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.

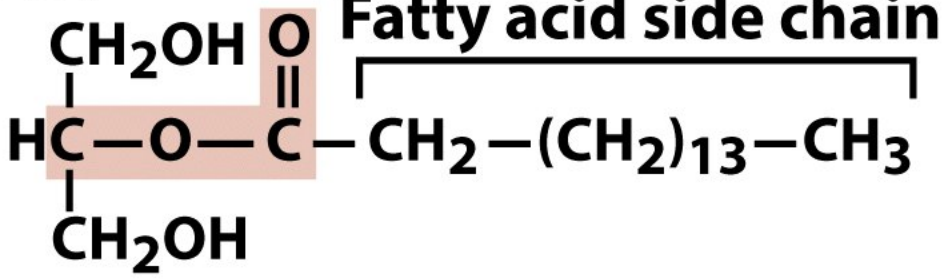
Slowly evolving molecules (e.g. rRNA) used for large-scale structure; "fast- clock" molecules for fine-structure.

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

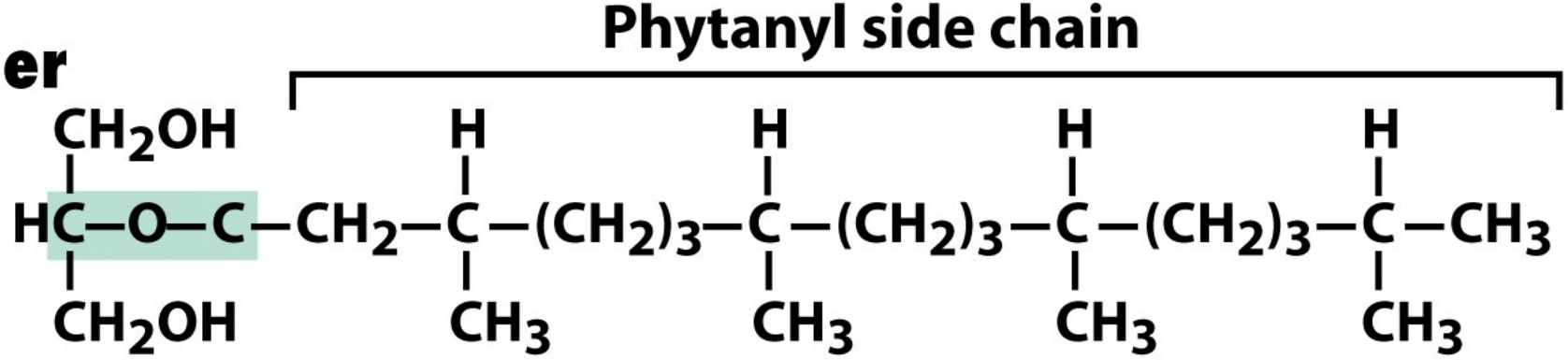
Major category	Components
I. Morphology	Shape; size; Gram reaction; arrangement of flagella, if present
II. Motility	Motile by flagella; motile by gliding; motile by gas vesicles; nonmotile
III. Nutrition and Physiology	Mechanism of energy conservation (phototroph, chemoorganotroph, chemolithotroph); 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

Ester



Bacteria, Eukarya

Ether



Archaea

ID of an enteric bacterium

I. Isolation and microscopy

Isolation → **Pure culture** → **Gram reaction/ morphology**

II. General physiology

Gram-negative rod → **Facultative** → **Ferments lactose to acid/gas**

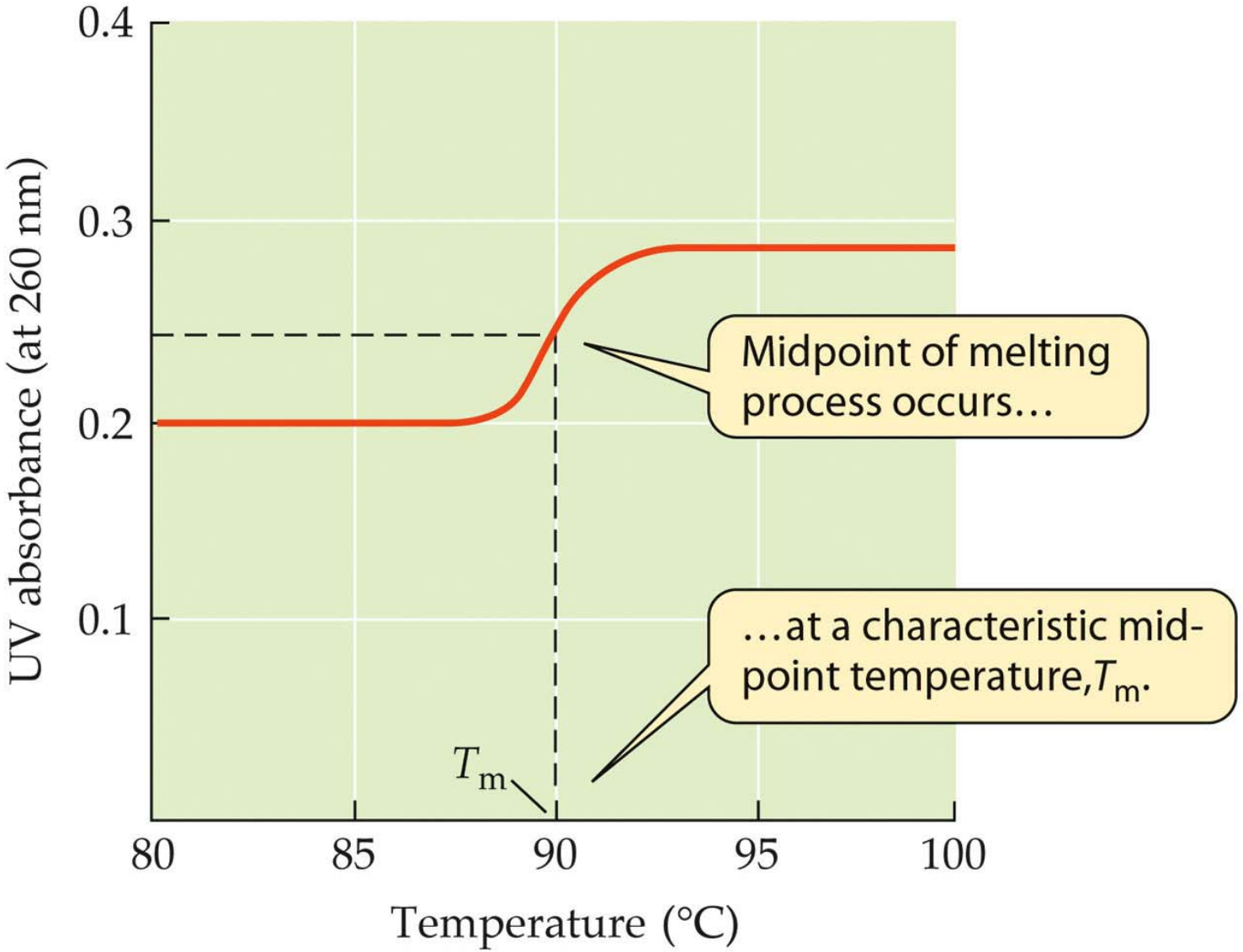
III. Detailed physiology

Facultative lactose fermenter → **Perform series of biochemical tests** → **Positive: indole, methyl red, mucate; Negative: citrate, Voges-Proskauer, H_2S**

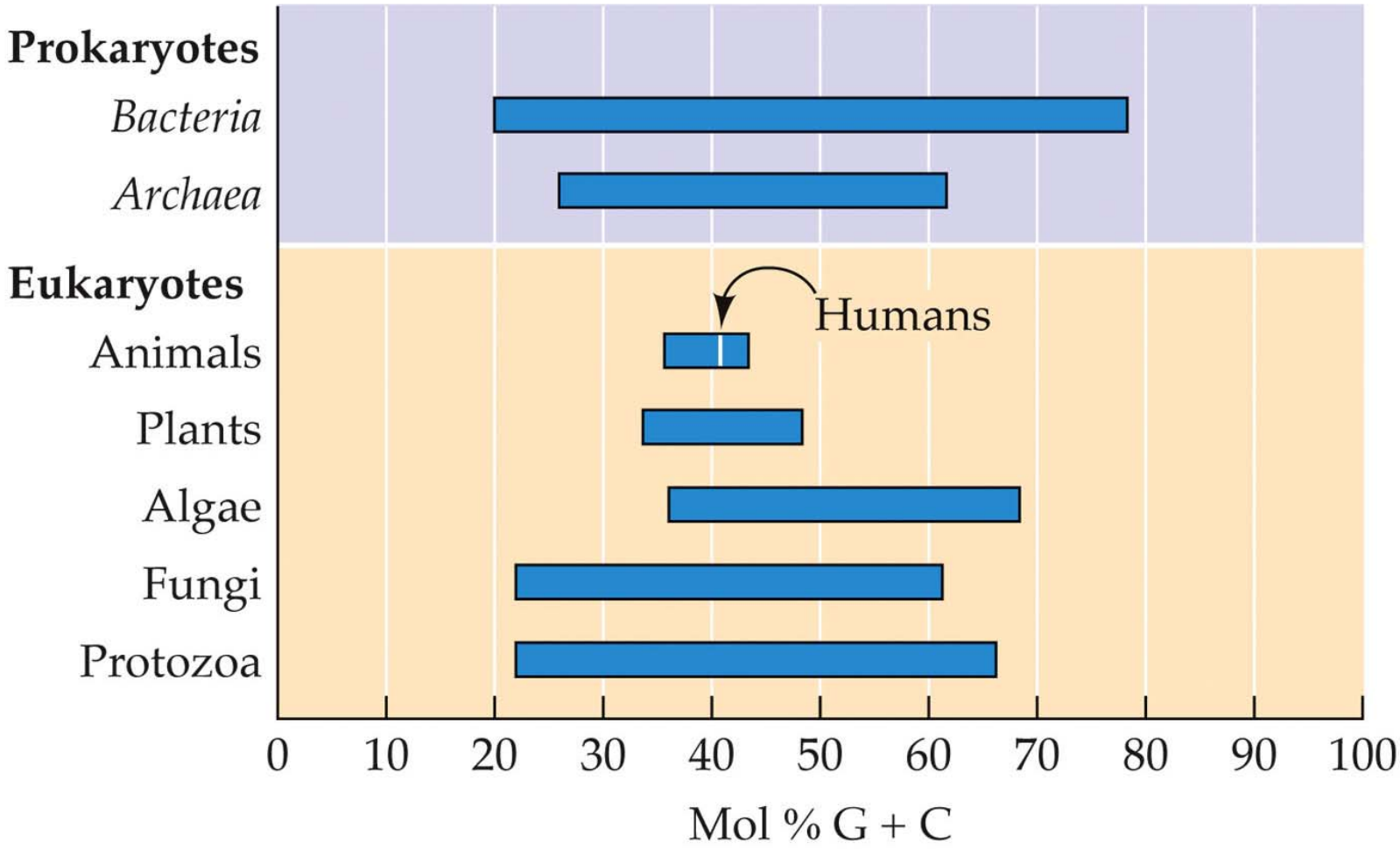
IV. Conclusion → ***Escherichia coli***

Note: requires isolation in pure culture!

