EVE 161:
Microbial Phylogenomics

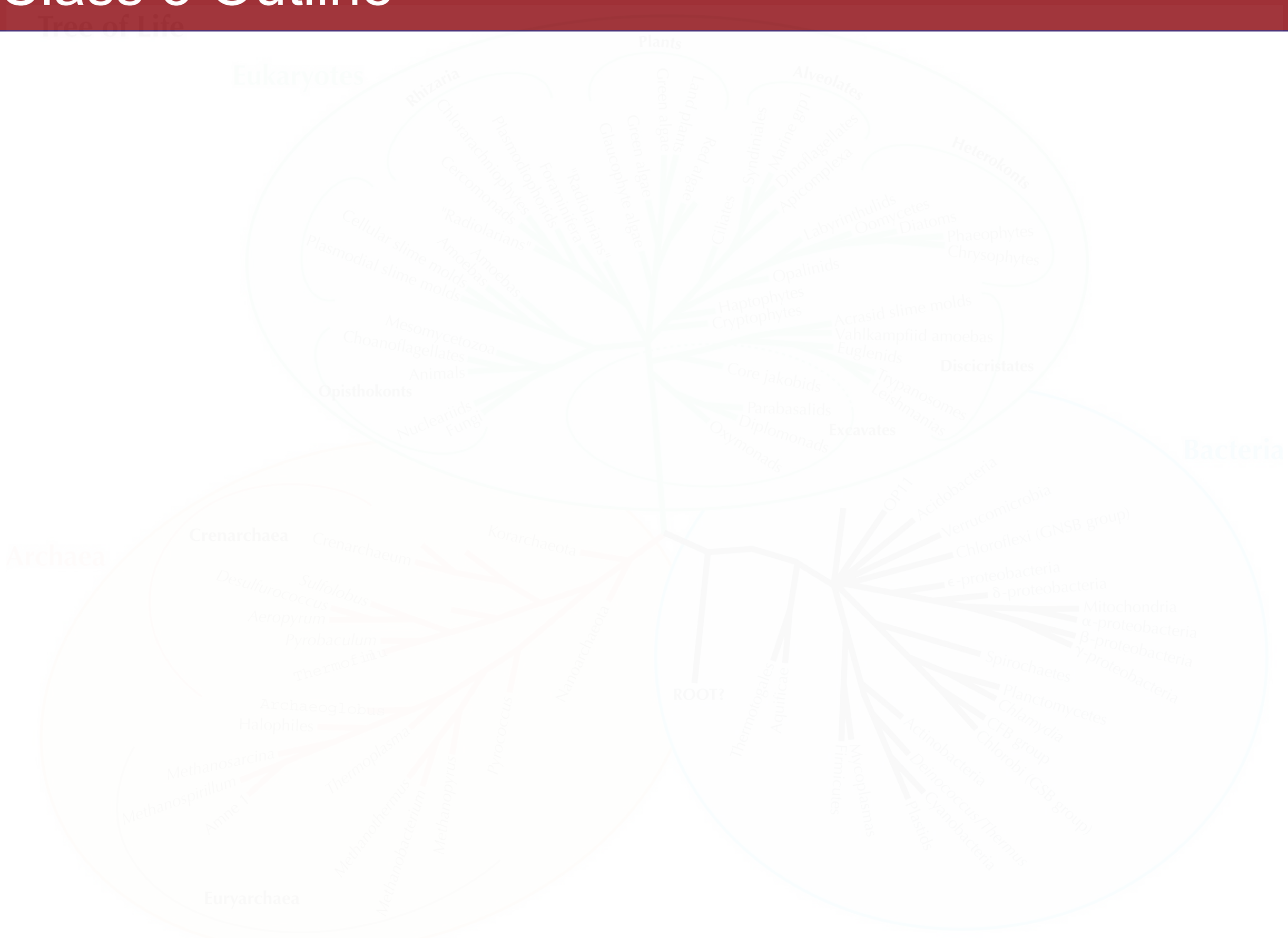
Class #3:
Era I: Woese and the Tree of Life

UC Davis, Winter 2018
Instructor: Jonathan Eisen
Teaching Assistant: Cassie Ettinger

Where we are going and where we have been

- Previous Class:
  - 2. Sequencing
- Current Class:
  - 3. Era I: Woese
- Next Class:
  - 4. Phylogeny

Class 3 Outline



Era I: rRNA Tree of Life

Era I:
rRNA Tree of Life

Two papers for today

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537–4541, October 1977
Evolution

Classification of methanogenic bacteria by 16S ribosomal RNA characterization

(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX*†, LINDA J. MAGRUM*, WILLIAM E. BALCH‡, RALPH S. WOLFE‡, AND CARL R. WOESE*‡

Departments of *Genetics and Development and †Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

Two papers for today

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537–4541, October 1977
Evolution

Classification of methanogenic bacteria by 16S ribosomal RNA characterization

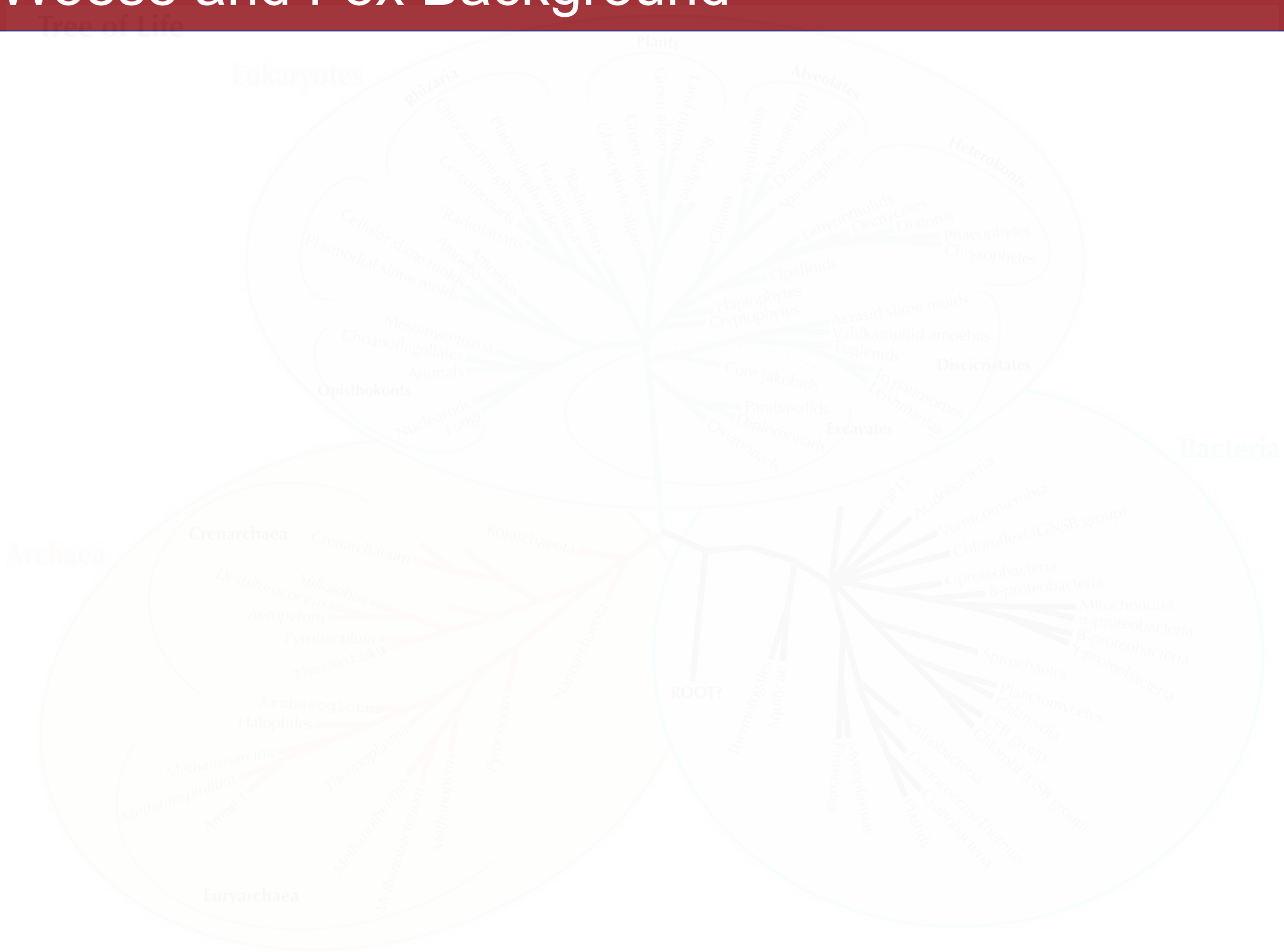
(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX*†, LINDA J. MAGRUM*, WILLIAM E. BALCH‡, RALPH S. WOLFE‡, AND CARL R. WOESE*‡

Departments of *Genetics and Development and †Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

Woese and Fox Background



Woese and Fox Background

The biologist has customarily structured his world in terms of certain basic dichotomies. Classically, what was not plant was animal. The discovery that bacteria, which initially had been considered plants, resembled both plants and animals less than plants and animals resembled one another led to a reformulation of the issue in terms of a yet more basic dichotomy, that of eukaryote versus prokaryote. The striking differences between eukaryotic and prokaryotic cells have now been documented in endless molecular detail. As a result, it is generally taken for granted that all extant life must be of these two basic types.

Woese and Fox Background

Thus, it appears that the biologist has solved the problem of the primary phylogenetic groupings. However, this is not the case. Dividing the living world into *Prokaryotae* and *Eukaryotae* has served, if anything, to obscure the problem of what extant groupings represent the various primeval branches from the common line of descent. The reason is that eukaryote/prokaryote is not primarily a phylogenetic distinction, although it is generally treated so. The eukaryotic cell is organized in a different and more complex way than is the prokaryote; this probably reflects the former's composite origin as a symbiotic collection of various simpler organisms (1–5). However striking, these organizational dissimilarities do not guarantee that eukaryote and prokaryote represent phylogenetic extremes.

Woese and Fox Background

The eukaryotic cell *per se* cannot be directly compared to the prokaryote. The composite nature of the eukaryotic cell makes it necessary that it first be conceptually reduced to its phylogenetically separate components, which arose from ancestors that were noncomposite and so individually are comparable to prokaryotes. In other words, the question of the primary phylogenetic groupings must be formulated solely in terms of relationships among "prokaryotes"—i.e., noncomposite entities. (Note that in this context there is no suggestion *a priori* that the living world is structured in a dichotomous way.)

Woese and Fox Background

The organizational differences between prokaryote and eukaryote and the composite nature of the latter indicate an important property of the evolutionary process: Evolution seems to progress in a "quantized" fashion. One level or domain of organization gives rise ultimately to a higher (more complex) one. What "prokaryote" and "eukaryote" actually represent are two such domains. Thus, although it is useful to define phylogenetic patterns within each domain, it is not meaningful to construct phylogenetic classifications between domains: Prokaryotic kingdoms are not comparable to eukaryotic ones. This should be recognized by an appropriate terminology. The highest phylogenetic unit in the prokaryotic domain we think should be called an "urkingdom"—or perhaps "primary kingdom." This would recognize the qualitative distinction between prokaryotic and eukaryotic kingdoms and emphasize that the former have primary evolutionary status.

