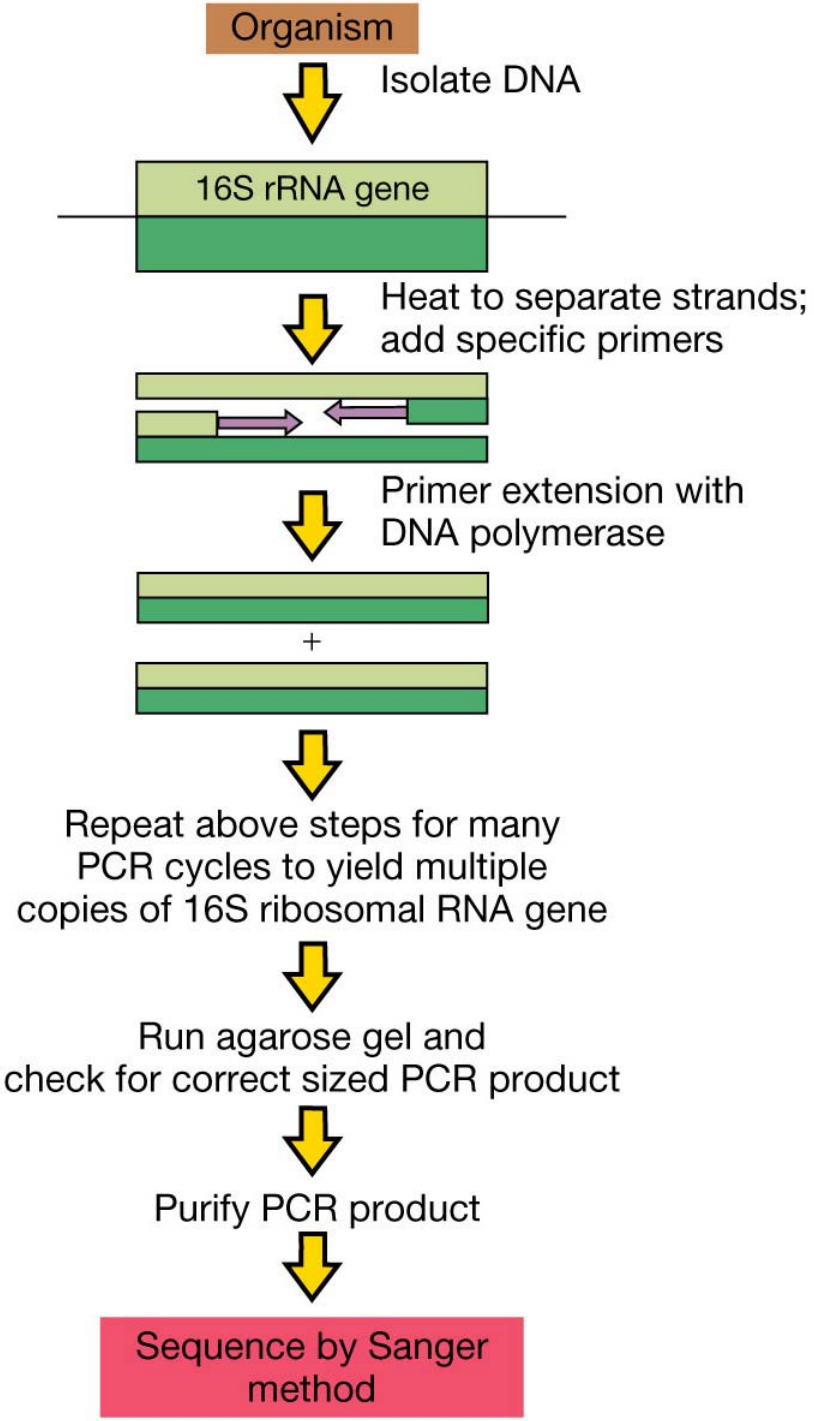
Ligation and transformation

↓

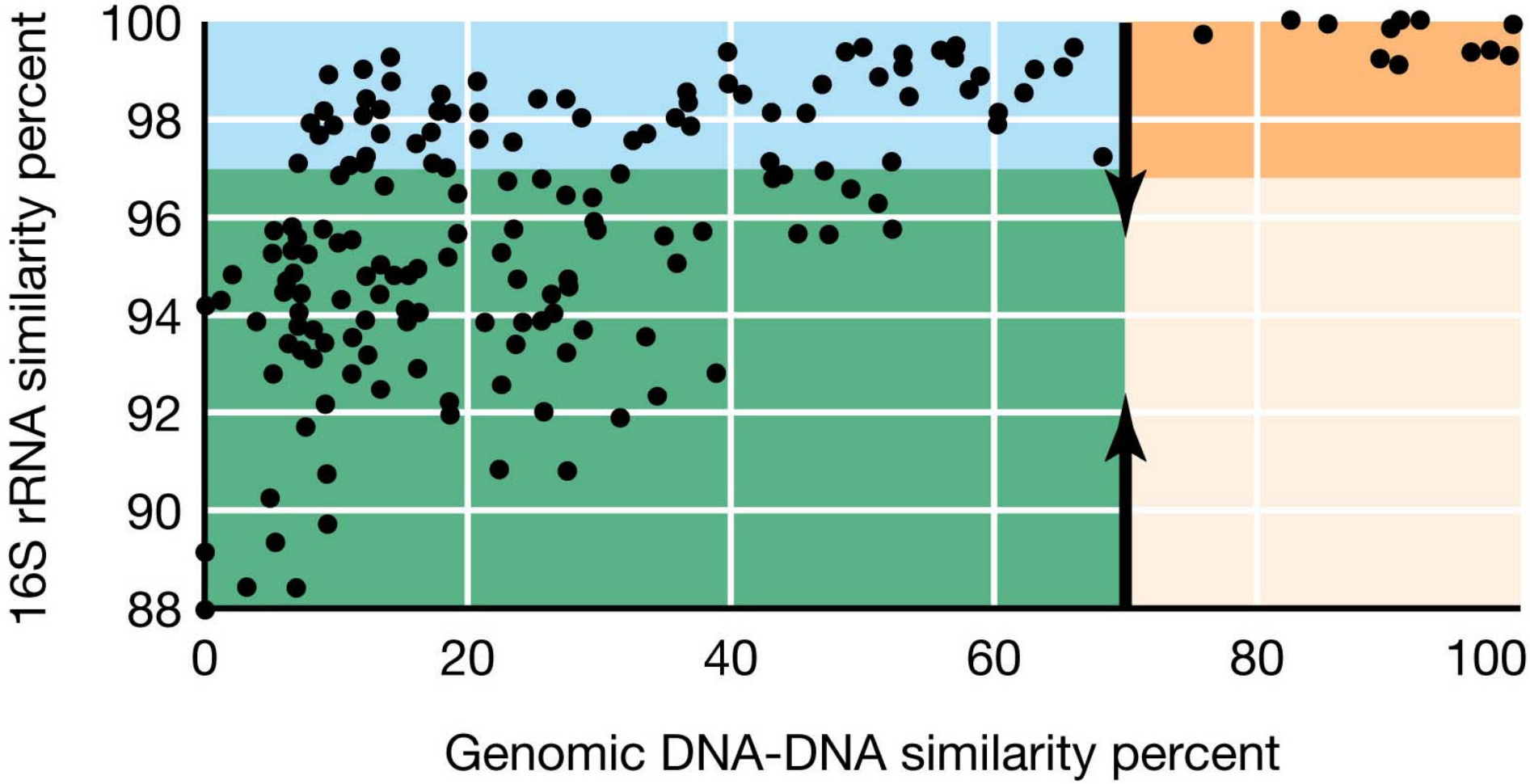
Clone selection and

plasmid purification

Pure culture SSU rRNA sequencing using PCR



Relationship between SSU rDNA and genomic DNA hybridization



What's up with the blue box???

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

Rank	*Bacteria*	*Archaea*	Total
Domains	1	1	2
Phyla	23	3[a]	26
Classes	32	8	40
Orders	77	12	89
Families	182	21	203
Genera	871	69	941
Species	5007	217	5224

[a]Numbers represent validly named genera and species of *Bacteria* and *Archaea* as of 2001. "Korarchaeota" is a provisional phylum.

Source: Garrity, G.M., Boone, D.R., and R.W. Castenholz (eds.). 2001. *Bergey's Manual of Systematic Bacteriology*, 2d ed., Vol. 1. Springer, New York.

