

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 microbes 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*

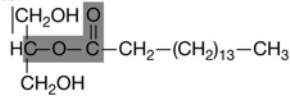
Taxon	Name
Domain	<i>Bacteria</i>
Phylum	<i>Spirochaetes</i> (vernacular name: spirochetes)
Class	<i>Spirochaetes</i>
Order	<i>Spirochaetales</i>
Family	<i>Spirochaetaceae</i>
Genus	<i>Spirochaeta</i>
Species	<i>plicatilis</i>

MICROBIAL LIFE, Table 17.1 © 2002 Sinauer Associates, Inc.

TABLE 11.4 Some phenotypic characteristics of taxonomic value

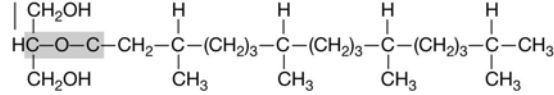
Major category	Components
I. Morphology	Shape; size; Gram reaction
II. Motility	Motile by flagella; motile by gliding; motile by gas vesicles; nonmotile
III. Nutrition and physiology	Mechanism of energy conservation (phototroph, chemoorganotroph, chemolithotroph); relationship to oxygen; temperature,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

Ester



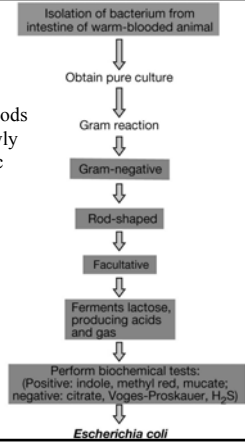
Bacteria, Eukarya

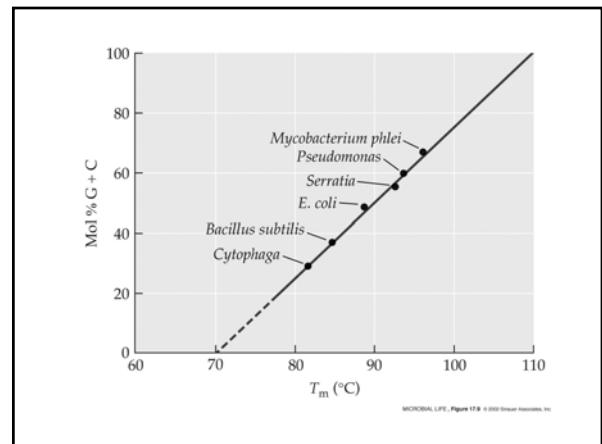
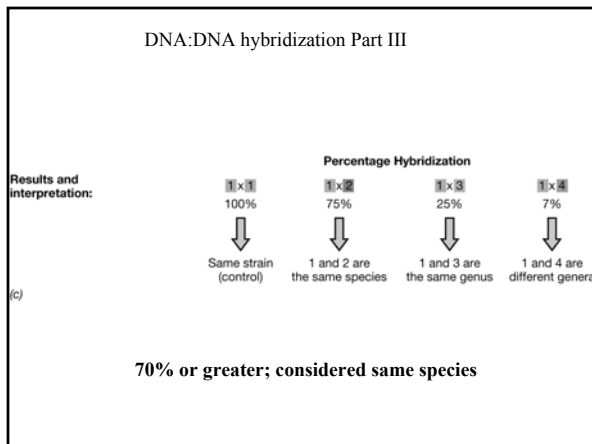
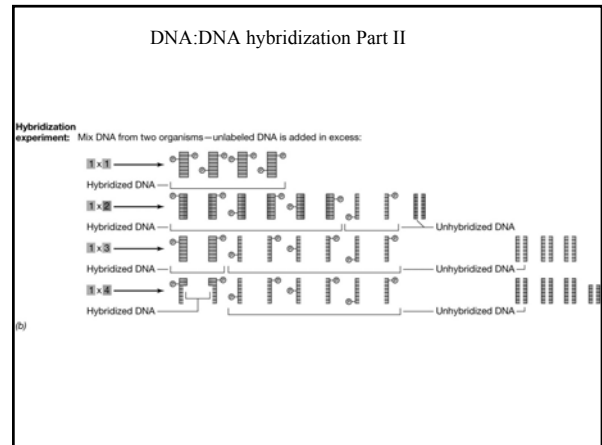
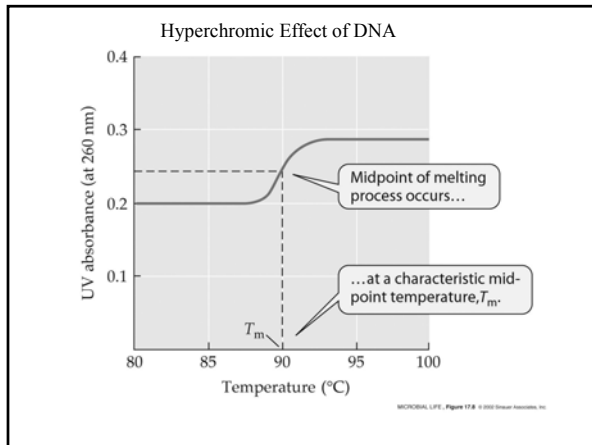
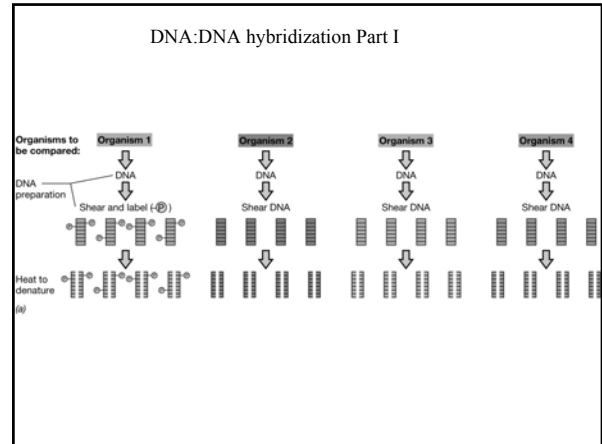
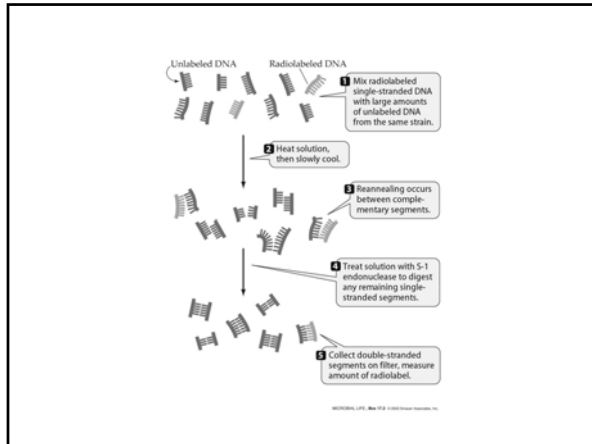
Ether



