J Mol Evol (1986) 24:152–166

Journal of
Molecular Evolution

Cladistic Analysis of 5S rRNA and 16S rRNA Secondary and Primary Structure—The Evolution of Eukaryotes and Their Relation to Archaebacteria*

Jörn Wolters and Volker A. Erdmann

Institut für Biochemie, Fachbereich Chemie, Freie Universität Berlin, Thielallee 63, D-1000 Berlin 33 (Dahlem), FRG

Summary. The secondary structure of 5S rRNA has been elucidated by a cladistic analysis resulting in minimal models for eukaryotes, eubacteria, and halophilic–methanogenic archaebacteria, as well as for an ur-5S rRNA. This ancestor of all present-day 5S rRNA molecules is compared with an ur-tRNA and can be fitted into a tRNA-like structure allowing tertiary-structure interactions at the equivalent positions. A phylogenetic analysis of eukaryotic 5SrRNA and 16S rRNA sequences confirms particular monophyletic taxa: rhodophytes (red algae), chlorobionts (green algae and plants), metazoans (multicellular animals), euglenozoans (euglenids and trypanosomatids), a group of zygomycetes (excluding Kickxellales), a group of ascomycetes (excluding Protomycetales), two distinct groups of basidiomycetes, and a group consisting of phaeophyceans (brown algae) and oomycetes (water molds). The Euglenozoa show a distinct relation to the Eumycota (true fungi) and Metazoa. An analysis of archaebacterial sequences substantiates the paraphyletic nature of this third urkingdom defining the eubacteria as a sister group of the halophile-methanogens and defining the eukaryotes as a sister group of a particular lineage of the eocytes/sulfur-dependents. The latter fact implies that even the eocytes/sulfur-dependent archaebacteria are paraphyletic.

Key words: 5S rRNA — 16S rRNA — Archaebacteria — Cladistics

Offprint requests to: J. Wolters

* Presented at the FEBS Symposium on Genome Organization and Evolution, held in Crete, Greece, September 1–5, 1986 Dedicated to the memory of Erik Huysmans who died on July 8, 1986, at the age of 29.

Introduction

Most sequence comparisons are usually based on the neutral theory of molecular evolution (Kimura 1968; King and Jukes 1969) and yield phenograms typical for numerical taxonomy.

Phylogenetic systematics sensu Hennig (1966), called cladistics in the Anglo-American literature, has for a long time been a tool for morphologists only and has hardly found entry into textbooks. Cladistic computer methods for sequence comparison such as the maximum parsimony method, which requires the reconstruction of ancestral sequences, are limited because the data cannot exceed a certain number of sequences and a certain number of nucleotides.

A pure cladistic analysis makes a distinction between plesiomorphic (primitive, ancient) and apomorphic (derived) characteristics. Application of cladistic methods to nucleic acid secondary and primary structure requires characters of very low variability: One of the alternatives should be easily definable as plesiomorphic. Such characters include insertions/deletions, odd base pairs (non-Watson–Crick base pairs that nonetheless do not disturb helical conformation), and so-called signature nucleotides. In this article we report a broad cladistic analysis based on 5S and 16S rRNA primary and secondary structure.

Materials and Methods

The published 5S rRNA sequences are compiled in the Berlin RNA Databank (Erdmann and Wolters 1986) and consist of 218 eukaryotic, 113 eubacterial, 11 plastid, 4 mitochondrial, and 15

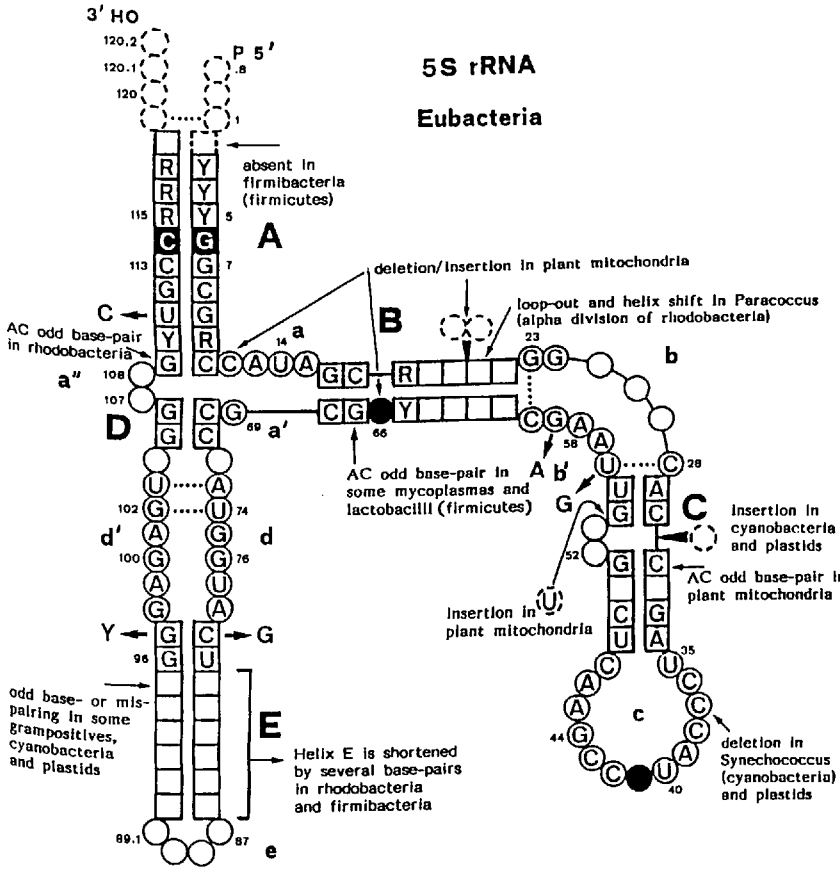

Fig. 1. Minimal model of eubacterial 5S rRNA secondary structure. Squares indicate conserved base pairing; circles, unpaired nucleotides. Dotted lines indicate possible helix extensions. Filled squares and circles indicate positions that are unique to this model. Where assignment was possible, plesiomorphic bases are indicated by letters at particular positions. Differing bases in the majority of groups are marked with arrows. Occurrences of AC and UU odd base pairs are also marked

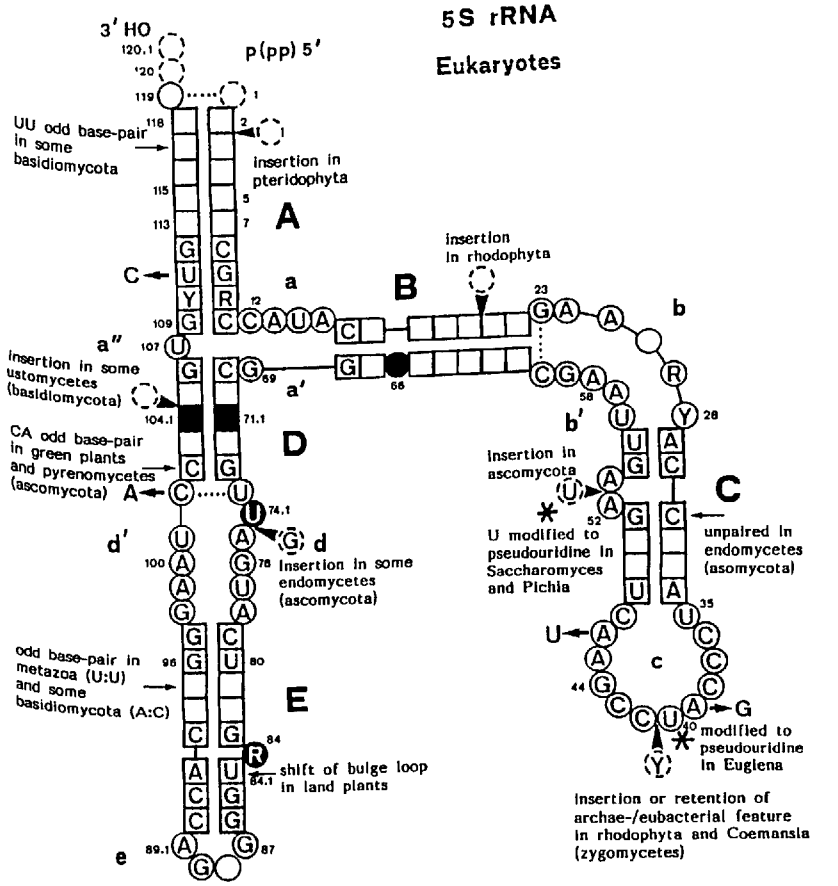

Fig. 2. Minimal model of eukaryotic 5S rRNA secondary structure. Symbols as in Fig. 1