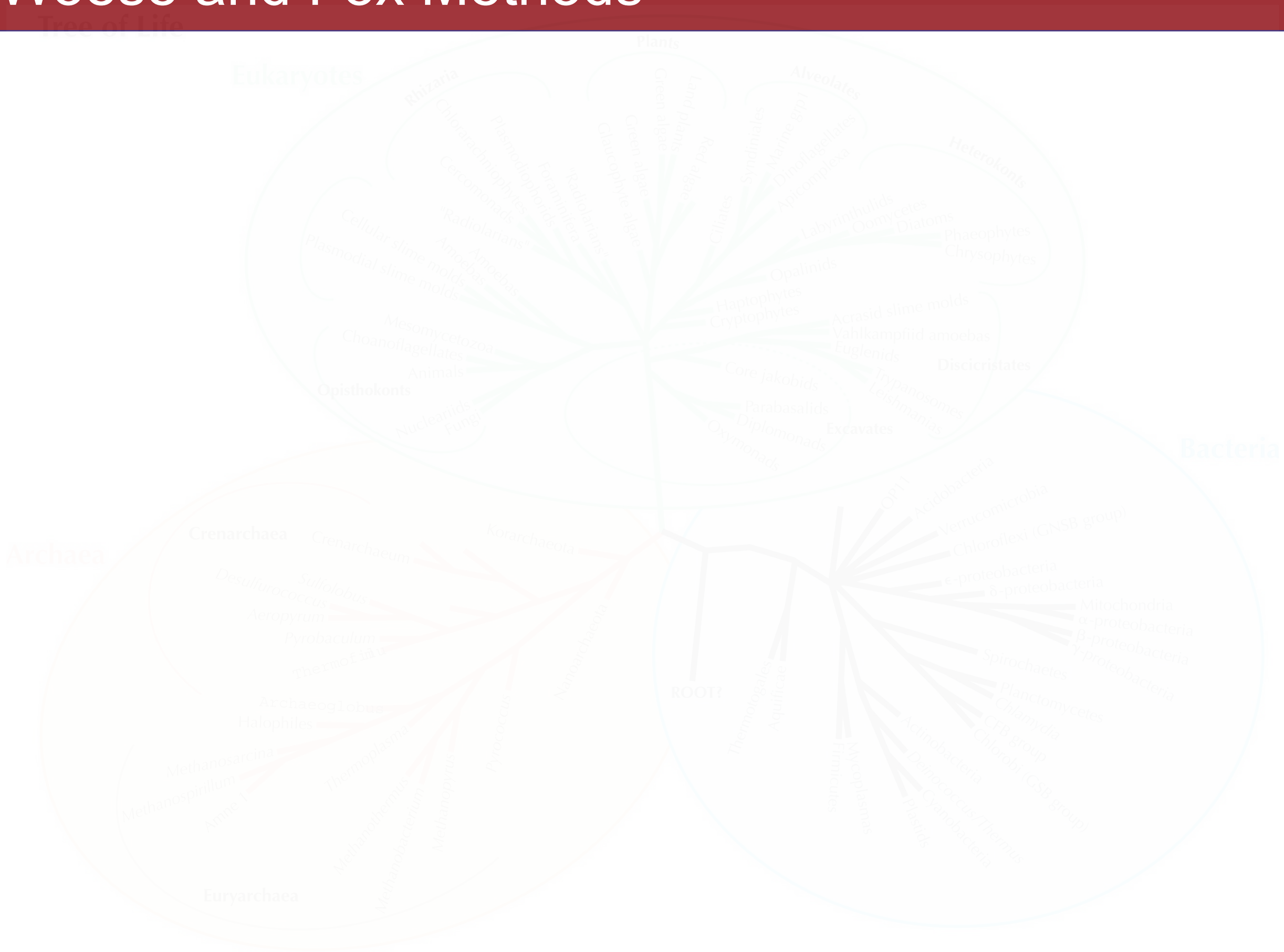
Woese and Fox Background

The passage from one domain to a higher one then becomes a central problem. Initially one would like to know whether this is a frequent or a rare (unique) evolutionary event. It is traditionally assumed—without evidence—that the eukaryotic domain has arisen but once; all extant eukaryotes stem from a common ancestor, itself eukaryotic (2). A similar prejudice holds for the prokaryotic domain (2). [We elsewhere argue (6) that a hypothetical domain of lower complexity, that of "progenotes," may have preceded and given rise to the prokaryotes.] The present communication is a discussion of recent findings that relate to the urkingdom structure of the prokaryotic domain and the question of its unique as opposed to multiple origin.

Woese and Fox Background

Phylogenetic relationships cannot be reliably established in terms of noncomparable properties (7). A comparative approach that can measure degree of difference in comparable structures is required. An organism's genome seems to be the ultimate record of its evolutionary history (8). Thus, comparative analysis of molecular sequences has become a powerful approach to determining evolutionary relationships (9, 10).

Woese and Fox Methods



Woese and Fox Methods

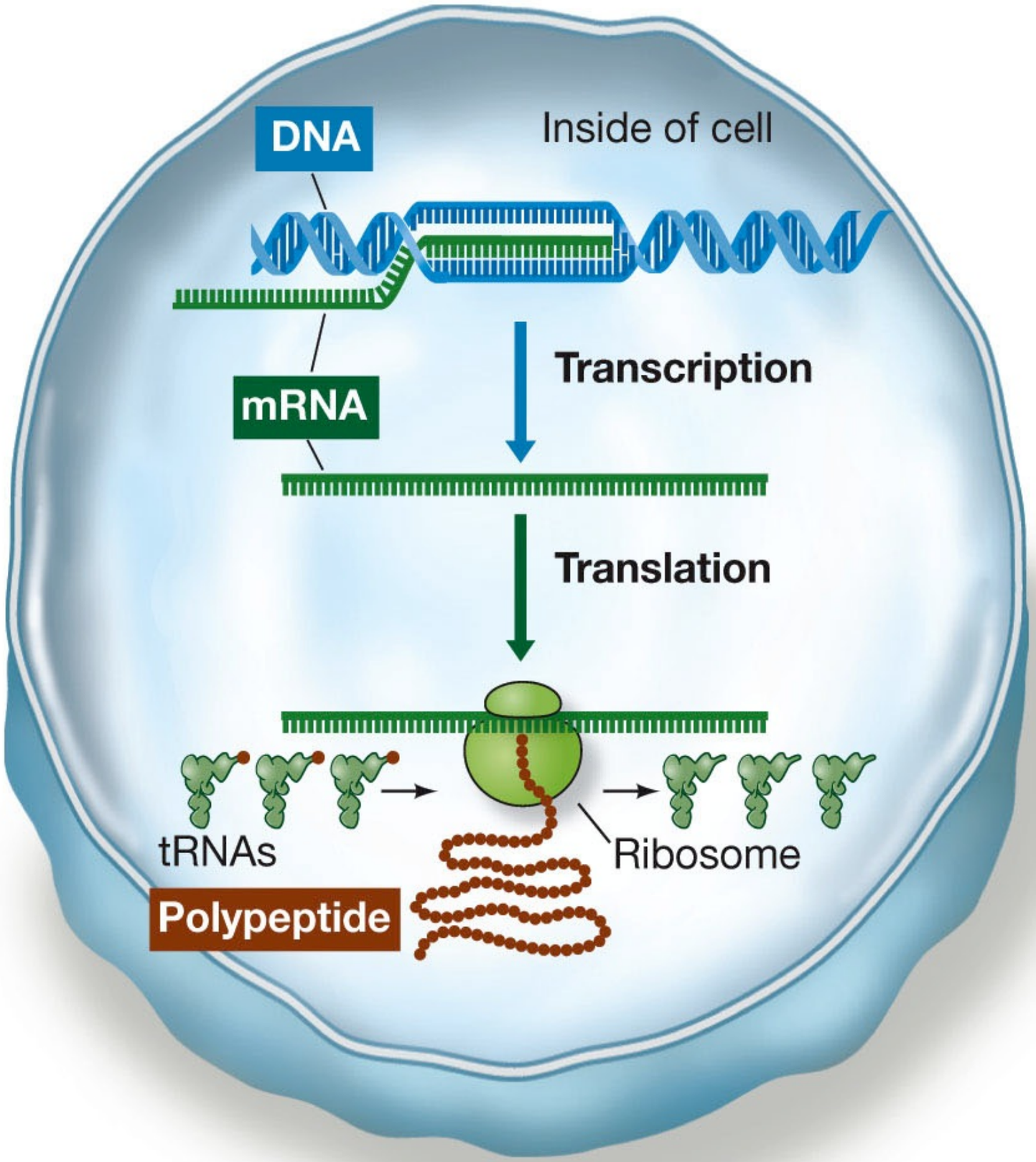
To determine relationships covering the entire spectrum of extant living systems, one optimally needs a molecule of appropriately broad distribution. None of the readily characterized proteins fits this requirement. However, ribosomal RNA does. It is a component of all self-replicating systems; it is readily isolated; and its sequence changes but slowly with time—permitting the detection of relatedness among very distant species (11–13). To date, the primary structure of the 16S (18S) ribosomal RNA has been characterized in a moderately large and varied collection of organisms and organelles, and the general phylogenetic structure of the prokaryotic domain is beginning to emerge.

Woese and Fox Methods

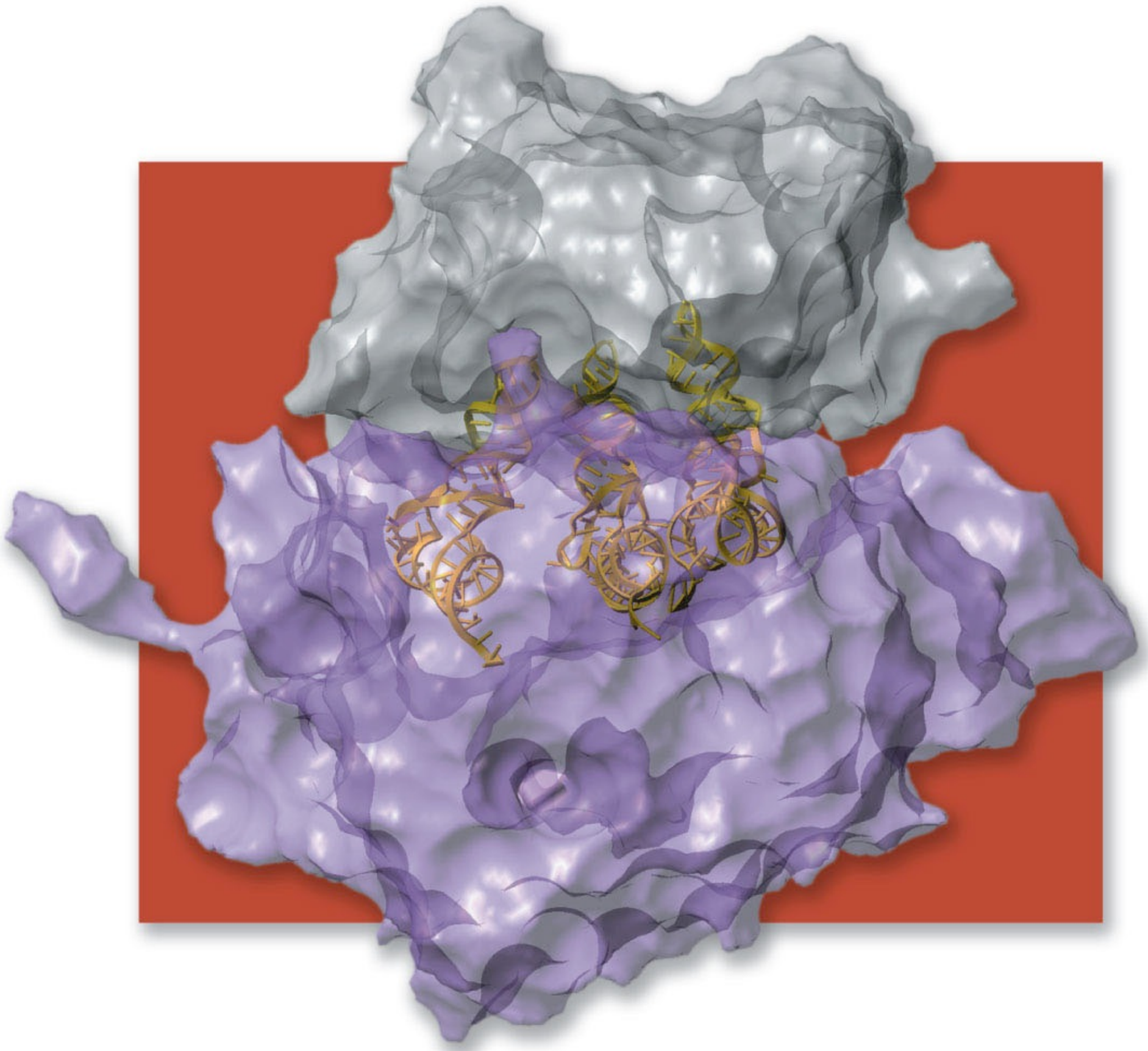
To determine relationships covering the entire spectrum of extant living systems, one optimally needs a molecule of appropriately broad distribution. None of the readily characterized proteins fits this requirement. However, ribosomal RNA does. It is a component of all self-replicating systems; it is readily isolated; and its sequence changes but slowly with time—permitting the detection of relatedness among very distant species (11–13). To date, the primary structure of the 16S (18S) ribosomal RNA has been characterized in a moderately large and varied collection of organisms and organelles, and the general phylogenetic structure of the prokaryotic domain is beginning to emerge.

What is ribosomal RNA?