Archaea

Example of methods used to ID a newly isolated enteric bacterium





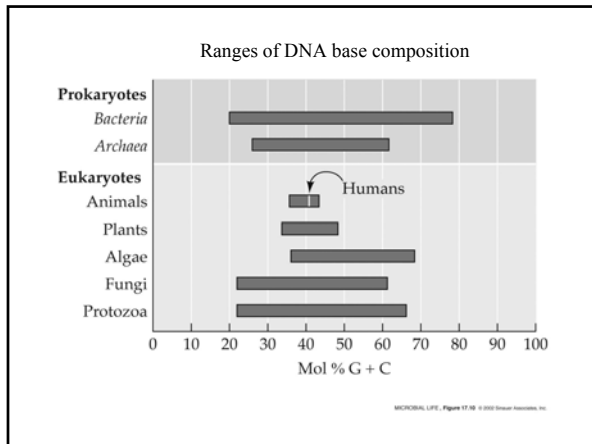


Table 17.2 Comparison of *E. coli* and its primate host species^a

Property	<i>E. coli</i>	<i>Homo sapiens</i>	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

^aAdapted from J. T. Staley, *ASM News*, 1999.
^bValue for all primates.
^cMouse 18S rRNA differs from humans by 16 bases.
^dComparison between *Homo sapiens* and chimpanzee.
^eComparison between *Homo sapiens* and lemurs.

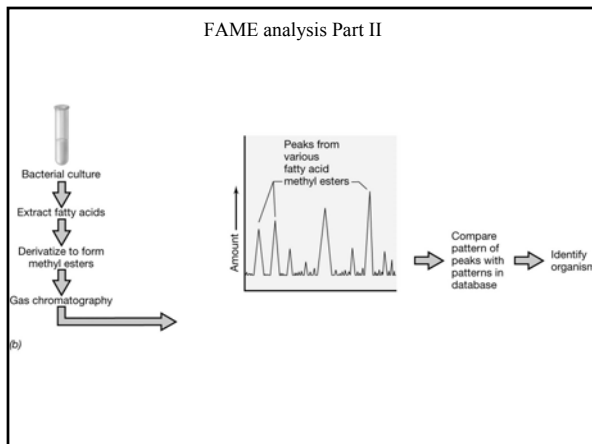
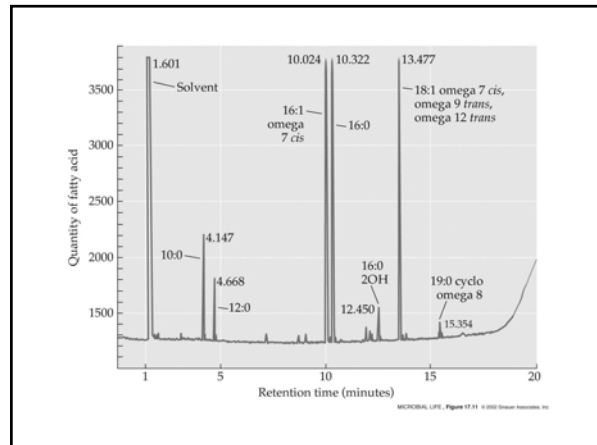
MICROBIAL LIFE, Table 17.2 © 2002 Sinauer Associates, Inc.