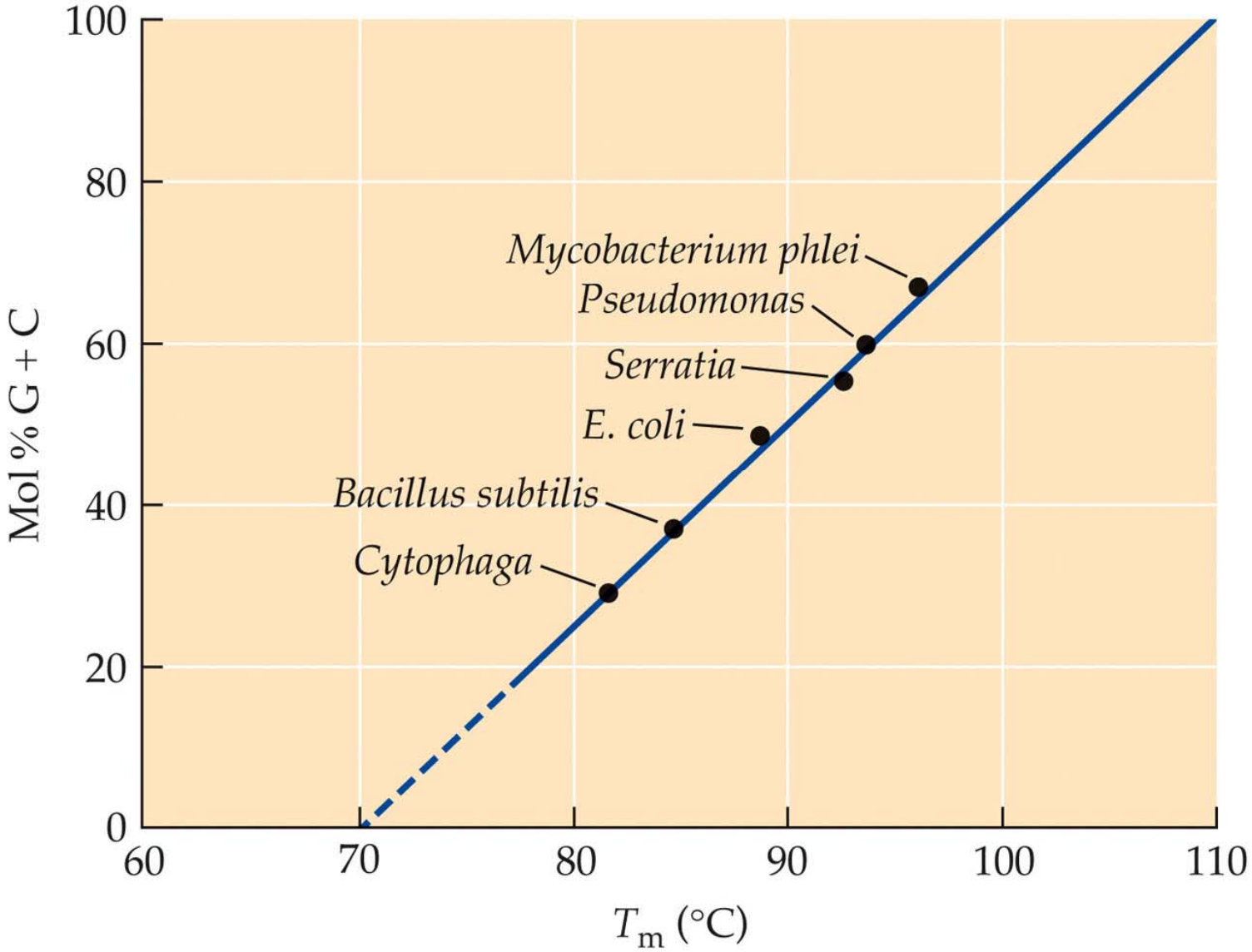
Hyperchromic Effect of DNA



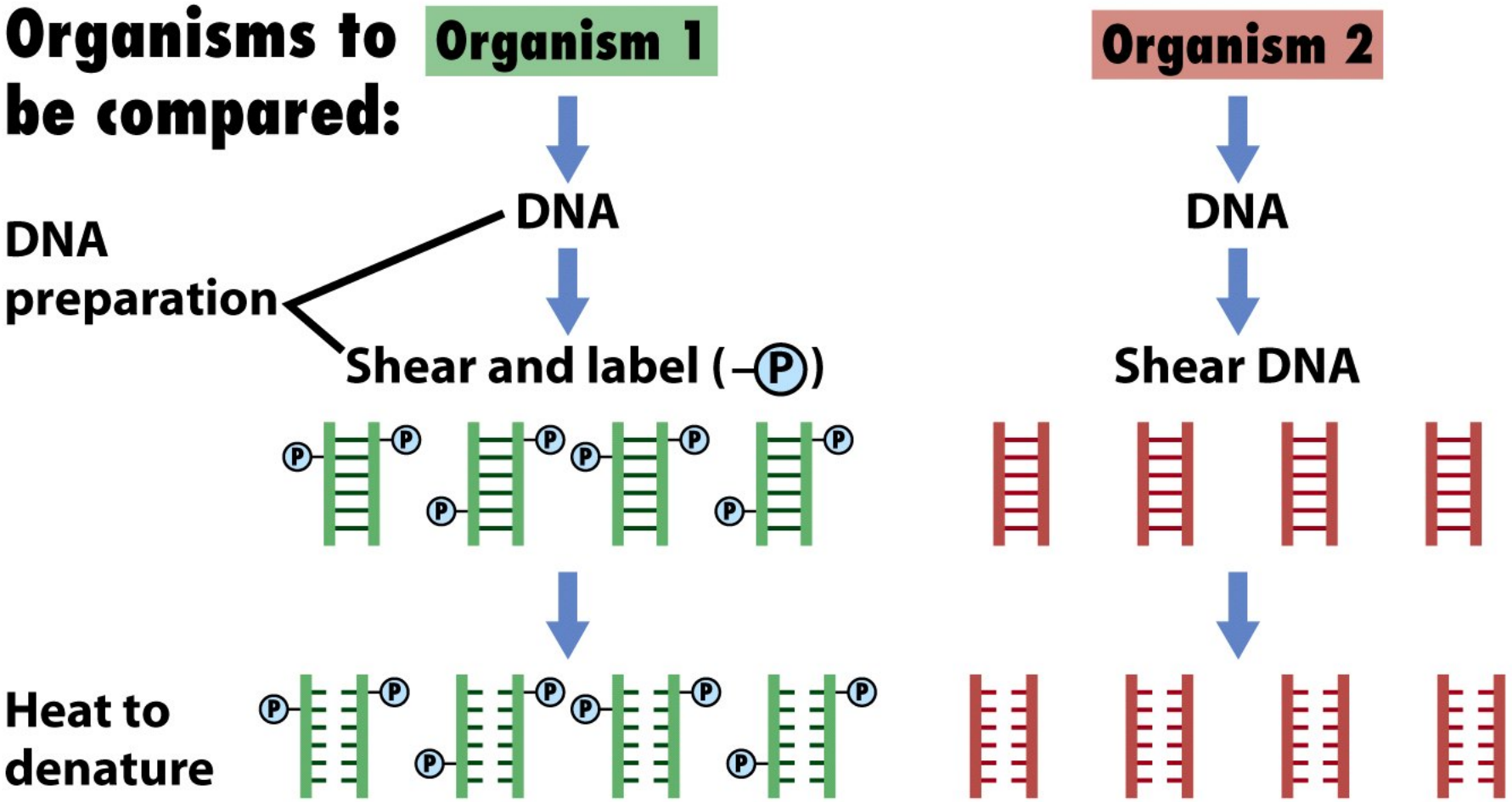
Ranges of DNA base composition



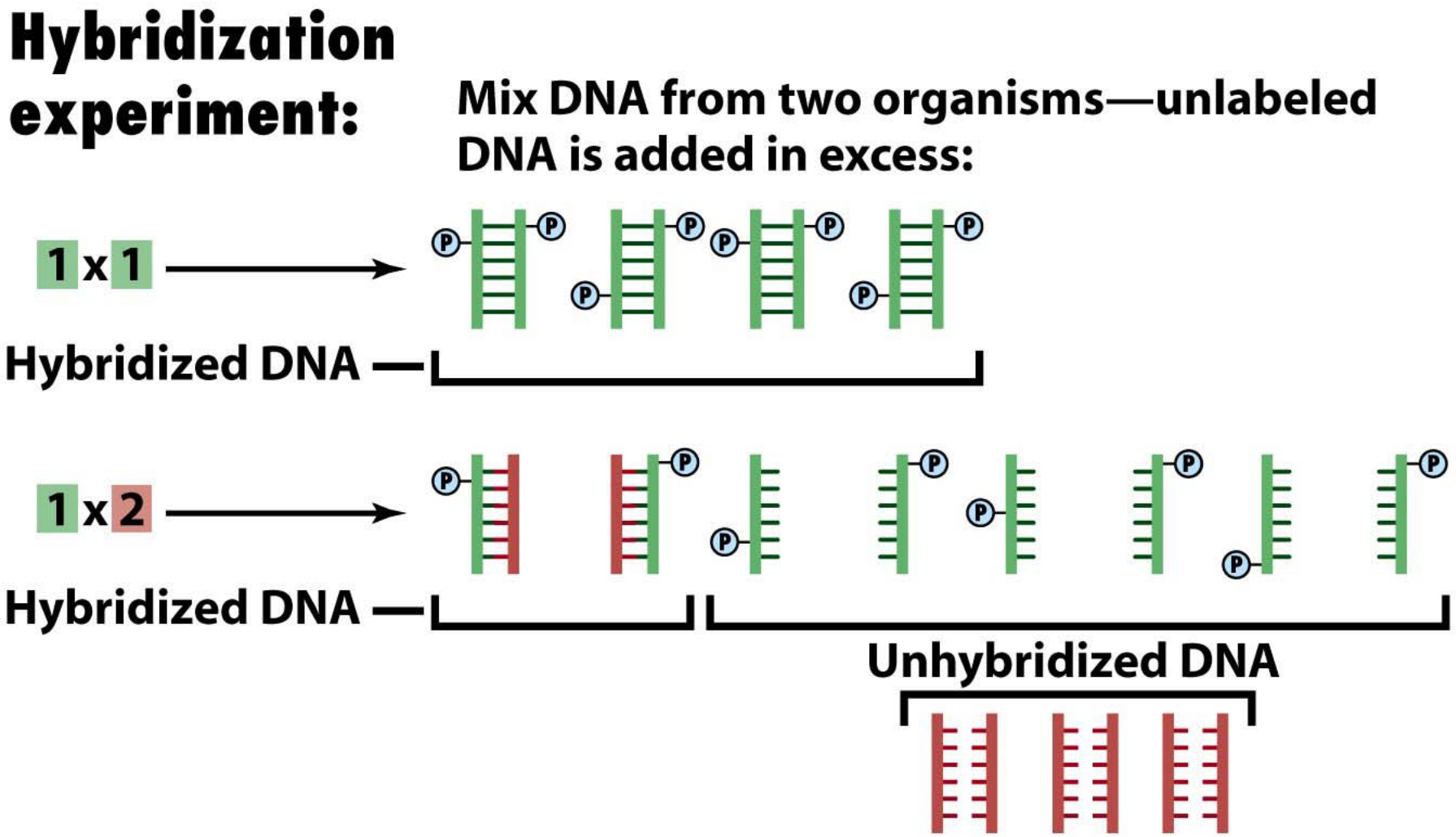
G+C Ratios


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

DNA:DNA hybridization

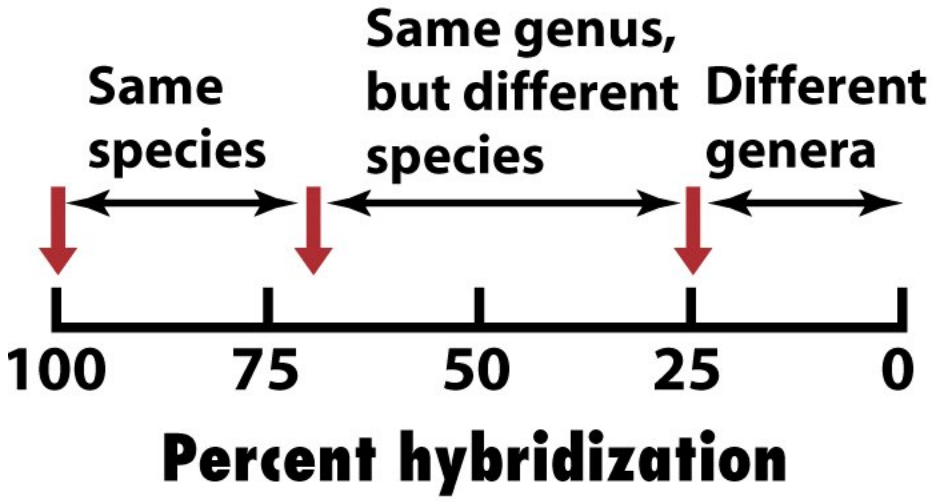


DNA:DNA hybridization



DNA:DNA hybridization

Results and interpretation:



1 x 1

100%

Same strain (control)

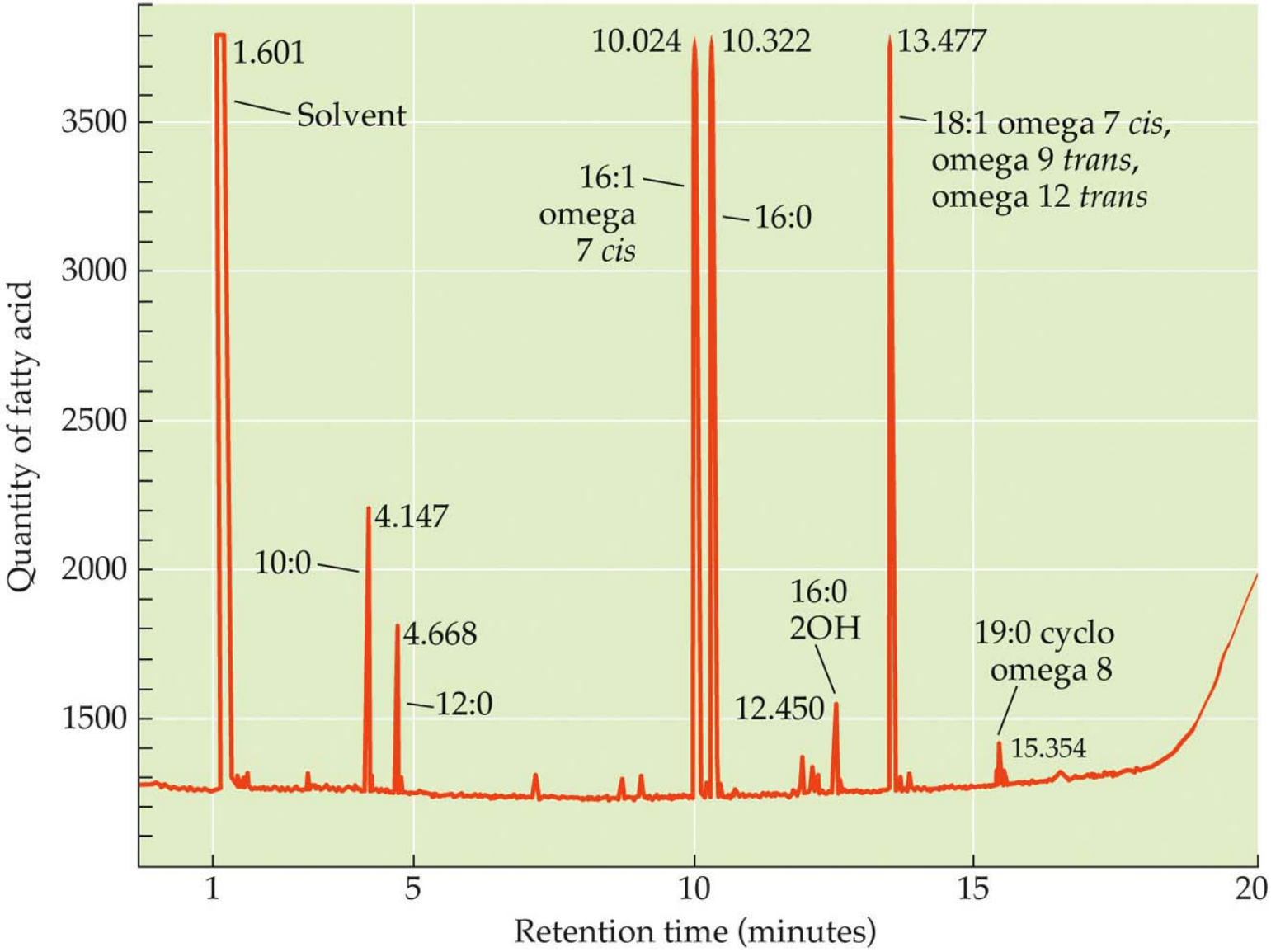
1 x 2

25%

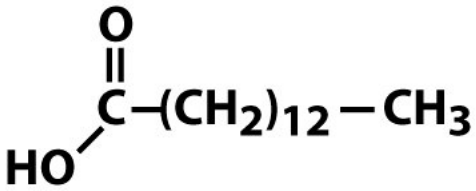
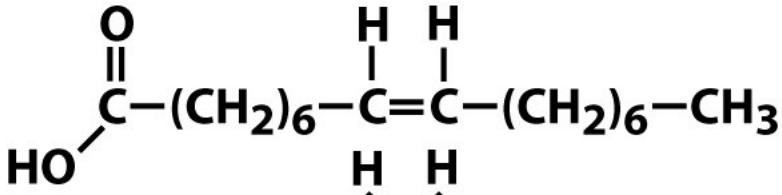
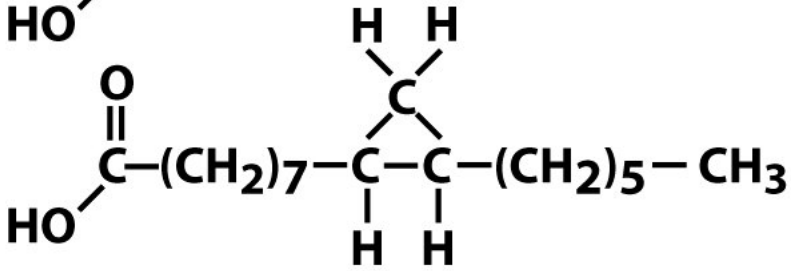
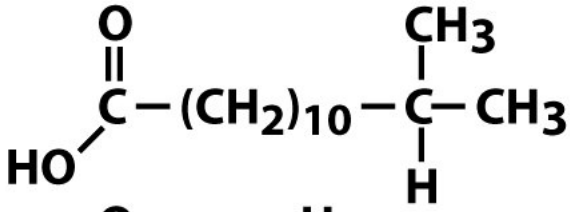
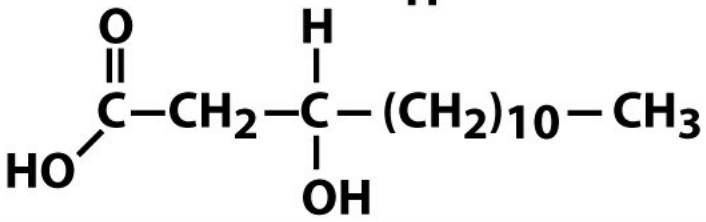
1 and 2 are likely different genera