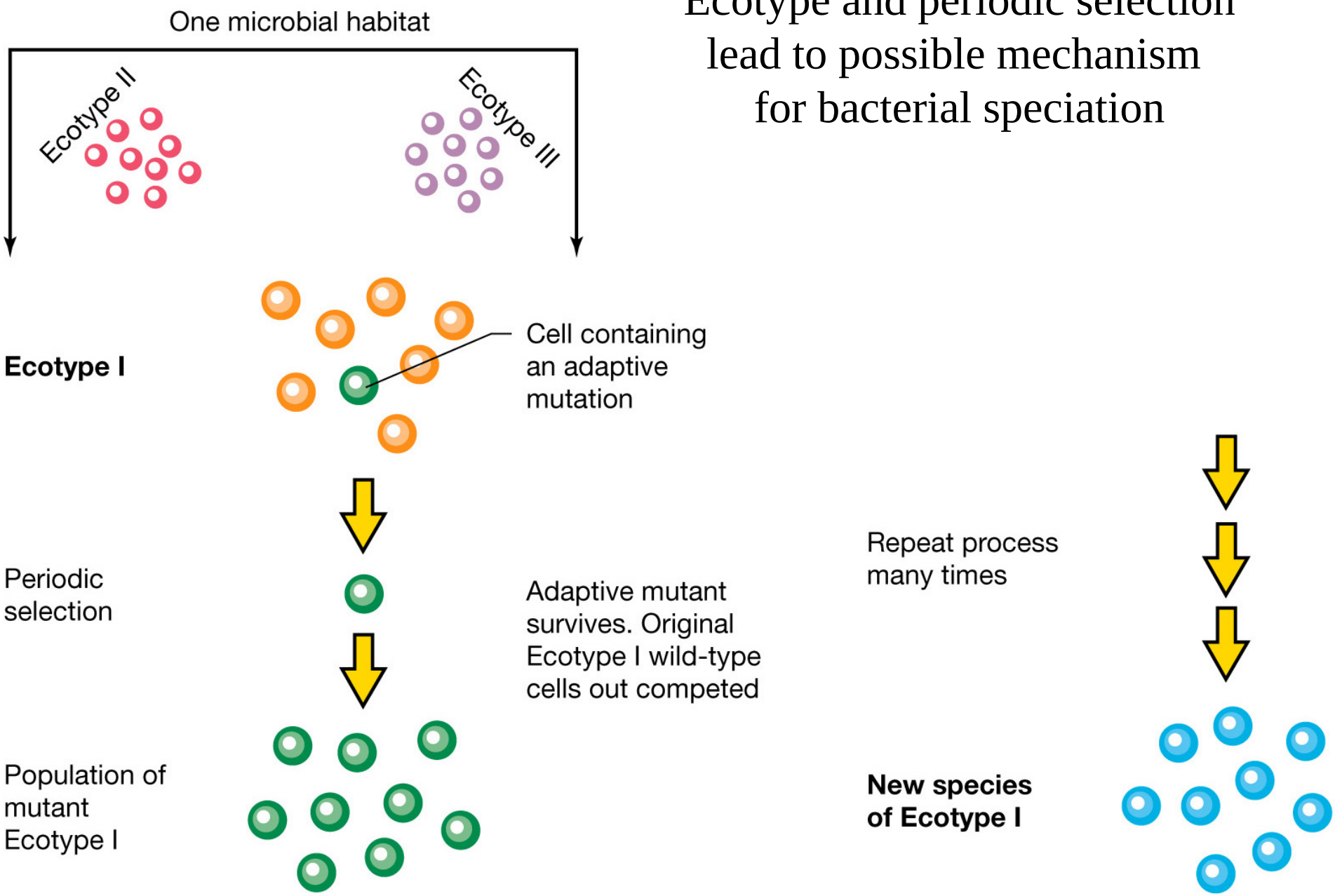
Bacterial species more like animal genus, order or family.

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.

Ecotype and periodic selection lead to possible mechanism for bacterial speciation



Why ribosomal RNAs?

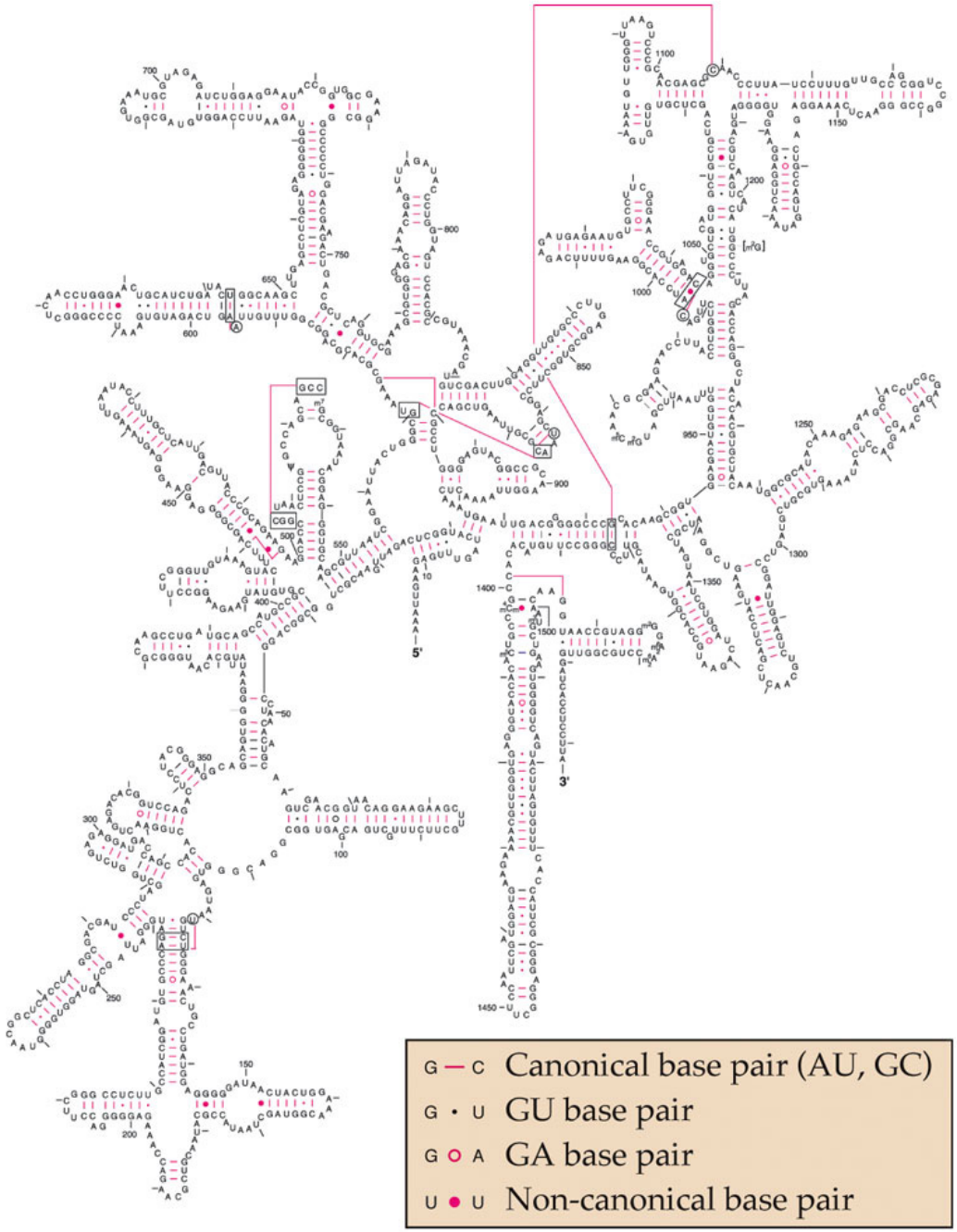
- Found among all living organisms (for 3.8 of the last 4.5 billion years). Integral part of protein synthesis machinery.