FAME analysis Part I

Classes of Fatty Acids in *Bacteria*

Class	Example	Structure of example
Saturated	tetradecanoic acid	<chem>CCCCCCCCCCCC(=O)O</chem>
Unsaturated	omega-7-cis hexadecanoic acid	<chem>CCCC=CCCCCCCC(=O)O</chem>
Cyclopropane	cis 7-8 methylene hexadecanoic acid	<chem>CCCC1=CCCCC1CCCC(=O)O</chem>
Branched	13-methyltetradecanoic acid	<chem>CCCC(C)CCCCCCCC(=O)O</chem>
Hydroxy	3-hydroxytetradecanoic acid	<chem>CCCC(O)CCCCCCCC(=O)O</chem>

(a)



BOX, ERIC, & REP PCR

Lanes represent: Strains RL1, ES1, & ES2

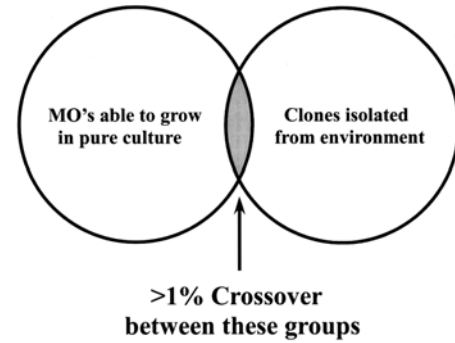
- Three different types of PCR based genomic fingerprinting methods. Collectively known as **REP PCR**.
- Minimal genetic variability shown among three strains of iron-oxidizing bacteria.

Taxonomy Summary

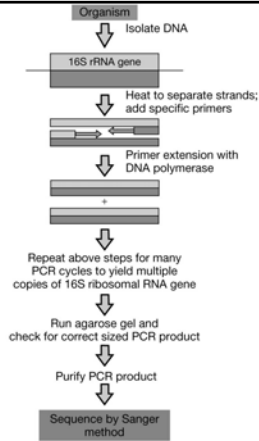
Classical physiological descriptions of microbes constitute a taxonomy, but do not provide relationships (except as might be inferred subjectively).

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.

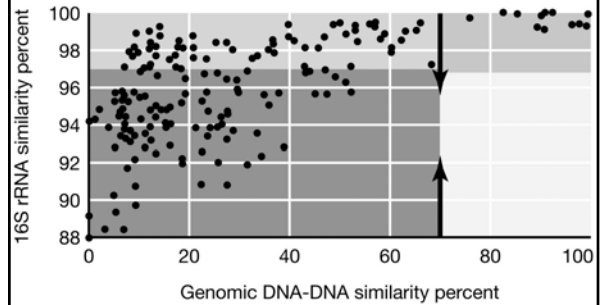
All these methods require pure-cultivation of organisms for characterization, but we can't cultivate much of what is out there.



Pure culture
SSU rRNA
sequencing
using PCR



Relationship between SSU rDNA and genomic DNA hybridization



Ecotype and periodic selection lead to possible mechanism for bacterial speciation

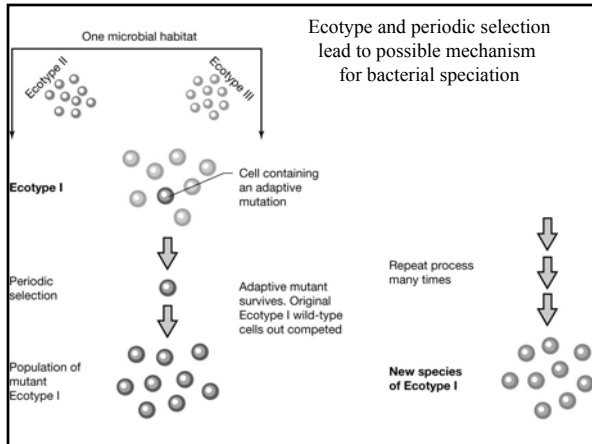
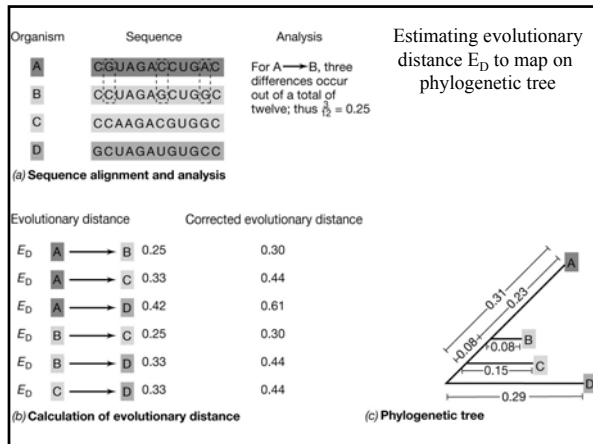


TABLE 11.6 Taxonomic ranks and numbers of known prokaryotic species^a

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

^aNumbers 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.



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** (~5x10⁶ bp genome), we are **complex** (~3x10⁹ bps); complexity has nothing to do with *evolutionary advancement*.