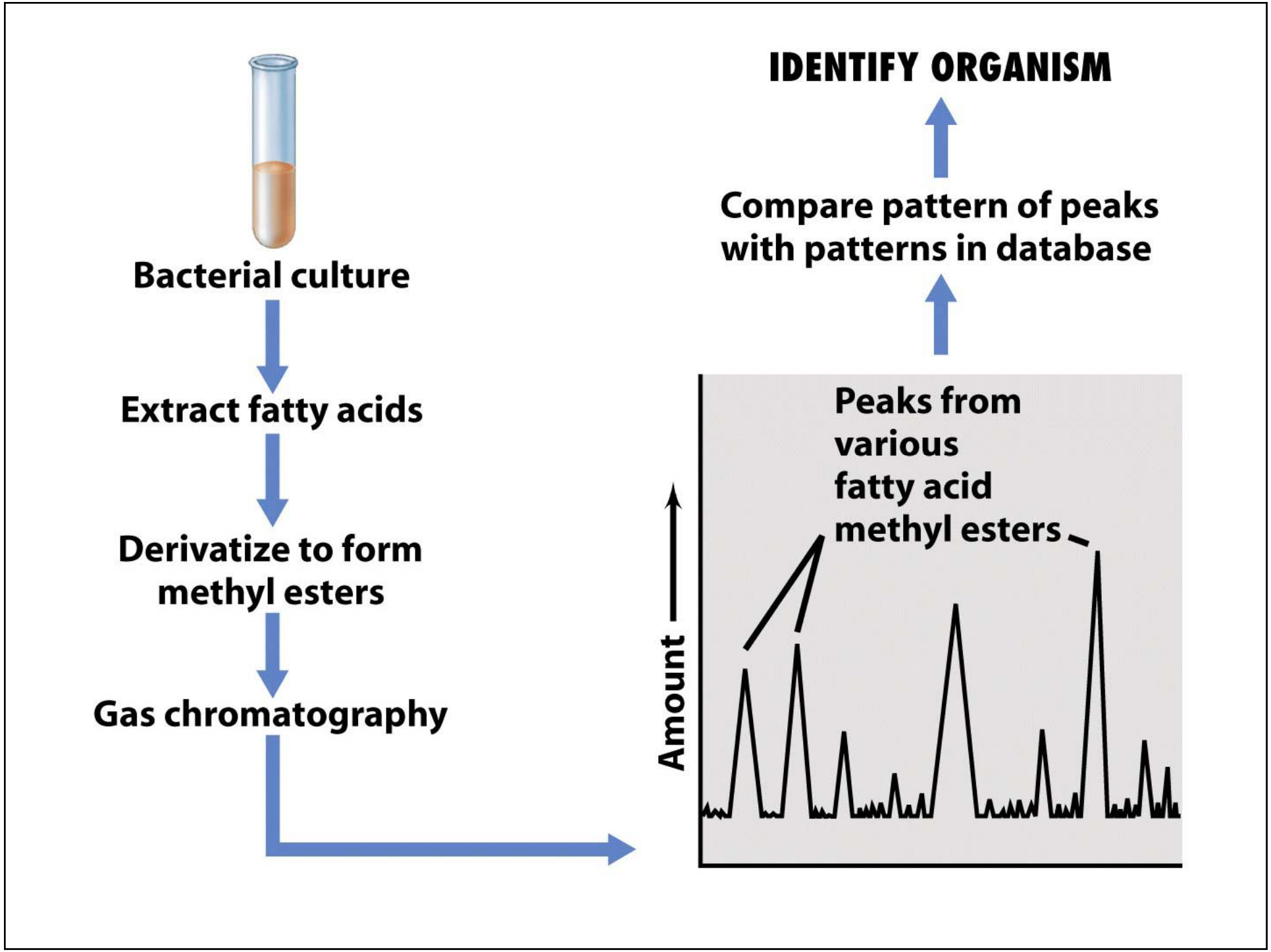
70% or greater; considered same species

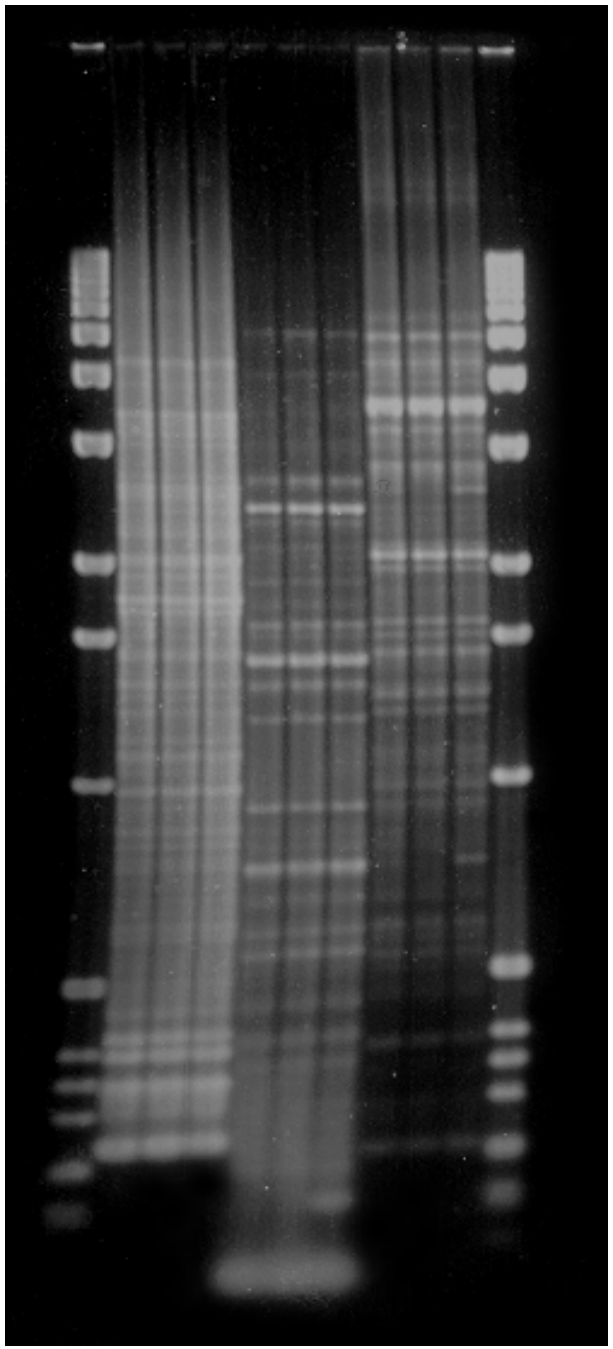
FAME analysis



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	


IDENTIFY ORGANISM
Compare pattern of peaks with patterns in database
Peaks from various fatty acid methyl esters
Amount
Bacterial culture
Extract fatty acids
Derivatize to form methyl esters
Gas chromatography



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 iron-oxidizing bacteria.

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

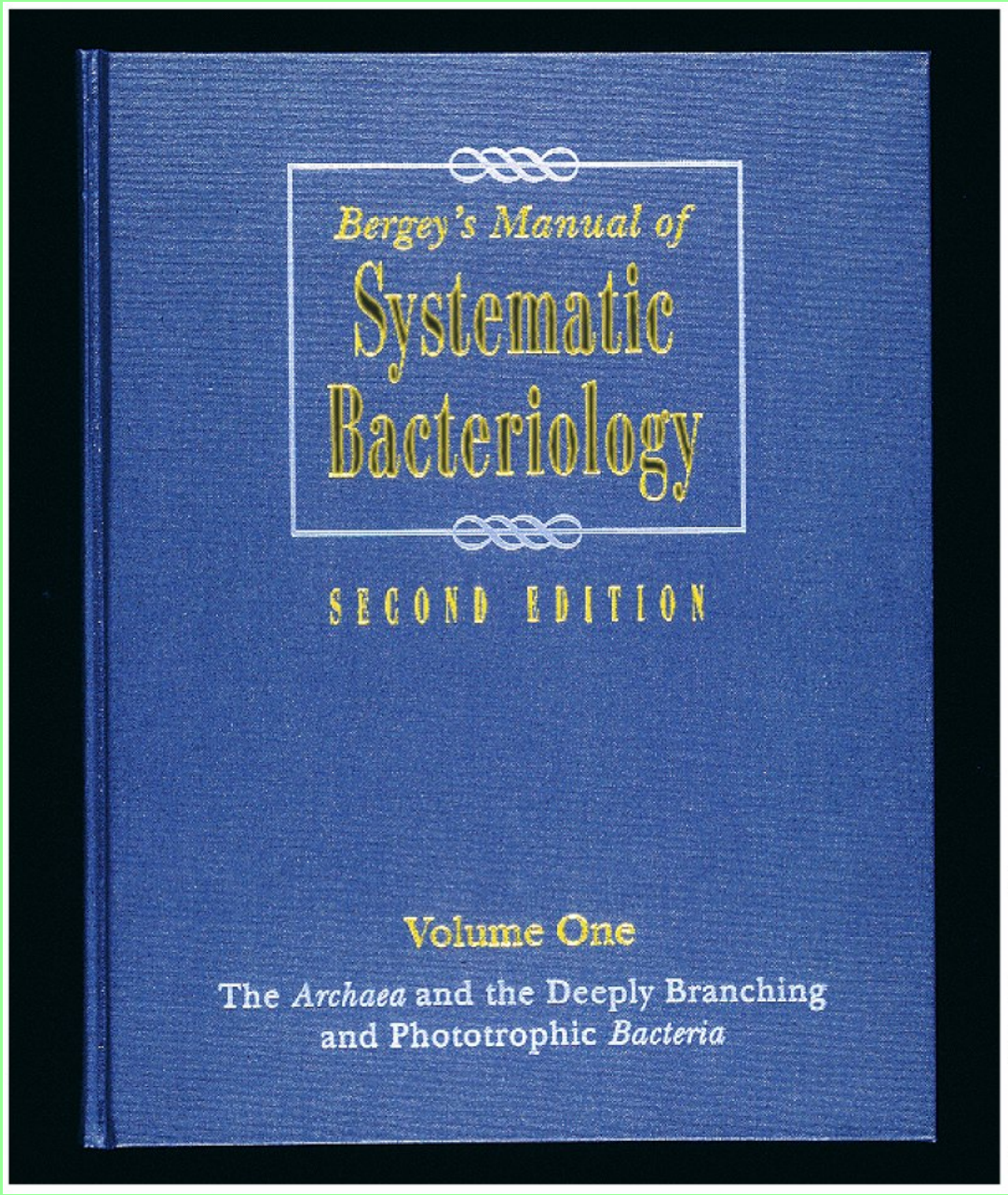
Taxon	Name
Domain	*Bacteria*
Phylum	*Spirochaetes* (vernacular name: spirochetes)
Class	*Spirochaetes*
Order	*Spirochaetales*
Family	*Spirochaetaceae*
Genus	*Spirochaeta*
Species	*plicatilis*

Table 11.6 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.


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

Taxonomy Summary

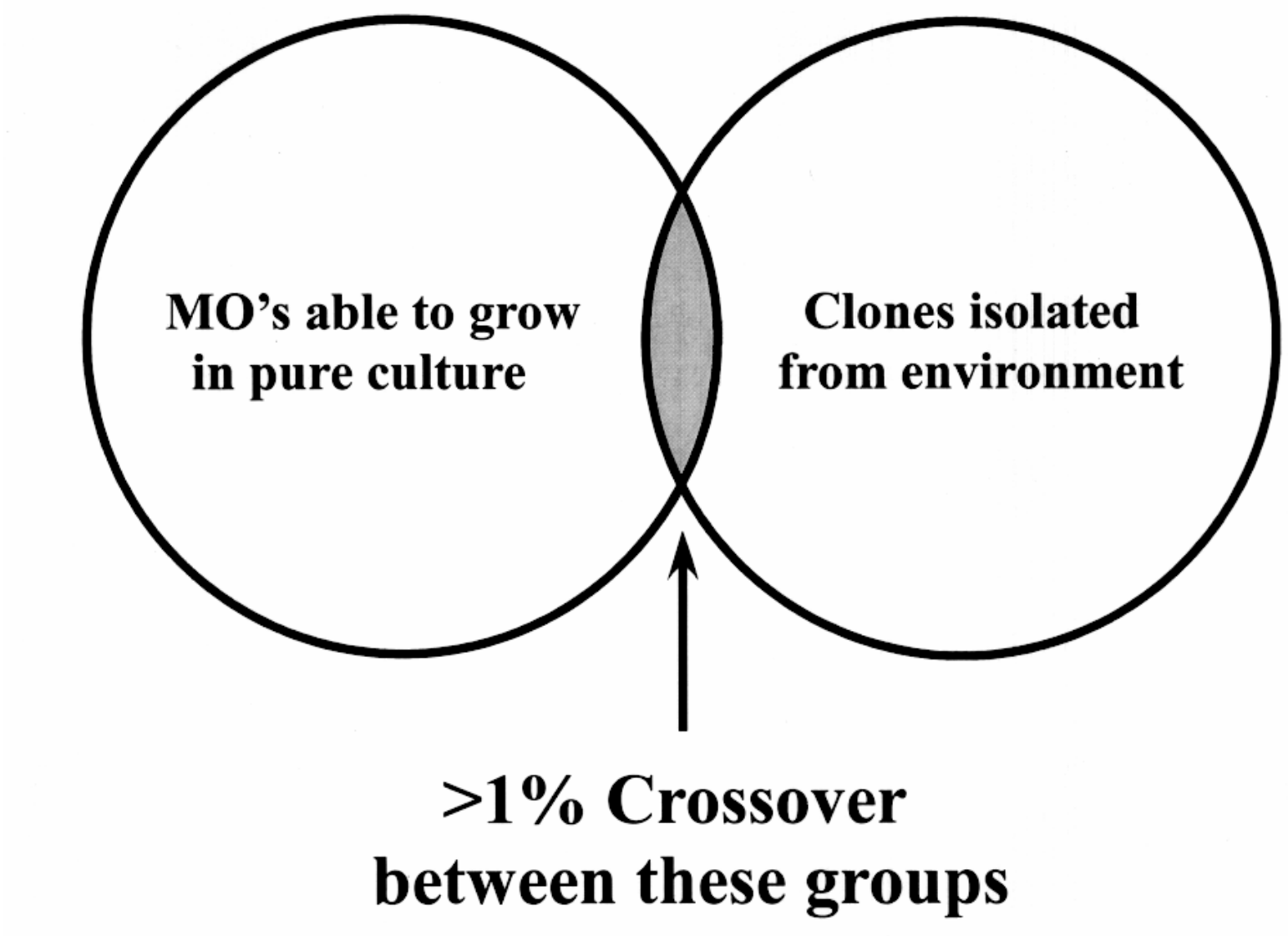
Classical physiological descriptions of microbes constitute a taxonomy, but do not provide relationships (except as might be inferred subjectively).

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.

Importance of a Molecular Biological Approach

- **Traditional culturing** techniques isolate ~1% of the total bacteria in marine ecosystems, thereby severely underestimating diversity and community structure.
- Because nutrient-rich **culture media** have been historically used during enrichment procedures, bacteria which may be dominant in natural communities are selected against in favor of copiotrophic (weedy) bacteria.
- **SSU rRNAs** and their respective genes are excellent descriptors of microbial taxa based on phylogeny.


MO's able to grow
in pure culture
Clones isolated
from environment
>1% Crossover
between these groups

Regarding Molecular Phylogeny

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

- Milestone #1: Zuckerkandl and Pauling (1965) “Semantides” (i.e., molecules as documents of evolutionary history).

- Milestone #2: Pace (1986) Applied phylogeny concept to microbial ecology's need to take a census.