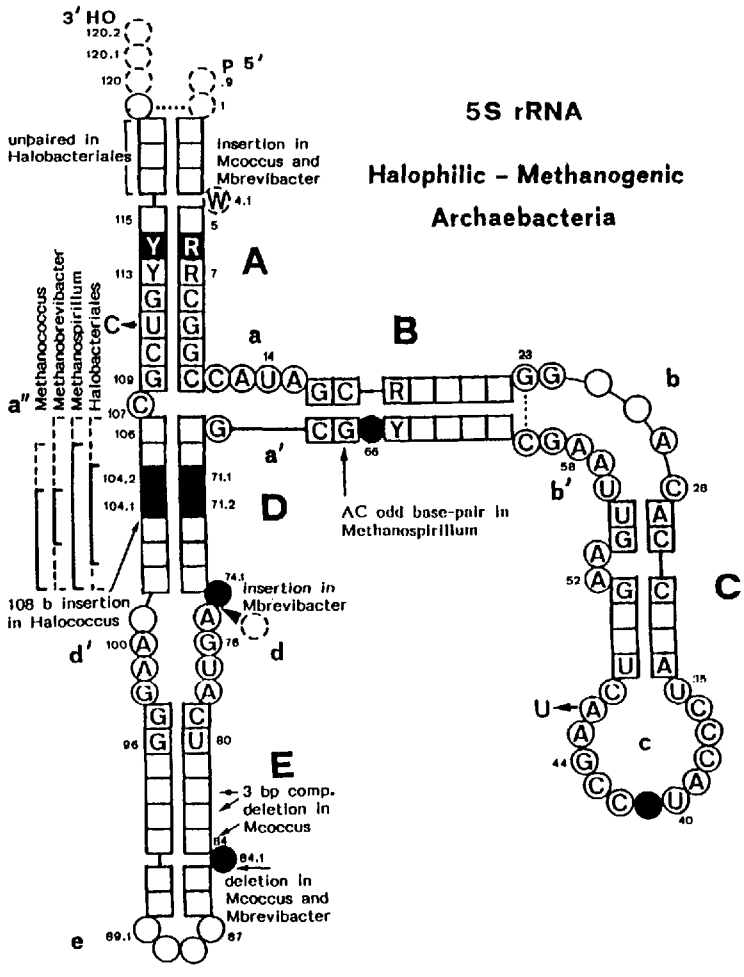

Fig. 3. Minimal model of 5S rRNA secondary structure of halophilic–methanogenic archaebacteria. Symbols as in Fig. 1

archaebacterial sequences. The cyanelle and cytoplasmic sequences of *Cyanophora paradoxa* are from Maxwell and Liu (1986). The 16S rRNA sequences are from the biannual compilation of Huysmans and DeWachter (1986a) and are supplemented with those of *Euplotes aediculatus* (Sogin et al. 1986a), *Crithidia fasciculata* (Schnare et al. 1986), *Paramecium tetra urelia* (Sogin and Elwood 1986), *Trypanosoma brucei* and *Euglena gracilis* (Sogin et al. 1986b), *Acanthamoeba castellanii* (Gunderson and Sogin 1986), and *Plasmodium berghei* (Gunderson et al. 1986).

Results and Discussion

5S rRNA Secondary Structure—The Minimal Model

A consensus 5S rRNA structure can be defined by five helical regions (A–E). Figures 1–3 show minimal models for eukaryotes, eubacteria, and halophilic–methanogenic archaebacteria, respectively. Terminus alterations leave the length of helix A somewhat unclear; a primitive helix A should consist of at least 10 bp. The outermost 6 bp of the helix [positions 2:119 to 7:113 according to the general numbering system of Delihas et al. (1984) and Erdmann and Wolters (1986)] are highly variable, while the innermost 4 bp (positions 8:112 to 11:109) are more conserved. The variable part is affected by several insertions in archaebacteria; a base-pair deletion reduces the total helix-A length to 9 bp in *Thermoplasma* and eukaryotes; and base pair 3:117 is a UU odd base pair in a group of Basidiomycota. A shorter helix A in firmibacteria (Gram-positives with low GC content) has led Hori and Osawa (1978) to define a 5S rRNA secondary structure called the 116N-type, but is actually due to different processing of the pre-5S rRNA (Stiekema et al. 1980). The inclusion of base pair 12:113 in the 116N type is not proven by compensating base changes. In the 5S rRNAs of most rhodobacteria (purple bacteria and relatives) of the beta and gamma subdivisions, base pair 11:109 ought to be a CA odd base pair, consistent with biochemical investigations (Digweed et al. 1986; Wolters et al. 1986). Not considering CA base pairing, Hori and Osawa (1978) call this secondary structure the 120N type. CA odd base pairs are frequently tolerated next to bulge loops—the A is positioned next to the loop (see also positions 17:67, 31:51, and 73:103). Taking this into account results in a uniform internal end for helix A.

Helix B is at least 7 bp long. The first 3 bp, which

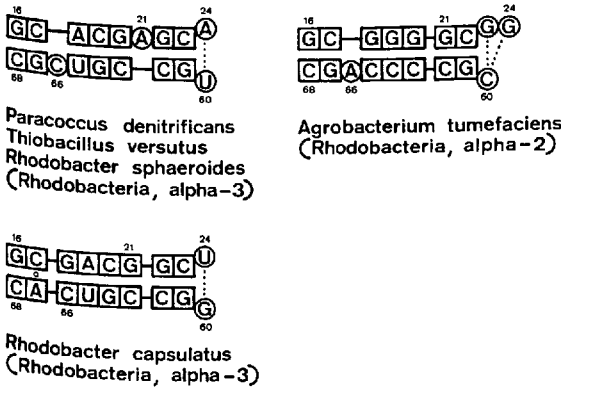

Fig. 4. Deviant base-pairing schemes adopted by helix B. On the left are deviant versions; on the right, the general version, as found in a close relative. Symbols as in Fig. 1

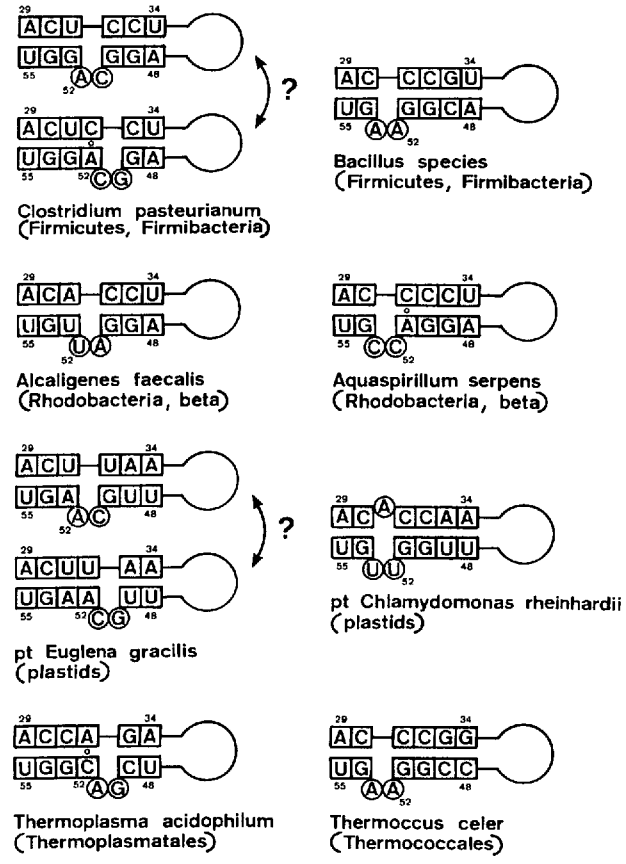

Fig. 5. Deviant base-pairing schemes adopted by helix C. On the left are deviant versions; on the right, the general versions, as found in close relatives. Symbols as in Fig. 1

embed an unpaired nucleotide, are more conserved, while the following 4 bp are highly variable. Additional base pairing of positions 23:60 (usually GC) is often proposed, but this fails to show compensating base changes and is not even possible (UC) in Charophyta and Embryophyta (land plants). The unpaired base is deleted in the archaebacteria *Thermoplasma* and *Sulfolobus*, while an insertion on the opposite site results in a continuous helix in another archaebacterium, Octopus Spring Isolate 1. The variable part bears an insertion in Rhodophyta. The base-pairing scheme is shifted by one base in some species of the alpha-3 subdivision of rhodobacteria (Fig. 4).

The most conserved helix is helix C with its two looped-out nucleotides. In some species of the beta subdivision of rhodobacteria a CA odd base pair is present at positions 31:51 adjacent to the bulge, whereas in some Endomycetidae these positions seem to contain unpaired bases. The base-pairing scheme is altered in some remote species, resulting in a shift of the bulge loop (Fig. 5). Some authors include a CG base pair in eubacteria at positions 28:56, but evidence from compensated base changes is again lacking.

The major difference among the 5S rRNAs of eubacteria, eukaryotes, and several archaebacterial branches is obviously in the number of base pairs in helix D, which ranges from two in eubacteria to seven (or ten) in the archaebacterium *Sulfolobus*. A detailed analysis has been presented in other publications from our laboratory (Erdmann et al. 1986, 1987; Wolters et al. 1986). In halophilic–methanogenic archaebacteria, UU odd base pairs appear to contribute to helix formation in this area. For eubacteria some authors propose additional base pairing at positions 73:103 (AU) and 74:102 (UG) (DeWachter et al. 1982; Delihas et al. 1984), but evidence from compensating base changes is lacking; moreover, in cyanobacteria, plastids, and mycoplasmas, base pairing is impossible. Delihas et al. (1984) even propose a helix continuous with helix E and containing several odd base pairs. In eukaryotes we consider helix D to be 5 bp long. There is no evidence from compensating base changes for GU pairing of bases 69:107; bases 74:102 can pair only in some Metazoa and some Embryophyta, and an AC odd base pair at positions 73:103 in Chlorobionta and a group of Ascomycota suggests an adjacent bulge that leaves positions 74:102 unpaired.