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

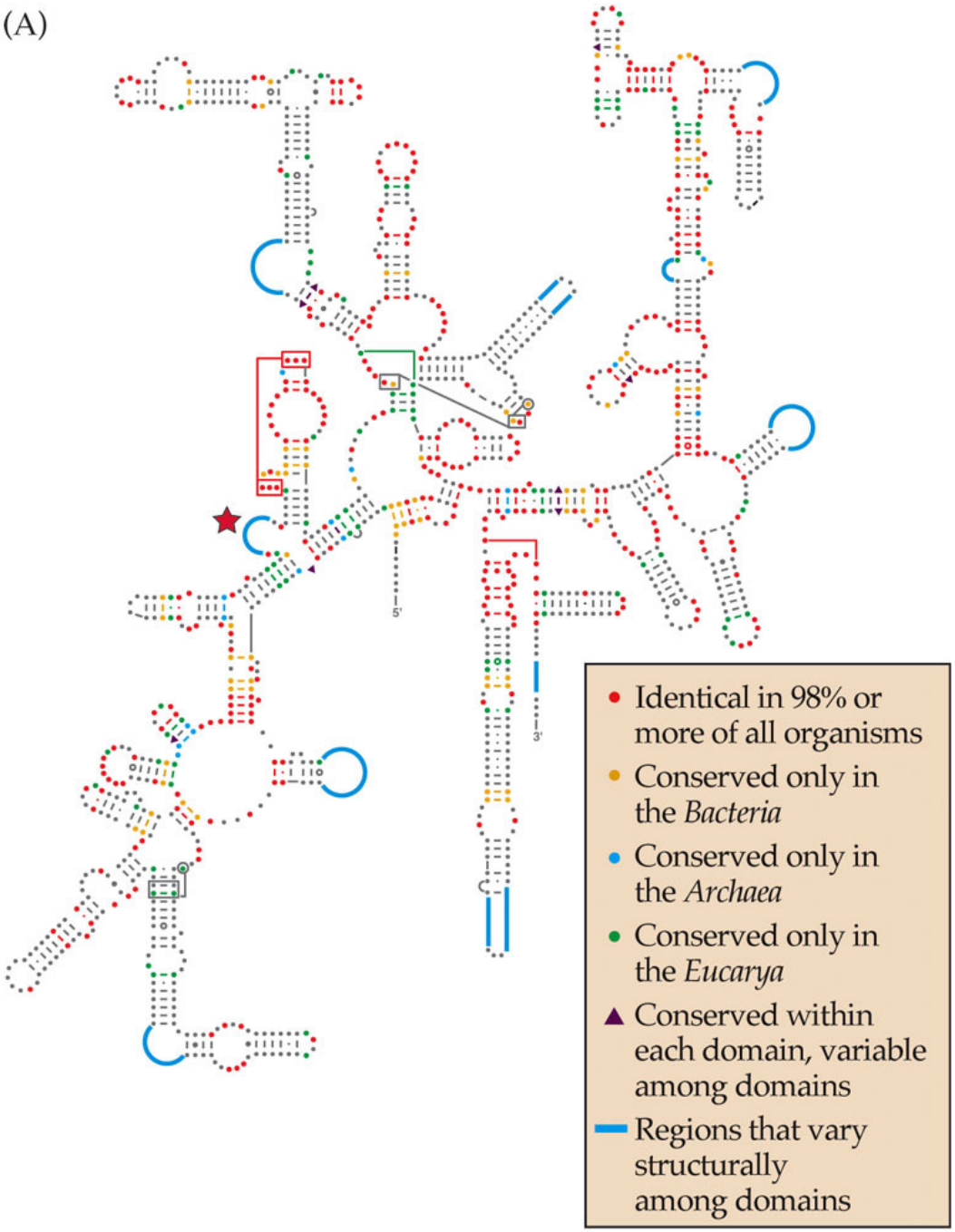
- rRNAs offer a type of sequence information that makes them excellent descriptors of an organism's evolutionary history.

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

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


G — C Canonical base pair (AU, GC)
G • U GU base pair
G ○ A GA base pair
U ● U Non-canonical base pair

(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

Similar Secondary Structures of SSU rRNA molecules

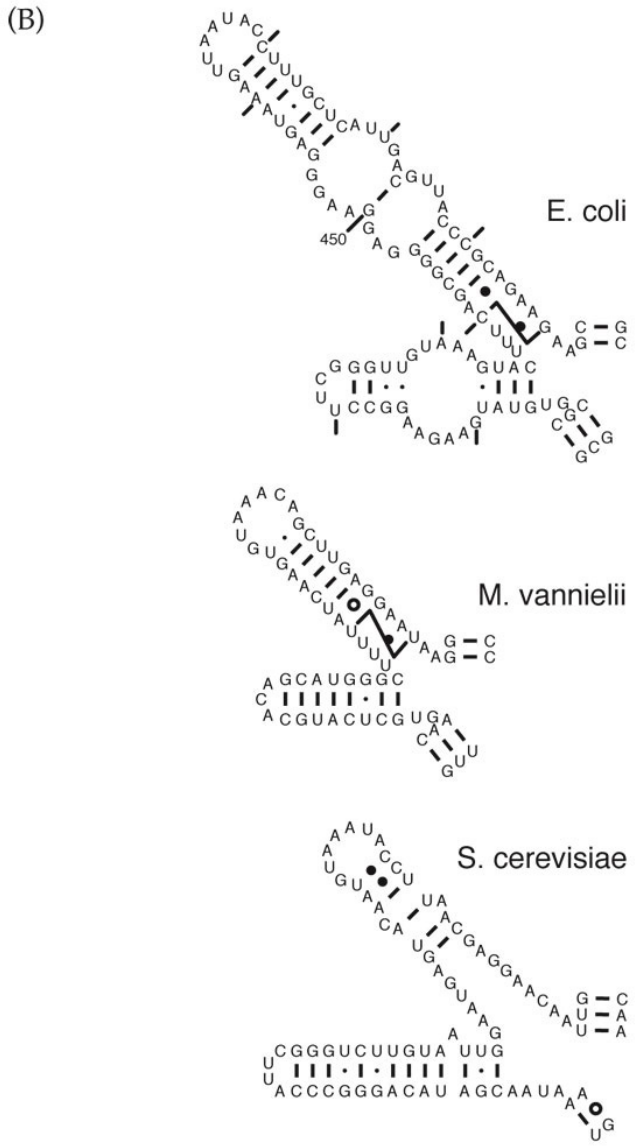


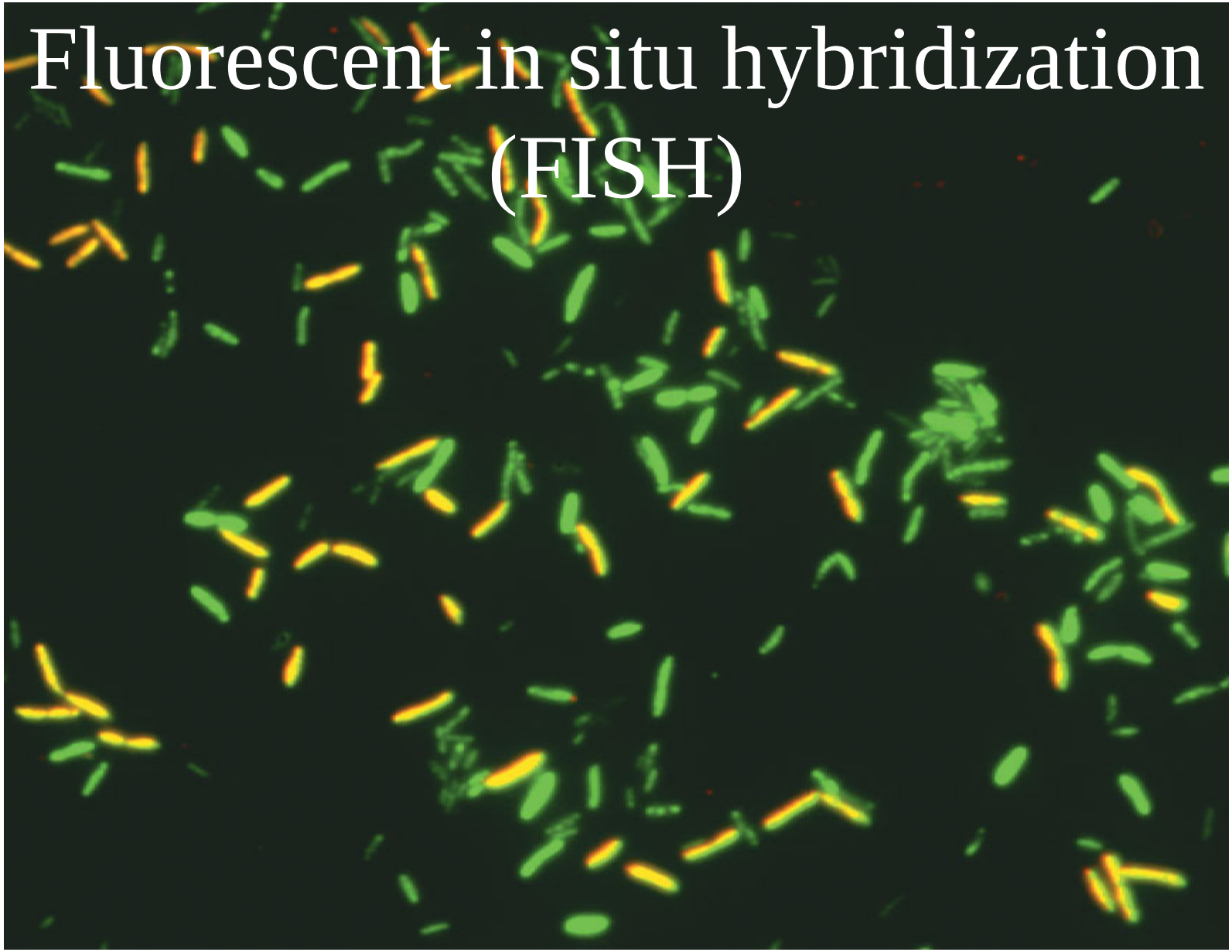
TABLE 11.1 Signature sequences from 16S or 18S rRNA defining the three domains of life

Oligonucleotide signatures[a]	Approximate position[b]	Occurrence among[c]		
		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 11.8*c* for numbering scheme of 16S rRNA.
[c] Occurrence refers to percentage of organisms examined in any domain that contain that sequence.