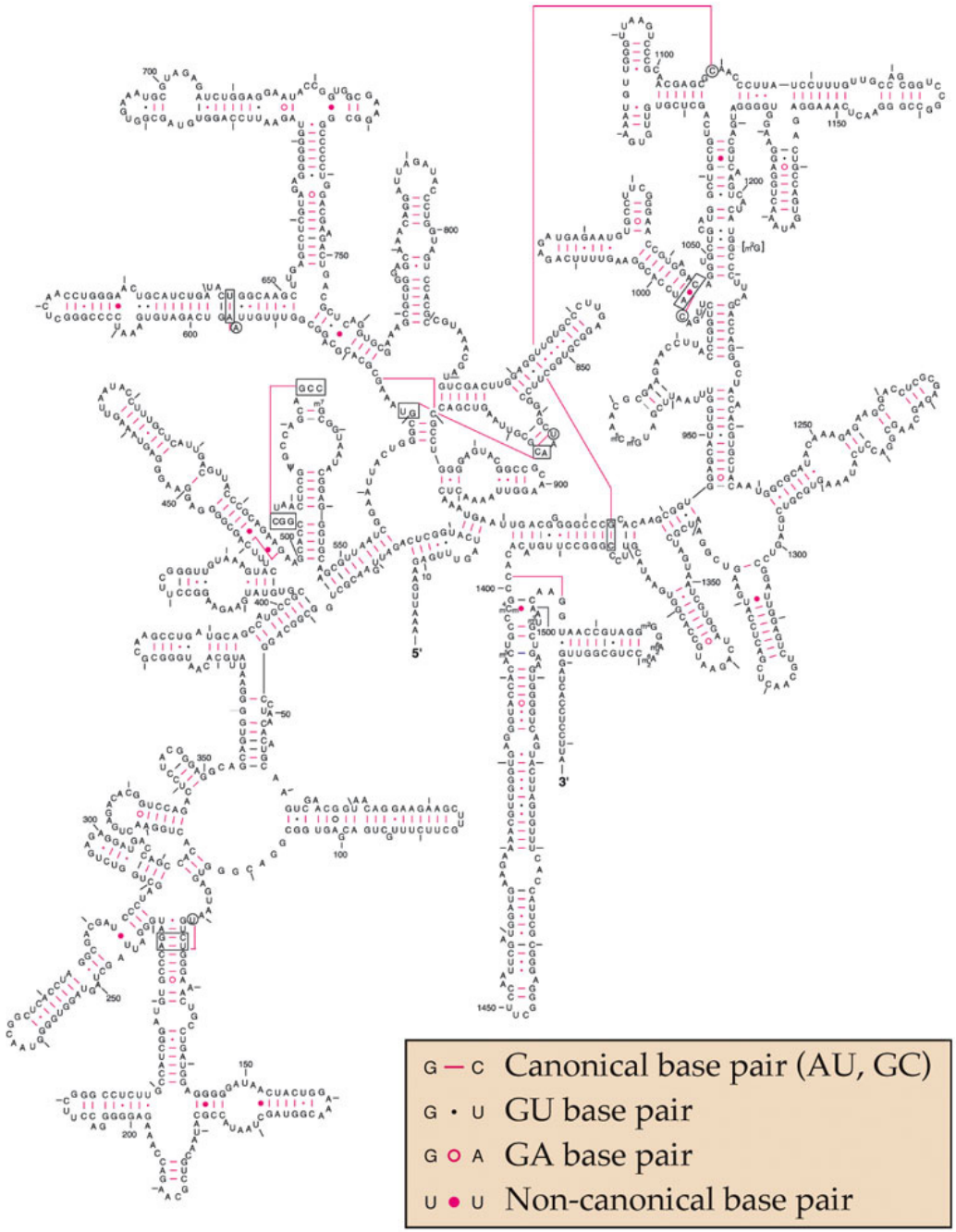
- Milestone #3: Woese (1987) Applied phylogeny concept to redefine microbial systematics or the need to understand microbial genealogy.

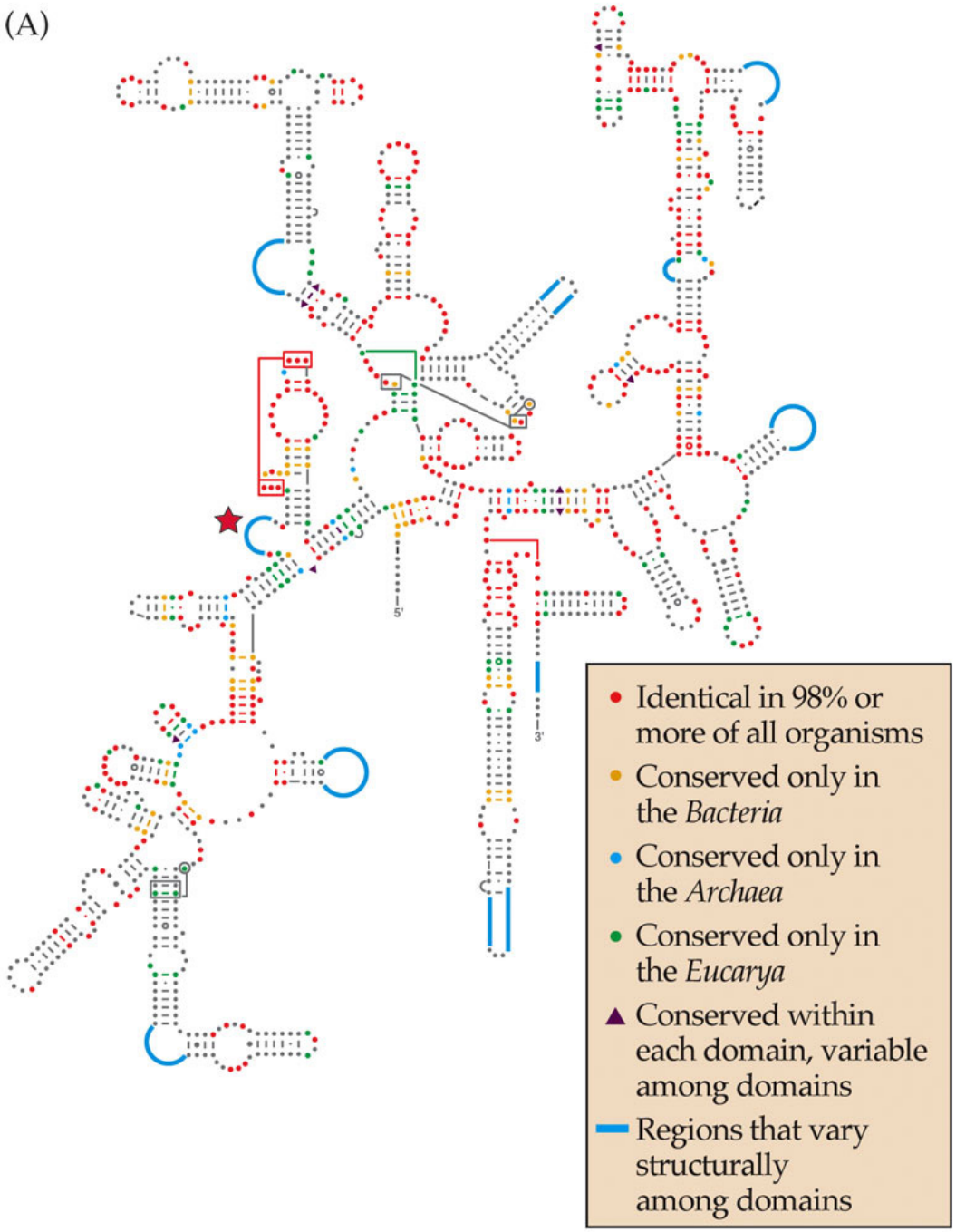
... the general course of evolution [for bacteria] will probably never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds... For these and other reasons, most modern taxonomists have explicitly abandoned the phylogenetic approach.

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

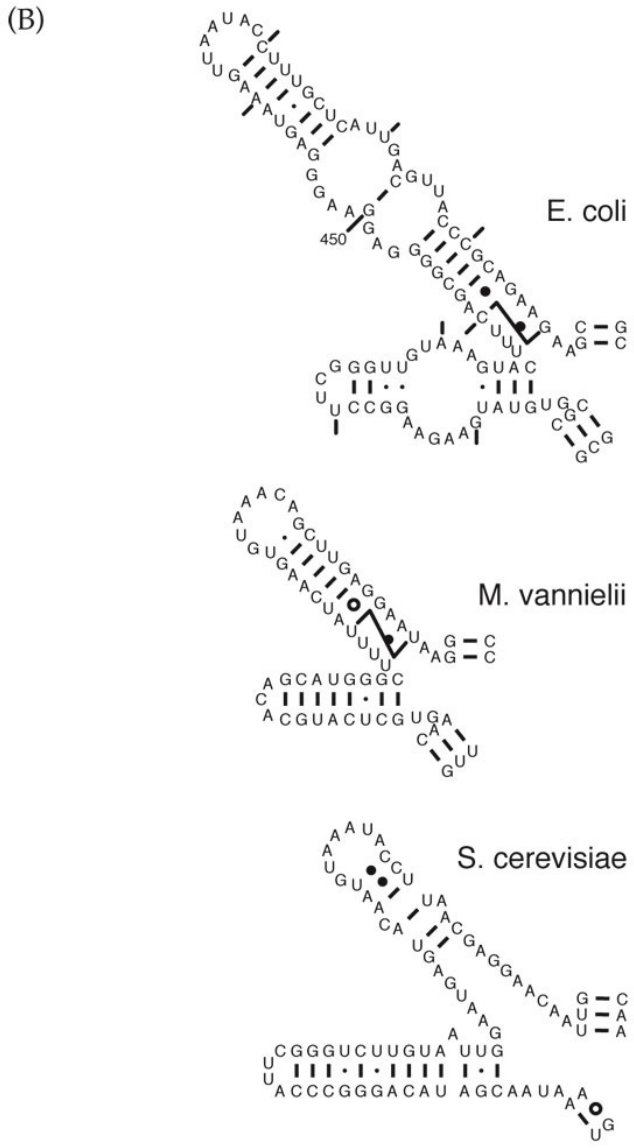
Why ribosomal RNAs?

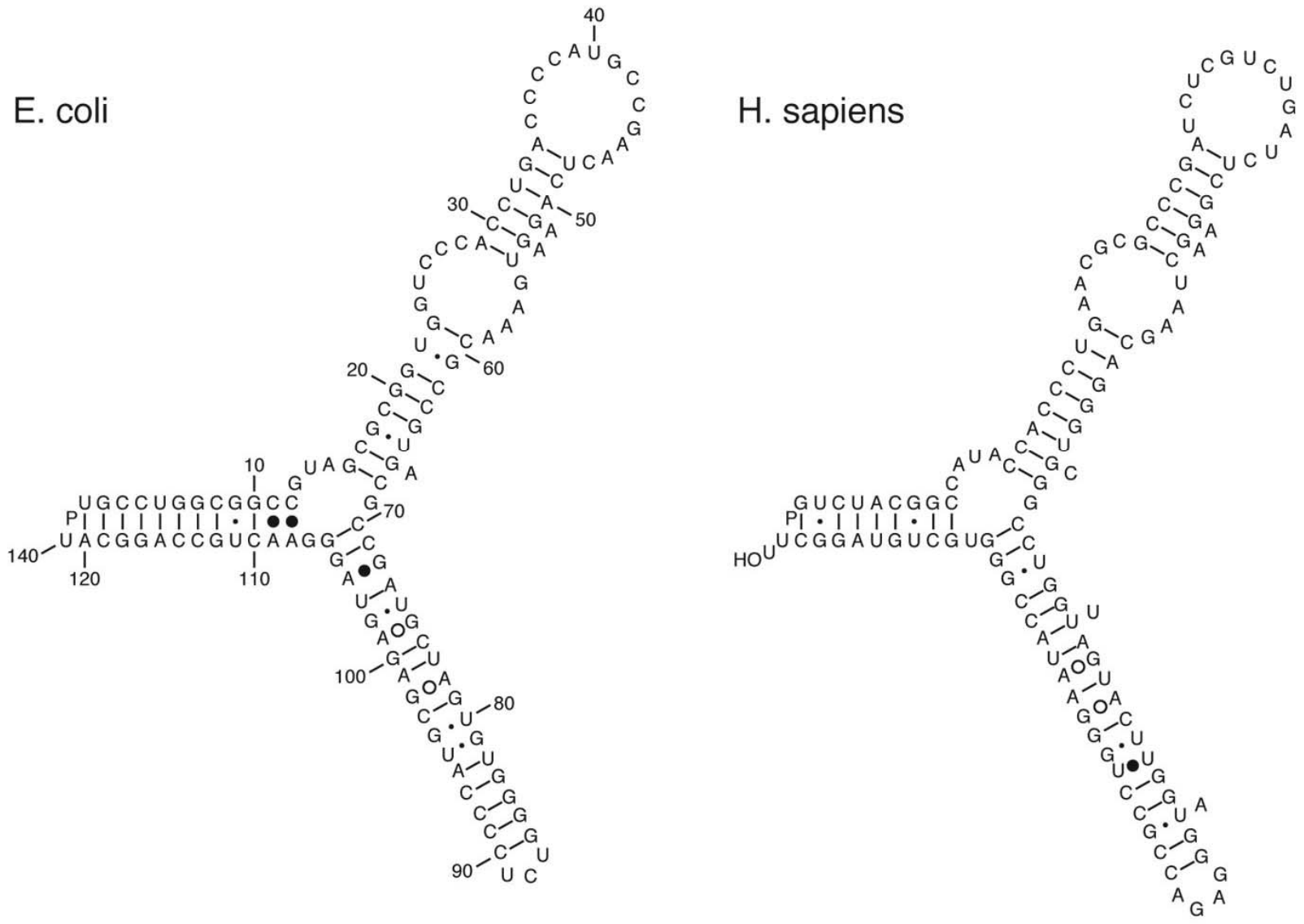
- Found among all living organisms (for 3.8 of the last 4.5 billion years). Integral part of protein synthesis machinery.
- Cell component analyses provide culture-independent means of investigating questions in microbial ecology (lack of morphology).
- rRNAs offer a type of sequence information that 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.


G — C Canonical base pair (AU, GC)
G • U GU base pair
G o 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

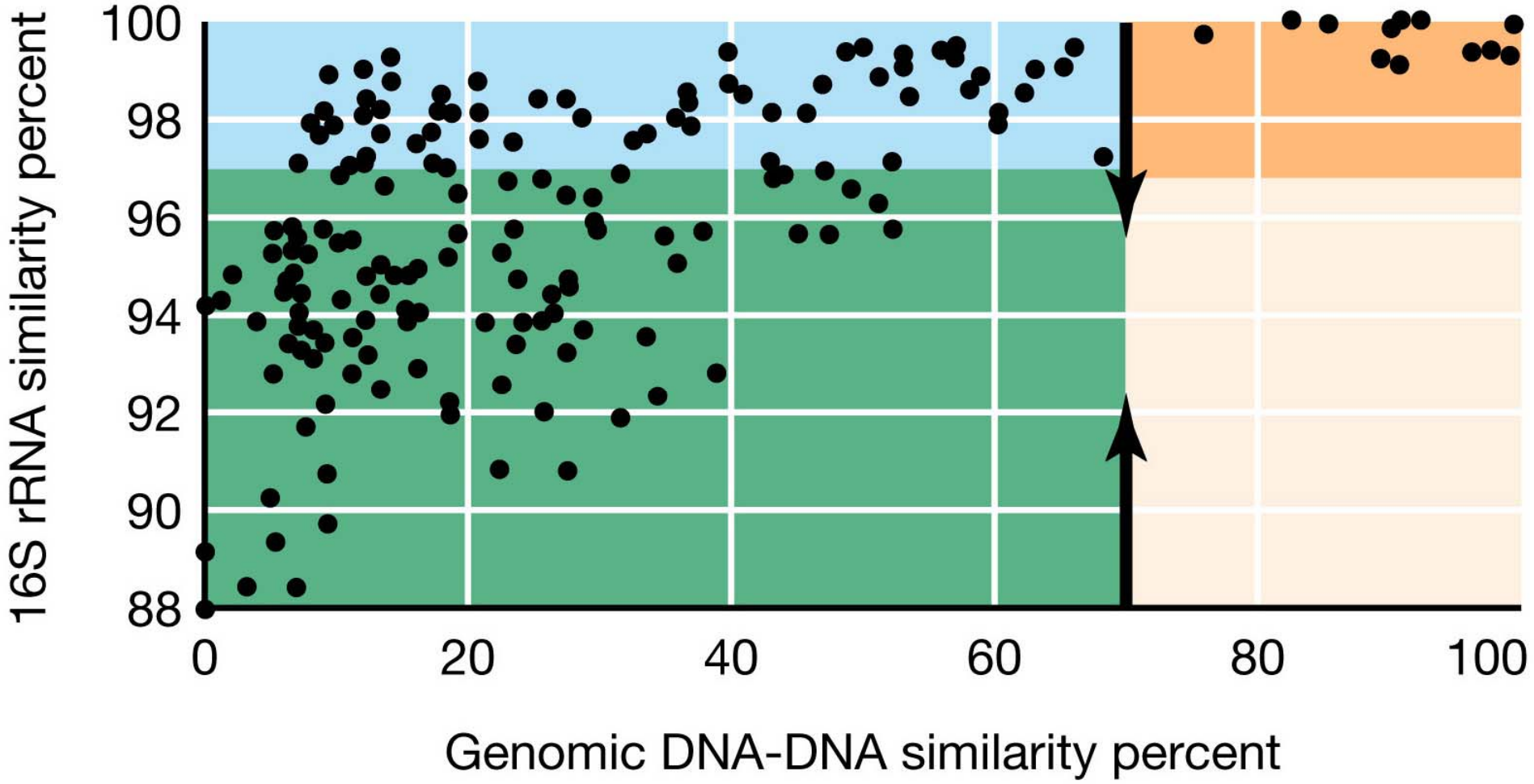
Secondary Structures of SSU rRNA show homology