Helix E originally consisted of eight base pairs and a looped-out U residue, but was subject to repeated deletions in eubacteria and the archaebacterial orders Methanobacteriales and Methanococcales. Base pair 81:95 turns to AC in a group of Basiodiomycota, UU in Mesozoa and Metazoa, and UU/YY in various eubacteria (18 of 25 firmicutes, 1 of 9 alpha rhodobacteria, 2 of 9 beta rhodobacteria, all 4 cyanobacteria, and 6 of 11 plastids).

5S rRNA Tertiary Structure—The tRNA Origin

Our cladistic analysis of 5S rRNA secondary structure has revealed an ur-5S rRNA, the ancestor of all present-day 5S rRNA molecules (Fig. 6). Where possible a plesiomorphic nucleotide has been as-

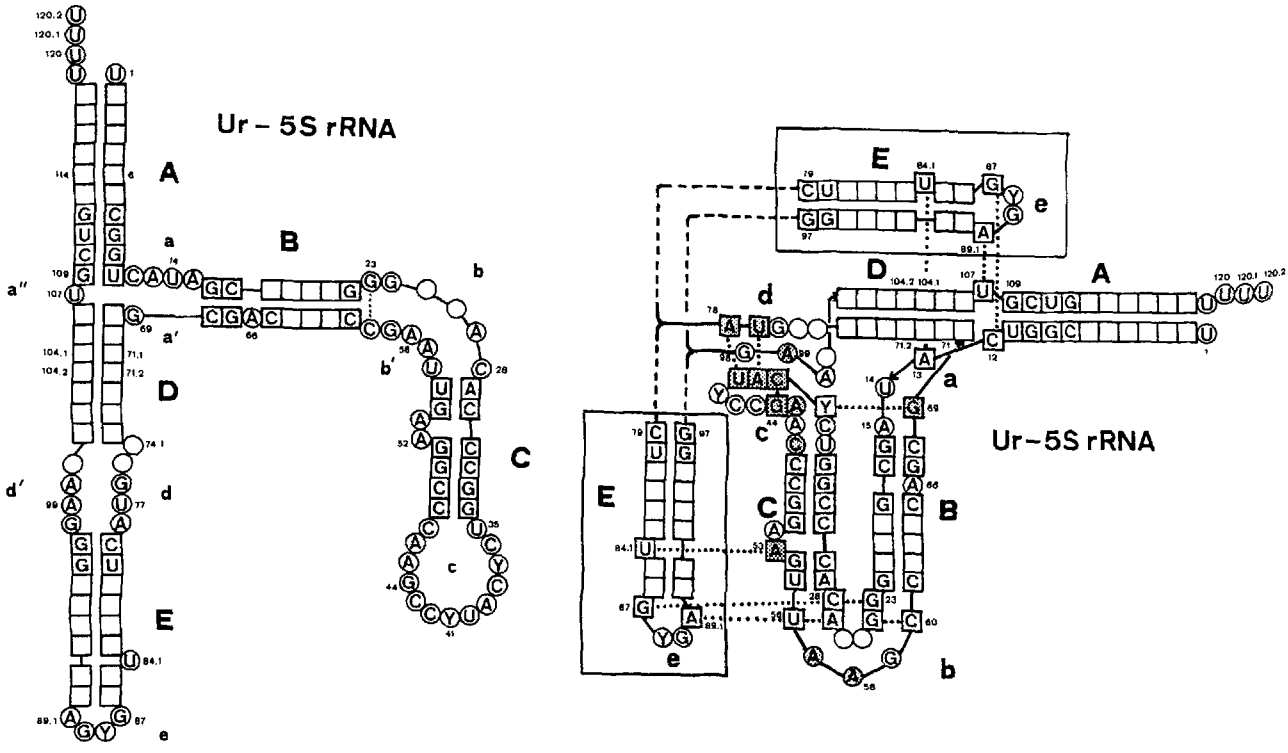

Fig. 6. Ur-5S rRNA, drawn as in Figs. 1–3 (left) and so as to maximize structural homology (right). Shaded positions have a low variability of 0–2 relative to all available sequence data. The proposed molecular switch resulting in alternative orientations of helix E in which loop e interacts either with loop b or with loop a is also indicated

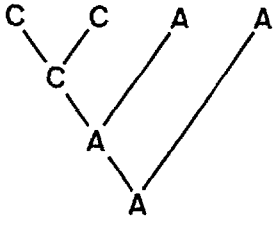

Fig. 7. Nucleotide A is plesiomorphic; nucleotide C, apomorphic. This diagram shows why only synapomorphic characteristics can define monophyletic groups

signed to some positions of medium and low variability. In the same way an ur-tRNA has been reconstructed. If one disregards parts of the molecules that are subject to deletions, namely helix E in 5S rRNA and the DHU stem and loop in tRNA, 5S rRNA can be fitted to a three-dimensional structure like that of tRNA, allowing tertiary interactions at the equivalent positions (Fig. 6) (Wolters and Erdmann 1986). Helix E can then backfold parallel to either helix C or helix D. Loop e and the looped-out U residue of helix E would then interact either with loop b or with loop a. The alternative orientations of helix E might be the result of a molecular switch triggered through the interaction of loop c with loop d, as proposed for tRNA (Moros et al. 1986). In eubacteria, helix E is freed for successive deletions possibly due to the lack of a special protein factor, and the molecular switch in these organisms might be triggered by a somewhat different mechanism: U103:A104 might replace U77:A78 in binding with loop c in one of the conformations.

Phylogenetic Considerations—The Cladistic Approach

Very few laboratories have produced phylogenetic trees derived from 5S rRNA sequence data that cover the whole range of organisms. Of those that have, the laboratories of Hori in Japan (Hori and Osawa 1986) and DeWachter in Belgium (Huysmans and DeWachter 1986b) used unweighted or weighted pairwise grouping; Walker (1985) in Nova Scotia, concentrating on fungi and more recently on general eukaryotic evolution, used the present-day-ancestor method. One glance at their trees shows that they contradict not only one another, but also those derived from 16S rRNA data (oligonucleotide catalogues or complete sequences) and basic general systematics. Besides inaccurate sequencing, especially of RNA, a major problem is the different pace of evolution (different fixation rates) in different organisms. One approach to this problem is excluding positions of high variability and assigning signature nucleotides, or even signature nucleotide combinations (Walker 1985), to specific taxa. But signature nucleotides are useful only if one also does a

cladistic evaluation. Consider, for example, a position at which one group of organisms has A, another C. A clustering procedure would create two clusters: one A and one C cluster. If one then concludes that these clusters represent monophyletic groups, one is absolutely misled (Fig. 7). Natural groups are defined only by synapomorphic (shared derived) characters. One has to decide the direction of the mutational event, A to C or C to A. If there is no outgroup for comparison, the only help one can get is from other biological information.

Archaebacteria are assumed to be the most ancient and primitive organisms adapted to environments also thought to be primitive to the early earth (Hahn and Haug 1986). Eukaryotes appeared latest in the fossil record; they represent a higher level of cellular organization embedding mitochondria, which are derived from eubacteria. This implies the existence of the respective eubacterial division prior to the endosymbiotic event. No sequences of a mitochondria-less nonparasite eukaryote, a candidate for the urkaryote, are available at present. This assumption is supported by the fact that the 5S rRNA is most diverse among archaebacteria and within that group among the thermoacidophilic phenotype. The plesiomorphic condition is therefore defined by the presence in most phyla of either eubacteria or eukaryotes *and* most orders of archaebacteria. Each position of 5S rRNA has been investigated (data not shown) for the following types of variability:

1. CON, the character is 100% CONserved.
2. SGL, a synapomorphic characteristic is present in only a SinGLe species or genus. This fact might have two reasons: (a) The species or genus is a member of a monophyletic taxon, which is only represented by a single species. The addition of related sequences would lead to a classification as CLU (see below); or (b) the isolated occurrence of the character is due to a sequencing error or cloning mutation; it should disappear after resequencing or reexamination of the sequence data.
3. CLU, the synapomorphic characters are CLUstered in one or few groups. There is a strong selection pressure on this character but unlike the CONserved ones, a mutation can be fixed. The probability of parallel mutations is very low; if they occur it is expected that they hit remote groups. The probability of back mutations only exists within a short period after the mutational event before it has been established.
4. SCA, the same apomorphic character is SCAttered throughout the various groups. Optimal functioning prefers one character but tolerates another one. A "leaky" selection pressure leads to numerous parallel mutations. A high degree of back mutations from the tolerated to the optimal character is observed. The phylogenetic significance is reduced to a lower systematic level.

Points 5, 6, and 7 refer only to bases:

5. SEM, the character consists of two nucleotides that are both selected for.
6. EXC, the character consists of three nucleotides that are selected for or, better, of one nucleotide that is selected against or EXCluded.
7. VAR, the character consists of VARiable nucleotides.