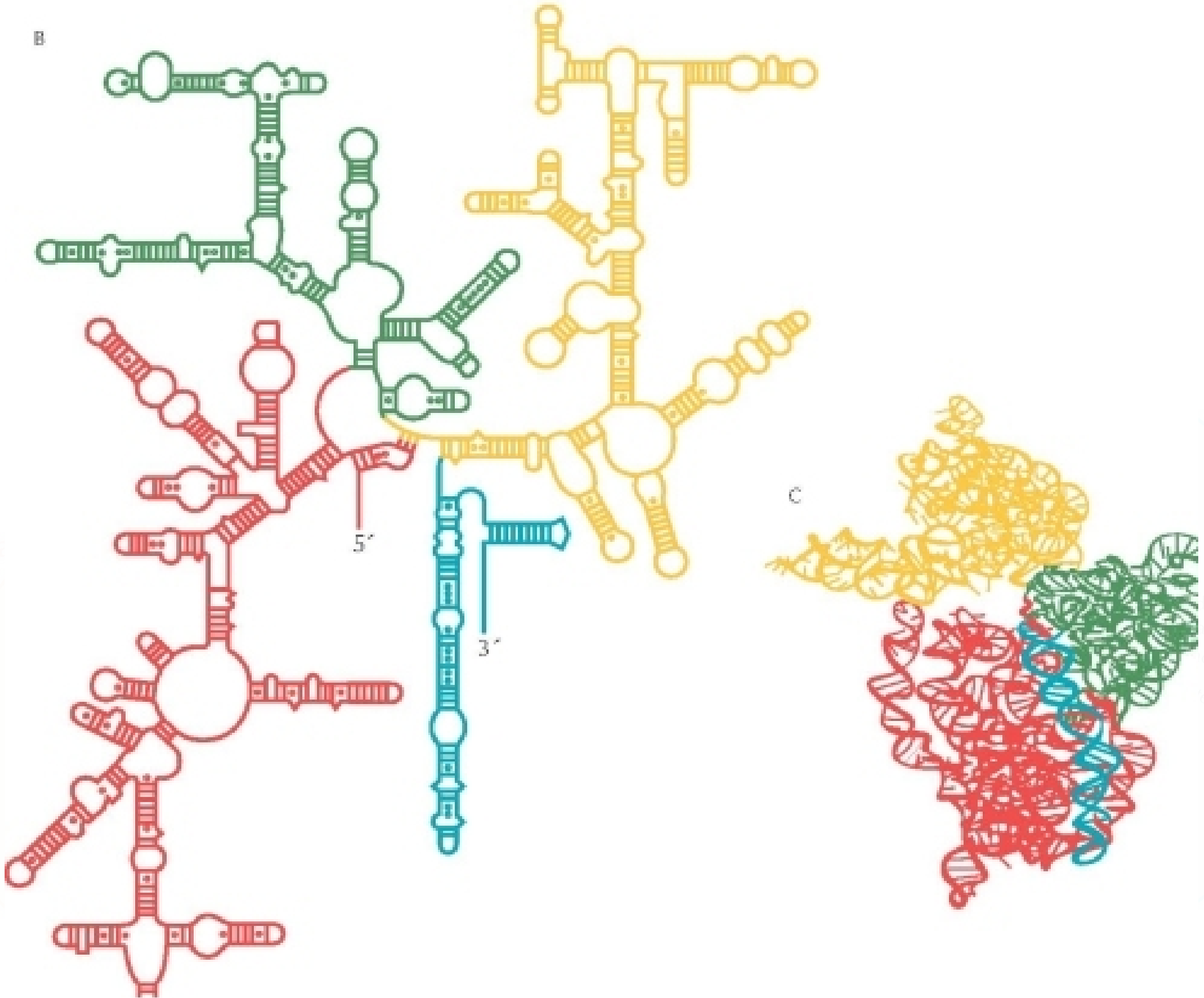
Central Dogma



The Ribosome



Ribosomal RNA structure



Woese and Fox Methods

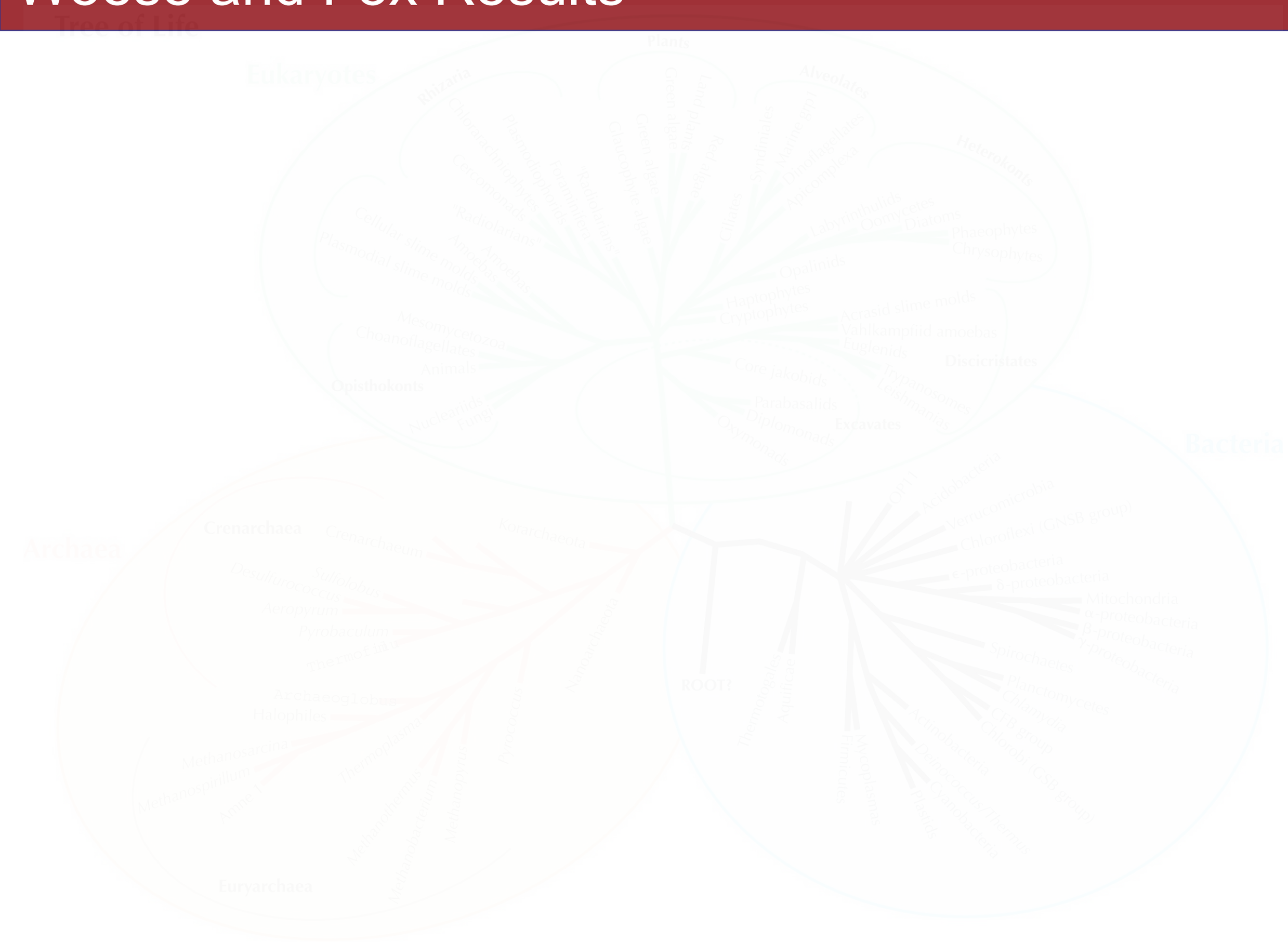
To determine relationships covering the entire spectrum of extant living systems, one optimally needs a molecule of appropriately broad distribution. None of the readily characterized proteins fits this requirement. However, ribosomal RNA does. It is a component of all self-replicating systems; it is readily isolated; and its sequence changes but slowly with time—permitting the detection of relatedness among very distant species (11–13). To date, the primary structure of the 16S (18S) ribosomal RNA has been characterized in a moderately large and varied collection of organisms and organelles, and the general phylogenetic structure of the prokaryotic domain is beginning to emerge.

Why Use ribosomal RNA for this?

Methods Question: Why Use rRNA for this?

- Universal
- Highly conserved functionally
- Evolves slowly
- Easy to extract and sequence

Woese and Fox Results



Woese and Fox Results

A comparative analysis of these data, summarized in Table 1, shows that the organisms clearly cluster into several primary kingdoms.

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

How Was This Calculated? And Why?

Two papers for today

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537–4541, October 1977
Evolution

Classification of methanogenic bacteria by 16S ribosomal RNA characterization

(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX* †, LINDA J. MAGRUM*, WILLIAM E. BALCH‡, RALPH S. WOLFE‡, AND CARL R. WOESE*‡

Departments of *Genetics and Development and †Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

Fox et al. Background

The methane-producing bacteria are a poorly studied collection of morphologically diverse organisms that share the common metabolic capacity to grow anaerobically by oxidizing hydrogen and reducing carbon dioxide to methane (1–3). Their relationships to one another and to other microbes remain virtually unknown. Protein and nucleic acid primary structures are perhaps the most reliable indicators of phylogenetic relationships (4–6). By using a molecule, such as the 16S ribosomal RNA, that is readily isolated, ubiquitous, and highly constrained in sequence (7), it is possible to relate even the most distant of microbial species. To date, approximately 60 bacterial species have been characterized in terms of their 16S ribosomal RNA primary structures (refs. 6–9, unpublished data). We present here results of a comparative study of the methanogens by this method, which shows their relationships to one another and to typical bacteria.

Fox et al. Methods

METHODS

Methanobacterium ruminantium strain *PS*, *Methanobacterium* strain *M.o.H.*, *Methanobacterium formicicum*, and *Methanosarcina barkeri* were provided by M. P. Bryant. *Methanobacterium arbophilicum* (10) was obtained from J. G. Zeikus. Two new marine isolates, Cariaco isolate JR-1 and Black Sea isolate JR-1, were provided by J. A. Romesser. *Methanospirillum hungatii* (11) and the above methanogens were cultivated in the following low-phosphate medium (values in g/liter): $(NH_4)_2SO_4$, 0.22; NaCl, 0.45; $MgSO_4 \cdot 7H_2O$, 0.09; $CaCl_2 \cdot H_2O$, 0.06; $FeSO_4 \cdot 7H_2O$, 0.002; resazurin, 0.001; sodium formate, 3.0; sodium acetate, 2.5; $NaHCO_3$, 6.0; trace mineral solution and vitamin solution (12), 10 ml each; and dephosphorylated yeast extract (Difco) and Trypticase (BBL), 2.0 each. For growth of marine isolates, NaCl was added to a final concentration of 15 g/liter. Procedures for preparation of media, growth of organisms, ^{32}P labeling, extraction of labeled 16S ribosomal RNA, and analysis of T_1 RNase digests of this RNA have been published (13–17).

Fox et al. Methods: rRNA Oligonucleotide Catalog

- Isolate rRNA
- Radioactively label
- Digest with T1 Rnase
- 2D electrophoresis
- Cut out fragments
- Sequence each fragment to get “oligonucleotide”
- Build catalog

Fox et al. Table 1. Oligonucleotide catalog

Table 1. Oligonucleotide catalogs for 16S rRNA of 10 methanogens

Oligonucleotide sequence	Present in organism number
5-mers	
CCCCG	1–10;1,5,8
CCCAG	6
CCACG	10
ACCCG	10
CCAAG	9
CACAG	9
CAACG	1–10;8–9
ACACG	7–9
ACCAG	7
AACCG	1–10;10
ACAAG	1–6;1,5
AAACG	7–9
AAAAG	1,6,9–10
CUCCG	4,7
CCCUG	9
UCCAG	6–8,10
CUCAG	1–10
CCAUG	1–10
UCACG	1–2,4–5
UACCG	1–6,8
ACCUG	4–5;5
ACUCG	6
AUCCG	9
UAACG	4–9
CAAUG	1–6;4
ACUAG	2–3,8–9
ACAUG	10
AUACG	7
AAUCG	10
UAAAG	2
AUAAG	3–10;3,6–9;7
AAAUG	4
UUCCG	1–6,8;4
CUUCG	5–6,8
UCCUG	1–6;4
CCUUG	1
CUCUG	6,8
UCUAG	7
UUCAG	5,7–9;9
CUAUG	5
UACUG	7–10;8–10
UAUCG	7–8
ACUUG	1–6,10
AUCUG	3–5,7–8
AUUCG	2–3,10
UUAAG	1–10;1–2,4,6,8,10
UAAUG	1–2,5,10;2
AUAUG	3–4,9
AAUUG	1–10;1–2,4–6,9
AUUAG	1–10;1–7,9;7
UUUCG	4,7,9
UUCUG	3
UCUUG	8–9
CUUUG	1–3,5,10
UUUAG	2,7
UUAUG	4,9;9
UUUUG	2,9
6-mers	
CCCCAG	4,6
CCCAAG	6,10
CAACCG	8–9
ACCACG	7–9
ACACCG	6–10
AAACCG	8–10
CCCUCG	5,8,10
CCUCAG	5
CUCCAG	5
UCCCAG	2,7–10
CCAUAG	4
CAUACG	1
ACACUG	4–5,7–9
AACCUG	1–6,10;1
AAUCCG	7,9–10
CUAAAG	7–9
UAAACG	1–6,8–10
ACUAAG	9
ACAAUG	1–10
AUAACG	10
AAUACG	1–6,10
AACAUG	10;10
AAACUG	1–10;8–9
AAAUCG	1–3,7
AAUAAG	1–2,4–6
CCCUUG	6,10
CCUCUG	7–9
UCCCUG	1–10
CCUUAG	4,7–8
CUCUAG	1–3
CUUCAG	9
UCCUAG	1–2
UUCCAG	1–6
CCUAUG	3
CUACUG	1–3,6
UCACUG	3,7–9
CUAUCG	7–10
UCAUCG	7,9
CAUCUG	7
ACUCUG	7–8
ACCUUG	4–6
AUCCUG	1–10
UCUAAG	7–8
UUACAG	8
UAUCAG	9
UAUACG	7
UAAUCG	1–10
AUACUG	3,7–8,10
ACAUUG	1
AACUUG	3
AAUCUG	5–9
UAAAUG	4
AUUAAG	1–8
AAUAUG	9
CCUUUG	1–2,5
CUUUCG	10
UCUCUG	1–2,4–6
UUCCUG	5
UCUUAG	5
CUAUUG	1–4,6
UUACUG	10
UAUUCG	3
AUUCUG	2,8–10
ACUUUG	2
UAUAUG	8
CUUUUG	1–5;1
UCUUUG	1,4
UUUUUG	7
7-mers	
ACCCACG	1–9
ACCACCG	7
AACCCCG	7
CCAACAG	7–8
CAACACG	1–2,5–6
CAAACCG	8–9
CCCUACG	1–10
CCCACUG	10
UCCACCG	4–6
CCACCUG	10
AUACCCG	1–10
AACCUCG	8
CCUAAAG	1–6
UAACACG	1–10
AUAACCG	7
AAUCCAG	8–10
AACAUCG	10
AAAUCCG	7–9
UAAAAAG	1,3–6
CCCUUAG	1,3–6
CAUCCUG	7–10
UACUCCG	7
AUCUCCG	8
ACCUUCG	9
UCCUAAG	7
UUACCAG	1–2,4–6,10
CUAACUG	3–4
UAACUCG	1–4,7–8,10
AUUCCAG	7
AUCAUCG	6
AAUCUCG	3
AACCUUG	6
UCUAAAG	10
CUUAAAG	7–9
CAAUAUG	10
AUACUAG	1
AAUCUAG	1–2,4,7–8
AAAUCUG	10
UAAAAUG	10
CUCCUUG	1–3,5–10
UCCCUUG	9
UUCUCCG	7
CUCUUAG	2
UACUUCG	8
UACUCUG	10
UCAUAUG	10
UAAUCUG	4
AAUUUAG	3
UUCUUCG	10
UCUCUUG	7–8
CUUUAUG	10
UUUAUCG	1
UAUUUCG	1
AUUAUUG	10
UUCUUUG	4–6
UAUUUUG	3
UUUUUUG	1–3
8-mers	
CCACAACG	1–3,5–6,9–10
ACCCCAAG	1,5
AAACCCCG	9
UCCACCAG	9
CCCACAUG	7–8
CUCAACCG	8
ACCCUCAG	7
ACCACCUG	1,3–6,8,10
UAACACCG	1–6,10
AUCCCAAG	2–3
AAAUCCCG	1
CCCUCAUG	1,3–4
UACUCCCG	4
AU(CCUC)CG	5
CCUAUCAG	10
CCUAACUG	5
CUUAACCG	4,7,9
UAAUCCCG	9
CUACAAUG	1–10