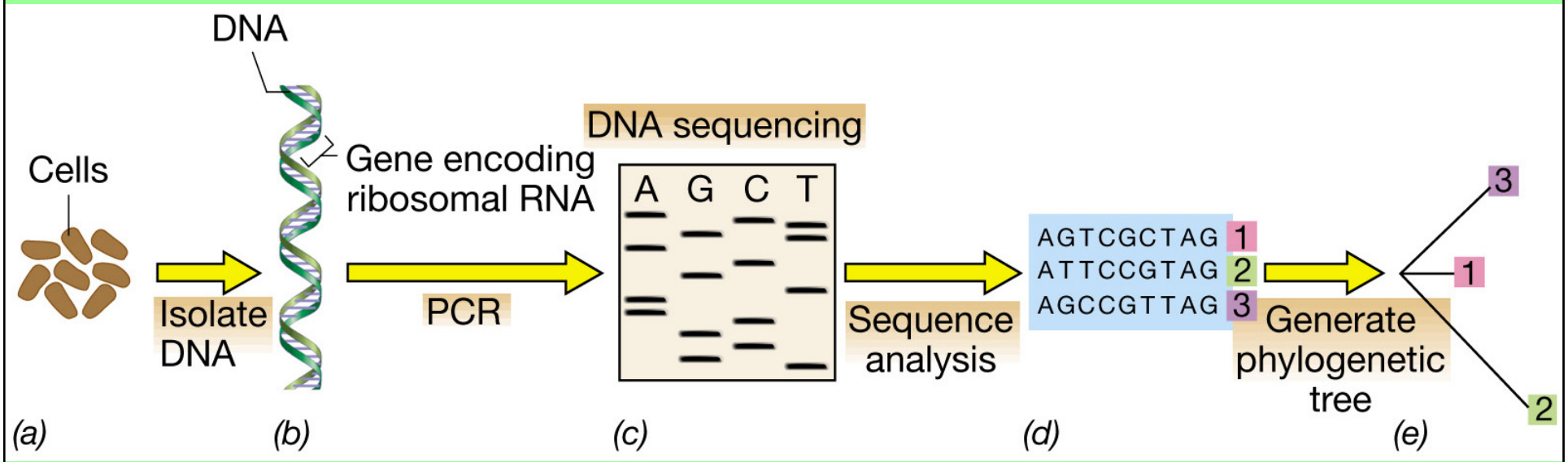


Secondary Structures of rRNAs show homology

Relationship between SSU rDNA and genomic DNA hybridization



Molecular Strategy Flow Chart



Note: Independent of pure culture isolation!

Estimating evolutionary distance E_D to map on phylogenetic tree

Organism	Sequence
A	CGUAGACCUGAC
B	CCUAGAGCUGGC
C	CCAAGACGUGGC
D	GCUAGAUGUGCC

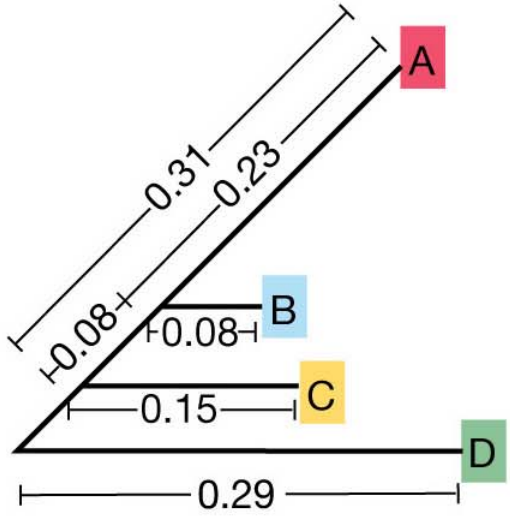
Analysis

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

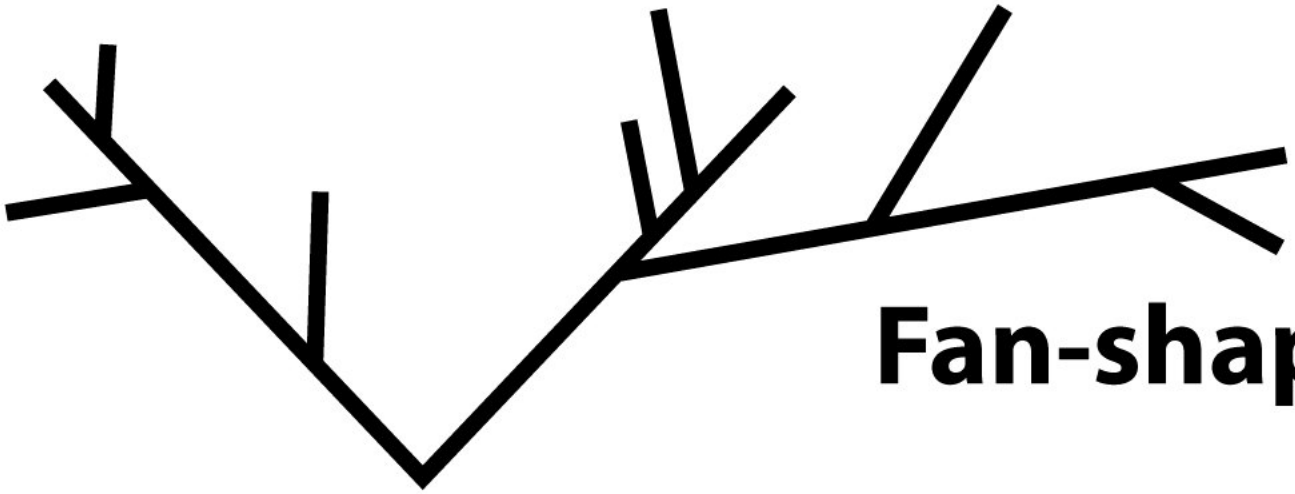
(a) **Sequence alignment and analysis**

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

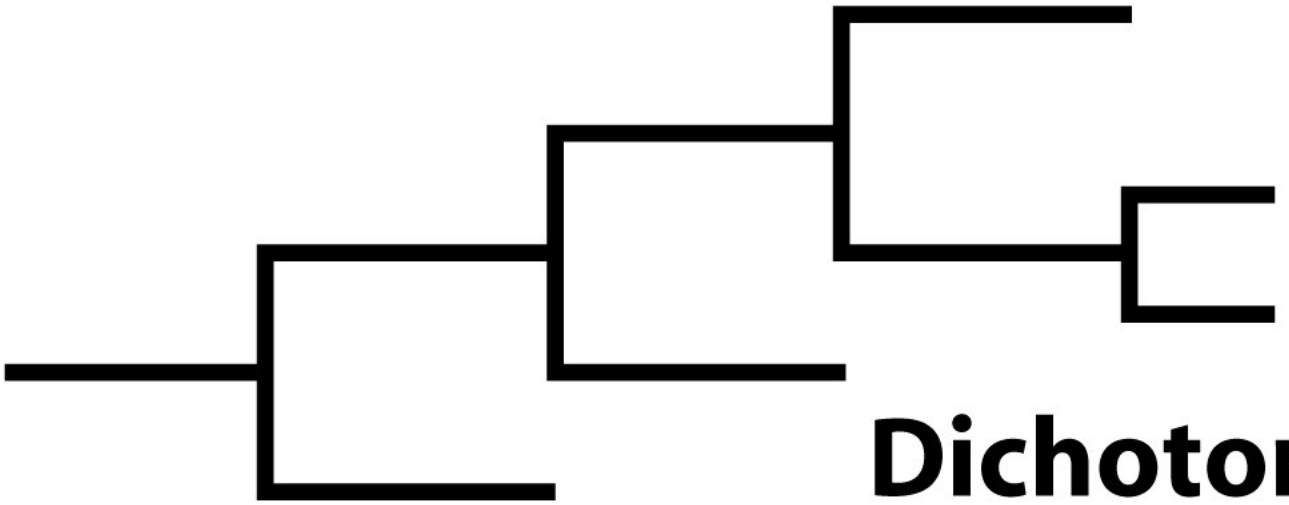
(b) **Calculation of evolutionary distance**