The neutral selection/random drift mechanism is only valid for the variable positions, in which no plesiomorphic condition could be defined. In the ur-5S rRNA the variable positions amount to only 49 of 128 positions, i.e., 38%, 90% of them appearing in double-stranded regions. With this method, synapomorphic conditions and monophyletic taxa can be defined first by insertions/deletions (Table 1), second by odd base pairs (Table 2), and third by signature nucleotides (Table 3) using the characters rated CLU. The mutational events leading to the synapomorphic conditions on the molecular level (CLU) appear synchronous with innovational mutations on the metabolic, cytological, and morphological level used to define monophyletic taxa for eukaryotes in the past. Recognized monophyletic taxa for eubacteria are taken from Woese et al. (1985), for archaebacteria from Schnabel et al. (1983) and Woese and Olsen (1986).

Phylogeny of Eukaryotes

Rhodophyta (3 sequences). Red algae are clearly defined by an insertion in helix B (position 20.1). Rhodophytes and the zygomycete *Coemansia* are the only eukaryotes sharing the presence of position 41 with eubacteria and archaebacteria. The appearance in a single species of the zygomyceteous order Kickxellales points to a reinsertion at least for *Coemansia*; whether this is also the case in rhodophytes or whether they retained the plesiomorphic condition cannot be decided yet, but the absence of position 41 in the archaebacterium Octopus Spring Isolate 1 points to a reinsertion in rhodophytes, too.

Chlorobiota (33 sequences). This grouping, also called the chlorophyte series (Taylor 1978) or Viridiplantae (Cavalier-Smith 1983) comprises Volvophyta, Chlorophyta (green algae), Charophyta (stoneworts), and Embryophyta (land plants). It is defined by an A in position 43 and an AC odd base

Table 1. Insertions and deletions in 5S rRNA

Position	Plesiomorphic character	Apomorphic character	Variability[a]	Monophyletic taxa[b]
1	N	DEL	3	1. *Thermoplasma* 2. Eukaryota
		INS		Three orders of Zygomycota (Mucorales, Entomophthorales, Harpellales) (8)
4.1	–	INS	2	1. Methanococcales and -bacteriales (3) 2. *Thermococcus*
5.1–5.2	–	INS	1	Octopus Spring Isolate 1 (arc)
6:114	NN	DEL	2	1. Eukaryota 2. *Thermoplasma*
7.1	–	INS	1	Octopus Spring Isolate 1 (arc)
20.1	–	INS	2	1. Rhodophyta (3) 2. mt Angiospermae (4)
30.1	–	INS	1	Cyanobacteria and pt (15)
36	C	DEL	1	*Synechococcus* and pt (13)
41	Y	DEL	3	Octopus Spring Isolate 1 (arc) and Eukaryota
		INS		1. Rhodophyta (3) 2. *Coemansia* (Zygomycota)
52.1	–	INS	1	Endomycetidae and Ascomycetidae (Plecto-, Pyreno-, Disco-, and Hyphomycetes) (31)
66	A	DEL	2	1. *Sulfolobus* 2. *Thermoplasma*
74.1	N	DEL	2	1. *Sulfolobus* 2. Eubacteria
74.2	–	INS	2	1. *Methanobrevibacter* 2. Saccharomycetales (Ascomycota) (7)
84.1	U	DEL	2	1. Methanococcales and -bacteriales 2. Eubacteria
104.2	N	DEL	1	Octopus Spring Isolate 1 (arc)
104.3	–	INS	1	A group of ustomycetes not yet named comprising *Microbotryum, Rhodosporidium, Aessosporon, Ustilago scabiosae, Sphacelotheca, Rhizoctonia crocorum,* and *R. iemalis, Pachnocybe,* and *Agaricostilbum* (10)
107	U	DEL	1	*Sulfolobus*
108	–	INS	2	Eubacteria
		DEL		*Vibrio marinus* (gamma ' rhodobacteria)
114.1	–	INS	1	*Sulfolobus*

arc, Archaebacterium; INS, insertion; DEL, deletion; N, variable position; pt, plastid; mt, mitochondrion

[a] The number of mutations per position of all extant organisms

[b] The number in parentheses following the taxon name refers to the number of different 5S rRNA sequences

Table 2. Odd base pairs in 5S rRNA[a]

Position	Plesiomorphic character	Apomorphic character	Variability	Monophyletic taxa
3:117	NN	UU	2	1. A group of Basidiomycota preliminarily named Doliporomycetes (28) 2. *Nadsonia* (Ascomycota)
73:103	NN	AC	2	1. Chlorobiota (33) 2. Pyrenomycetes, *Monilinia, Aureobasidium* (12)
81:95	NN	UU	?	1. Metazoa (68) 2. *Dicyema* (Mesozoa)
		UU/YY		1. Gram-positive bacteria (18 of 25) 2. Alpha rhodobacteria (1 of 9) 3. Beta rhodobacteria (2 of 9) 4. Cyanobacteria and plastids (10 of 15)
		AC		A group of higher basidiomycetes including rusts and not yet named (13)

[a] Odd base pairs are defined as those that are able to replace Watson–Crick base pairs without disturbing the helical conformation, so that double-strand-specific nuclease will still cleave. This has been demonstrated for AC and UU. For abbreviations and further explanation, see Table 1

pair at positions 73:103. A detailed analysis of this group (Fig. 8) agrees with the cluster analysis of Hori et al. (1985).

Metazoa (68 sequences) and Mesozoa (1 sequence). Multicellular animals and the single mesozoan species *Dicyema* share a UU odd base pair at positions 81:95, unique in eukaryotes but frequent in several eubacterial groups. This fact increases the variability of these positions so that parallel events should not be excluded for the two groups (compare Ohama et al. 1984). Mesozoa is considered to be a sister group of Metazoa, although its mitochondrial cristae are tubular whereas metazoans have lamellar cristae.

Eumycota (87 sequences). The uniflagellate Chytridiomycota (2 sequences) and the nonflagellate Zygomycota (11 sequences), Ascomycota (35 sequences), and Basidiomycota (39 sequences) are generally thought to comprise a monophyletic lineage because of the presence of chitin in their cell walls and the amino-adipic acid (AAA) pathway in lysine synthesis (LéJohn 1974). In 5S rRNA no sig-

Table 3. Bases of low variability in 5S rRNA[a]

Position	Plesiomorphic character	Apomorphic character	Variability	Monophyletic taxa
37	U/C	A	2	1. pt *Euglena* (3) 2. *Streptomyces* (Actinobacteria)
39	A	G	3–5	1. Methanobacteriales 2. *Thermococcus* 3–5. Eukaryotic group H[b] (164)
43	C	A	4	1. Octopus Spring Isolate 1 (arc) 2. *Dipsacomyces* and *Linderina* (Zygomycota) (2) 3. *Artemia* (Crustacea) 4. Chlorobiota (33)
45	A	C	3	1. Chrysophyta (2) and oomycetes (2) 2. Asomycetidae (Plecto-, Pyreno-, Disco-, and Hyphomycetes) (22) 3. *Chlamydomonas* (2)
46	A	U	2–4	1. Methanococcales and -bacteriales (3) 2–4. Eukaryotic group H[b] (164)
47	C	U	2	1. Kinetoplastida (2) and *Euglena* (2) 2. Porphyra (2)
69	G	U	1	pt *Euglena* (3)
77	U	C	1	*Sulfolobus*
78	A	C	1	*Sulfolobus*
99	A	G	1	*Sulfolobus*

[a] For abbreviations and explanation, see Table 1

[b] Eukaryotic group H comprises (a) Chrysophyta (2) and oomycetes (2); (b) *Dictyostelium, Physarum,* and *Amoebidium*; (c) Cryptophyta; (d) Chytridio- (2), Zygo- (11), Asco- (33), and Basidiomycota (39); (e) Kinetoplastida (2) and *Euglena* (2); and (f) Metazoa except *Haliclona* (Porifera) (67)

nature nucleotides are found to support the monophyly of the four eumycotan phyla. Terminus alterations can be used to elucidate the phylogeny of the Zygomycota, defining a subphylum consisting of the orders Mucorales, Entomophthorales, and Harpellales (Fig. 9).

Two ascomycete genera, *Protomyces* and *Schizosaccharomyces*, exhibit plesiomorphic characteristics while the two subphyla Endomycetidae and Ascomycetidae share an insertion at position 52.1 (Fig. 10). The Ascomycetidae comprising plectomycetes, pyrenomycetes, the two discomycetes *Monilinia* and *Trichophaea*, and the hyphomycete *Aureobasidium* are clearly defined by a mutation from A to C in position 45; a position of very low variability.