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

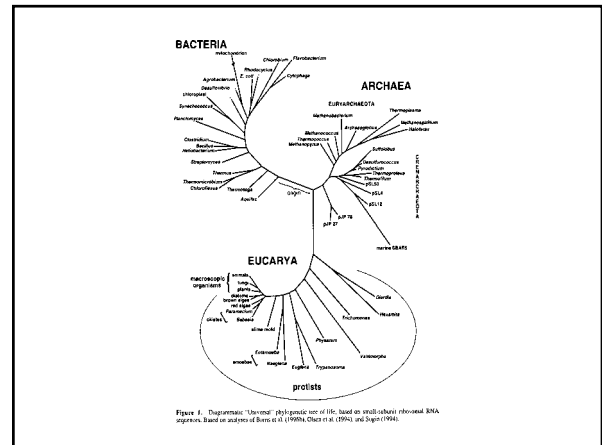
D. There is no such thing as a *primitive* organism alive today. **Simple**, yes, but still a finely honed product of ~4 billion years under the selective hammer of the niches that it and its progenitors have occupied.

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 (Rb)
<i>Navicula pelliculosa</i> (diatom)	35,000
<i>Drosophila melanogaster</i> (fruitfly)	180,000
<i>Paramecium aurelia</i> (ciliate)	190,000
<i>Gallus domesticus</i> (chicken)	1,200,000
<i>Eryngium cichoracearum</i> (fungus)	1,500,000
<i>Cyprinus carpio</i> (carp)	1,700,000
<i>Lamprolaima planeri</i> (lamprey)	1,900,000
<i>Bass constrictor</i> (snake)	2,100,000
<i>Parascaris equorum</i> (roundworm)	2,500,000
<i>Carcharias obscurus</i> (shark)	2,700,000
<i>Rattus norvegicus</i> (rat)	2,900,000
<i>Xenopus laevis</i> (toad)	3,100,000
<i>Homo sapiens</i> (human)	3,400,000
<i>Nicotiana glauca</i> (tobacco)	3,800,000
<i>Paramecium caudatum</i> (ciliate)	8,600,000
<i>Schistocerca gregaria</i> (locust)	9,300,000
<i>Allium cepa</i> (onion)	18,000,000
<i>Cocconeis asteroides</i> (diatom)	25,000,000
<i>Lilium formosanum</i> (lily)	36,000,000
<i>Pinus resinosa</i> (pine)	68,000,000
<i>Ampelisca maesi</i> (newt)	84,000,000
<i>Protoperus aethiopicus</i> (lungfish)	140,000,000
<i>Ophioglossum petiolatum</i> (fern)	160,000,000
<i>Amoeba proteus</i> (amoeba)	290,000,000
<i>Amoeba dubia</i> (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.



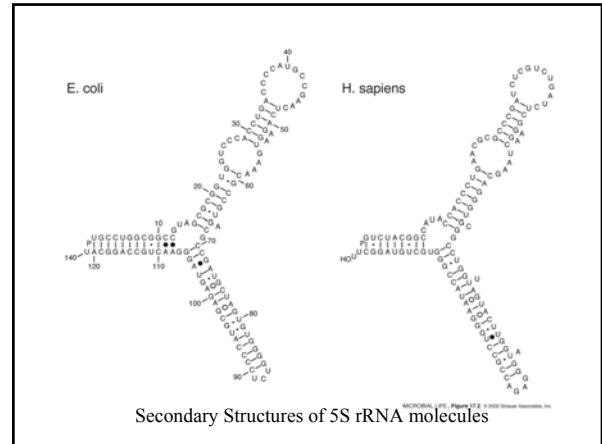
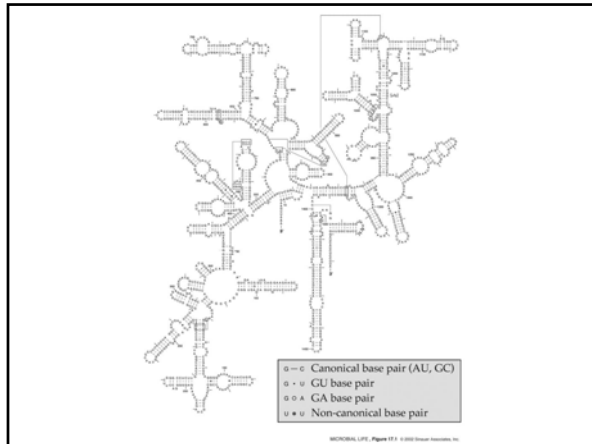
Some Lessons from the BIG TREE: Map of the Biological Record