Signature sequences can be obtained at any level of taxonomic hierarchy

Fluorescent in situ hybridization (FISH)



T-RFLP FLOWCHART

Extract genomic DNA

PCR w/ fluorescent primers

CGGTACCGGGCAGCTGAATTACG

CGGAGCTCTAAGCTAACCGGACG

CGTCCGGCAGCTCTAACGTTACG

CGGCCGGCTAAGCAGCTGTTACG

Environmental Sample

Genomic DNA

Community of PCR amplicons

Cut with *AluI*

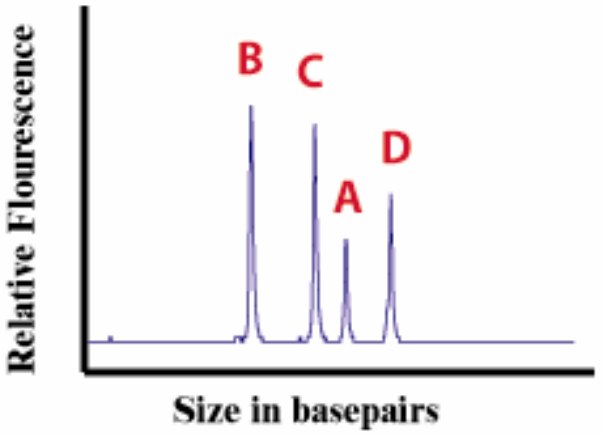
A CGGTACCGGGCAG E CTGAATTACG

B CGGAG F CTCTAAGCTAACCGGACG

C CGTCCGGCAG G CTCTAACGTTACG

D CGGCCGGCTAAGCAG H CTGTTACG

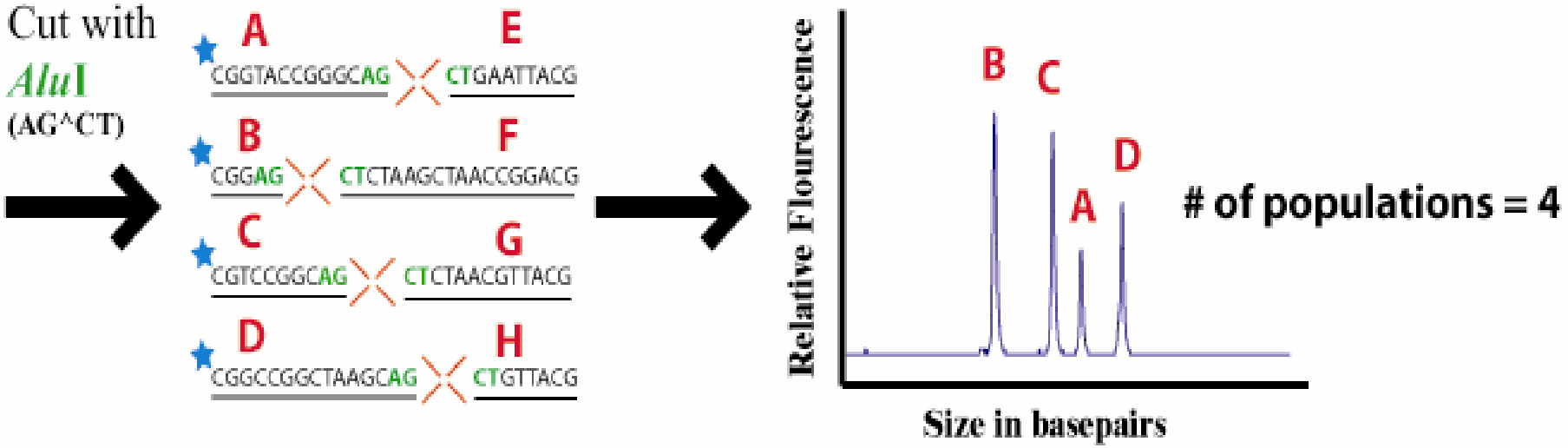
LIF-CE