Analysis of 5S rRNA revises basidiomycete systematics deduced from morphology and creates a new phylogeny with new groupings to which hardly

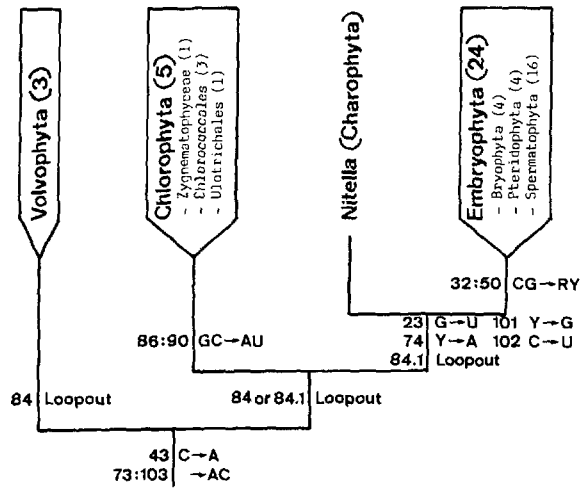

Fig. 8. Cladogram of Chlorobiota as deduced from 5S rRNA secondary and primary structure. Position numbering is according to the general system of Erdmann and Wolters (1986). The numbers in parentheses indicate the numbers of species sequenced

any conventional name can be assigned (Fig. 11). The phylum Basidiomycota is split into two groups, one having an insertion at position 104.3 and comprising saprophytes and ustomycetes that attack dicotyledons, the other exhibiting a UU odd base pair at positions 3:117 and comprising ustomycetes that attack monocotyledons and hymenomycetes (higher basidiomycetes). Within the latter grouping the hymenomycetes form a monophyletic group defined by GC instead of UG at positions 80:96. An AC odd base pair at positions 81:95 defines a subgroup of hymenomycetes not yet named.

Euglenozoa (4 sequences). Phototrophic and phagotrophic euglenoids (Euglenida) and trypanosomes (Kinetoplastida), collectively called Euglenozoa by Cavalier-Smith (1983), and the cyanelle-containing *Cyanophora* represent a distinct group sharing a C to U mutation at position 47. Cytological investigations have led to a proposal for a bodonid-like ancestor for this group (Willey and Wibel 1985). Plastidless euglenoids are therefore primitive, and the acquisition of plastids an apomorphy. Present data from 5S rRNA sequences cannot resolve the question whether phototrophic euglenoids evolved by an uptake of a cyanobacterium (*Cyanophora* stage) or of a eukaryotic volvophytean algae (Gibbs 1978).

Chrysobionta (8 sequences). Also called the chromophyte series (Taylor 1978) or the Chromista (Cavalier-Smith 1983), this lineage is represented by only a single cryptophyte, a single dinophyte, and two closely related phaeophyceans (brown algae) belonging to the phylum Chrysophyta. Position 45 is changed from A to C in brown algae and oomycetes (water molds), proving a close relationship between

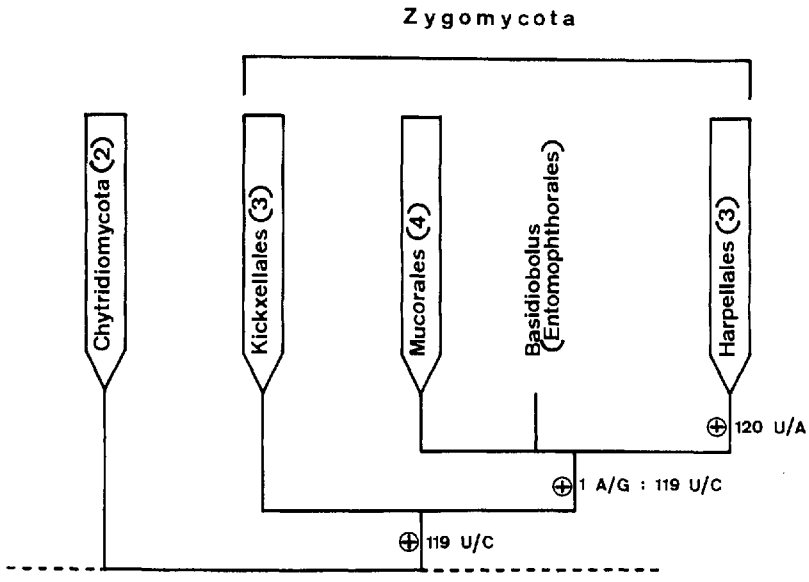

Fig. 9. Cladogram of Chytridiomycota and Zygomycota as deduced from 5S rRNA secondary and primary structure. A plus sign indicates a terminus alteration. Other conventions are as in Fig. 8

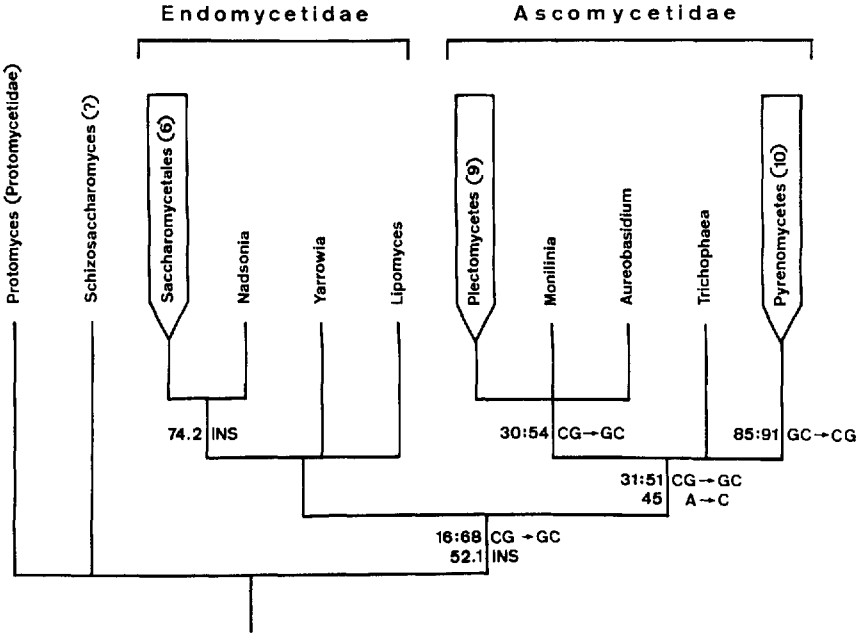

Fig. 10. Cladogram of Ascomycota as deduced from 5S rRNA secondary and primary structure. "INS" indicates an insertion. Other conventions are as in Fig. 8

oomycetes and the phylum Chrysophyta, both of which have typical heterokont flagella. In 1858, Pringsheim had already suggested that water molds are phylogenetically linked to the Xanthophyceae (formerly Tribophyceae). No other signature is found to define this lineage.

Figure 12 shows a phylogeny of the Euglenozoa and Chrysobionta, the groups with hairy flagella, based on cytological and biochemical criteria. Clustering procedures group the dinophyte species with the Ciliophora, but a phylogenetic analysis reveals no specific relationship. Thraustomycetes (thraustochytridiomycetes) exhibit typical heterokont flagella and are generally classified with the oomycetes and labyrinthulomycetes. Their 5S rRNA, however, is totally different from the oomycetean type, and even the RNAs of the two examined species are quite different from each other. Thraustomycetes show plesiomorphic characteristics, so their origin remains unclear.

Ciliophora (10 sequences). The ciliates cannot be defined by any signature nucleotide but remain a distinct cluster in all treeing procedures.

Rhizopoda (4 sequences). This group is represented by *Acanthamoeba* (Amoebina), *Dictyostelium* (Acrasea), *Physarum* (Myxogastria), and