Evolutionary "clock" is NOT constant between different lineages

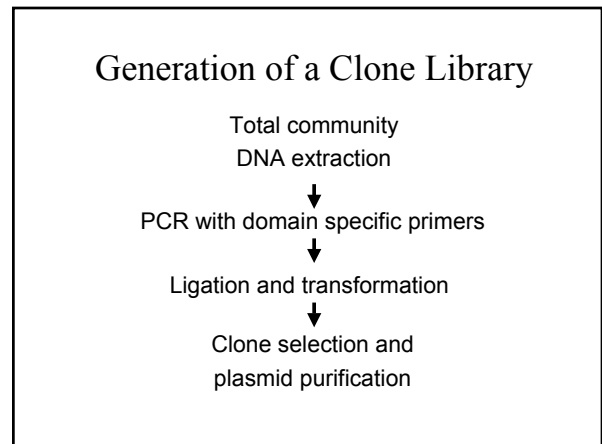
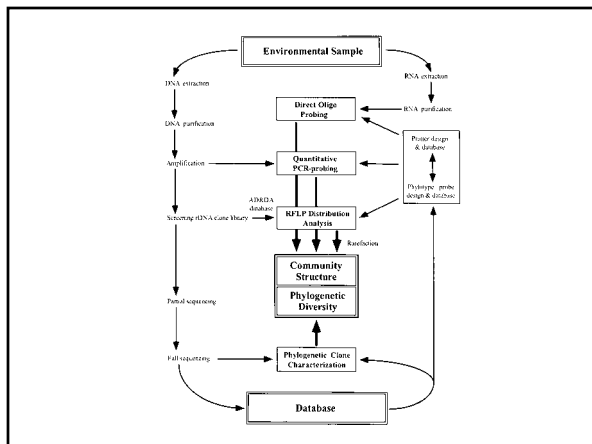
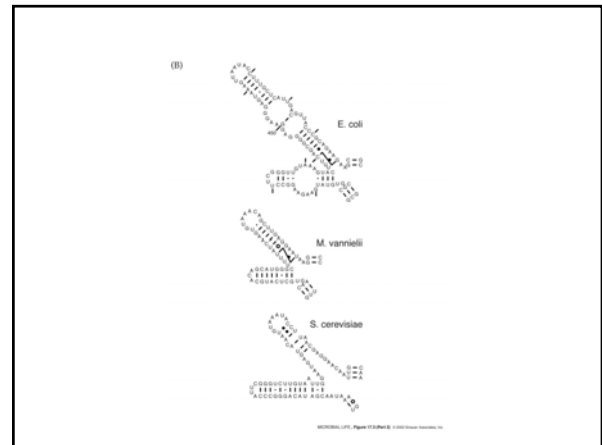
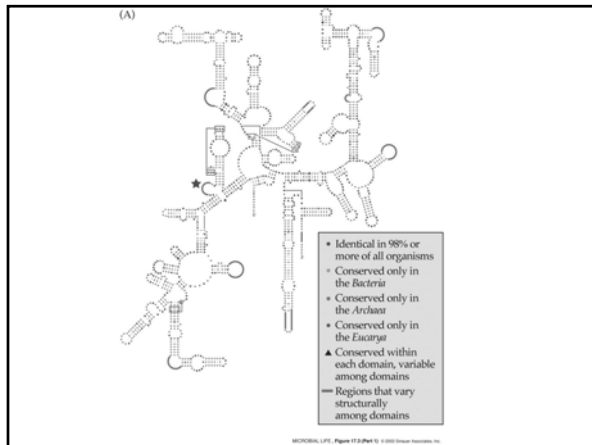
- Terminal nodes NOT all the same length, so not constant for all organisms either!
- Endosymbionts sped up very fast (semi-autonomous organelles).
- Eucarya – Fast clocks
- Archaea – Slow clocks
- Bacteria – Intermediate

Why ribosomal RNAs?

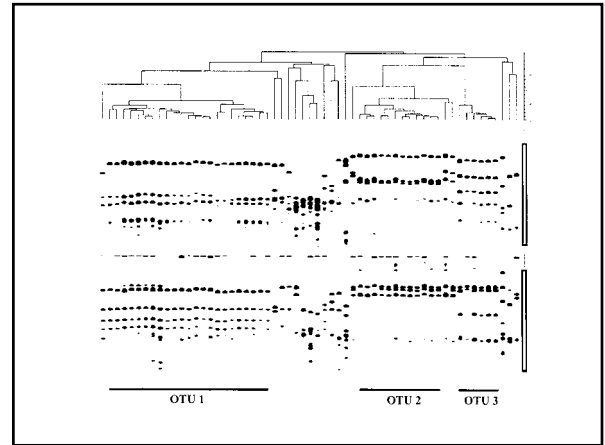
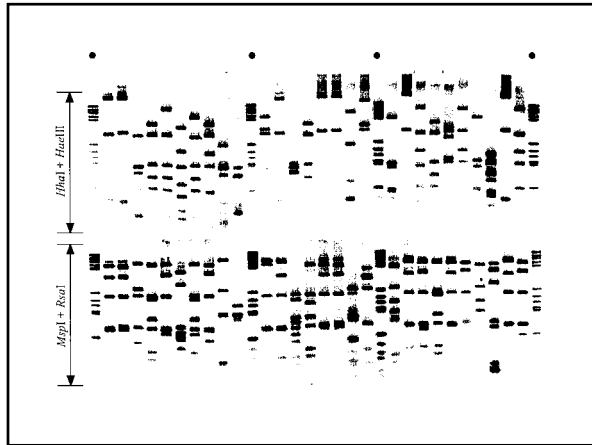
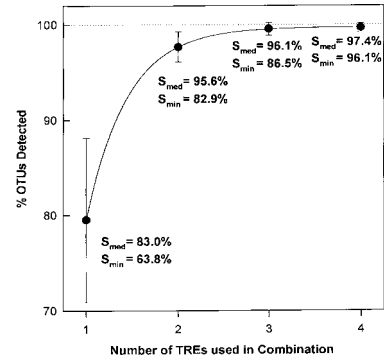
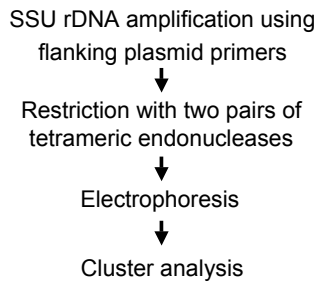
- Found among all living organisms (for 3.8 of the last 4.5 billion years). Integral part of protein synthesis machinery.
- Cell component analyses provide culture-independent means of investigating questions in microbial ecology (lack of morphology).
- rRNAs offer a type of sequence information that 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 ~52,000 SSU rRNAs.