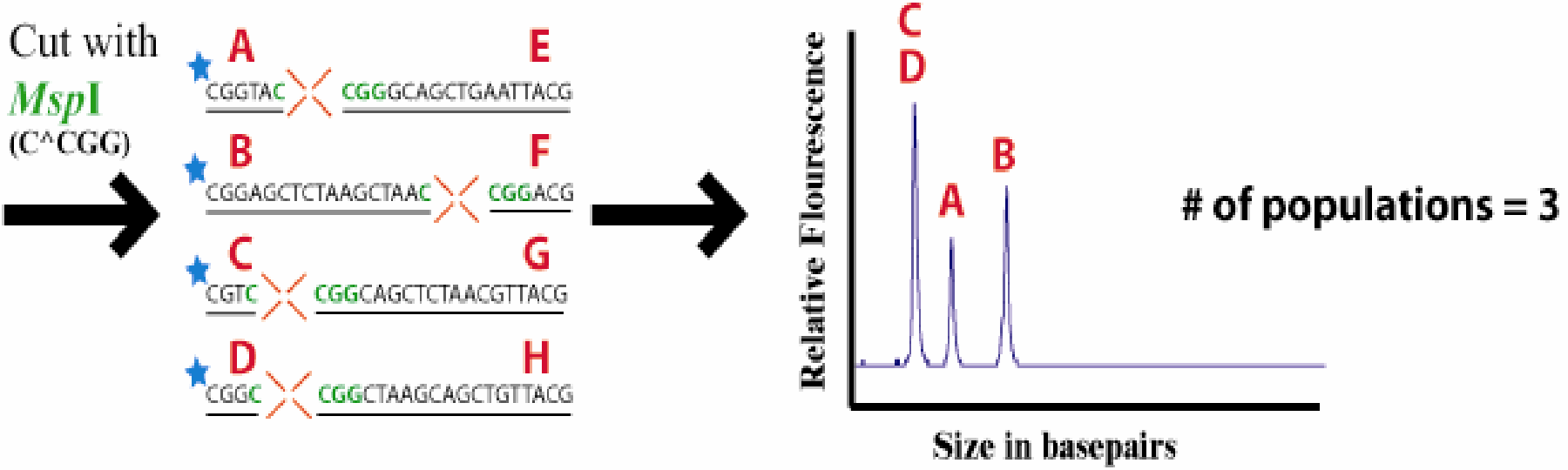
Community of RFs

Separated fragments

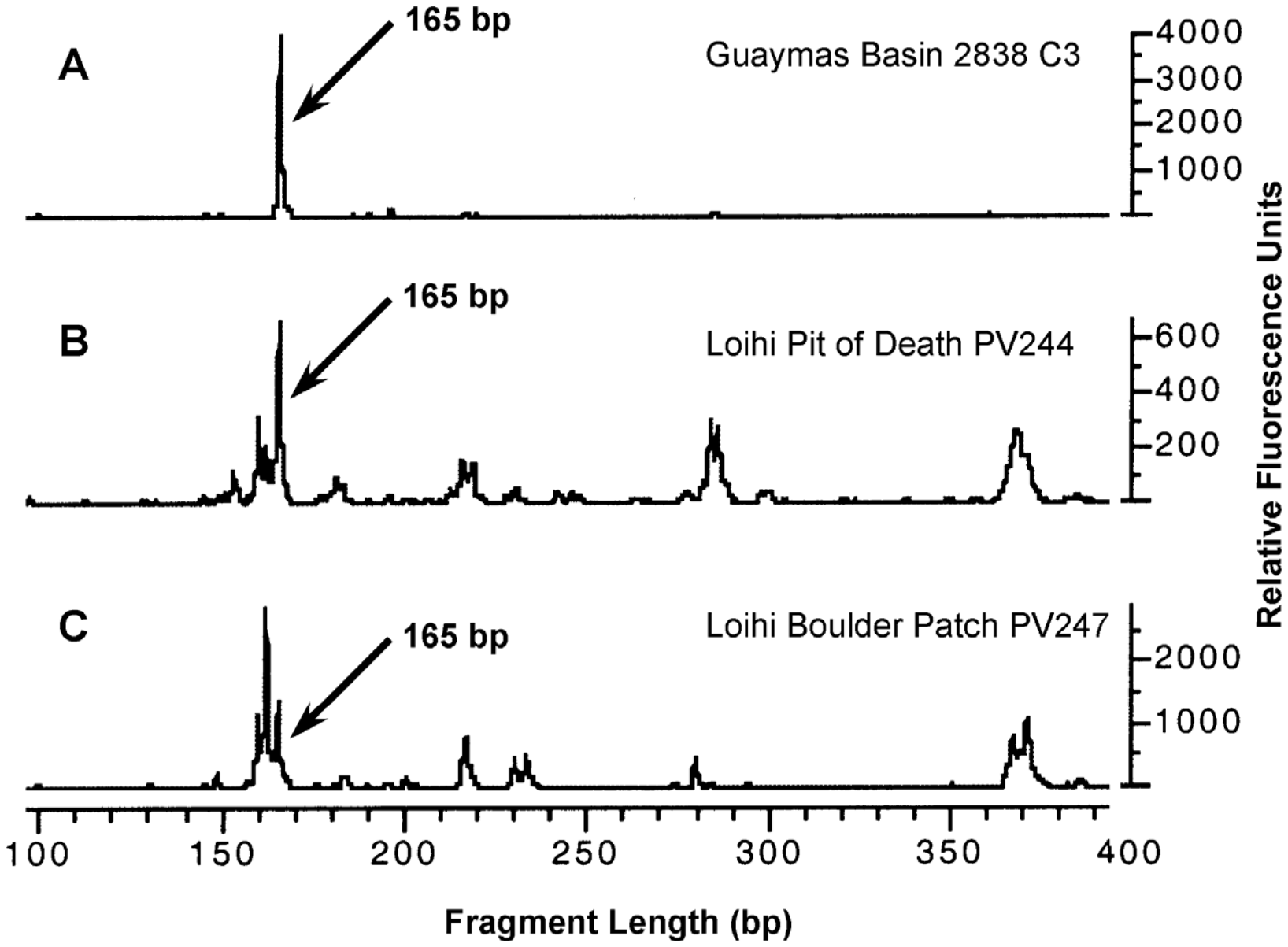
Chromatogram of peak heights



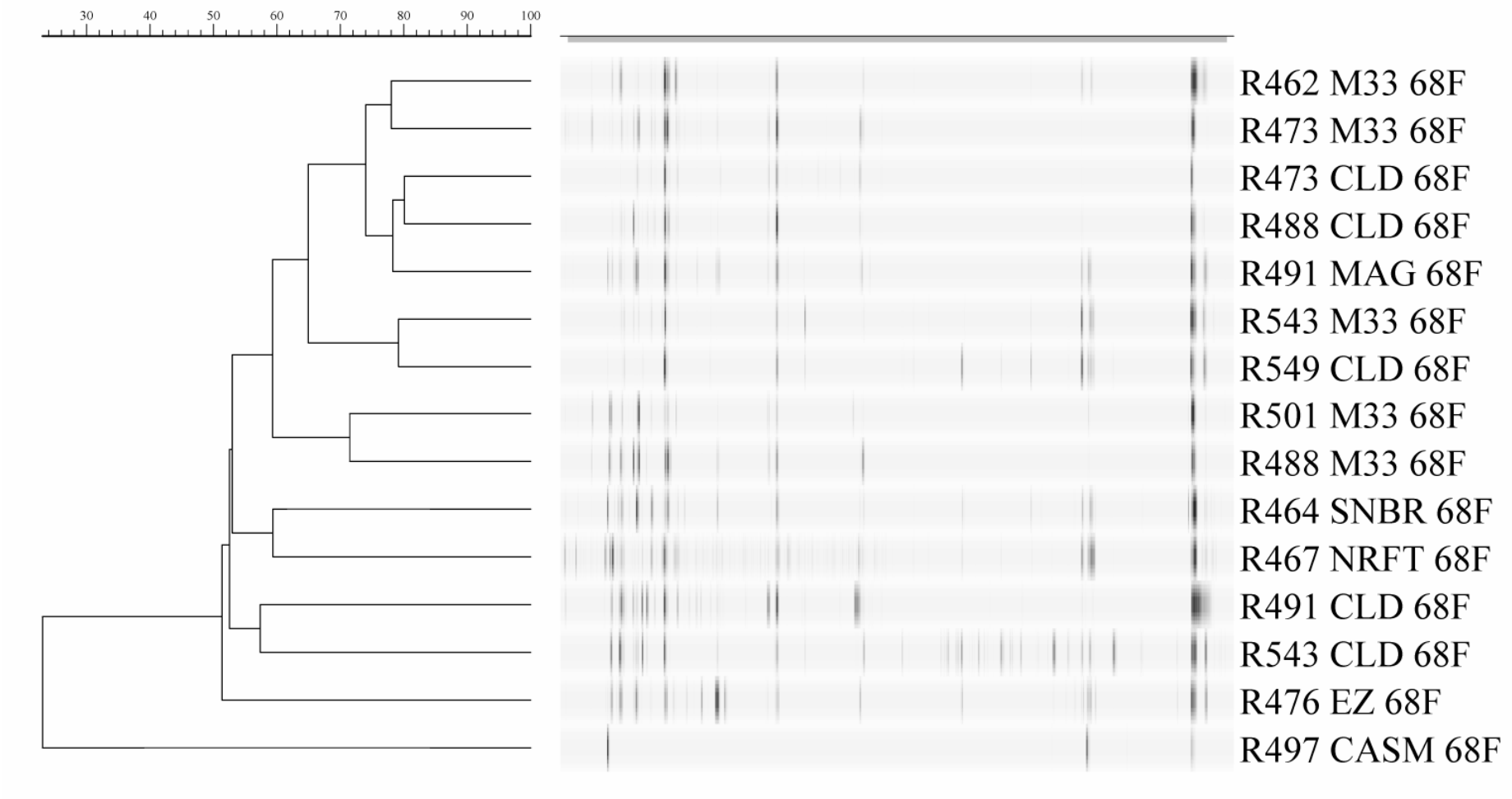
Size is limited to 50-500 basepairs



T-RFLP profiles from Iron-rich Hydrothermal Vents



Cluster Analysis of T-RFLP Data



Estimating evolutionary distance E_D to map on phylogenetic tree

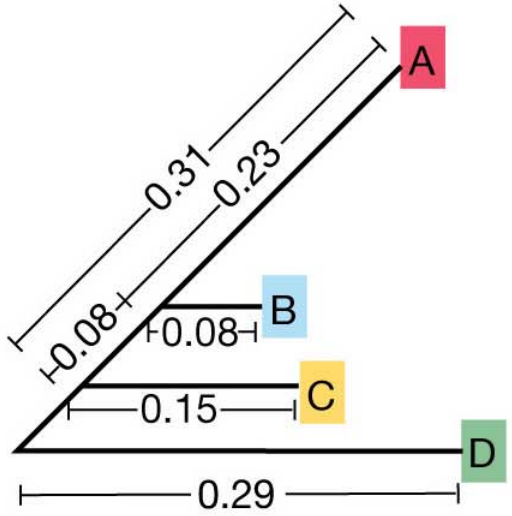
Organism	Sequence
A	CGUAGACCUGAC
B	CCUAGAGCUGGC
C	CCAAGACGUGGC
D	GCUAGAUGUGCC

Analysis: For A ⟶ B, three differences occur out of a total of twelve; thus $\frac{3}{12} = 0.25$