Table 1. (continued)

Oligonucleotide sequence	Present in organism number
ACAAUCUG	9
AAAUCCUG	1–2,6–9
AUAAACUG	3–6
AUAAAUAG	2
(CU,CCUU)CG	4
AUCCUUCG	4
UCUAACUG	1
CUUAACUG	2–3,5–6
UAAUCCUG	1–3,6
UCUAAAUG	1
UUAAAUCG	10
CAUAAUAG	10
AAAUCUUG	10
AAAUUCUG	2–3
AUAAAUUG	1
CUUUUCAG	6
UUCUCAUG	2
UUUAAUCG	9
UAUCAUUG	9
UUUAAAUG	2–3
UUUAAUUG	1–8,10
UUUUUUCG	2–3
UUUUAUUG	1
9-mers	
CCCACCAAG	4–5
CACACACCG	1–10
(CCA,CAA)CAG	8
CCCAACAAG	7–9
AACCCCAAG	6
AAACCCAAG	4
CCUCACCAG	8
CCUACCAAG	6
CCUACAACG	10
AUAACCCCG	6,8,10
AAACCUCCG	1–6
CACACUAAG	1–6
AUAAACCCG	6
UACUCCCAG	1–3,5–6
UAAUCCCCG	7
AAUCCCCUG	1,3–6
CUUACCAAG	1–3
(UC)ACACAUG	3
(UC)ACAAUCG	2–3
UCAUAACCG	4
CUAAUACCG	3
ACCCUUAAG	7
AUAAUCCCG	9
AUAACCCUG	1–5
AUAAUACCG	4–5
AUAUACAAG	9
UCUUACCAG	10
UCACUAUCG	6
UAAUCCCUG	10
UAAUCCUCG	8
AAUUUCCCG	10
AAUCCUCUG	2
UCAUAAUCG	1,5
CUAAUACUG	1
CAUCAUAUG	10
AUAAUUCCG	10
AAUUAUCCG	7–9
UUUAAAACG	7
UAAACUAUG	7
AUAAUACUG	2
CUAUUACUG	9
UUAAAUUCG	1
UUUAAUAAG	2
UUAUAUUCG	2
UAUUUCUAG	9
UUUAUUAAG	1
CUUUUAUUG	6
UUUUUAUUG	2,4
UUUUUUUCG	1
10-mers	
AAUAACCCCG	7
ACCACCUAUG	9
AAUCUCACCG	8
AAAUCUCACG	4
UAACUCAAAG	8
AAACUUAAAG	1–10
ACCUUACCUG	10
UUACCAUCAG	3
UACCUACUAG	10
AAUCACUUCG	5
AACCCUUAUG	6
UAAAUAACUG	9
UUCUUCACCG	6
ACUCUACUUG	9
CUUAACUAUG	1
AUACUAUUAG	2,4–5
UUCCCUAUUG	4
UCUUCUUAAG	4
AUUUUUUUCG	1
UUUUCUUUUG	5
11-mers	
ACAACUCACCG	10
AAAUCCCACAG	6
CAUCUCACCAG	7,9
UAACUCACCCG	9
AAAUCUCACCG	7,9
AAACACCUUCG	6
AAAUCCCAUAG	5
UCCCUCCCCUG	10
CAUAUCCUCCG	10
AAAUCCUAUAG	3
UUUCAACAUAG	7,9
A(UA,UCA,CUA)UG	6
UUUCAAUAUAG	10
CUUUUCUUAAG	1,3
CUUUUCAUUAG	2
UUCUUUAAUCG	7
UUUUUUUCCUG	1
UUUUUUUUAAG	2
12-mers	
CCACCCAAAAAG	1–2,4,6
UCAAACCACCCG	8–10
UCAAACCAUCCG	7
ACAUCUCACCAG	1–6
CCACUCUUAACG	4–6
CCAUUCUUAACG	1–3
CUCAACUAUUAG	10
CCACUAUUAUUG	7
CAAUUAUUCCUG	2
CCACUUUUAUUG	8
CCAUUUUUAUUG	5
(CUA,CUUUUA)UUG	3
13-mers	
UAAACUACACCUG	10
(CAA,CCA)CAUUCUG	6
UAAUACUCCAUAG	9
UUUCAAAAUAACG	8
AUAAUUUUUCCUG	3
(UUU,CUU,CU)AAAUG	5
14-mers	
AAAACUUUACCAUG	9
AAAACUUUACAAUG	7–8,10
AUUUUU(CCU,CU)UUG	2
15-mers	
UCUAAAACACACCUG	8
AUAACCUACCCUUAG	1–3
AUAACCUAACCUUAG	4
AAUAAUACCCUAUAG	8
AAUAAUACUCCAUAG	7
AUAAUCUACCCUUAG	5
16-mers	
UAAUCCCCUAAACCAG	6
AAAUCCUAUAAUCCCG	4
AAUCUCCUAAACAUAG	5
CAAUCUCUUAAACCUG	7
UAAUCUCCUAAACCUG	4
AAAUCCUAUAAUCCUG	5
17-mers	
CAAUCUUUUAAACCUAG	3
UAAU(CCU,CU)AAACUUAG	1–2
AUAAU(CCU,CU)AAACCUG	9
18-mer	
AACAAUCUCCUAAACCUG	8
24-mer	
(AAACA,UAAUCUCA)—CCCAUCCUUAG	10
termini	
5′ end	
pAG	4,6
pAAUCCG	5
pAAUCUG	1,3
pAUUCUG	2,7–10
3′ end	
AUCACCUCCU$_{OH}$	1–10

First column is oligonucleotide sequence; second column shows organisms in which that sequence is found. Organisms are designated by number

What Do You Do With This?

Fox et al. Methods: S_{AB}

The resulting oligonucleotide catalogs were examined with standard clustering techniques (18). An association coefficient for each binary couple is defined as follows: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total number of residues represented by hexamers and larger in catalog A and in catalog B and their overlap of common sequences, respectively. The association coefficient, S_{AB}, so defined provides what is generally an underestimate of the true degree of homology between two catalogs because related but nonidentical oligomers are not considered.

Fox et al. Methods: S_{AB}

Association constant

$$S_{AB} = 2N_{AB}/(N_A+N_B)$$

- N_A = total number of nucleotides in catalog for A
- N_B = total number in nucleotides in catalog for B
- N_{AB} = total number in the shared catalog

Fox et al. Table 3. S_{AB} Scores

Table 3. S_{AB} values for each indicated binary comparison

Organism	Organism 1	2	3	4	5	6	7	8	9	10	11	12	13
1. *M. arbophilicum*	—												
2. *M. ruminantium PS*	.66	—											
3. *M. ruminantium M-1*	.60	.60	—										
4. *M. formicicum*	.50	.48	.49	—									
5. *M.* sp. *M.o.H.*	.53	.49	.51	.60	—								
6. *M. thermoautotrophicum*	.52	.49	.51	.54	.60	—							
7. Cariaco isolate JR-1	.25	.27	.25	.26	.23	.25	—						
8. Black Sea isolate JR-1	.26	.28	.26	.28	.27	.29	.59	—					
9. *Methanospirillum hungatii*	.20	.24	.21	.23	.23	.22	.51	.52	—				
10. *Methanosarcina barkeri*	.29	.26	.24	.24	.26	.25	.33	.41	.34	—			
11. Enteric-vibrio sp.	.08	.08	.11	.09	.09	.10	.05	.06	.07	.10	—		
12. *Bacillus* sp.	.10	.10	.14	.11	.11	.12	.08	.10	.10	.08	.27	—	
13. Blue-green sp.	.10	.10	.10	.10	.10	.11	.08	.09	.08	.11	.24	.26	—

The values given for enteric-vibrio sp., *Bacillus* sp., and blue-green sp. represent averages obtained from 11 (9), 7 (6), and 4 (23) individual species, respectively.

Two papers for today

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537–4541, October 1977
Evolution

Classification of methanogenic bacteria by 16S ribosomal RNA characterization

(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX*†, LINDA J. MAGRUM*, WILLIAM E. BALCH‡, RALPH S. WOLFE‡, AND CARL R. WOESE*‡

Departments of *Genetics and Development and †Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

Woese and Fox Results

A comparative analysis of these data, summarized in Table 1, shows that the organisms clearly cluster into several primary kingdoms.

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: Kingdom #1

A comparative analysis of these data, summarized in Table 1, shows that the organisms clearly cluster into several primary kingdoms. The first of these contains all of the typical bacteria so far characterized, including the genera *Acetobacterium*, *Acinetobacter*, *Acholeplasma*, *Aeromonas*, *Alcaligenes*, *Anacystis*, *Aphanocapsa*, *Bacillus*, *Bdellovibrio*, *Chlorobium*, *Chromatium*, *Clostridium*, *Corynebacterium*, *Escherichia*, *Eubacterium*, *Lactobacillus*, *Leptospira*, *Micrococcus*, *Mycoplasma*, *Paracoccus*, *Photobacterium*, *Propionibacterium*, *Pseudomonas*, *Rhodopseudomonas*, *Rhodospirillum*, *Spirochaeta*, *Spiroplasma*, *Streptococcus*, and *Vibrio* (refs. 13–17; unpublished data). The group has three major subdivisions, the blue-green bacteria and chloroplasts, the "Gram-positive" bacteria, and a broad "Gram-negative" subdivision (refs. 3, 4, 13–17; unpublished data). It is appropriate to call this urkingdom the *eubacteria*.

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: Kingdom 2

Tree of Life

A second group is defined by the 18S rRNAs of the eukaryotic cytoplasm—animal, plant, fungal, and slime mold (unpublished data).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: Kingdom 2

A second group is defined by the 18S rRNAs of the eukaryotic cytoplasm—animal, plant, fungal, and slime mold (unpublished data). It is uncertain what ancestral organism in the symbiosis that produced the eukaryotic cell this RNA represents. If there had been an "engulfing species" (1) in relation to which all the other organisms were endosymbionts, then it seems likely that 18S rRNA represents that species. This hypothetical group of organisms, in one sense the major ancestors of eukaryotic cells, might appropriately be called *urkaryotes*. Detailed study of anaerobic amoebae and the like (18), which seem not to contain mitochondria and in general are cytologically simpler than customary examples of eukaryotes, might help to resolve this question.

Woese and Fox Results: Kingdom 3

Eubacteria and urkaryotes correspond approximately to the conventional categories "prokaryote" and "eukaryote" when they are used in a phylogenetic sense. However, they do not constitute a dichotomy; they do not collectively exhaust the class of living systems. There exists a third kingdom which, to date, is represented solely by the methanogenic bacteria, a relatively unknown class of anaerobes that possess a unique metabolism based on the reduction of carbon dioxide to methane (19–21). *These "bacteria" appear to be no more related to typical bacteria than they are to eukaryotic cytoplasms.*

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: Kingdom 3

Eubacteria and urkaryotes correspond approximately to the conventional categories "prokaryote" and "eukaryote" when they are used in a phylogenetic sense. However, they do not constitute a dichotomy; they do not collectively exhaust the class of living systems. There exists a third kingdom which, to date, is represented solely by the methanogenic bacteria, a relatively unknown class of anaerobes that possess a unique metabolism based on the reduction of carbon dioxide to methane (19–21). *These "bacteria" appear to be no more related to typical bacteria than they are to eukaryotic cytoplasms.* Although the two divisions of this kingdom appear as remote from one another as blue-green algae are from other eubacteria, they nevertheless correspond to the *same* biochemical phenotype. The apparent antiquity of the methanogenic phenotype plus the fact that it seems well suited to the type of environment presumed to exist on earth 3–4 billion years ago lead us tentatively to name this urkingdom the *archaebacteria*. Whether or not other biochemically distinct phenotypes exist in this kingdom is clearly an important question upon which may turn our concept of the nature and ancestry of the first prokaryotes.

Woese and Fox Results: Three Kingdoms

Table 1 shows the three urkingdoms to be equidistant from one another. Because the distances measured are actually proportional to numbers of mutations and not necessarily to time, it cannot be proven that the three lines of descent branched from the common ancestral line at about the same time. One of the three may represent a far earlier bifurcation than the other two, making there in effect only two urkingdoms. Of the three possible unequal branching patterns the case for which the initial bifurcation defines urkaryotes vs. all bacteria requires further comment because, as we have seen, there is a predilection to accept such a dichotomy.

Woese and Fox Results: Three Kingdoms

Table 1 shows the three urkingdoms to be equidistant from one another. Because the distances measured are actually proportional to numbers of mutations and not necessarily to time, it cannot be proven that the three lines of descent branched from the common ancestral line at about the same time. One of the three may represent a far earlier bifurcation than the other two, making there in effect only two urkingdoms. Of the three possible unequal branching patterns the case for which the initial bifurcation defines urkaryotes vs. all bacteria requires further comment because, as we have seen, there is a predilection to accept such a dichotomy.