Secondary Structures of 5S rRNA molecules

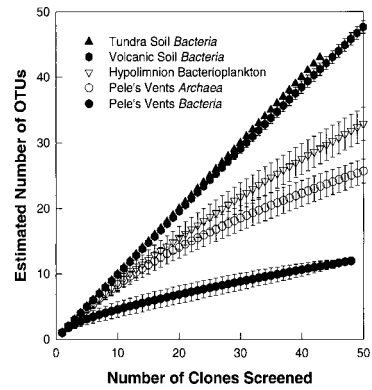


Amplified Ribosomal DNA Restriction Analysis



$$E_s = \sum_{i=1}^s \left[1 - \left(\frac{N - n_i}{N} \right)^n \right] \left(\frac{N}{n} \right)^{-1}$$

Equation 1. The rarefaction equation allows for unequal sample size and estimates the number of phylotypes (E_s) in a random sample of n clones sampled without replacement from a finite parent collection of N clones, where n_i is the number of clones of the i^{th} phylotype and S is the number of phylotypes in the parent collection.



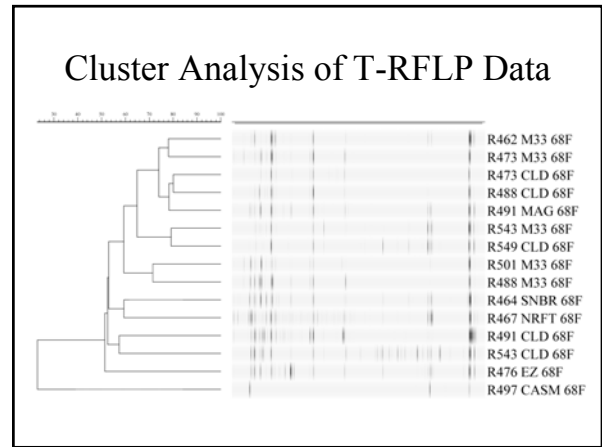
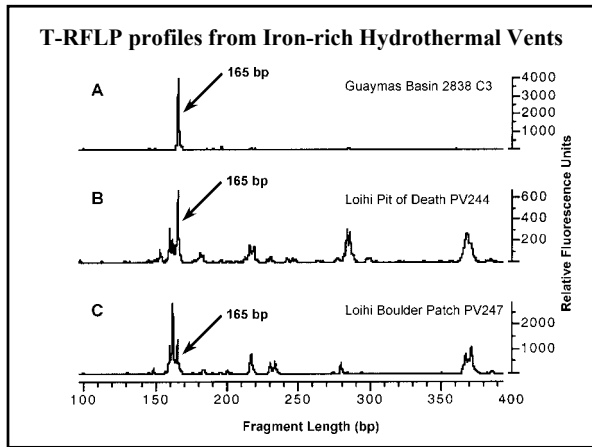
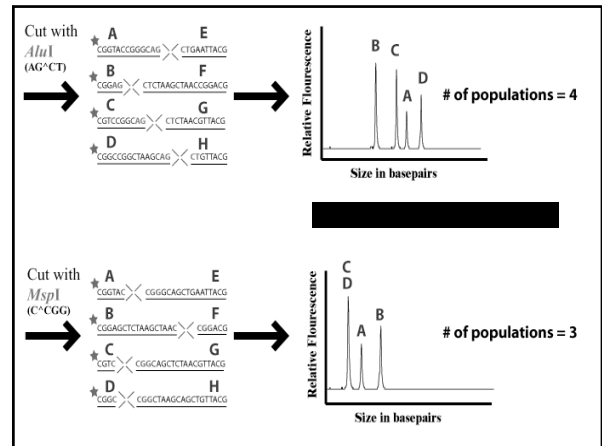
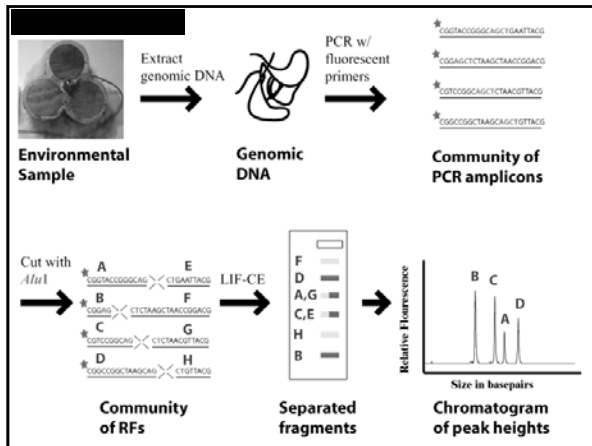
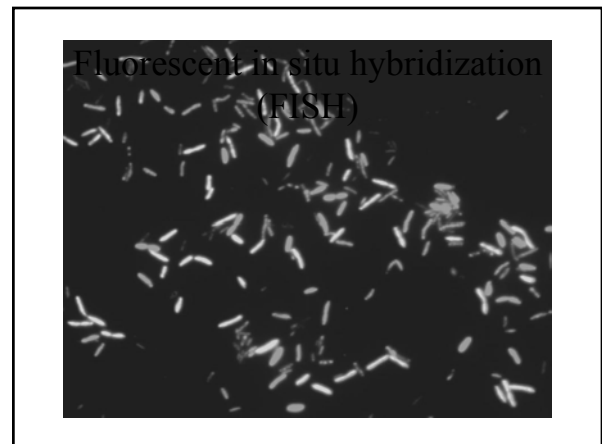