(a) **Sequence alignment and analysis**

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**

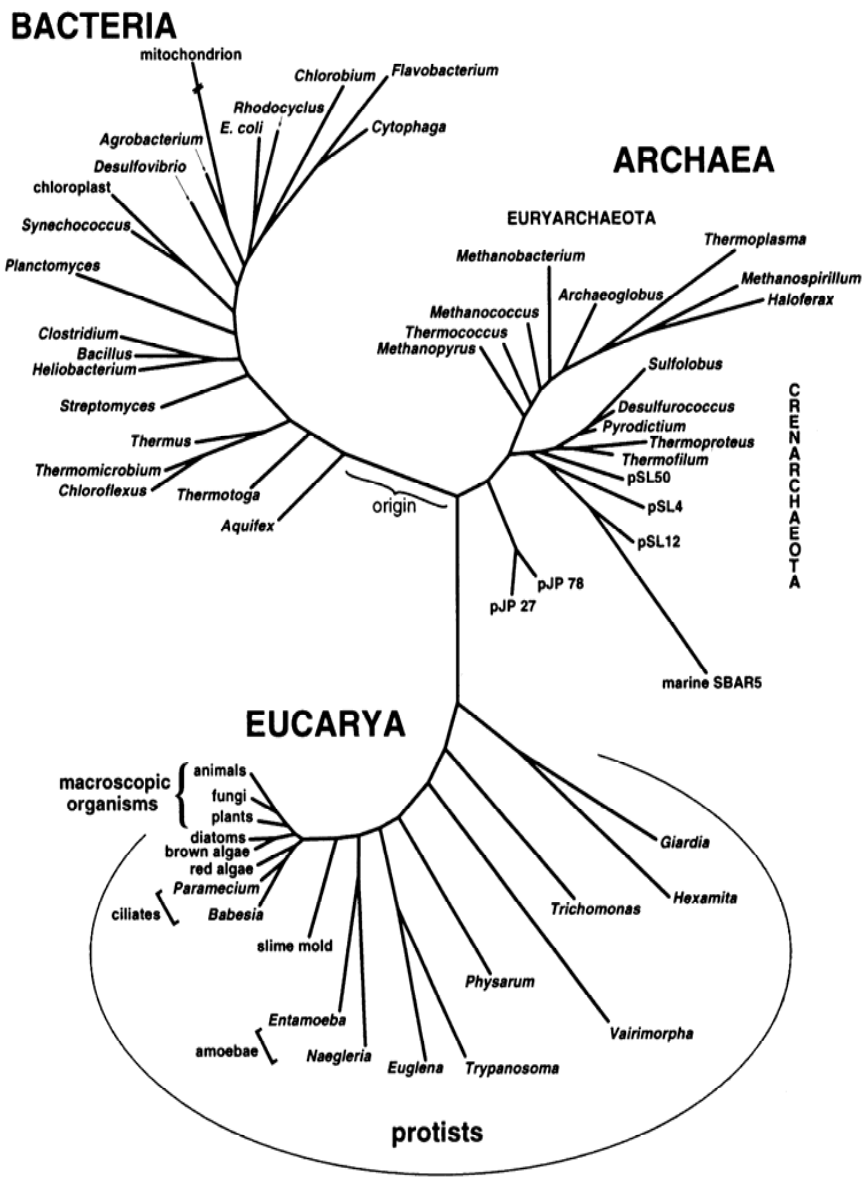
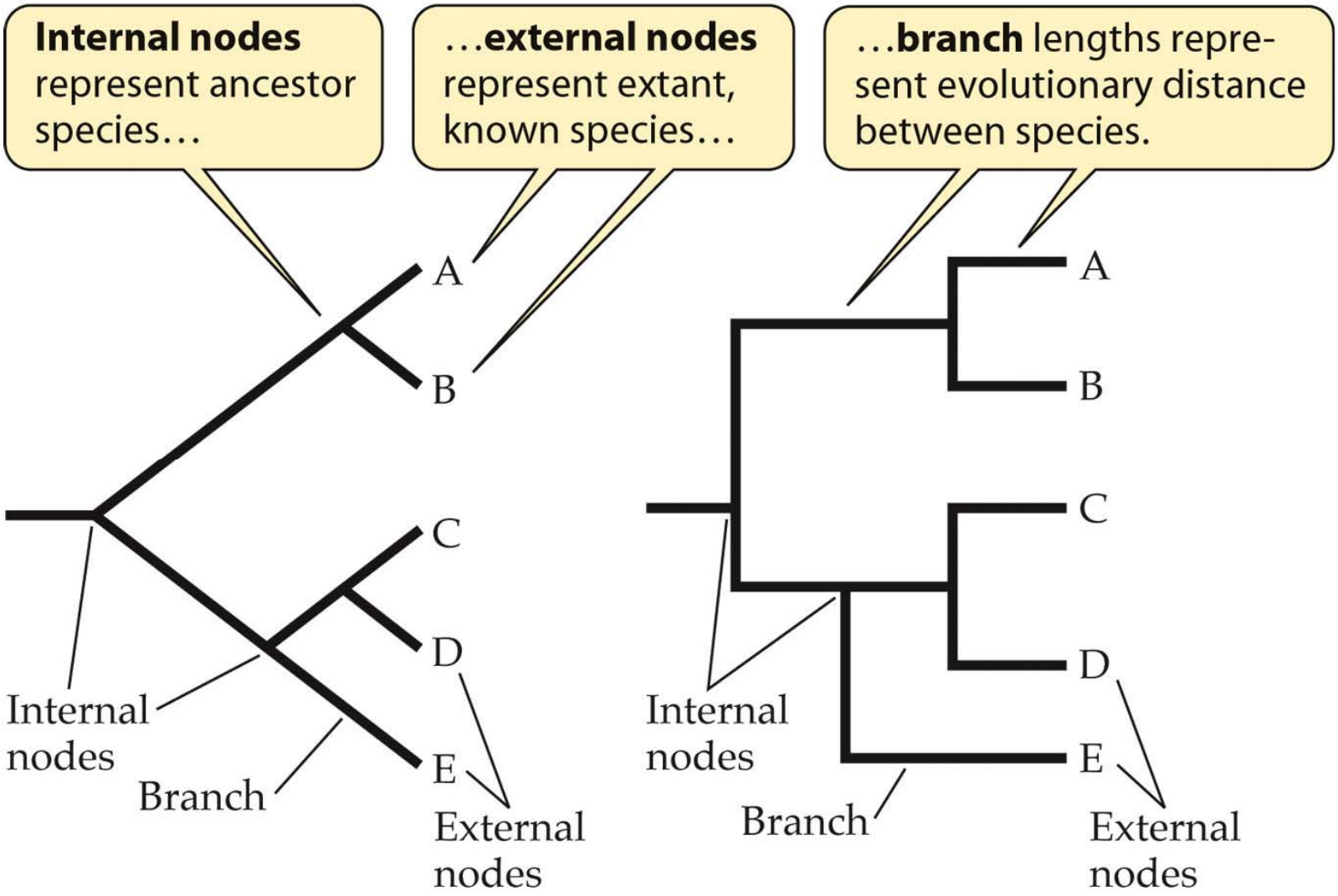


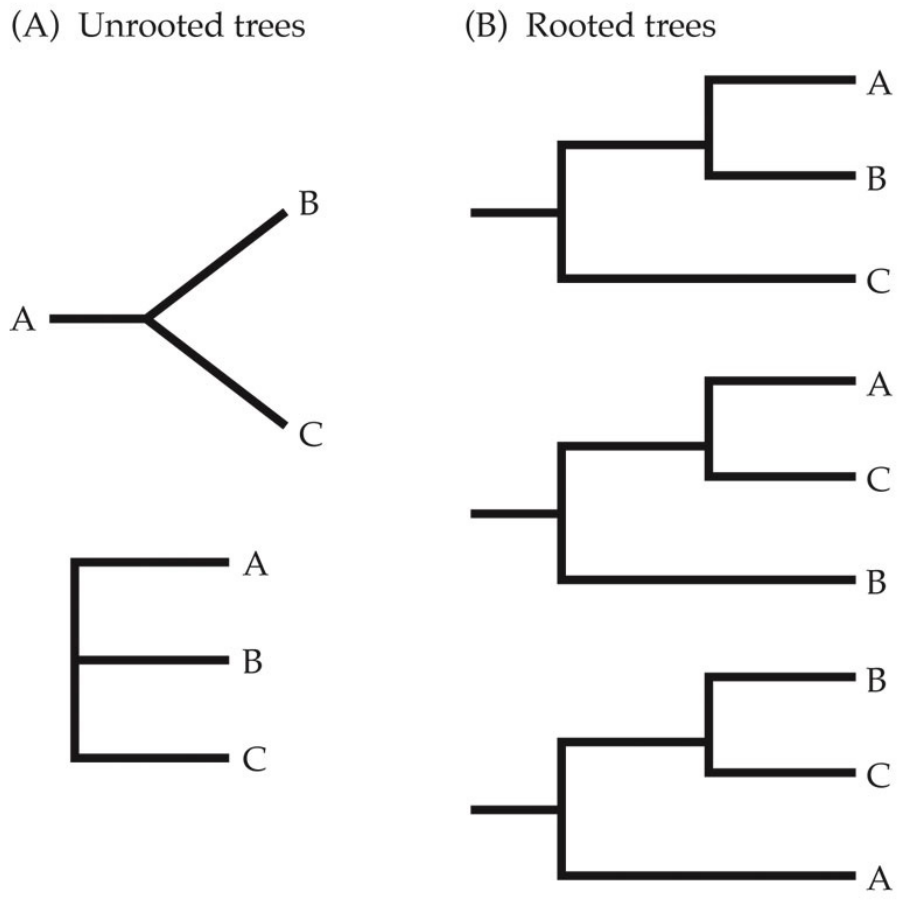
Figure 1. Diagrammatic "Universal" phylogenetic tree of life, based on small-subunit ribosomal RNA sequences. Based on analyses of Barns et al. (1996b), Olsen et al. (1994), and Sogin (1994).