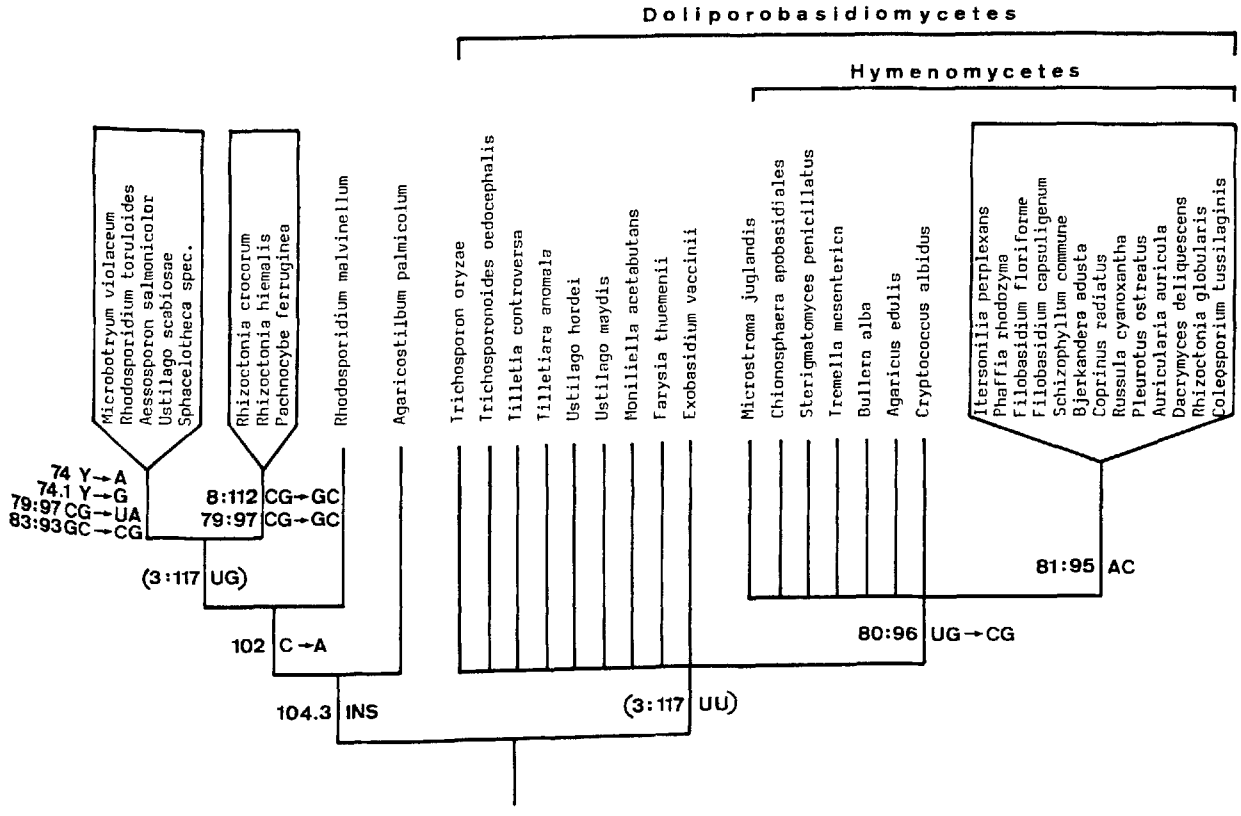

Fig. 11. Cladogram of Basidiomycota as deduced from 5S rRNA secondary and primary structure. "INS" indicates an insertion

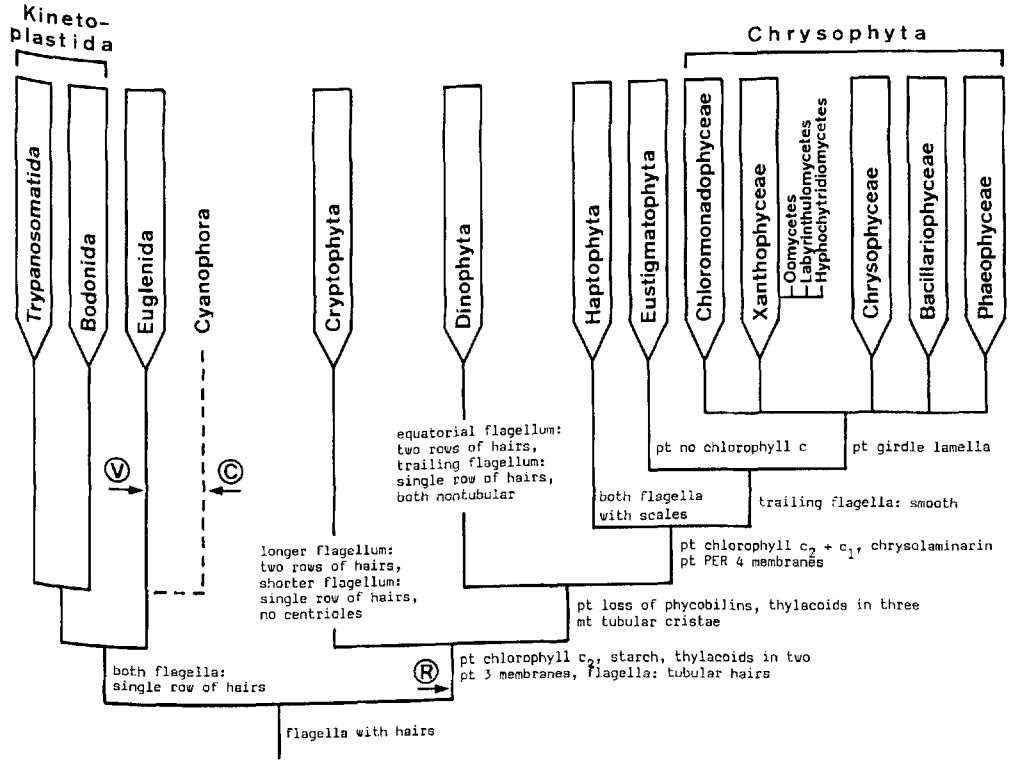

Fig. 12. Cladogram of Euglenozoa and Chrysobiota based on cytological and biochemical criteria. pt, Plastid; mt, mitochondrion; PER, plastidic endoplasmic reticulum. Arrows indicate possible endosymbiotic acquisitions of a rhodophyte (circled R), a volvophyte (circled V), or a cyanobacterium (circled C)

Table 4. Bases of low variability that resolve the phylogeny of eukaryotes

Taxa	Position										
	194	521	568	570	600	1235	1464	1593	2040	2085	2120: 2250
Eubacteria											
Rhodobacteria (purple) (4)	R	G	U	C	A	A	C	G	G	A	CG
Sulfur-dependent & Myxobacteria (2)	R	G	U	C	A	A	C	G	G	A	CG
Cyanobacteria (1) & plastids (4)	R	G	U	C	A	A	C	G[a]1	G	A	CG
Firmicutes (Gram-positives) (4)	R	G	U	C	A	A	C	G	G	A	CG
Bacteroides/Flavobact./Cytophaga (2)	R	G	U	C	A	A	C	G[b]2	G	A	AU
Archaebacteria											
Halobacteriales (4)	A	G	U	C	C	C	U	G	G	A	AU
Methanomicrobiales (1)	A	G	U	C	A	A	U	G	G	A	AU
Methanococcales (1)	A	G	U	C	A	C	U	G	G	A	CG
Methanobacteriales (1)	A	G	U	C	A	C	U	G	G	A	CG
Sulfolobales (1)	A	G	U	C	A	A	U	G	G	A	CG
Thermoproteales (1)	A	G	U	C	A	A	C	G	G	A	CG
Plesiomorphic character	A	G	U	C	A	A	Y	G	G	A	CG
Eukaryota											
				A(3)			A(4)	U(4)		U(4)	
Ciliophora (5)	A	U	U	C(2)	A	A	U(1)	A(1)	G	A(1)	CG
Plasmodium (Apicomplexa, Sporozoa)	A	U	U	A	A	A	A	A	G	C	CG
Dictyostelium (Acrasea) 2 refs	A	U	U	C	A	A	A	A	A	A	CG
Acanthamoeba (Amoebina)	U	U	U	A	A	A	A	U	C	U	CG
Spermatophyta (3)	U	U	C	A	A	C	A	U	C	C	CG
Kinetoplastida (2)	U	U	U	C	A	A	C	U	C	U	UA
Euglena (Euglenida)	U	C	C	A	A	A	U	A	G	U	UA
Saccharomyces (Ascomycota)	U	A	C	A	A	A	A	A	C	U	CA
Metazoa (7)	U	C	C	A	C	U	C	A	G	U	CA
Positions used in Fig. 12	x		x		x	x					xx

[a] A pt *Chlamydomonas*
[b] A *Bacteroides*

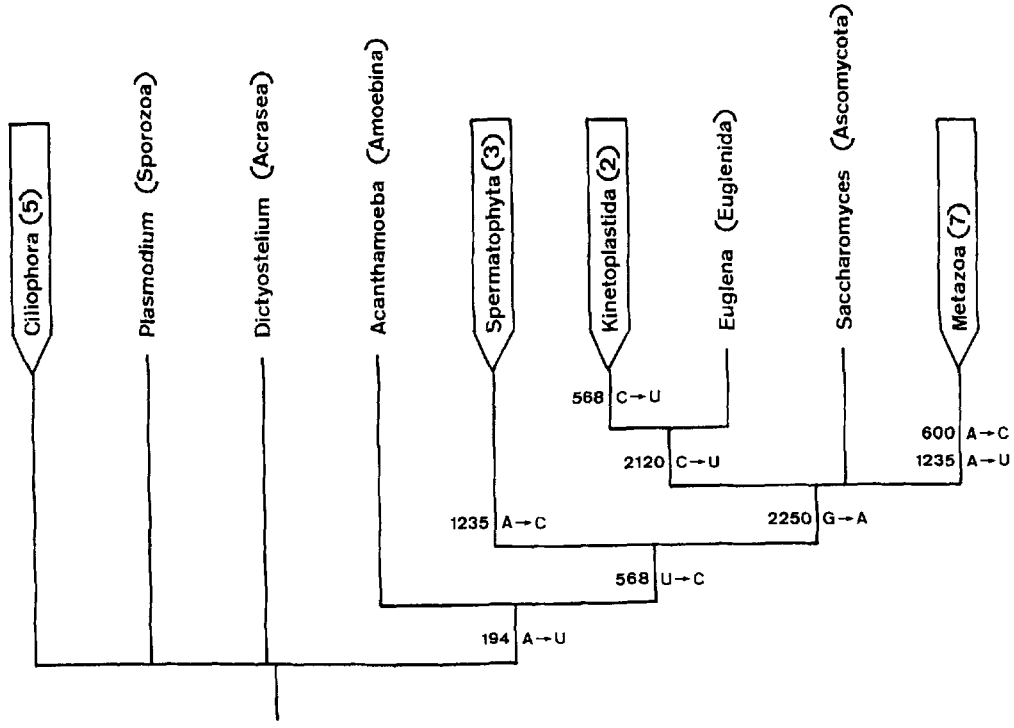

Fig. 13. Cladogram of Eukaryota as deduced from complete 16S rRNA sequences. Position numbering is done according to the system used in the biannual compilation of Huysmans and DeWachter (1986a). Other conventions are as in Fig. 8