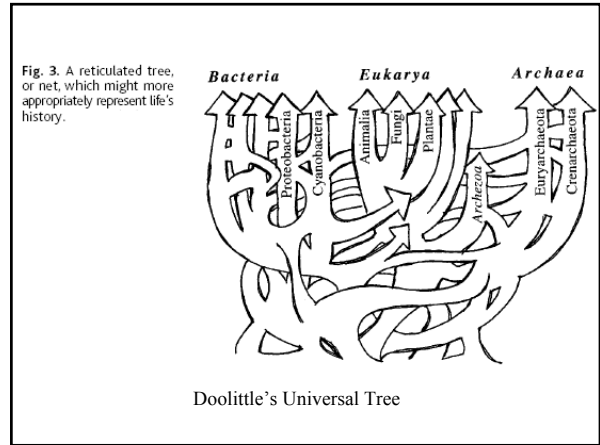
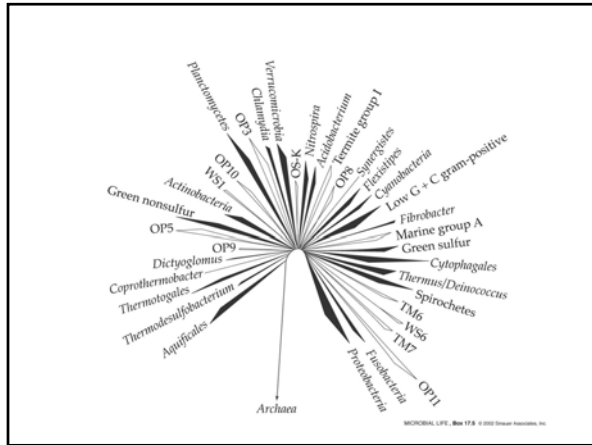
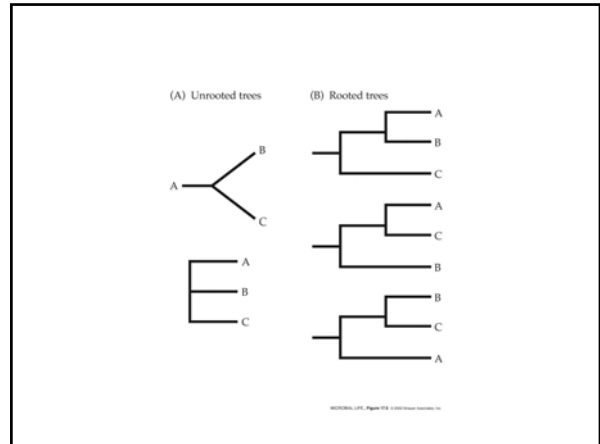
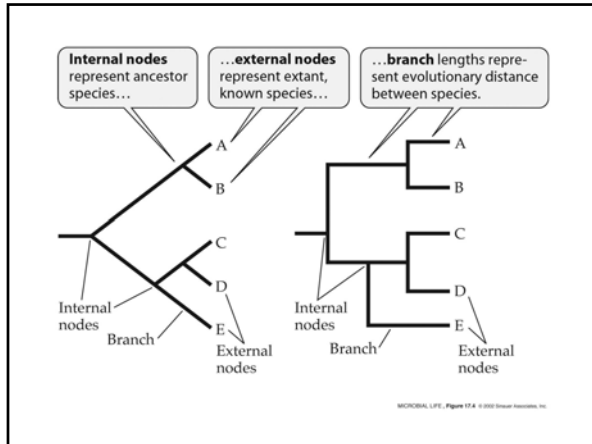
TABLE 11.1 Signature sequences from 16S or 18S rRNA defining the three domains of life