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 organelles).
- Eucarya – Fast clocks
- Archaea – Slow clocks
- Bacteria – Intermediate


Internal nodes represent ancestor species…
…**external nodes** represent extant, known species…
…**branch** lengths represent evolutionary distance between species.
A
B
C
D
E
Internal nodes
Branch
External nodes
A
B
C
D
E
Internal nodes
Branch
External nodes


(A) Unrooted trees
B
A
C
A
B
C
(B) Rooted trees
A
B
C
A
C
B
B
C
A

36 ~Phyla, 23 with isolates

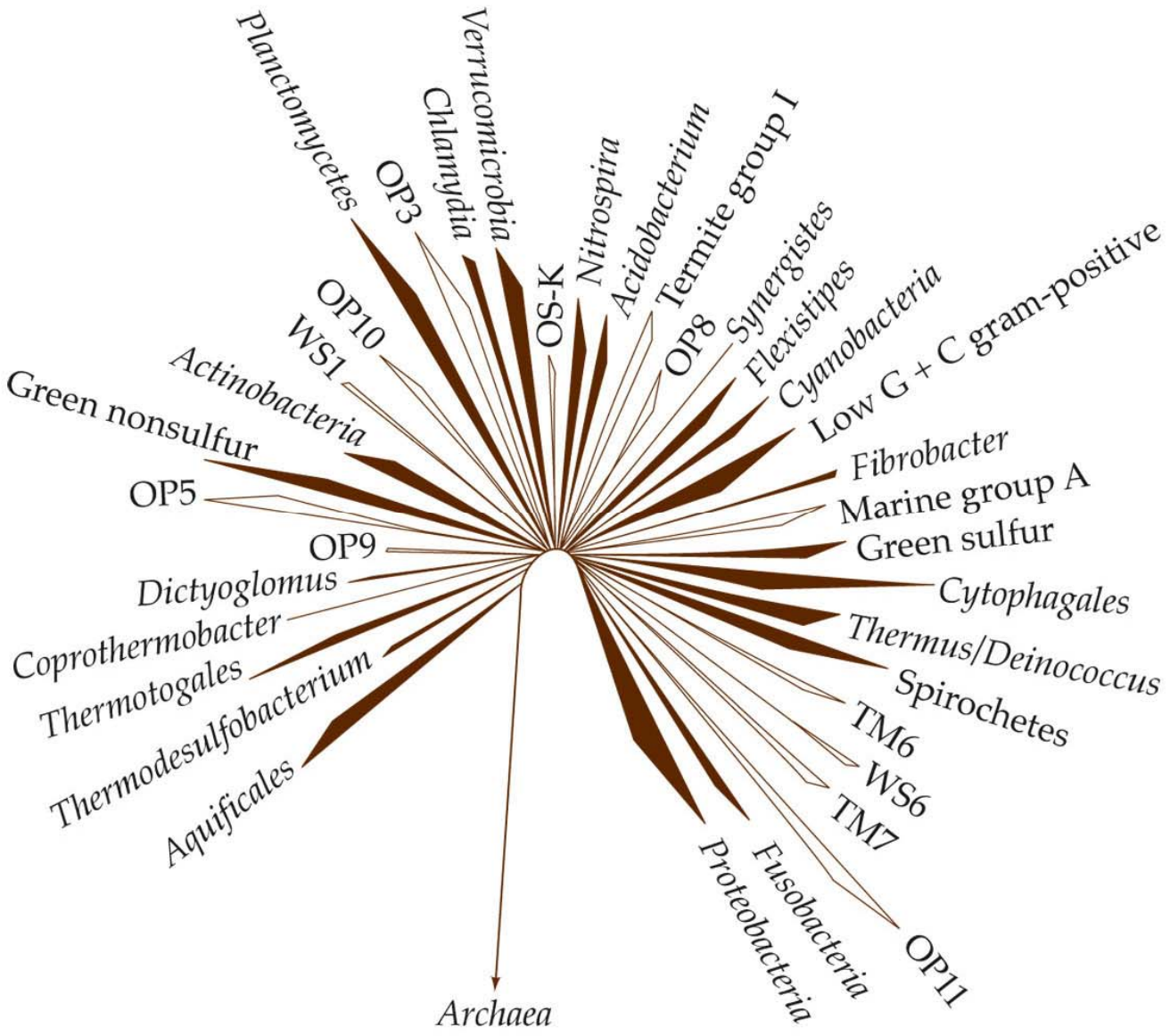
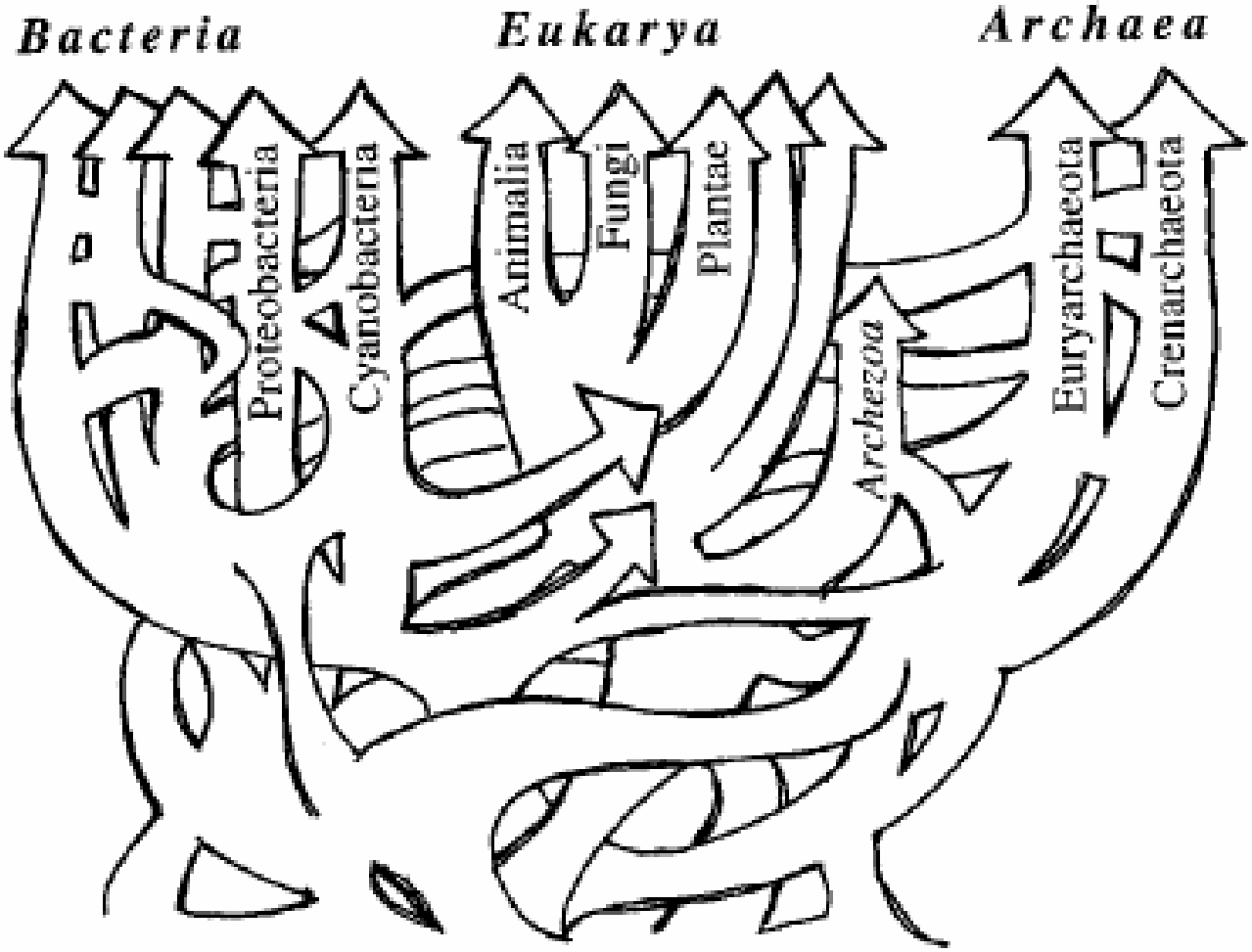


Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history.