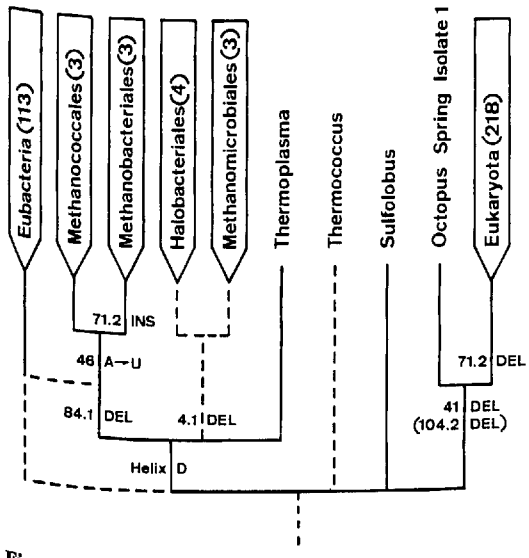

Fig. 14. Cladogram of archaebacteria and the origins of eubacteria and eukaryotes as deduced from 5S rRNA sequences. "DEL" indicates a deletion. Other conventions are as in Figs. 8 and 10

Amoebidium (Amoebidiales). Cytological investigations have led to the conclusion that the obligate amoeboid stage in protostelids (amoeboflagellates) has evolved independently several times (Spiegel and Feldman 1985). Therefore solitary amoeba are probably a polyphyletic assemblage, but might nevertheless have their origin in one particular group.

Major Divisions Among Eukaryotes. There are two positions in 5S rRNA, both in loop c, that are of special interest. In archaebacteria and eubacteria A39 and A46 are rather conserved. This plesiomorphic configuration also exists in some eukaryotic lineages, namely Rhodophyta, Chlorobiota, thraustomycetes, Ciliophora, and *Acanthamoeba,* whereas G39 and U46 are found in Metazoa (except one poriferan species), Eumycota, Euglenozoa, Phaeophyceae and oomycetes, the cryptophyte species, and the rhizopods *Dictyostelium, Physarum,* and *Amoebidium* (but not *Acanthamoeba*). The dinophytean and mesozoan characters are in some way intermediate. They both exhibit a C at position 46; the dinophyte has a U at position 39, as in some archae- and eubacteria, and the mesozoan has the plesiomorphic A39. Does the second lineage defined by G39 and U46 represent a monophyletic group? Since only one of three poriferan species that form a distinct cluster shows the plesiomorphic character, it is best explained by coordinate back mutations, and therefore the phylogenetic value of this signature is quite reduced.

The cladistic approach should be as easily applied to complete 16S rRNA sequences even if the number and range of species investigated are still relatively small. A detailed analysis reveals 12 positions

Table 5. Bases of low variability in 16S rRNA that show the paraphyletic character of archaebacteria

Position[a]	Plesiomorphic character	Apomorphic character	Variability	Monophyletic taxa
1413:2098	AU	GC CG	2	*Thermoproteus* and eukaryotes, eubacteria
1440:1923	CG	UA	1	Halophilic–methanogenic archaebacteria and eubacteria
2012	G	C	1	Halophilic–methanogenic archaebacteria and eubacteria
2043	G	C	1	Halophilic–methanogenic archaebacteria and eubacteria
2044	G	C	?	Halophilic–methanogenic archaebacteria
		Y		*Bacillus* and *Mycoplasma* (Gram-positives); rhodobacteria (purple bacteria); and *Desulfovibrio, Bacteroides,* and *Flavobacterium*
		G		1. *Heliobacterium* (Gram-positives?) 2. *Myxococcus* (Myxobacteria) 3. *Anacystis* (Cyanobacteria)

[a] Numbering is according to the annual compilation of 16S rRNA sequences (Huysmans and De Wachter 1986a)

that are rather conserved in archae- and eubacteria but for which eukaryotes show some variation due to their higher rate of evolution (Table 4). Positions 570, 1464, 1593, and 2085 (according to the biannual compilation of Huysmans and DeWachter 1986a) show variation even within the phylum Ciliophora and are of limited value. Figure 13 shows a tentative cladogram derived from 6 of the 12 positions. The tree resembles that derived by a variation of the Fitch–Margoliash matrix method (Sogin et al. 1986b) with one exception: The origin of the euglenid and kinetoplastid flagellates is considerably more ancient in the Fitch–Margoliash-type analysis.

The cladistic analysis of both 5S rRNA and complete 16S rRNA defines a monophyletic group that consists at least of the euglenids and kinetoplastids. This monophyletic group shows a close relationship to the Eumycota and Metazoa. This is in harmony with biological reasoning: Euglenida and Eumycota have been proposed to be sister groups (Whittaker 1969) because both use the apomorphic aminoadipic acid pathway for lysine synthesis, whereas archaebacteria, eubacteria, and other eukaryotes use the diamino-pimelic acid pathway. Metazoa cannot

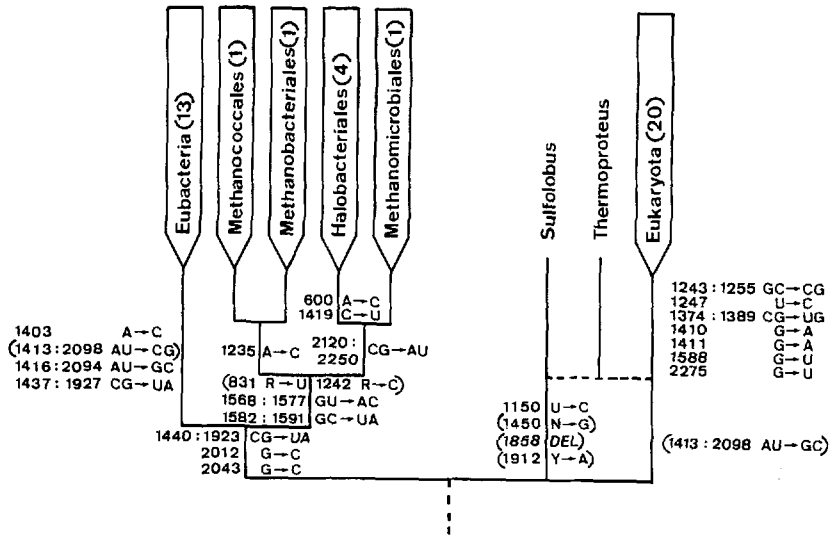

Fig. 15. Cladogram of archaebacteria and the origins of eubacteria and eukaryotes as deduced from complete 16S rRNA sequences. Conventions are as in Fig. 8

synthesize lysine and must acquire it heterotrophically. The fact that Chlorobiota are more primitive also perfectly fits with the phylogeny of deviant mitochondrial codes (Wallace 1983) and the absence of a mitochondrial 5S rRNA as in Eumycota and Metazoa. The different positioning obtained with the matrix method might be the result of the higher rates of evolution observed in parasites such as trypanosomes, *Plasmodium* (Sporozoa), Mesozoa, nematodes, and mycoplasmas. The deeper branching of the Eumycota obtained using a clustering method and 5S rRNA sequences (Huysmans and DeWachter 1986b) might be due to the same effect.

The Archaebacterial Dogma

Within the archaebacteria the 5S rRNAs of thermoacidophilic organisms (*Sulfolobus,* Octopus Spring Isolate 1, *Thermococcus,* and *Thermoplasma*) show the highest degree of structural diversity, while those of the halophilic–methanogenic types show consistent features. Therefore the thermoacidophilic phenotype is considered primitive. Different thermoacidophilic species share different characteristics with eukaryotes; for example, there is a deletion of base pair 6:114 (helix A) in *Thermoplasma,* but the number of bases in helix D is identical to that found in halophiles and methanogens. A deletion of position 41 (loop c) in Octopus Spring Isolate 1 is also shared with eukaryotes, where it apparently has reinserted twice: once in rhodophytes and once in the zygomycete *Coemansia.* A deletion in the 3′ segment of helix D in Octopus Spring Isolate 1, *Sulfolobus* and *Thermococcus* to be a step towards the eukaryotic condition (but also towards the eubacterial condition, in which even more deletions have occurred in this region). *Sulfolobus* and *Thermococcus* could adopt similar base-pairing schemes (the most primitive?) while that of Octopus Spring Isolate 1 resembles that of eukaryotes. The latter species might indeed be a candidate for a sister taxon of eukaryotes (Fig. 14), implying that even eocytes/sulfur-dependents are paraphyletic.

A cladistic analysis of 16S rRNA reveals three positions in which eukaryotes share a nucleotide with the eocytes/sulfur-dependents while at these positions eubacteria share another with halophile-methanogens, all at the 100% level (Table 5, Fig. 15). This supports the proposal of Zillig et al. (1982) and Lake et al. (1984) that eukaryotes are derived from eocytes/sulfur-dependents and that eubacteria are derived from halophile-methanogens, i.e., archaebacteria are a paraphyletic group.