(c) **Phylogenetic tree**

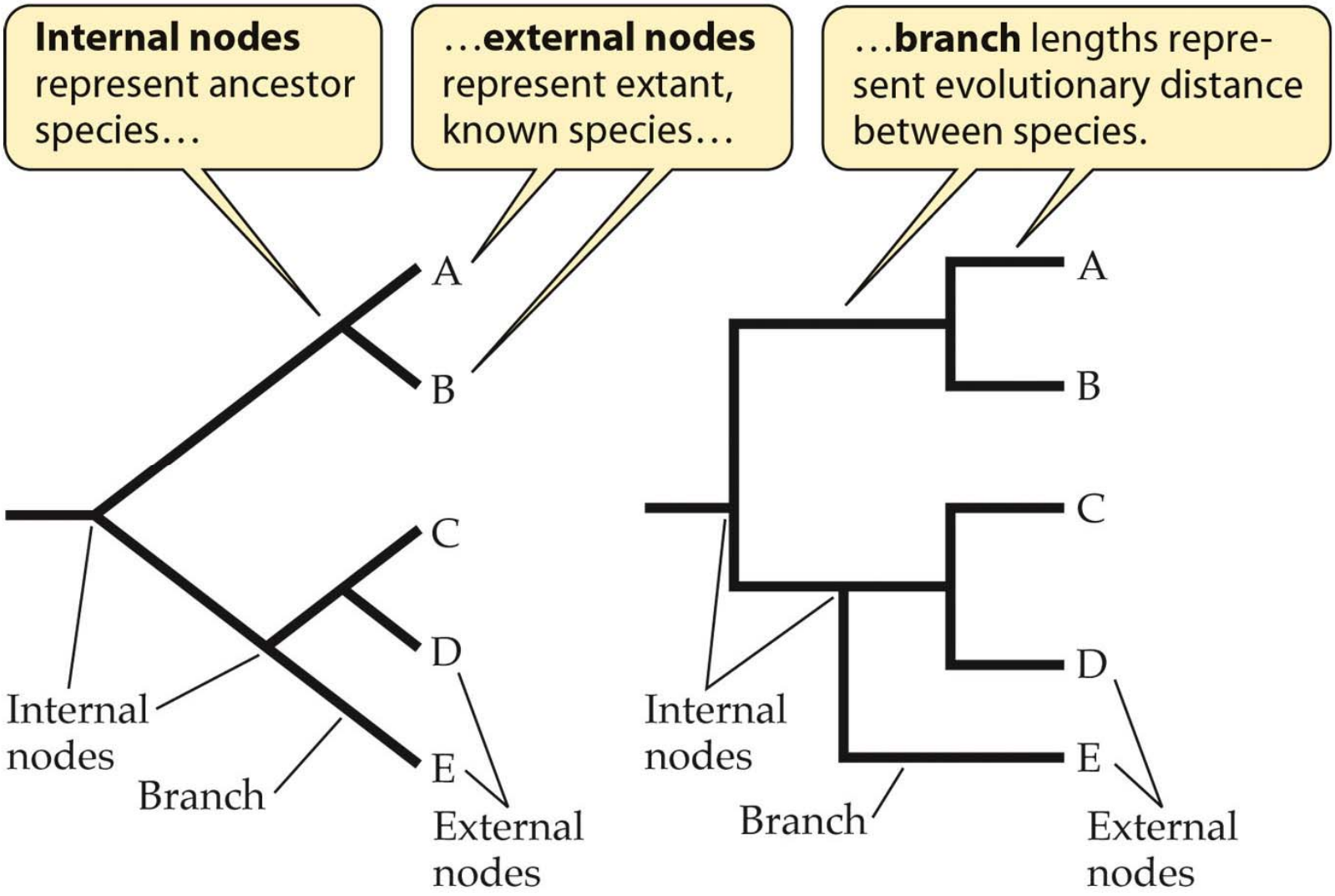


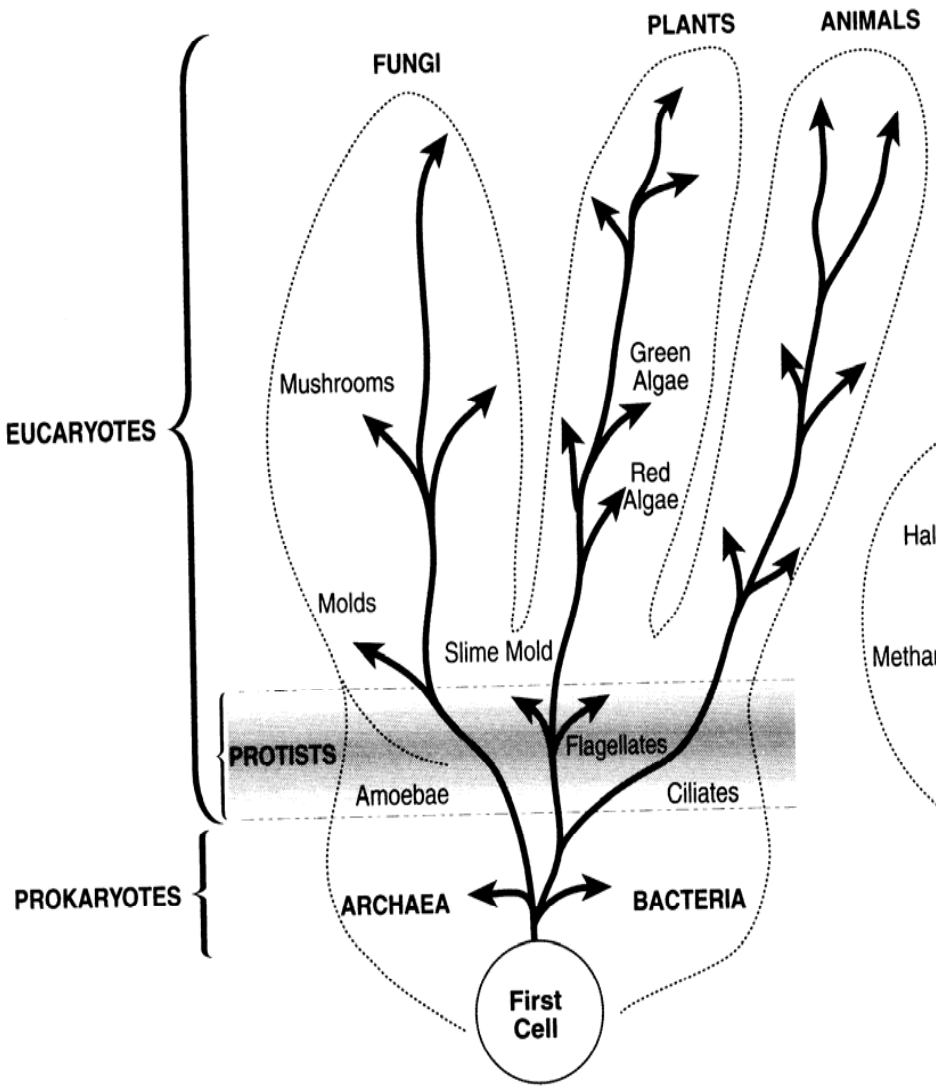
Fan-shaped



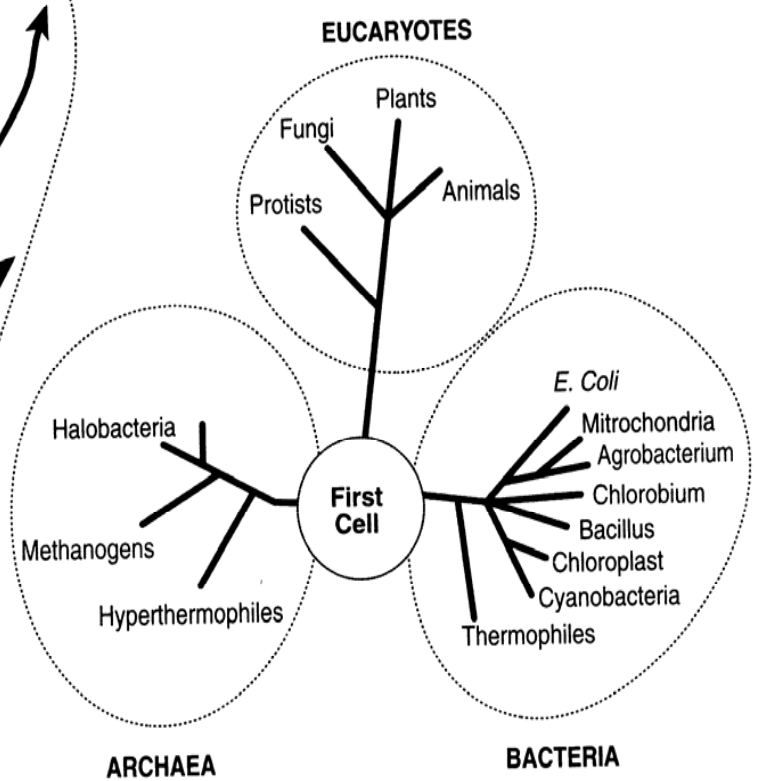
Dichotomous

Tree topologies


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



"CLASSICAL" PHYLOGENY



MOLECULAR PHYLOGENY

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*

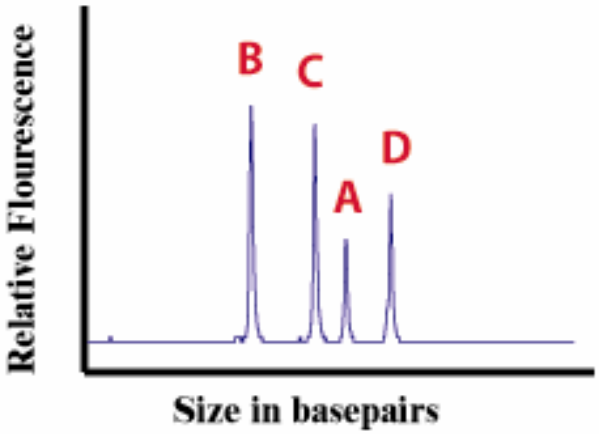
A CGGTACCGGGCAG E CTGAATTACG

B CGGAG F CTCTAAGCTAACCGGACG

C CGTCCGGCAG G CTCTAACGTTACG

D CGGCCGGCTAAGCAG H CTGTTACG

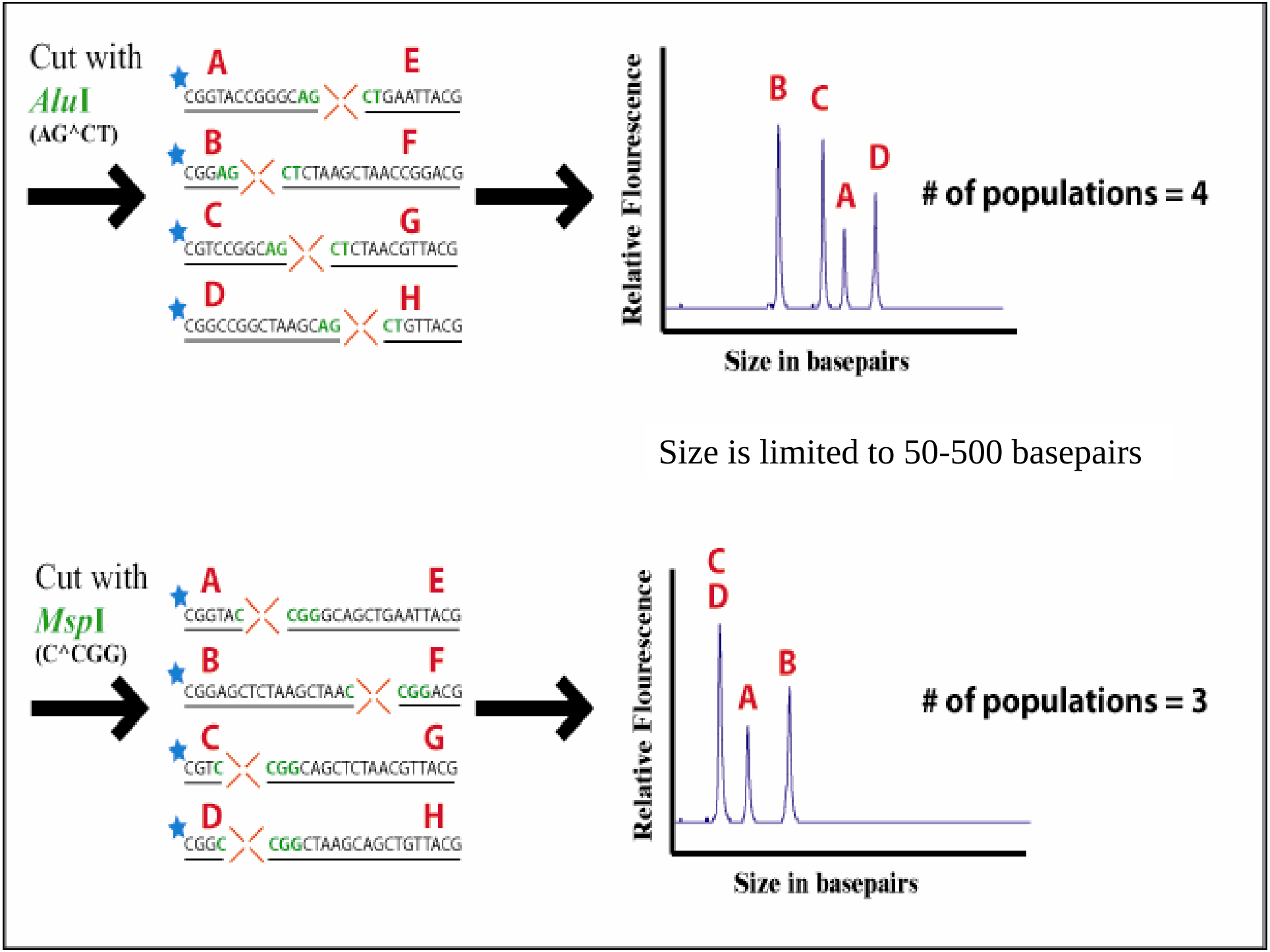
LIF-CE



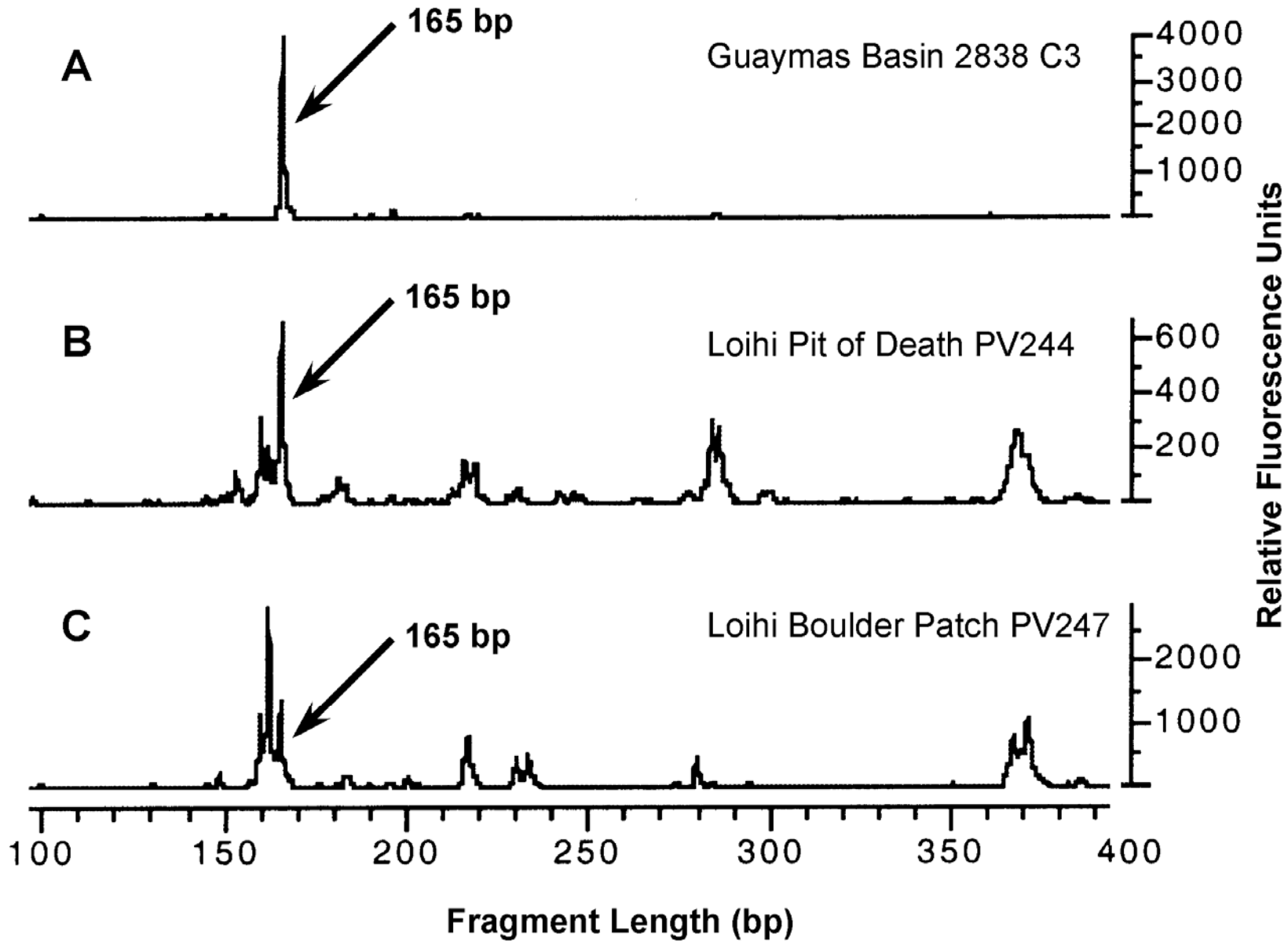
Community of RFs

Separated fragments

Chromatogram of peak heights


Cut with
AluI
(AG^CT)
A
CGGTACCGGGCAG
E
CTGAATTACG
B
CGGAG
F
CTCTAAGCTAACCGGACG
C
CGTCCGGCAG
G
CTCTAACGTTACG
D
CGGCCGGCTAAGCAG
H
CTGTTACG
Relative Flourescence
B
C
A
D
# of populations = 4
Size in basepairs
Size is limited to 50-500 basepairs
Cut with
MspI
(C^CGG)
A
CGGTAC
E
CGGGCAGCTGAATTACG
B
CGGAGCTCTAAGCTAAC
F
CGGACG
C
CGTC
G
CGGCAGCTCTAACGTTACG
D
CGGC
H
CGGCTAAGCAGCTGTTACG
Relative Flourescence
C
D
A
B
# of populations = 3
Size in basepairs