Doolittle's Universal Tree

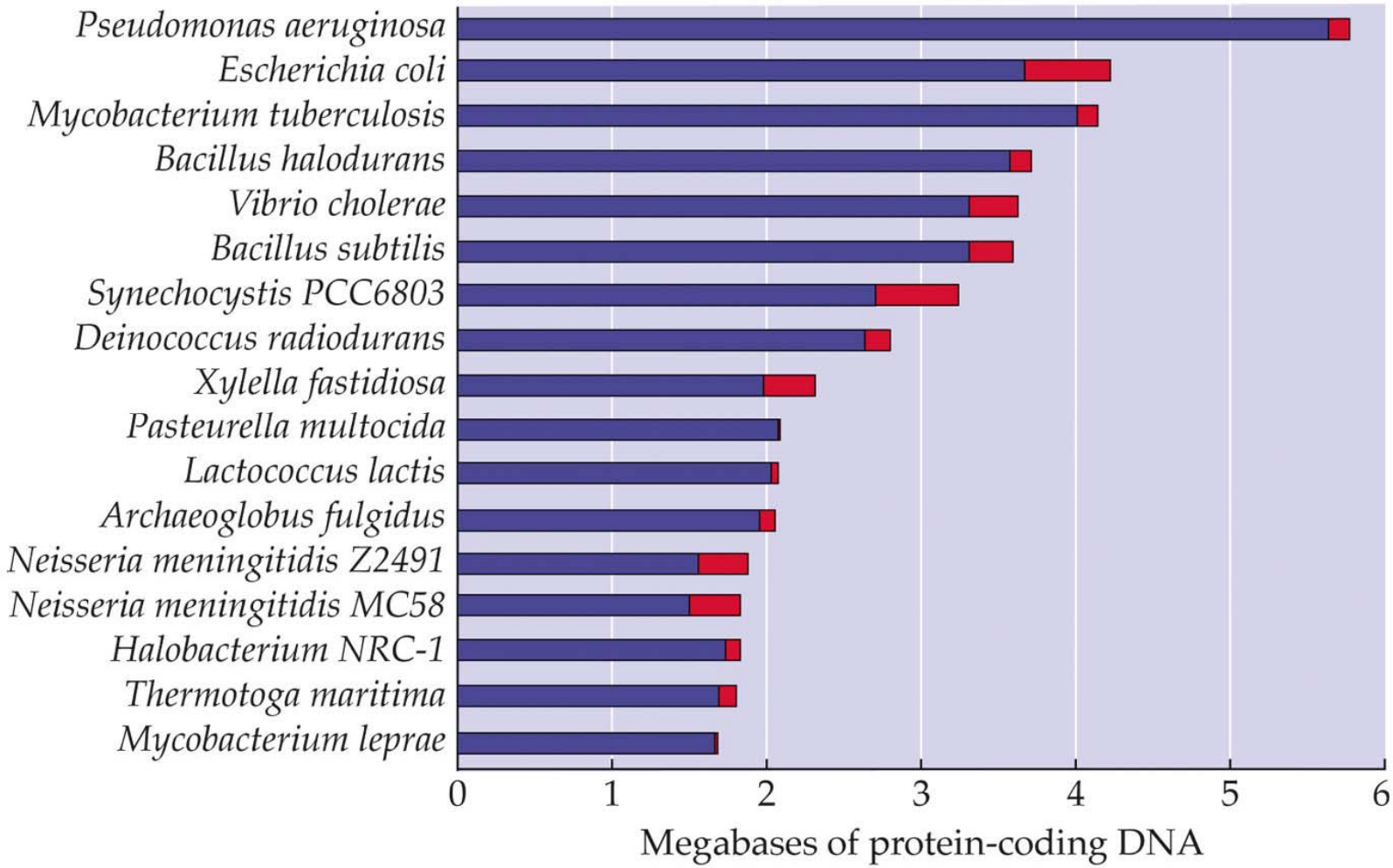
... the general course of evolution [for bacteria] will probably never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds... For these and other reasons, most modem taxonomists have explicitly abandoned the phylogenetic approach.

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

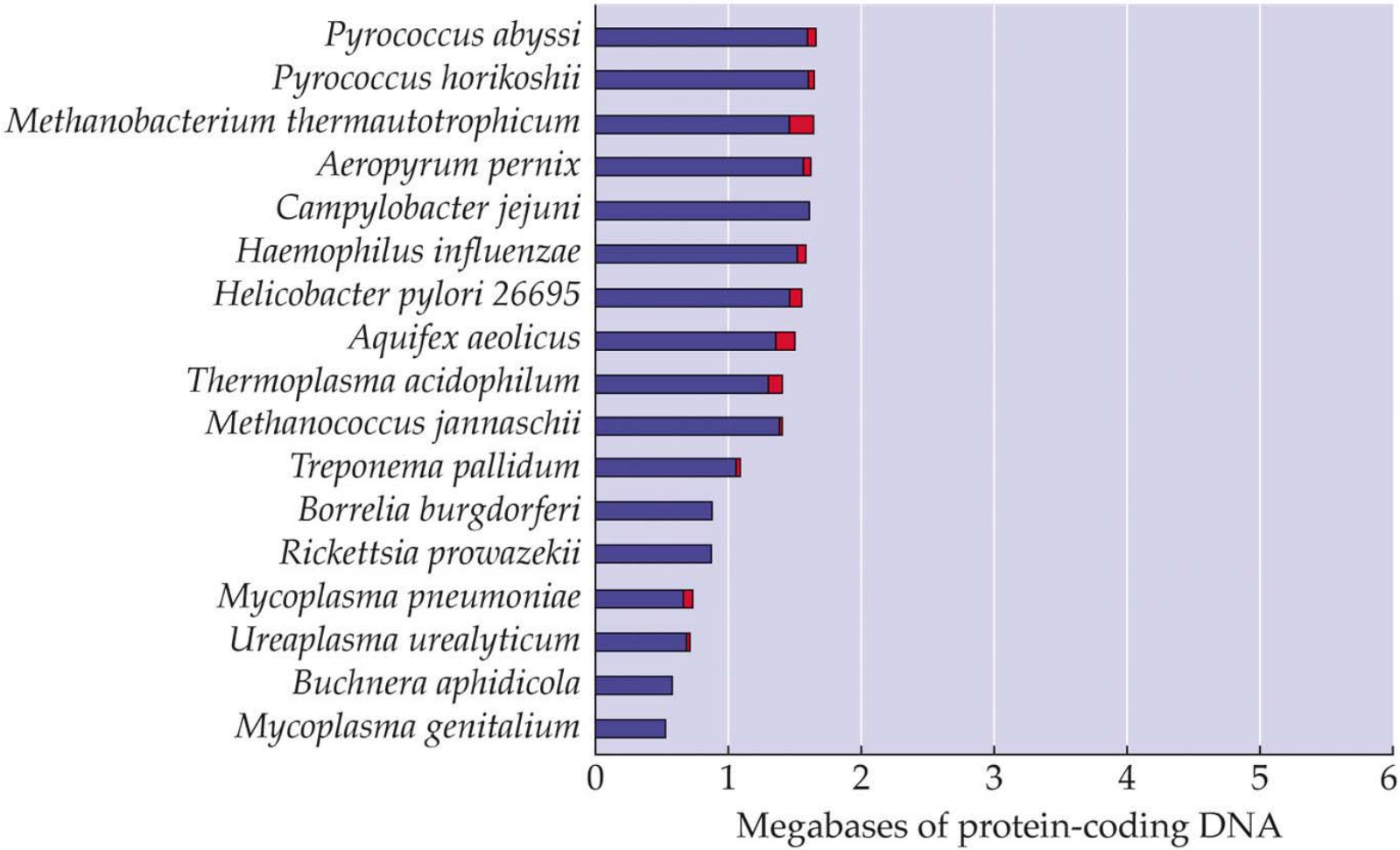
Molecular phylogeneticists will have failed to find the "true tree," not because their methods are inadequate or because they have chosen the wrong genes, but because the history of life cannot property be represented as a tree."

(W. F. Doolittle, *1999*)

Relative importance of horizontal gene transfer



Relative importance of horizontal gene transfer



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

What does genome sequencing and study of functional genomics add to our perspective?

- The central information processing machinery encompasses core genome.
- Metabolic functions, that's when relationships get murky.
- Endosymbiosis involved more than organelles, i.e., two-way transfer of genes with most going to the nucleus.
- Mitochondria have been at it much longer than chloroplasts.

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.

Take Home Message

- Phylogeny is right or wrong, we try to infer it the best we can.
- Taxonomy is useful or not, depending upon your point of view.
- Phylogeny allows us to ask testable questions, e.g., hypothesis testing.