Oligonucleotide signatures ^a	Approximate position ^b	Occurrence among ^c		
		Archaea	Bacteria	Eukarya
CACYYG	315	0	> 95	0
AAACDCAAA	910	3	100	0
AAACCTAAAG	910	100	0	100
YUYAAUUG	960	100	< 1	100
CAACCYCR	1110	0	> 95	0
UCCCUUG	1380	> 95	0	100
UACACACCG	1400	0	> 99	100
CACACACCG	1400	100	0	0

^aY, any pyrimidine; R, any purine.
^bRefer to Figure 11.8c for numbering scheme of 16S rRNA.
^cOccurrence refers to percentage of organisms examined in any domain that contain that sequence.

Signature sequences can be obtained at any level of taxonomic hierarchy



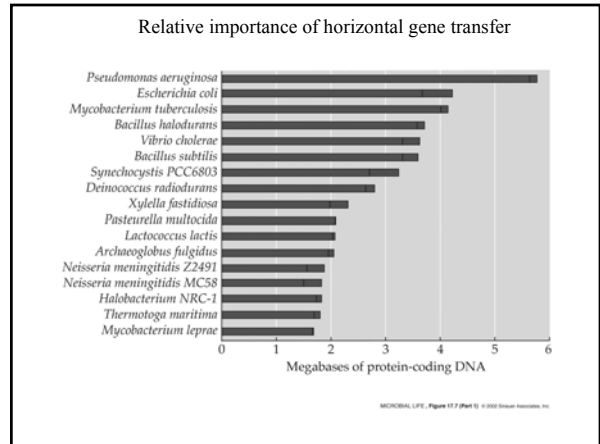


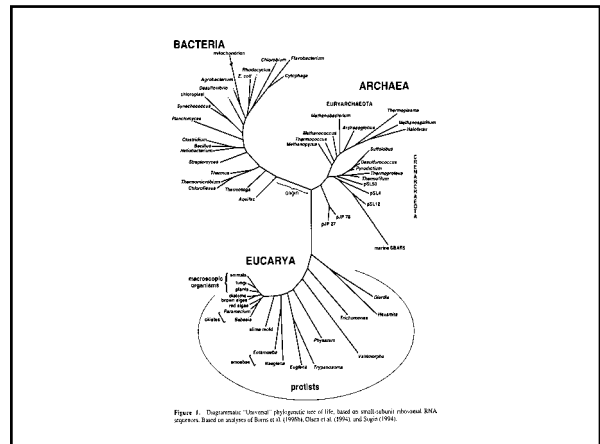
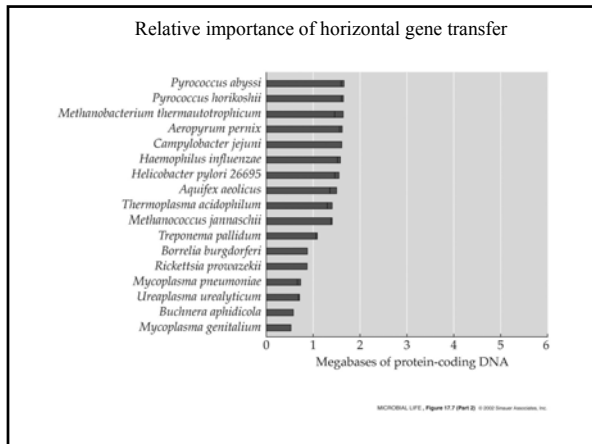
... the general course of evolution [for bacteria] will probably never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds... For these and other reasons, most modern taxonomists have explicitly abandoned the phylogenetic approach.

(Stanier et al., 1976)

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 properly be represented as a tree.

(W. F. Doolittle, 1999)





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.

TABLE 11.3 Summary of major differential features among Bacteria, Archaea, and Eukarya*

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	No	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 11.16)	One (4 subunits)	Several (8–12 subunits each)	Three (12–14 subunits each)
Transcription factors required (see Section 7.10)	No	Yes	Yes
Promoter structure (see Section 7.9)	–10 and –35 sequences (Pribnow box)	TATA box	TATA box
Sensitivity to chloramphenicol, streptomycin, and kanamycin	Yes	No	No

TABLE 11.3 Summary of major differential features among Bacteria, Archaea, and Eukarya*

Characteristic	Bacteria	Archaea	Eukarya
Physiological			
Methanogenesis	No	Yes	No
Disimilative reduction of S ⁰ or SO ₄ ²⁻ to H ₂ S, or Fe ³⁺ to Fe ²⁺	Yes	Yes	No
Nitrification	Yes	No	No
Denitrification	Yes	Yes	No
Nitrogen fixation	Yes	Yes	No
Chlorophyll-based photosynthesis	Yes	No	Yes (in chloroplasts)
Rhodospin-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

*Note that for many features only particular representatives within a domain show the property

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 microorganisms and their genes
 - inference dynamics of sequence change (Rolex or Timex)