The thermoacidophilic genus *Thermoplasma* clearly clusters with the halophile-methanogens according to Woese and Olsen (1986); their 5S rRNA exhibits the same base-pairing scheme in helix D.

Methanobacteriales and Methanococcales are definite sister groups defined by an A to U mutation at position 46 in 5S rRNA and an A to C mutation at position 1235 in 16S rRNA. Halobacteriales and Methanomicrobiales are also definite sister groups defined by a CG to AU mutation at positions 2120: 2250 in 16S rRNA. No specific relationship of eubacteria to the Halobacteriales as proposed by Lake et al. (1985) could be detected; the 16S rRNA cladistic analysis shows a more remote affiliation with the halophilic–methanogenic archaebacteria, which are clearly defined as monophyletic by two characteristics. The occurrence of a sulfur metabolism in some eubacterial divisions might be a primitive characteristic—an affiliation with a novel thermoacidophile would not be an unexpected future discovery.

Acknowledgment. We would like to thank Angela Schreiber for the elaborate drawings, Martin Digweed for the help with the English text, and Mitchell Sogin and E. Maxwell for the communication of sequences prior to publication. Jörn Wolters is financed by the Deutsche Forschungsgemeinschaft (Sfb-9/B5). The financial support by the Fonds der Chemischen Industrie e.V. is gratefully acknowledged.

References

Cavalier-Smith T (1983) A 6-kingdom classification and a unified phylogeny. In: Schenk HEA, Schwemmler W (eds) Endocytobiology, vol II. deGruyter, Berlin, pp 1027–1034

Delihas N, Andersen J, Singhal RP (1984) Structure, function and evolution of the 5S ribosomal RNAs. Progr Nucl Acid Res Mol Biol 31:161–190

DeWachter R, Chen MW, Vandenberghe A (1982) Conservation of secondary structure in 5S ribosomal RNA: a uniform model for eukaryotic, eubacterial, archaebacterial and organelle sequences is energetically favourable. Biochimie 64:311–329

Digweed M, Pieler T, Kluwe D, Schuster L, Walker R, Erdmann VA (1986) Improved procedure for the isolation of a double-strand-specific ribonuclease and its application to structural analysis of various 5S rRNAs and tRNAs. Eur J Biochem 154:31–39

Erdmann VA, Wolters J (1986) Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. Nucl Acids Res 14:r1–r59

Erdmann VA, Pieler T, Wolters J, Digweed M, Vogel D, Hartmann R (1986) Comparative structural and functional studies on small ribosomal RNAs. In: Hardesty B, Kramer G (eds) Structure, function and genetics of ribosomes. Springer, New York, pp 164–183

Erdmann VA, Wolters JW, Pieler T, Digweed M, Specht T, Ulbrich N (1987) Evolution of organisms and organelles as studied by comparative computer and biochemical analyses of ribosomal 5S RNA structure. Ann NY Acad Sci, in press

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56:2883–2889

Gunderson JH, Sogin ML (1986) Gene 44:63–70

Gunderson JH, McCutchan T, Sogin ML (1986) J Protozool, in press

Hahn J, Haug P (1986) Traces of archaebacteria in ancient sediments. System Appl Microbiol 7:178–183

Hennig W (1966) Phylogenetic systematics. University of Illinois Press, Urbana

Hori H, Osawa S (1978) Evolutionary change in 5S rRNA secondary structure and a phylogenetic tree of 54 5S RNA species. Proc Natl Acad Sci USA 76:381–385

Hori H, Osawa S (1986) Evolutionary change in 5S rRNA secondary structure and a phylogenetic tree of 352 5S rRNA species. in press

Hori H, Lim BL, Osawa S (1985) Evolution of green plants as deduced from 5S rRNA sequences. Proc Natl Acad Sci USA 82:820–823

Huysmans E, DeWachter R (1986a) Compilation of small ribosomal subunit RNA sequences. Nucl Acids Res 14:r73–r118

Huysmans E, DeWachter R (1986b) The distribution of 5S ribosomal RNA sequences in phenetic hyperspace. Implications for eubacterial, eukaryotic, archaebacterial and early biotic evolution. Endocyt Cell Res 3:133–155

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164:788–798

Lake JA, Henderson E, Oakes M, Clark MW (1984) Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci USA 81:3786–3790

Lake JA, Clark MW, Henderson E, Fay SP, Oakes M, Scheinman A, Thornber JP, Mah RA (1985) Eubacteria, halobacteria, and the origin of photosynthesis: the photocytes. Proc Natl Acad Sci USA 82:3716–3720

LéJohn HB (1974) Biochemical parameters of fungal phylogenies. Evol Biol 7:79–125

Lim BL, Kawai H, Hori H, Osawa S (1986) Molecular evolution of 5S ribosomal RNA from red and brown algae. Jpn J Genet 61:169–176

Maxwell ES, Liu J (1986) Nucleotide sequence of *Cyanophora paradoxa* cytoplasmic and cyanelle 5S ribosomal RNAs, in press

Moras D, Dock AC, Dumas P, Westhof E, Romby P, Ebel JP, Giegé R (1986) Anticodon-anticodon interaction induces conformational changes in tRNA: yeast $tRNA^{asp}$, a model for tRNA recognition. Proc Natl Acad Sci USA 83:932–936

Ohama T, Kumazaki T, Hori H, Osawa S (1984) Evolution of multicellular animals as deduced from 5S rRNA sequences: a possible early emergence of the Mesozoa. Nucl Acids Res 12:5101–5108

Pringsheim N (1858) Beiträge zur Morphologie und Systematik der Algen. II. Die Saprolegnien. Jahrb Wiss Bot 1:284–304

Schnabel R, Huet J, Thomm M, Zillig W, Sentenac A, Stetter KO (1983) Phylogeny of the archaebacteria and eukaryotes: homology of the DNA-dependent RNA polymerases. In: Schwemmler W and Schenk HEA (eds) Endocytobiology, vol II de Gruyter, Berlin, pp 895–912

Schnare MN, Collings JC, Gray MW (1986) Structure and evolution of the small subunit ribosomal RNA gene of *Crithidia fasciculata*. Curr Genet 10:405–410

Sogin ML, Elwood HJ (1986) Primary structure of the *Paramecium tetraurelia* small-subunit rRNA coding region: phylogenetic relationships within the Ciliophora. J Mol Evol 23: 53–60

Sogin ML, Swanton MT, Gunderson JH, Elwood HJ (1986a) Sequence of the small subunit ribosomal RNA gene from the hypotrichous ciliate *Euplotes aediculatus*. J Protozool 33: 26–29

Sogin ML, Elwood HJ, Gunderson JH (1986b) Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc Natl Acad Sci USA 83:1383–1387

Spiegel FW, Feldman J (1985) Obligate amoebae of the protostelids: significance for the concept of eumycetozoa. Biosystems 18:377–386

Stiekema WJ, Raué HA, Planta RJ (1980) Sequence analysis and in vitro maturation of five precursor 5S RNAs from *Bacillus* Q. Nucl Acids Res 8:2193–2211

Taylor FJR (1978) Problems in the development of an explicit hypothetical phylogeny of the lower eukaryotes. BioSystems 10:67–89

Walker WF (1985) 5S and 5.8S ribosomal RNA sequences and protist phylogenetics. BioSystems 18:269–278

Wallace DC (1983) Structure and evolution of organelle DNAs. In: Schenk HEA, Schwemmler W (eds) Endocytobiology, vol II. deGruyter, Berlin, pp 87–100

Whittaker RH (1969) New concepts of kingdoms of organisms. Science 163:150–160

Willey R, Wibel RG (1985) A cytostome/cytopharynx in green euglenoid flagellates (Euglenales) and its phylogenetic implications. BioSystems 18:369–376

Woese CR, Olsen GJ (1986) Archaebacterial phylogeny: perspectives on the urkingdoms. System Appl Microbiol 7:161–177

Woese CR, Stackebrandt E, Macke TJ, Fox GE (1985) A phylogenetic definition of the major eubacterial taxa. System Appl Microbiol 6:143–151

Wolters J, Erdmann VA (1986) A 5S rRNA tertiary structural model inspired by the known tRNA structure. Endocyt Cell Res 3:157–166

Wolters J, Pieler T, Digweed M, Erdmann VA (1986) Reconciliation of comparative computer analysis and biochemical investigations of 5S rRNA secondary structure. In: Kandler O. Zillig W (eds) Archaebacteria 85. Fischer, Stuttgart, pp 414–416

Zillig W, Schnabel R, Tu J (1982) The phylogeny of archaebacteria, including novel anaerobic thermoacidophiles in the light of RNA polymerase structure. Naturwissenschaften 69: 197–204

Note added in proof. After the submission of the manuscript additional 5S rRNA sequence data of red and brown algae (Lim et al. 1986) came to our attention, substantiating our analysis. Most interestingly, the signature shared by brown algae (Phaeophyceae) and oomycetes is also shared by a diatom (Bacillariophyceae) but not by a golden-yellow algae (Chrysophyceae).