What Was This Based On?

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Woese and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Is There Another Way to View This?

Two papers for today

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Communicated by T. M. Sonneborn, August 18, 1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537–4541, October 1977
Evolution

Classification of methanogenic bacteria by 16S ribosomal RNA characterization

(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX* †, LINDA J. MAGRUM*, WILLIAM E. BALCH‡, RALPH S. WOLFE‡, AND CARL R. WOESE*‡

Departments of *Genetics and Development and †Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

Fox et al. 1977. Table 3. S_{AB} Scores

Table 3. S_{AB} values for each indicated binary comparison

Organism	Organism 1	2	3	4	5	6	7	8	9	10	11	12	13
1. *M. arbophilicum*	—												
2. *M. ruminantium PS*	.66	—											
3. *M. ruminantium M-1*	.60	.60	—										
4. *M. formicicum*	.50	.48	.49	—									
5. *M.* sp. *M.o.H.*	.53	.49	.51	.60	—								
6. *M. thermoautotrophicum*	.52	.49	.51	.54	.60	—							
7. Cariaco isolate JR-1	.25	.27	.25	.26	.23	.25	—						
8. Black Sea isolate JR-1	.26	.28	.26	.28	.27	.29	.59	—					
9. *Methanospirillum hungatii*	.20	.24	.21	.23	.23	.22	.51	.52	—				
10. *Methanosarcina barkeri*	.29	.26	.24	.24	.26	.25	.33	.41	.34	—			
11. Enteric-vibrio sp.	.08	.08	.11	.09	.09	.10	.05	.06	.07	.10	—		
12. *Bacillus* sp.	.10	.10	.14	.11	.11	.12	.08	.10	.10	.08	.27	—	
13. Blue-green sp.	.10	.10	.10	.10	.10	.11	.08	.09	.08	.11	.24	.26	—

The values given for enteric-vibrio sp., *Bacillus* sp., and blue-green sp. represent averages obtained from 11 (9), 7 (6), and 4 (23) individual species, respectively.

What Else Did They Do With This?

Fox et al. Methods

The resulting oligonucleotide catalogs were examined with standard clustering techniques (18). An association coefficient for each binary couple is defined as follows: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total number of residues represented by hexamers and larger in catalog A and in catalog B and their overlap of common sequences, respectively. The association coefficient, S_{AB}, so defined provides what is generally an underestimate of the true degree of homology between two catalogs because related but nonidentical oligomers are not considered. The matrix of S_{AB} values for each binary comparison among the members of a given set of organisms is used to generate a dendrogram by average linkage (between the merged groups) clustering. The resulting dendrogram is, strictly speaking, phyletic because no "ancestral catalog" has been postulated. However, it is clear from the molecular nature of the data that the topology of this dendrogram would closely resemble, if not be identical to, that of a phylogenetic tree based upon such ancestral catalogs.

Fox et al. 1977. Figure 1. Dendrogram.

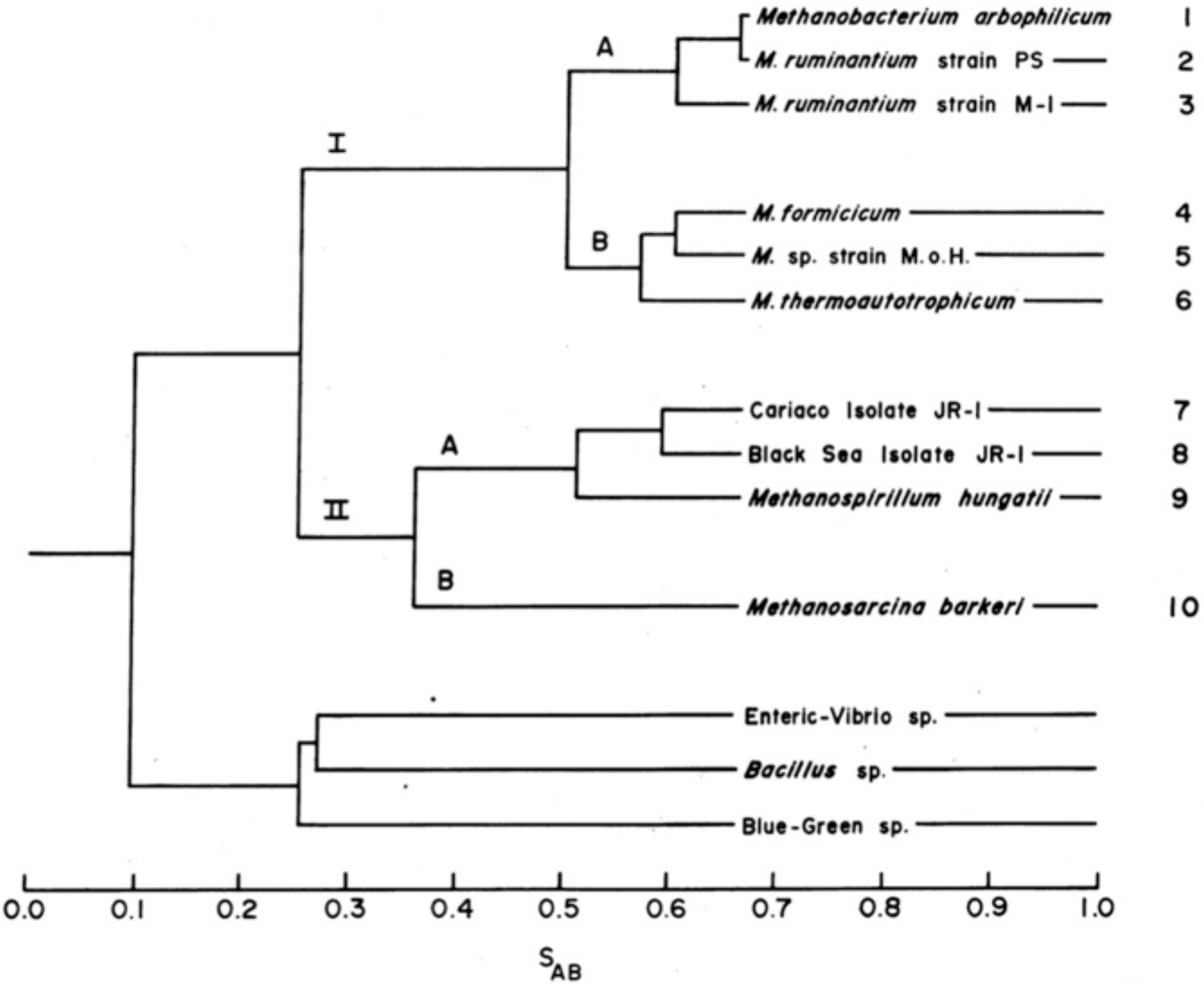


FIG. 1. Dendrogram of relationships of methanogens and typical bacteria. The figure was constructed by average linkage clustering (between the merged groups) from the S_{AB} values given in Table 3.

Worse and Fox Results: S_{AB} Table For 13 Species

Proc. Natl. Acad. Sci. USA 74 (1977)

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Worse and Fox Results: S_{AB} Table For 13 Species

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *Saccharomyces cerevisiae*, 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. *Lemna minor*, 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. *Escherichia coli*	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. *Chlorobium vibrioforme*	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. *Bacillus firmus*	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. *Corynebacterium diphtheriae*	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. *Aphanocapsa* 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (*Lemna*)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. *Methanobacterium thermoautotrophicum*	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. *M. ruminantium* strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. *Methanobacterium* sp., Cariaco isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. *Methanosarcina barkeri*	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

The 16S (18S) ribosomal RNA from the organisms (organelles) listed were digested with T1 RNase and the resulting digests were subjected to two-dimensional electrophoretic separation to produce an oligonucleotide fingerprint. The individual oligonucleotides on each fingerprint were then sequenced by established procedures (13, 14) to produce an oligonucleotide catalog characteristic of the given organism (3, 4, 13–17, 22, 23; unpublished data). Comparisons of all possible pairs of such catalogs defines a set of association coefficients (S_{AB}) given by: $S_{AB} = 2N_{AB}/(N_A + N_B)$, in which N_A, N_B, and N_{AB} are the total numbers of nucleotides in sequences of hexamers or larger in the catalog for organism A, in that for organism B, and in the interreaction of the two catalogs, respectively (13, 23).

Why No Tree?

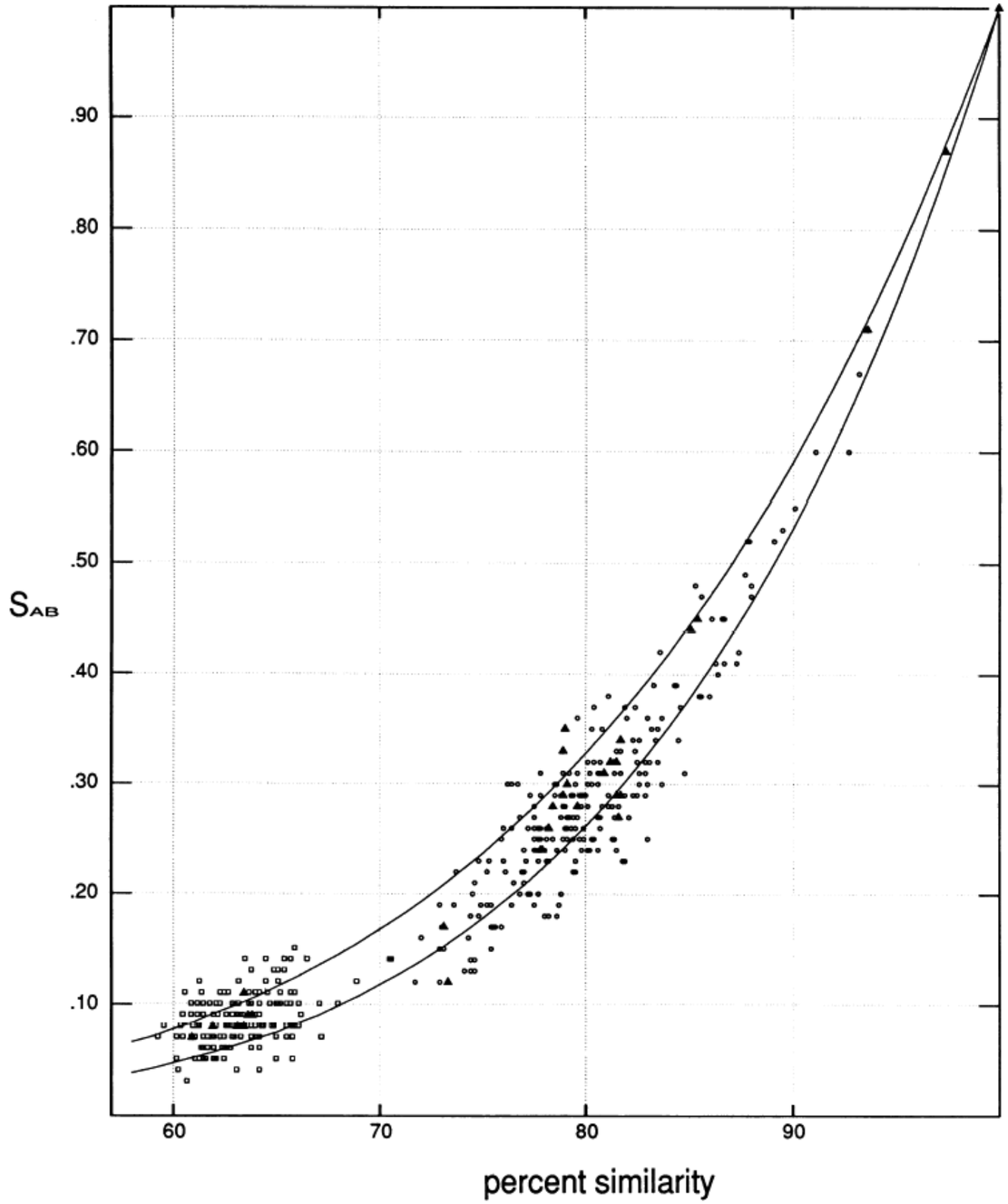
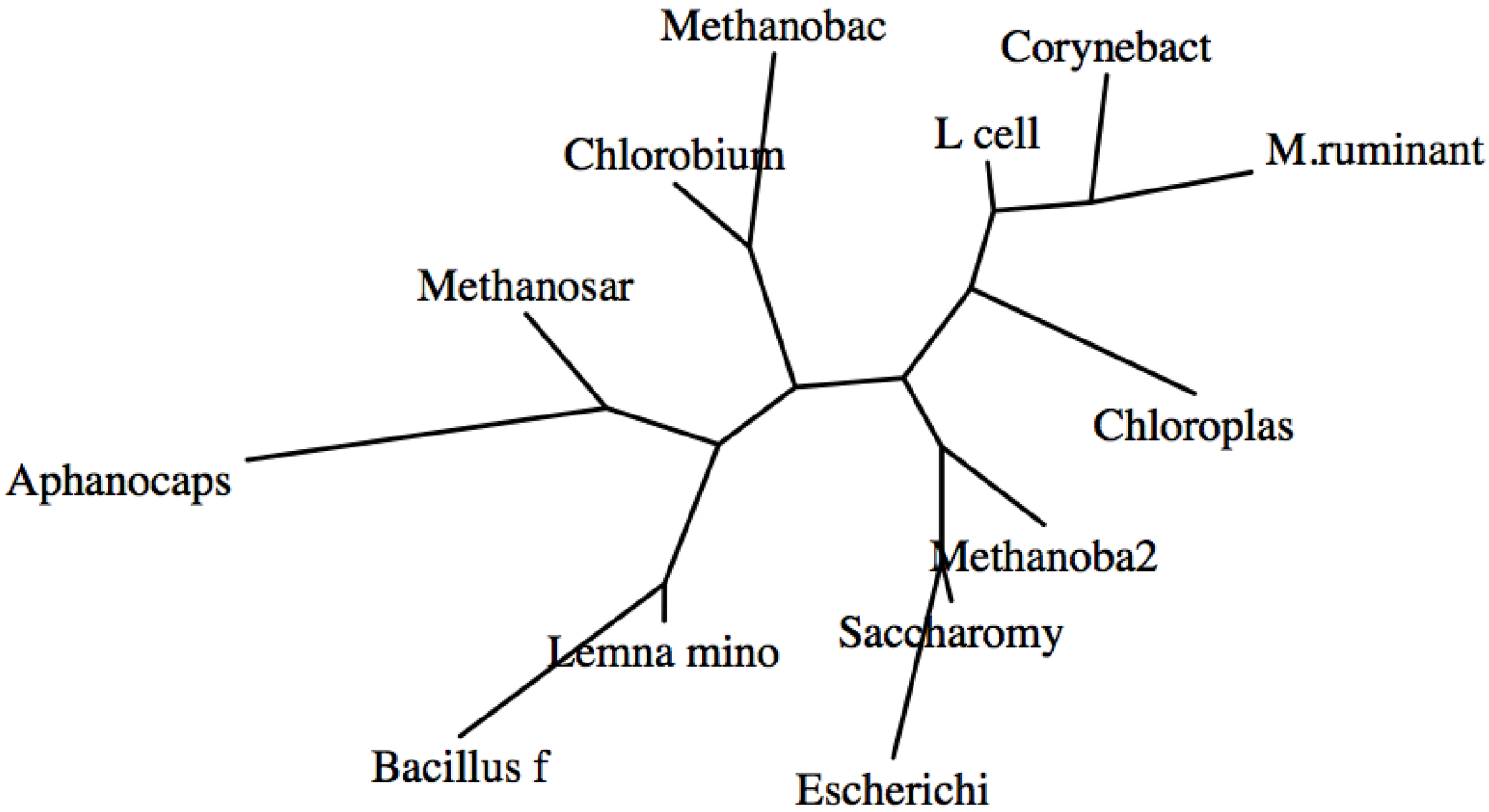