T-RFLP profiles from Iron-rich Hydrothermal Vents



Cluster Analysis of T-RFLP Data

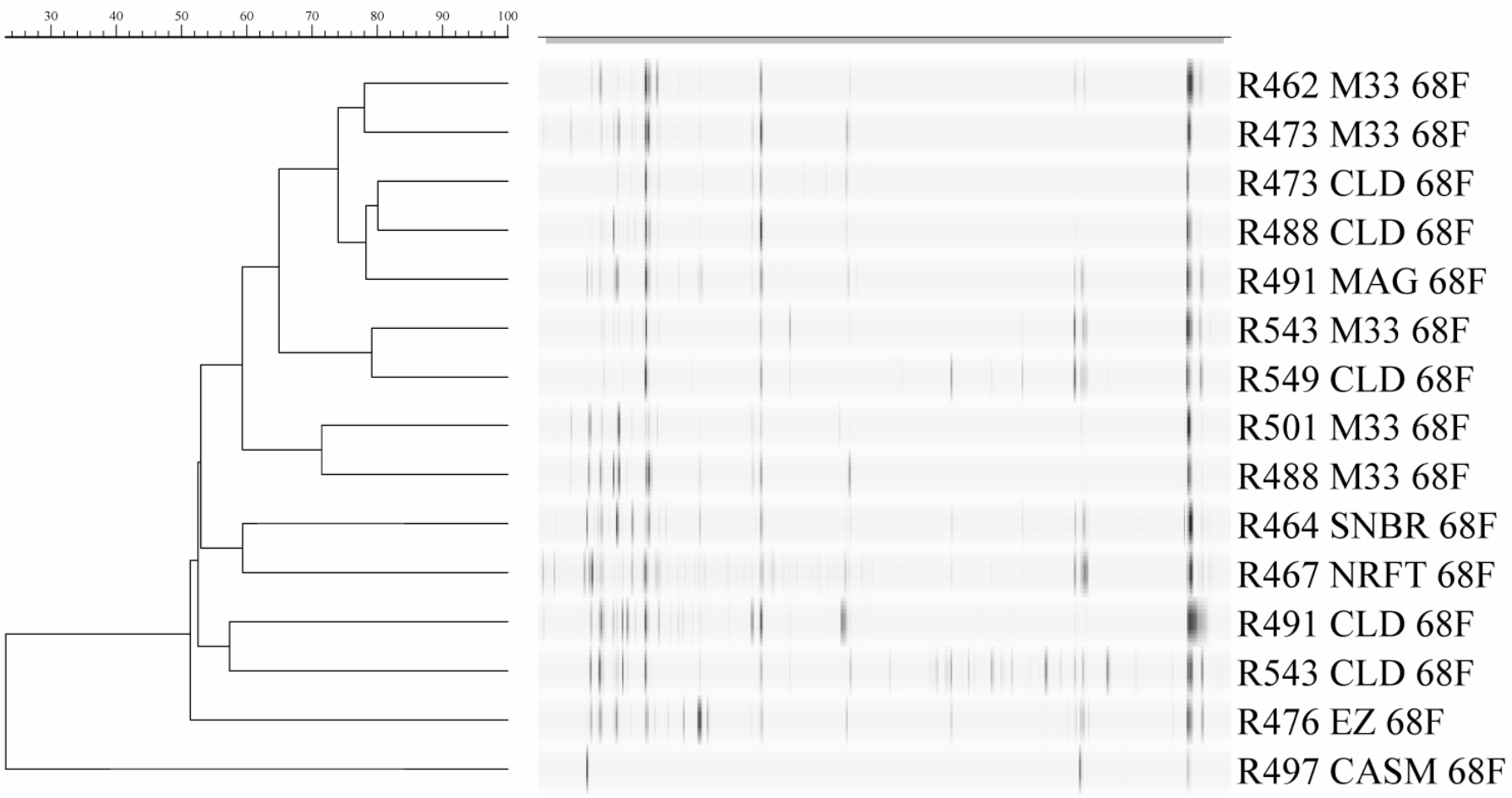


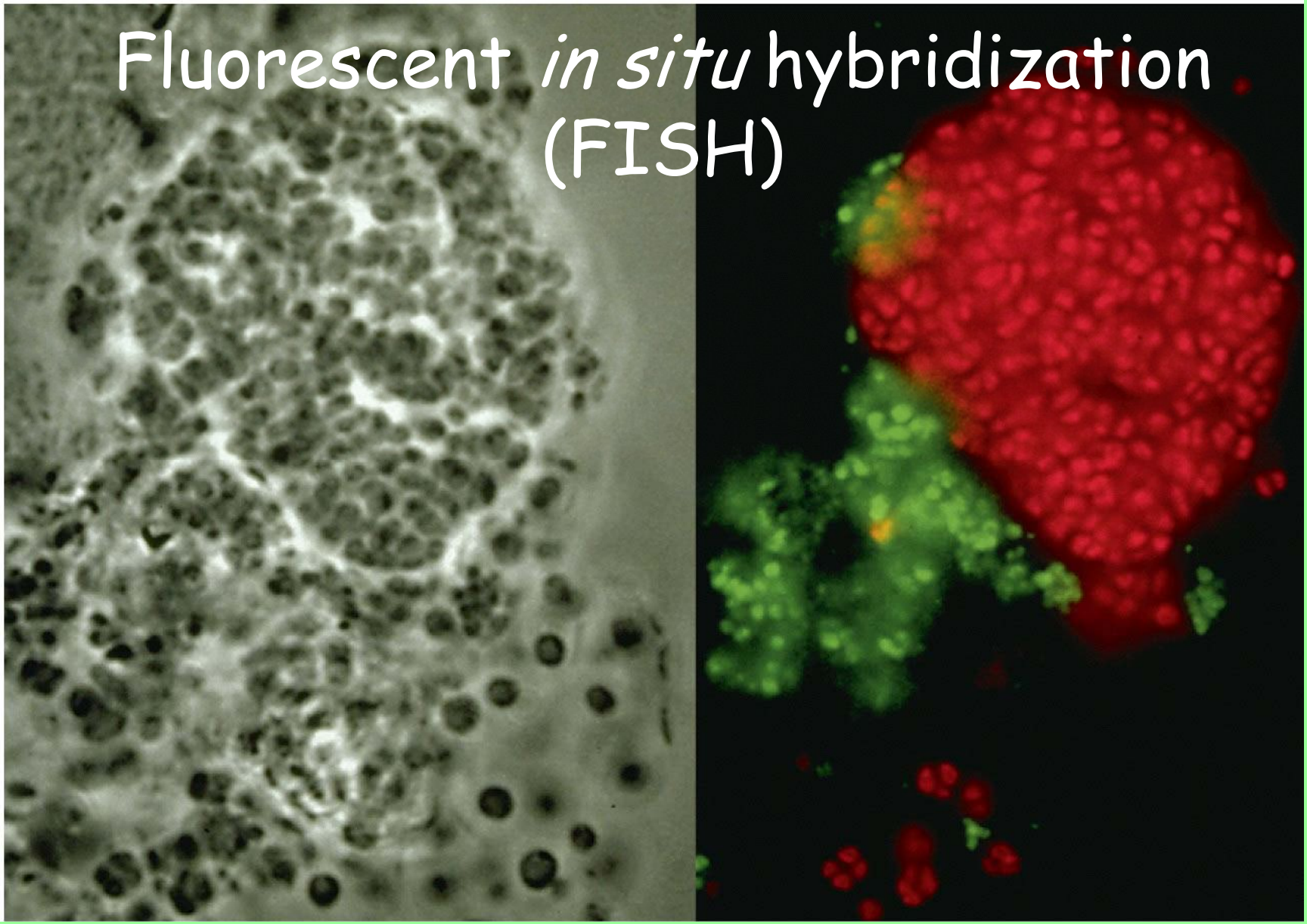
Table 11.1 **Signature sequences from 16S or 18S rRNA defining the three domains of life**

		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.

Signature sequences can be obtained at any level of taxonomic hierarchy

Fluorescent *in situ* hybridization (FISH)



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.

  - microbial ecology relationships can now be truly examined
  - relationships between MOs and their genes can be studied
  - infer dynamics of sequence change (Rolex vs Timex)

Table 11.3 Summary of major differential features among *Bacteria*, *Archaea*, and *Eukarya*[a]

Characteristic	*Bacteria*	*Archaea*	*Eukarya*
Morphological and Genetic			
Prokaryotic cell structure	Yes	Yes	No
DNA present in covalently closed and circular form	Yes	Yes	No
Histone proteins present	No	Yes	Yes
Membrane-enclosed nucleus	Absent	Absent	Present
Cell wall	Muramic acid present	Muramic acid absent	Muramic acid absent
Membrane lipids	Ester-linked	Ether-linked	Ester-linked
Ribosomes (mass)	70S	70S	80S
Initiator tRNA	Formylmethionine	Methionine	Methionine
Introns in most genes	No	No	Yes
Operons	Yes	Yes	No
Capping and poly-A tailing of mRNA	No	No	Yes
Plasmids	Yes	Yes	Rare
Ribosome sensitivity to diphtheria toxin	No	Yes	Yes
RNA polymerases (see Figure 11.19)	One (4 subunits)	Several (8–12 subunits each)	Three (12–14 subunits each)
Transcription factors required (Section 7.11)	No	Yes	Yes
Promoter structure (Sections 7.10 and 7.11)	−10 and −35 sequences (Pribnow box)	TATA box	TATA box
Sensitivity to chloramphenicol, streptomycin, and kanamycin	Yes	No	No

[a] Note that for many features only particular representatives within a domain show the property.

[b] Environmental genomics studies of prokaryotes in marine waters strongly suggest that nitrifying *Archaea* exist (Section 18.6).

Table 11.3 Summary of major differential features among *Bacteria*, *Archaea*, and *Eukarya*[a]

Characteristic	*Bacteria*	*Archaea*	*Eukarya*
Physiological/Special Structures			
Methanogenesis	No	Yes	No
Dissimilative reduction of S^0 or SO_4^{2-} to H_2S, or Fe^{3+} to Fe^{2+}	Yes	Yes	No
Nitrification	Yes	No[b]	No
Denitrification	Yes	Yes	No
Nitrogen fixation	Yes	Yes	No
Chlorophyll-based photosynthesis	Yes	No	Yes (in chloroplasts)
Rhodopsin-based energy metabolism	Yes	Yes	No
Chemolithotrophy (Fe, S, H_2)	Yes	Yes	No
Gas vesicles	Yes	Yes	No
Synthesis of carbon storage granules composed of poly-β-hydroxyalkanoates	Yes	Yes	No
Growth above 80° C	Yes	Yes	No
Growth above 100°C	No	Yes	No

[a] Note that for many features only particular representatives within a domain show the property.

[b] Environmental genomics studies of prokaryotes in marine waters strongly suggest that nitrifying *Archaea* exist (Section 18.6).

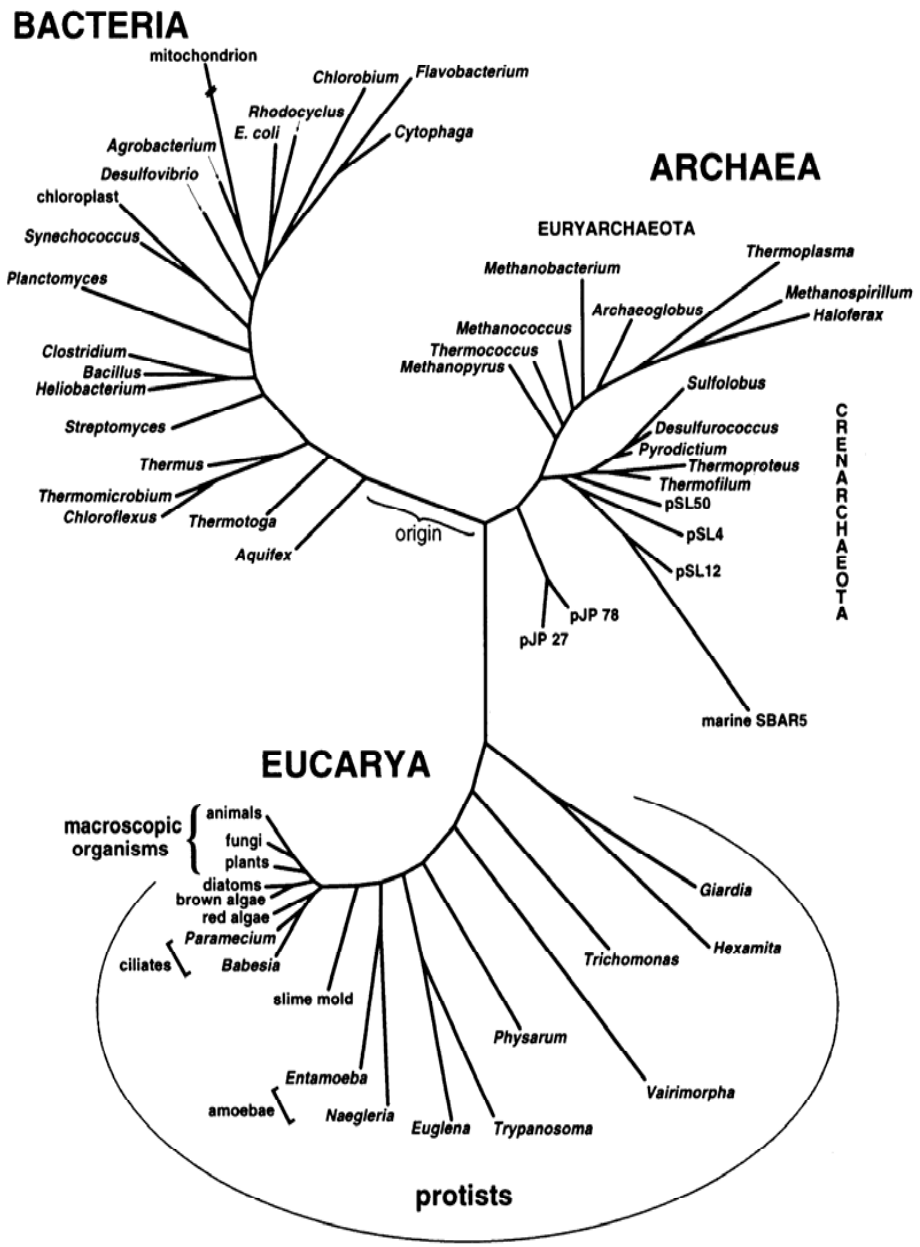


Figure 1. Diagrammatic "Universal" phylogenetic tree of life, based on small-subunit ribosomal RNA sequences. Based on analyses of Barns et al. (1996b), Olsen et al. (1994), and Sogin (1994).

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

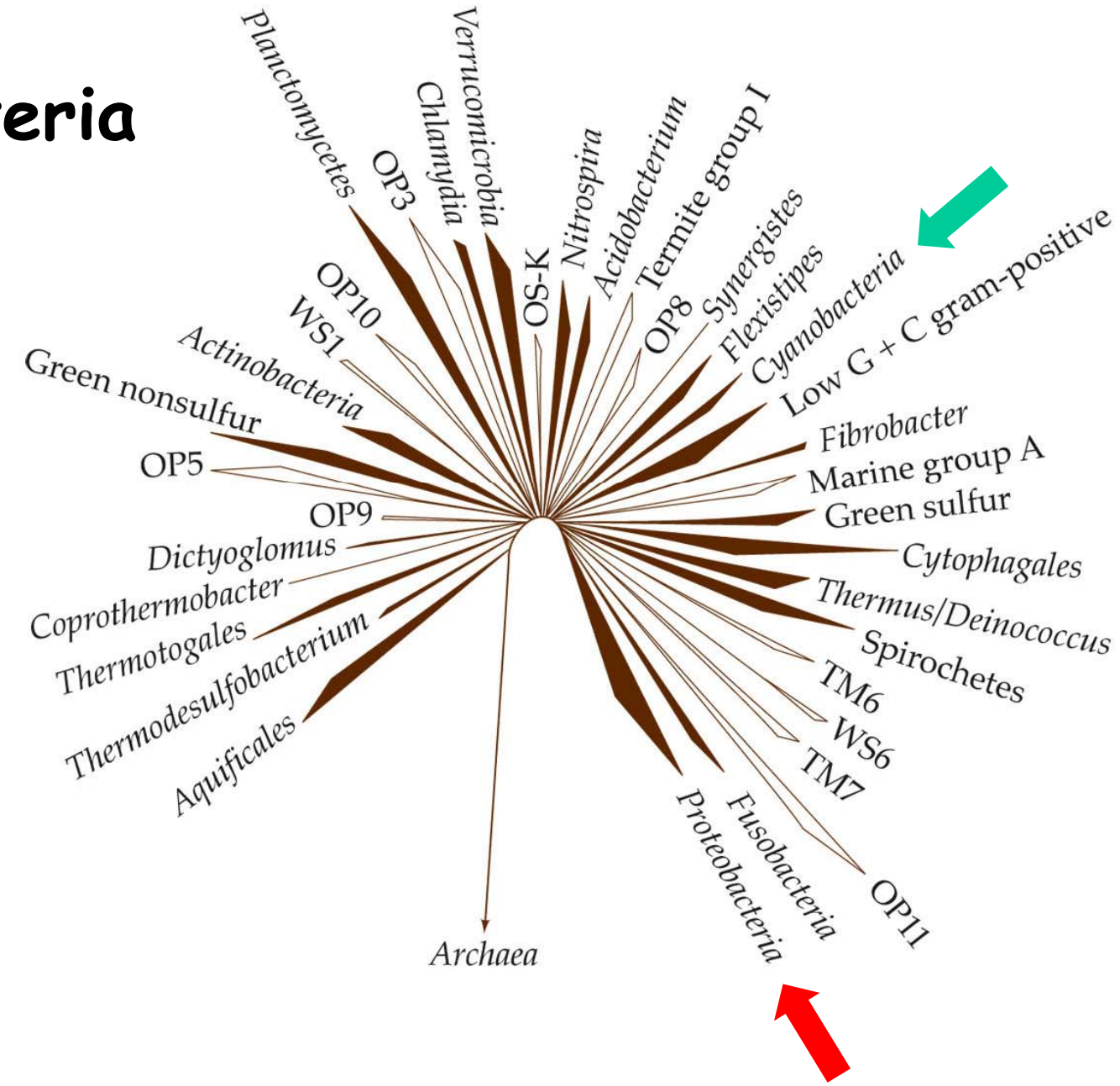
Single origin for all life on Earth...

- 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 Eucarya "*nuclear*" lineage almost as old as other two.
- Prokaryotes split between *Bacteria* and *Archaea*.
- Shown for only a limited number of representative org's.
- Mitochondria and chloroplasts proven to be of bacterial origin.

Bacteria


Planctomycetes
OP3
Chlamydia
Verrucomicrobia
OS-K
Nitrospira
Acidobacterium
Termite group I
OP8
Synergistes
Flexistipes
Cyanobacteria
Low G + C gram-positive
Fibrobacter
Marine group A
Green sulfur
Cytophagales
Thermus/Deinococcus
Spirochetes
TM6
WS6
TM7
OP11
Fusobacteria
Proteobacteria
Archaea
OP10
WS1
Actinobacteria
Green nonsulfur
OP5
OP9
Dictyoglomus
Coprothermobacter
Thermotogales
Thermodesulfobacterium
Aquificales

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

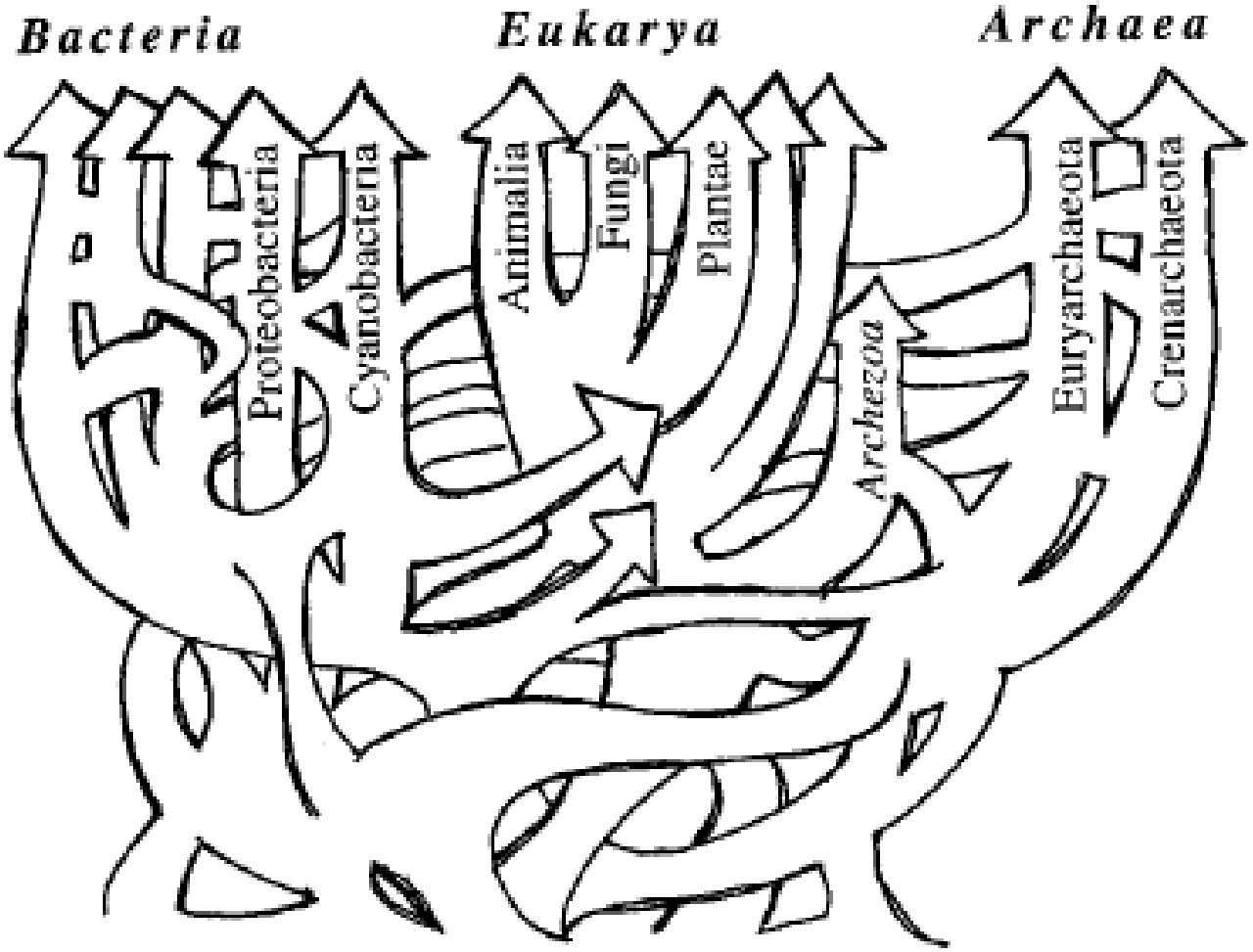
Horizontal gene transfer

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

conjugation, transduction, and transformation.

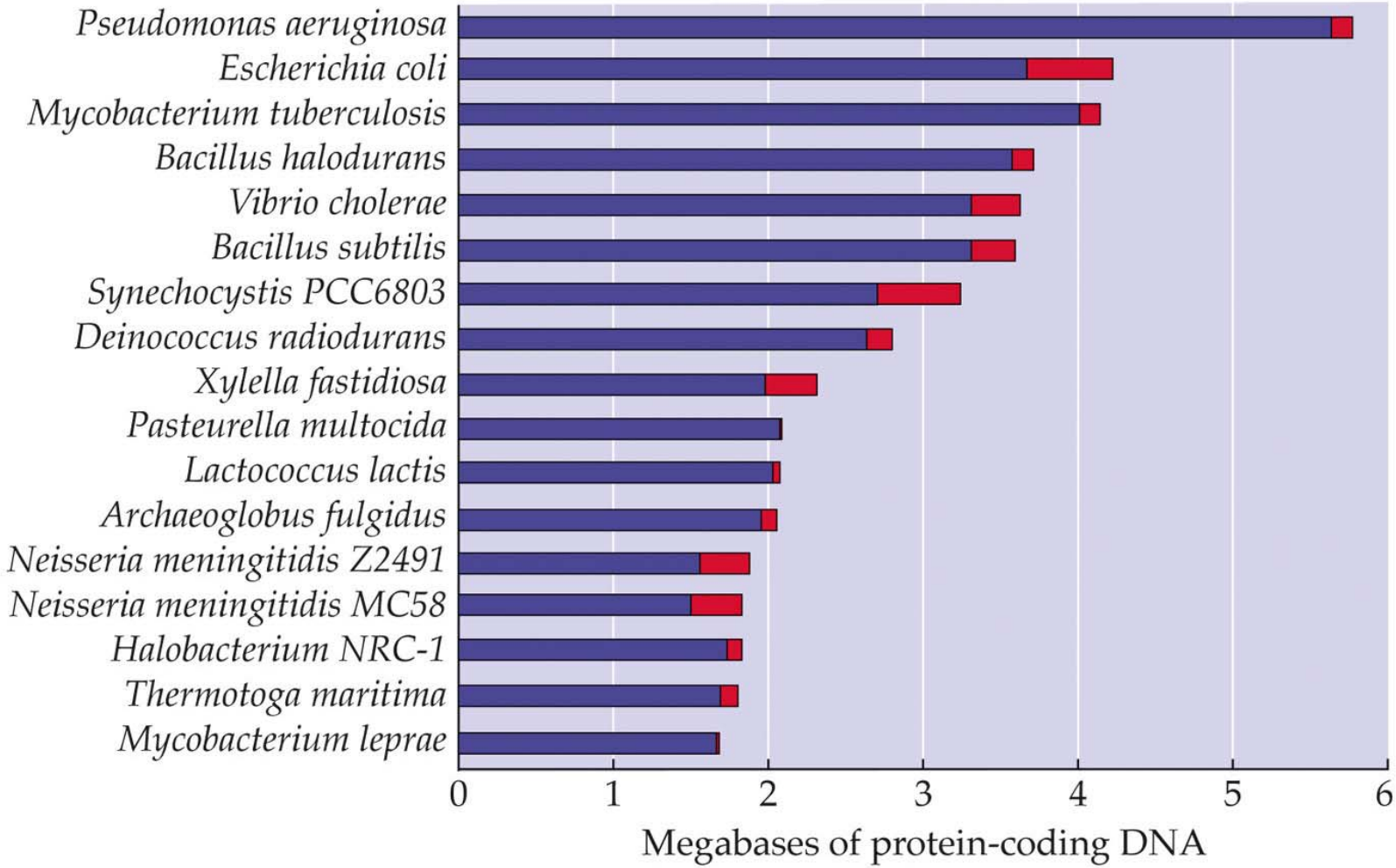
Rem: These are one-way processes!

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

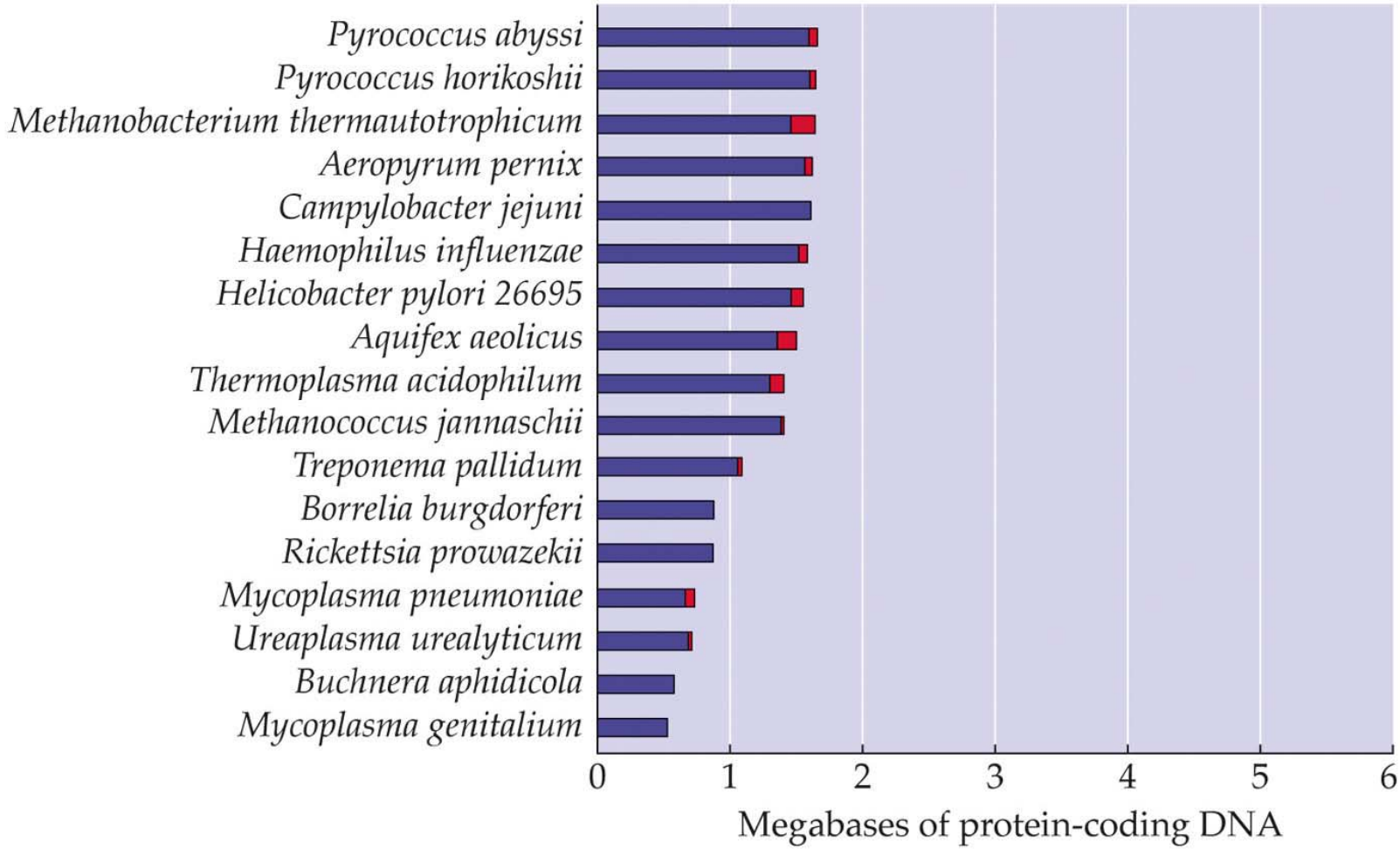


Doolittle's Universal Tree (1999)

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 involves more than simply 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.

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

TABLE 13.3 C values from eukaryotic organisms ranked by size

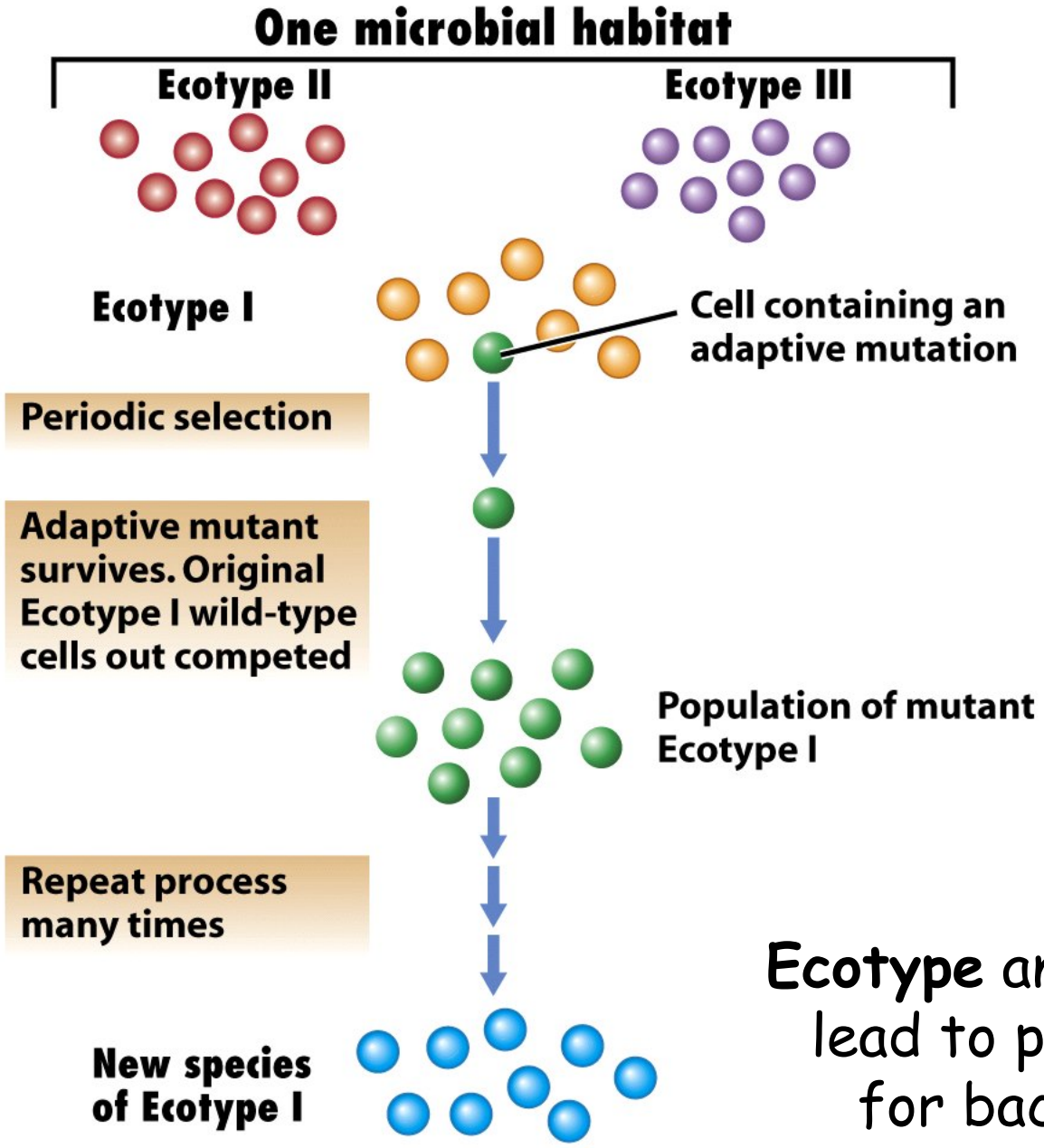
Species	C value (kb)
Navicola pelliculosa (diatom)	35,000
Drosophila melanogaster (fruitfly)	180,000
Paramecium aurelia (ciliate)	190,000
Gallus domesticus (chicken)	1,200,000
Erysiphe cichoracearum (fungus)	1,500,000
Cyprinus carpio (carp)	1,700,000
Lampreta planeri (lamprey)	1,900,000
Boa constrictor (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.



Ecotype and **periodic selection** lead to possible mechanism for bacterial speciation.