FIG. 2. Plot of percent sequence similarity versus binary association coefficient (S_{AB} value [55]), for a representative sampling of eubacterial and archaebacterial sequences (unpublished analysis). The two theoretical curves are X^5 (upper curve) and X^6 (lower curve), where X = percent similarity. Symbols: ○, values for pairs of eubacterial sequences; □, values for eubacteria with archaebacteria; ▲, values for *E. coli* with either eubacteria or archaebacteria.

S_{AB}

	13												
Saccharomy	0.00	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
Lemna_mino	0.29	0.00	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
L_cell	0.33	0.36	0.00	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
Escherichi	0.05	0.10	0.06	0.00	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
Chlorobium	0.06	0.05	0.06	0.24	0.00	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
Bacillus_f	0.08	0.06	0.07	0.25	0.22	0.00	0.34	0.26	0.20	0.11	0.13	0.06	0.12
Corynebact	0.09	0.10	0.07	0.28	0.22	0.34	0.00	0.23	0.21	0.12	0.12	0.09	0.10
Aphanocaps	0.11	0.09	0.09	0.26	0.20	0.26	0.23	0.00	0.31	0.11	0.11	0.10	0.10
Chloroplas	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	0.00	0.14	0.12	0.10	0.12
Methanobac	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	0.00	0.51	0.25	0.30
M.ruminanti	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	0.00	0.25	0.24
Methanoba2	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	0.00	0.32
Methanosar	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	0.00

http://bioweb.pasteur.fr/seqanal/interfaces/neighbor.html



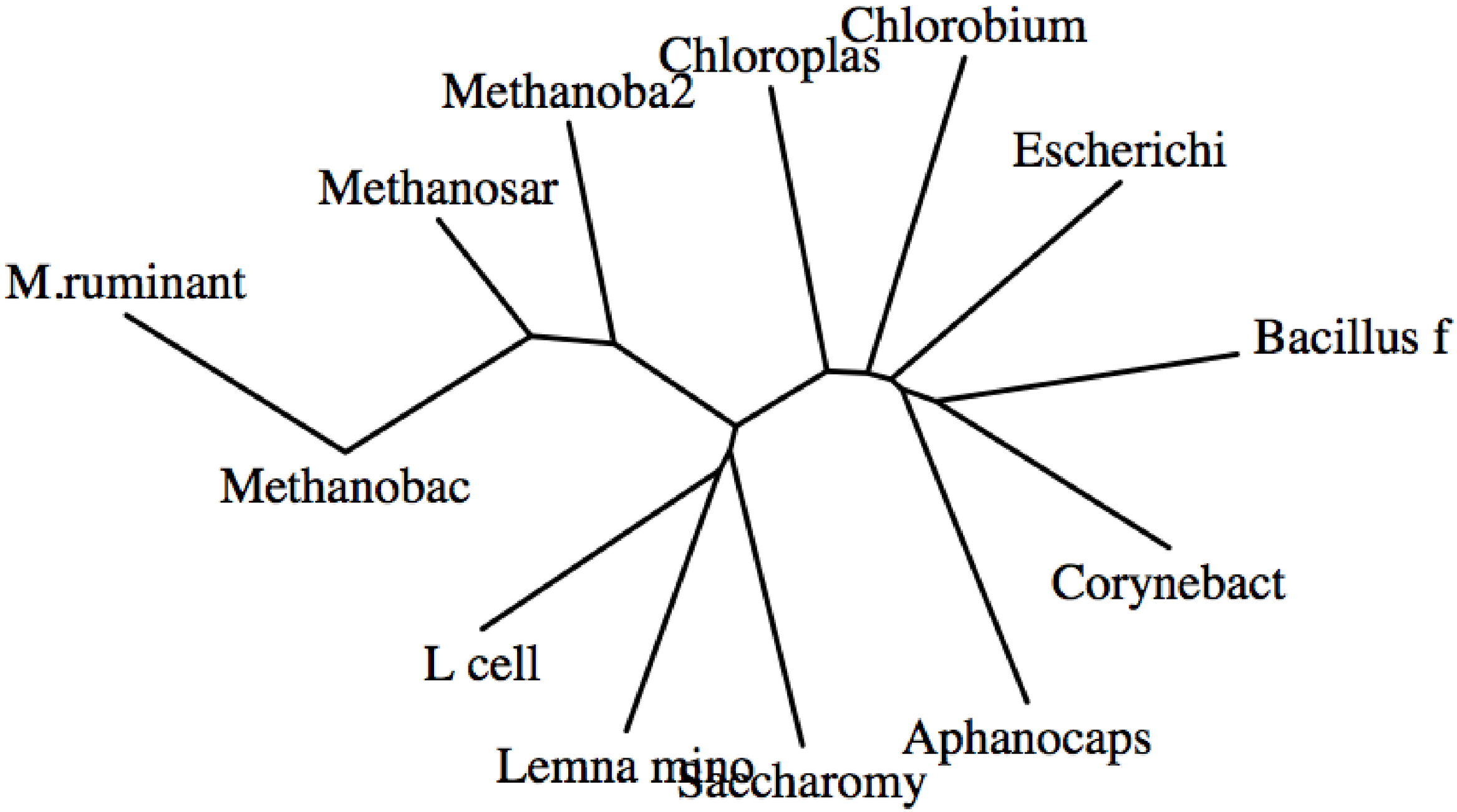
S_{AB}

	13												
Saccharomy	0.00	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
Lemna_mino	0.29	0.00	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
L_cell	0.33	0.36	0.00	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
Escherichi	0.05	0.10	0.06	0.00	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
Chlorobium	0.06	0.05	0.06	0.24	0.00	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
Bacillus_f	0.08	0.06	0.07	0.25	0.22	0.00	0.34	0.26	0.20	0.11	0.13	0.06	0.12
Corynebact	0.09	0.10	0.07	0.28	0.22	0.34	0.00	0.23	0.21	0.12	0.12	0.09	0.10
Aphanocaps	0.11	0.09	0.09	0.26	0.20	0.26	0.23	0.00	0.31	0.11	0.11	0.10	0.10
Chloroplas	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	0.00	0.14	0.12	0.10	0.12
Methanobac	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	0.00	0.51	0.25	0.30
M.ruminanti	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	0.00	0.25	0.24
Methanoba2	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	0.00	0.32
Methanosar	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	0.00

$$D_{AB} = (1 - S_{AB})$$

	1	3										
Saccharomy	0	0.71	0.67	0.95	0.94	0.92	0.91	0.89	0.92	0.89	0.89	0.92	0.92
Lemna_mino	0.71	0	0.64	0.9	0.95	0.94	0.9	0.91	0.89	0.9	0.9	0.87	0.93
L_cell	0.67	0.64	0	0.94	0.94	0.93	0.93	0.91	0.94	0.9	0.9	0.91	0.93
Escherichi	0.95	0.9	0.94	0	0.76	0.75	0.72	0.74	0.79	0.89	0.88	0.93	0.88
Chlorobium	0.94	0.95	0.94	0.76	0	0.78	0.78	0.8	0.81	0.94	0.93	0.94	0.91
Bacillus_f	0.92	0.94	0.93	0.75	0.78	0	0.66	0.74	0.8	0.89	0.87	0.94	0.88
Corynebact	0.91	0.9	0.93	0.72	0.78	0.66	0	0.77	0.79	0.88	0.88	0.91	0.9
Aphanocaps	0.89	0.91	0.91	0.74	0.8	0.74	0.77	0	0.69	0.89	0.89	0.9	0.9
Chloroplas	0.92	0.89	0.94	0.79	0.81	0.8	0.79	0.69	0	0.86	0.88	0.9	0.88
Methanobac	0.89	0.9	0.9	0.89	0.94	0.89	0.88	0.89	0.86	0	0.49	0.75	0.7
M.ruminanti	0.89	0.9	0.9	0.88	0.93	0.87	0.88	0.89	0.88	0.49	0	0.75	0.76
Methanoba2	0.92	0.87	0.91	0.93	0.94	0.94	0.91	0.9	0.9	0.75	0.75	0	0.68
Methanosar	0.92	0.93	0.93	0.88	0.91	0.88	0.9	0.9	0.88	0.7	0.76	0.68	0

Tree from Woese and Fox data



http://bioweb.pasteur.fr/seqanal/interfaces/neighbor.html

Woese 1987

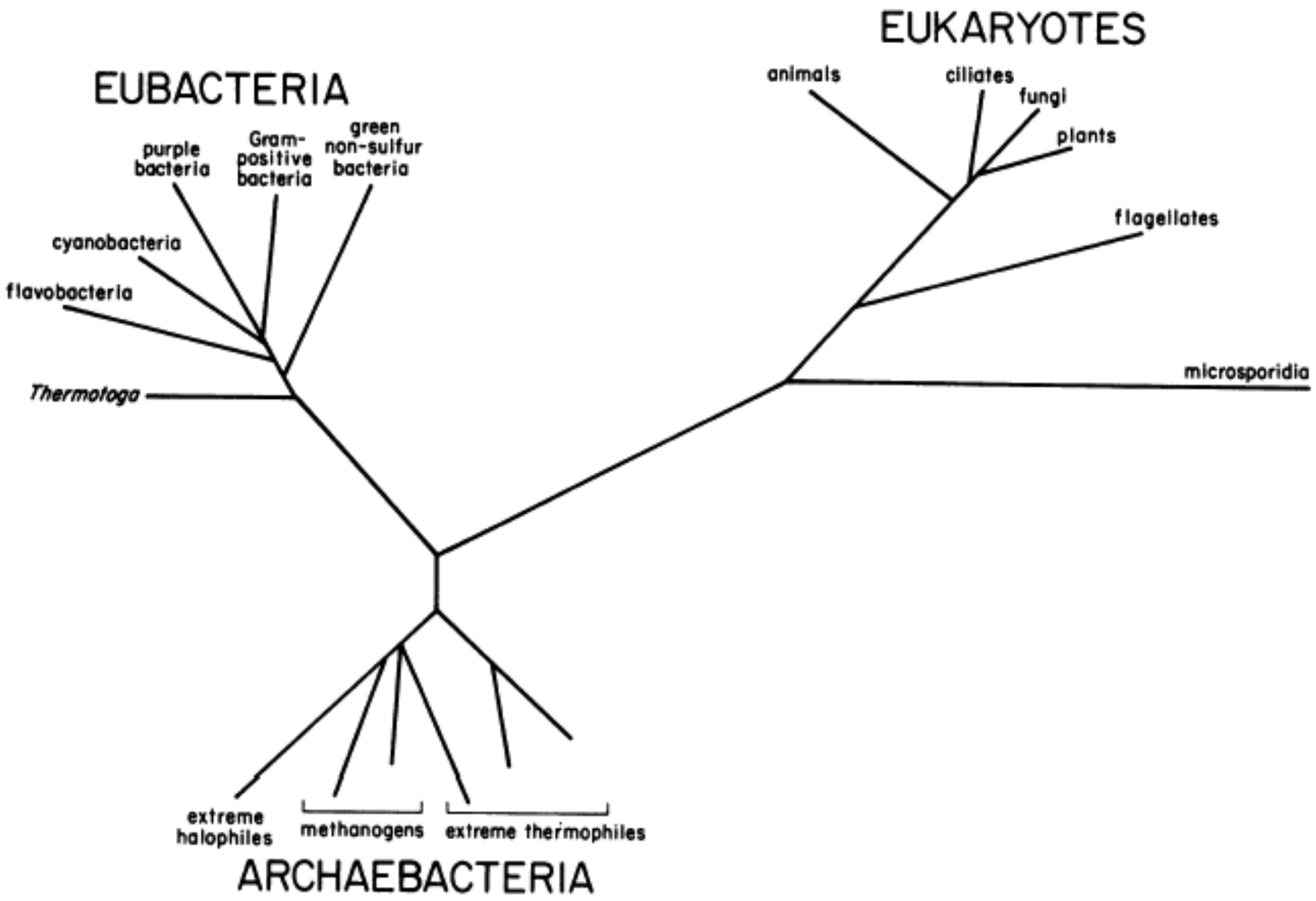
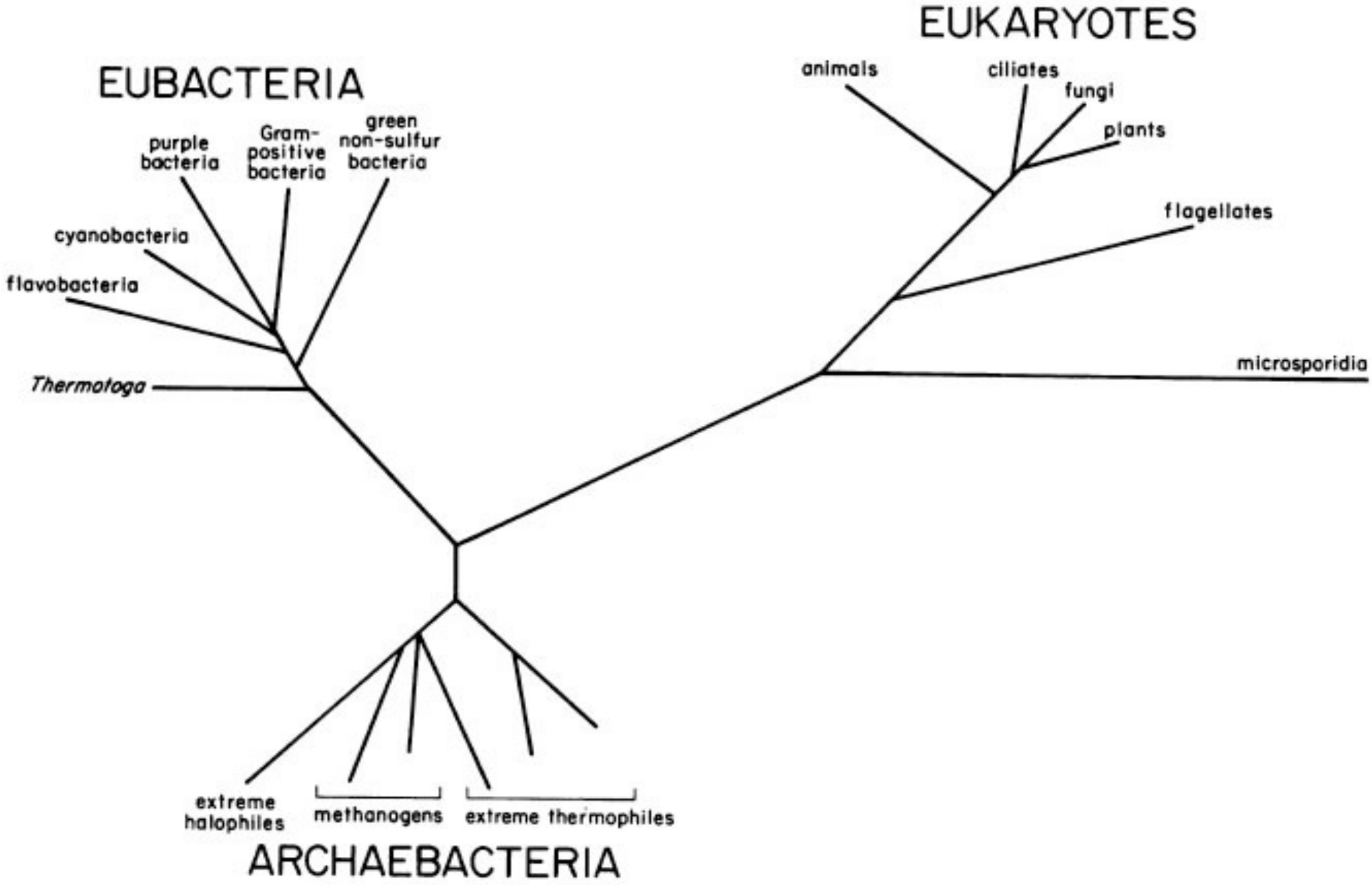
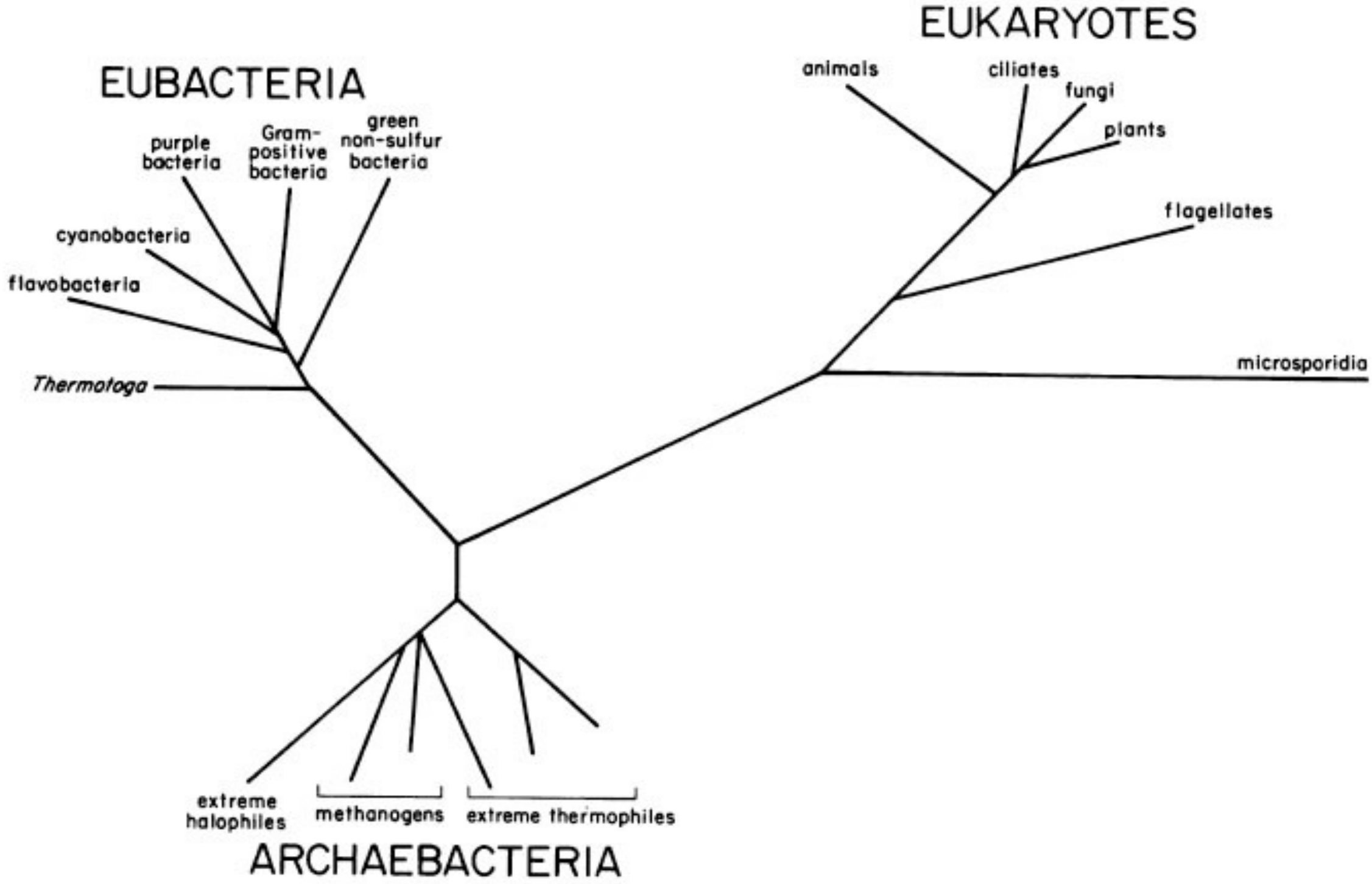


FIG. 4. Universal phylogenetic tree determined from rRNA sequence comparisons. A matrix of evolutionary distances (99) was calculated from an alignment (260) of representative 16S rRNA sequences from each of the three urkingdoms. This was used to construct a distance tree (36), based upon those positions represented in all sequences in the alignment in homologous secondary structural elements (75). Line lengths on the tree are proportional to calculated distances. The alignment includes the following eubacterial sequences: *Thermotoga maritima* (1); green non-sulfur bacteria, *Thermomicrobium roseum* (162); flavobacteria, *Flavobacterium heparinum* (234); cyanobacteria, *Anacystis nidulans* (224); gram-positive bacteria, *Bacillus subtilis* (68); and purple bacteria, *Escherichia coli* (19); the following archaebacterial sequences: extreme halophiles, *Halobacterium volcanii* (72); methanogens, *Methanococcus vannielii* (96) and *Methanobacterium formicicum* (124); and extreme thermophiles, *Thermococcus celer* (Woese et al., unpublished data), *Desulfurococcus mobilis* (R. Garrett, personal communication), and *Thermoproteus tenax* (126); and the following eucaryotic sequences: microsporidia, *Vairimorpha necatrix* (226a); flagellates, *Euglena gracilis* (196); cellular slime molds, *Dictyostelium discoideum* (145); ciliates, *Paramecium tetraurelia* (195); fungi, *Saccharomyces cerevisiae* (179); plants, *Zea mays* (147); and animals, *Xenopus laevis* (181). Branching order within each kingdom is correct to a first approximation only. See the trees for the individual kingdoms for precise branching orders.

Woese 1987

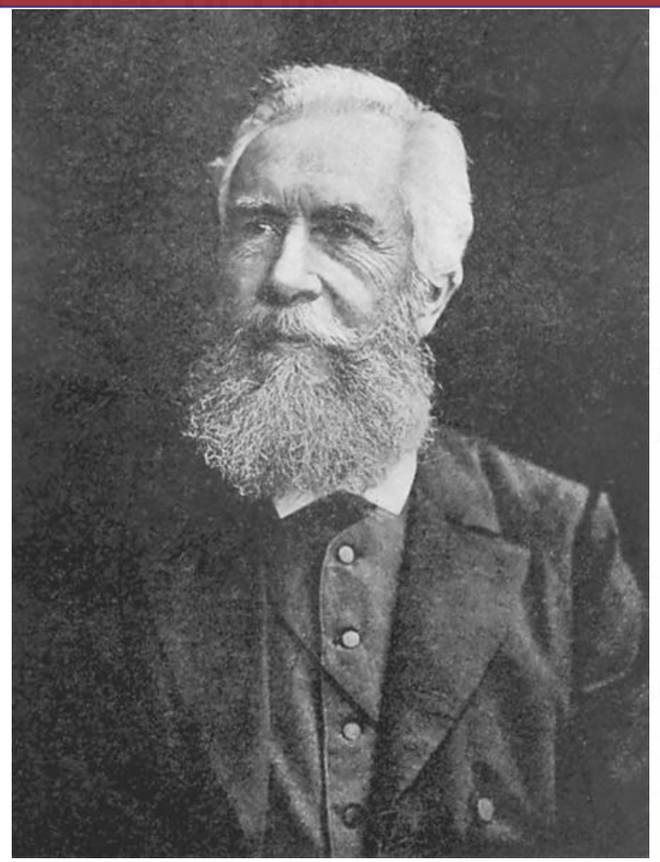

EUBACTERIA
purple bacteria
Gram-positive bacteria
green non-sulfur bacteria
cyanobacteria
flavobacteria
Thermotoga
EUKARYOTES
animals
ciliates
fungi
plants
flagellates
microsporidia
extreme halophiles
methanogens
extreme thermophiles
ARCHAEBACTERIA

Woese 1987

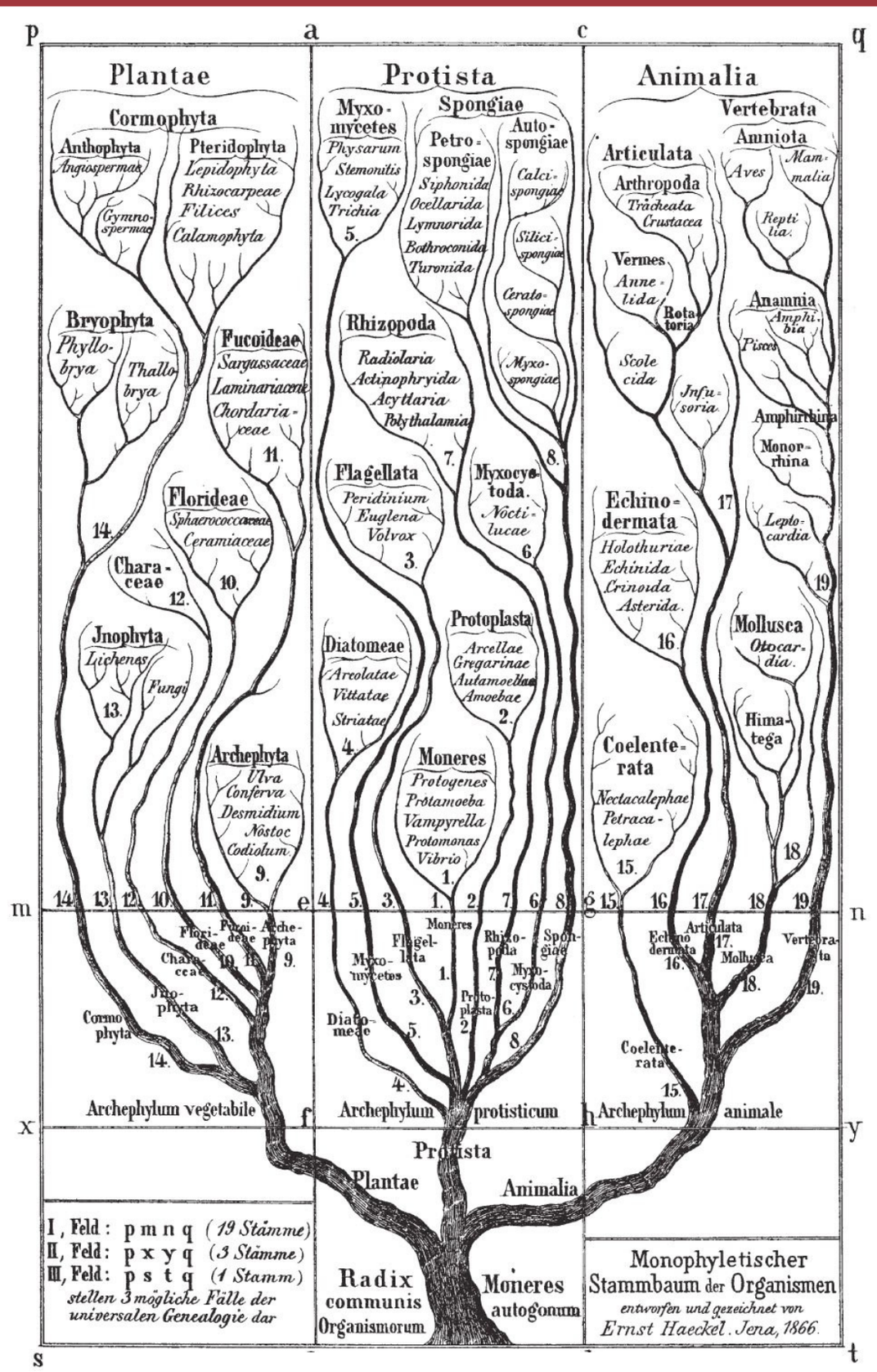


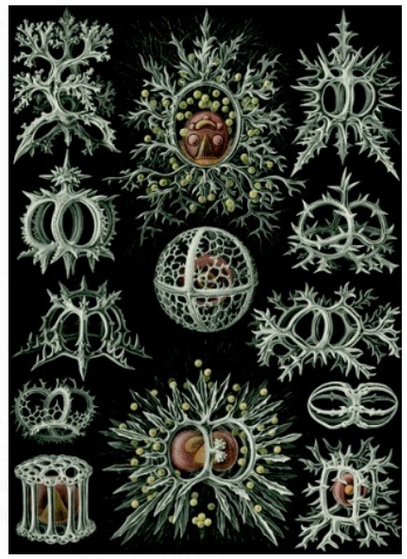
Why Is This Interesting?

Ernst Haeckel 1866



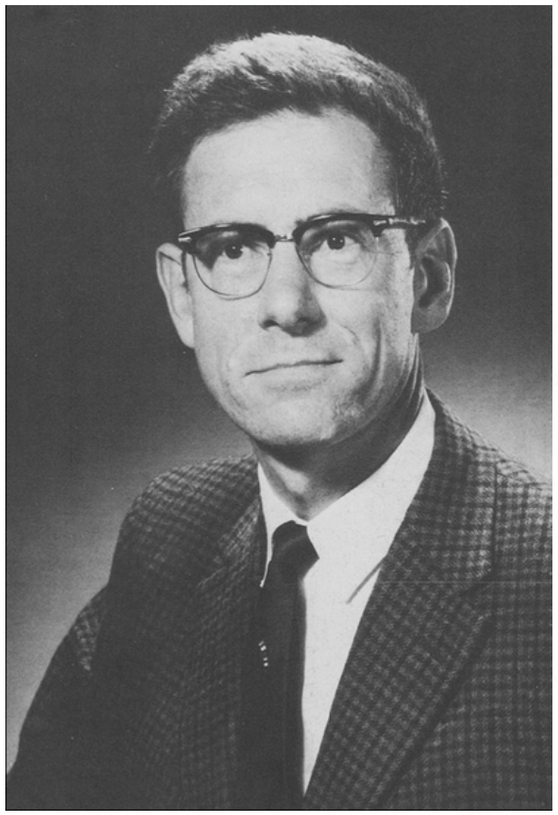
Plantae
Protista
Animalia



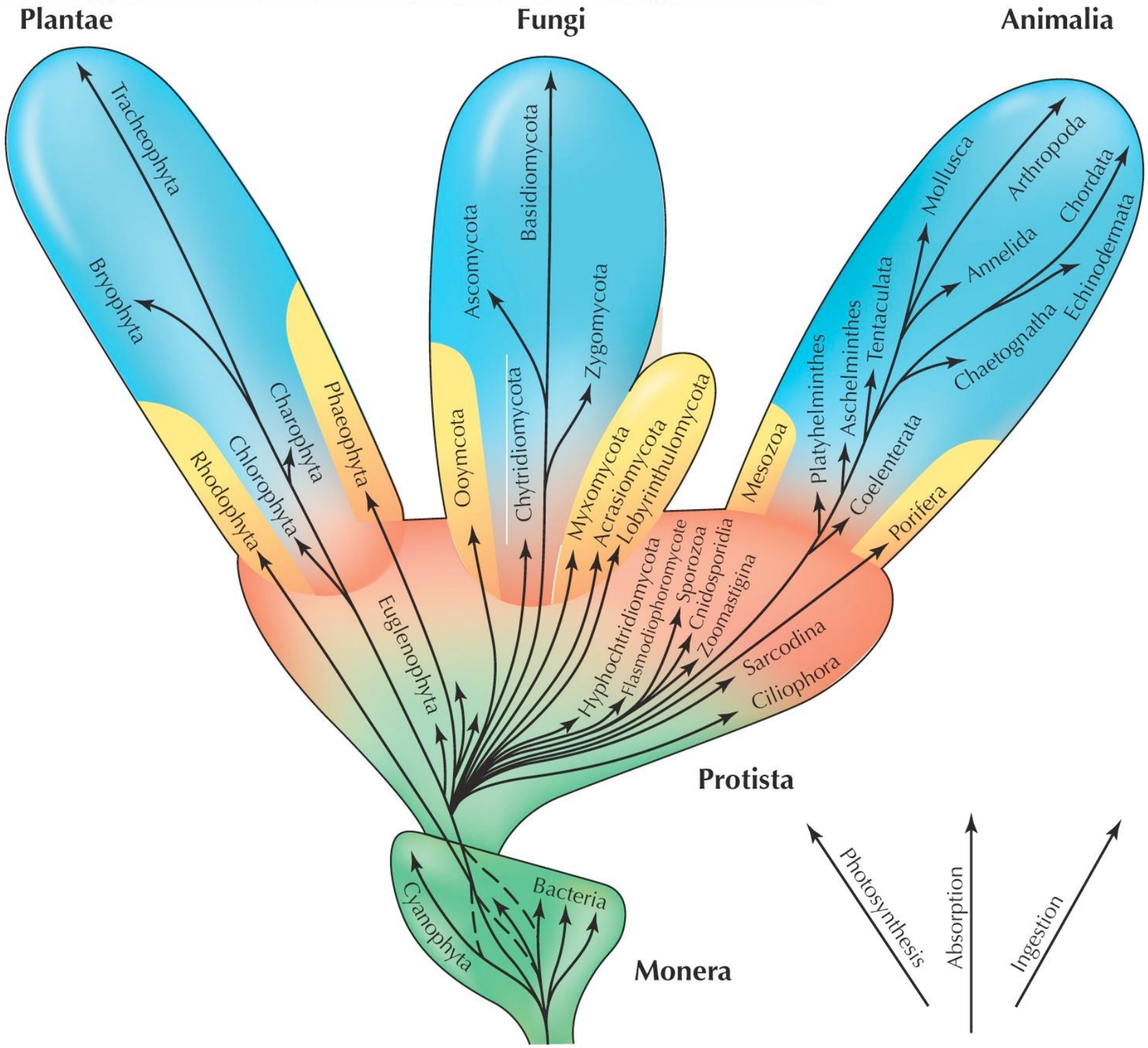


Evolution © 2007 Cold Spring Harbor Laboratory Press

Whittaker – Five Kingdoms 1969



Monera
Protista
Plantae
Fungi
Animalia



Woese and Fox Discussion

The phenotype of the methanogens, although ostensibly "bacterial," on close scrutiny gives no indication of a specific phylogenetic resemblance to the eubacteria. For example, methanogens do have cell walls, but these do not contain peptidoglycan (24). The biochemistry of methane formation appears to involve totally unique coenzymes (23, 25, 26). The methanogen rRNAs are comparable in size to their eubacterial counterparts, but resemble the latter specifically in neither sequence (Table 1) nor in their pattern of base modification (23). The tRNAs from eubacteria and eukaryotes are characterized by a common modified sequence, TΨCG; methanogens modify this tRNA sequence in a quite different and unique way (23). It must be recognized that very little is known of the general biochemistry of the methanogens—and almost nothing is known regarding their molecular biology. Hence, although the above points are few in number, they represent most of what is now known. There is no reason at present to consider methanogens as any closer to eubacteria than to the "cytoplasmic component" of the eukaryote. Both in terms of rRNA sequence measurement and in terms of general phenotypic differences, then, the three groupings appear to be distinct urkingdoms.

Woese and Fox Discussion

If a third urkingdom exists, does this suggest that many more such will be found among yet to be characterized organisms? We think not, although the matter clearly requires an exhaustive search. As seen above, the number of species that can be classified as eubacteria is moderately large. To this list can be added *Spirillum* and *Desulfovibrio*, whose rRNAs appear typically eubacterial by nucleic acid hybridization measurements (27). Because the list is also phenotypically diverse, it seems unlikely that many, if any, of the yet uncharacterized

prokaryotic groups will be shown to have coequal status with the present three. Conceivably the halophiles whose cell walls contain no peptidoglycan, are candidates for this distinction (28, 29).

Woese and Fox Discussion

Eukaryotic organelles, however, could be a different matter. There can be no doubt that the chloroplast is of specific eubacterial origin (3, 4). A question arises with the remaining organelles and structures. Mitochondria, for example, do not conform well to a "typically prokaryotic" phenotype, which has led some to conclude that they could not have arisen as endosymbionts (30). By using "prokaryote" in a phylogenetic sense, this formulation of the issue does not recognize a third alternative—that the organelle in question arose endosymbiotically from a separate line of descent whose phenotype is not "typically prokaryotic" (i.e., eubacterial). It is thus conceivable that some endosymbiotically formed structures represent still other major phylogenetic groups; some could even be the only extant representation thereof.

Woese and Fox Discussion

The question that remains to be answered is whether the common ancestor of all three major lines of descent was itself a prokaryote. If not, each urkingdom represents an independent evolution of the prokaryotic level of organization. Obviously, much more needs to be known about the general properties of all the urkingdoms before this matter can be definitely settled. At present we can point to two arguments suggesting that each urkingdom does represent a separate evolution of the prokaryotic level of organization.

Woese and Fox Discussion

The first argument concerns the stability of the general phenotypes. The general eubacterial phenotype has been stable for at least 3 billion years—i.e., the apparent age of blue-green algae (31). The methanogenic phenotype seems to be at least this old in that branchings within the two urkingdoms are comparably deep (see Table 1). The time available to form each phenotype (from their common ancestor) is then short by comparison, which seems paradoxical in that the two phenotypes are so fundamentally different. We think that this ostensible paradox implies that the common ancestor in this case was not a prokaryote. It was a far simpler entity; it probably did not evolve at the "slow" rate characteristic of prokaryotes; it did not possess many of the features possessed by prokaryotes, and so these evolved independently and differently in separate lines of descent.

Woese and Fox Discussion

The second argument concerns the quality of the differences in the three general phenotypes. It seems highly unlikely, for example, that differences in general patterns of base modification in rRNAs and tRNAs are related to the niches that organisms occupy. Rather, differences of this nature imply independent evolution of the properties in question. It has been argued elsewhere that features such as RNA base modification generally represent the final stage in the evolution of translation (32). If these features have evolved separately in two lines of descent, their common ancestor, lacking them, had a more rudimentary version of the translation mechanism and consequently, could not have been as complex as a prokaryote (6).

Woese and Fox Discussion

With the identification and characterization of the urkingdoms we are for the first time beginning to see the overall phylogenetic structure of the living world. It is not structured in a bipartite way along the lines of the organizationally dissimilar prokaryote and eukaryote. Rather, it is (at least) tripartite, comprising (*i*) the typical bacteria, (*ii*) the line of descent manifested in eukaryotic cytoplasms, and (*iii*) a little explored grouping, represented so far only by methanogenic bacteria.

Woese and Fox Abstract

- Abstract: A phylogenetic analysis based upon ribosomal RNA sequence characterization reveals that living systems represent one of three aboriginal lines of descent: (i) the eubacteria, comprising all typical bacteria; (ii) the archaebacteria, containing methanogenic bacteria; and (iii) the urkaryotes, now represented in the cytoplasmic component of eukaryotic cells.