Proc. Natl. Acad. Sci. USA
Vol. 86, 4569–4573, June 1989
Evolution

Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome

(phylogeny/taxonomy/transcription/subunit/gene organization)

GABRIELA PÜHLER*, HENRIK LEFFERS†, FELIX GROPP*, PETER PALM*, HANS-PETER KLENK*, FRIEDRICH LOTTSPEICH*, ROGER A. GARRETT†, AND WOLFRAM ZILLIG*

*Max-Planck-Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany; and †Biostructural Chemistry, Kemisk Institut, Aarhus University, DK-8000 Aarhus C, Denmark

Communicated by Carl R. Woese, January 30, 1989

ABSTRACT **Genes for DNA-dependent RNA polymerase components B, A, and C from the archaebacterium *Sulfolobus acidocaldarius* and for components B'', B', A, and C from the archaebacterium *Halobacterium halobium* were cloned and sequenced. They are organized in gene clusters in the order above, which corresponds to the order of the homologous *rpoB* and *rpoC* genes in the corresponding operon of the *Escherichia coli* genome. Derived amino acid sequences of archaebacterial components A and C were aligned with each other and with the sequences of corresponding (largest) subunits from the archaebacterium *Methanobacterium thermoautotrophicum*, with sequences of various eukaryotic nuclear RNA polymerases I, II, and III, and with the sequence of the β' component from *E. coli* polymerase. The archaebacterial genes for component A are homologous to about the first two-thirds of genes for the eukaryotic component A and the eubacterial component β', and the archaebacterial genes for component C are homologous to the last third of the genes for the eukaryotic component A and the eubacterial component β'. Unrooted phylogenetic dendrograms derived from both distance matrix and parsimony analyses show the archaebacteria are a coherent group closely related to the eukaryotic nuclear RNA polymerase II and/or III lineages. The eukaryotic polymerase I lineage appears to arise separately from a bifurcation with the eubacterial β' component lineage.**

Phylogenetic relations between the primary kingdoms of life have been established using rRNA as an evolutionary chronometer (1, 2). RNA polymerase appears to be an appropriate alternative because it is ubiquitous, highly conserved, and a complex macromolecule. We have cloned and sequenced the genes for the large RNA polymerase components of two phylogenetically distant archaebacteria: the sulfur-dependent extreme thermophile *Sulfolobus acidocaldarius* (*Su. acidocaldarius*) and the extreme halophile *Halobacterium halobium*. Eukaryotes possess three specialized nuclear RNA polymerases (pol I, II, and III) thought to be derived from one ancestral "urkaryotic" enzyme by diversification (3). Archaebacteria should provide an independent outside reference, in addition to eubacteria, for elucidating the origin of the eukaryotic lineage.

DNA-dependent RNA polymerases from archaebacteria are of two types (4, 5). The BAC type (largest subunits arranged in order of decreasing size) is found in the sulfur-dependent extreme thermophiles (6), including the *Thermococcales* (7), and in the genus *Thermoplasma*, which belongs to the second major branch of the archaebacterial kingdom (1). The AB'B''C type, characterized by a split B component, is common to the methanogens and extreme halophiles, other representatives of the second major branch. Immunochemical cross-reactivity indicates that the two (or three) largest components of eubacterial, eukaryotic nuclear, and archaebacterial RNA polymerases are homologous (4, 8). Interkingdom homologies between smaller components, including the archaebacterial C subunits, have not yet been ascertained.

The complex component patterns of archaebacterial RNA polymerases, comprising around 10 subunits each, resemble those of the nuclear eukaryotic rather than the more streamlined eubacterial enzymes. Accordingly, immunochemical cross-reactivity was larger between archaebacterial and eukaryotic components than between archaebacterial and eubacterial components (8), and the consensus sequences of putative promoters of archaebacterial stable RNA and protein genes (9, 10) are strikingly similar to those of eukaryotic promoters for RNA polymerase II (11).

Here we compare the derived amino acid sequences of components A and C from three archaebacteria with each other and with the sequences of a eukaryotic A subunit (largest) and a eubacterial β' subunit. We present unrooted phylogenetic trees constructed from the sequence distance values (12, 13) and from parsimony analyses (13, 14) and discuss implications for the genesis of the eukaryotic chimera. The complete nucleotide sequences for the B'', B', A, and C genes of *H. halobium* (15), for the B, A, and C genes of *Su. acidocaldarius* (G. P., unpublished results), and for the adjacent reading frames, the organization of these sequences into transcription units, and their expression will be described separately.

MATERIALS AND METHODS

Material. Deoxyadenosine [α-[^{35}S]thio]triphosphate was from Amersham. Restriction enzymes, T4 DNA ligase, and *E. coli* DNA polymerase I (Klenow fragment) were from Pharmacia and the BAL-31 exonuclease was from BRL.

Bacterial Strains and Culture Conditions. *Su. acidocaldarius* DSM 639 was obtained from the Deutsche Sammlung von Mikroorganismen, and *H. halobium* strain R_1 was a gift from D. Oesterhelt (Martinsried). They were grown as described by Huet *et al.* (8) and Oesterhelt and Stoeckenius (16).

Isolation of the RNA Polymerases and Preparation of the Antibodies. The RNA polymerases were isolated as described by Huet *et al.* (8) and Madon and Zillig (17). The preparation of antibodies was as described by Huet *et al.* (8).

Cloning and Screening of the Expression Library and Genomic Libraries. *Bam*HI (*Su. acidocaldarius*) or partial *Hae* III (*H. halobium*, fragments in the size range 0.5–3.0 kilobases ligated to *Eco*RI linkers) digested chromosomal DNA was "shotgun" cloned into expression vectors PIN III_{1-3} (18)

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: pol I, II, and III, RNA polymerases I, II, and III, respectively.

or λgt11 (19), respectively, as described by Maniatis *et al.* (20). The clones were screened with polyclonal rabbit antisera against the C subunit of the RNA polymerase as described by Konheiser *et al.* (21). Positive clones were detected by using peroxidase-coupled rabbit antibodies (Sigma) and a color reaction with 4-chloro-1-naphthol. Inserts in the positive clones were then used as probes to screen genomic libraries constructed in λ-EMBL4 (22) (*Su. acidocaldarius*) or λL47 (20) (*H. halobium*) to obtain clones containing the contiguous DNA regions, including the genes for the other polymerase subunits.

Sequencing. Both strands of restriction fragments and BAL-31 exonuclease deletion derivatives of the polymerase gene subclones (23) were cloned into the single-strand vectors M13tg130/31 (24) or M13mp18/19 (25). They were sequenced by the dideoxy-nucleotide chain-termination method (26). The N-terminal protein sequence analyses were performed as described by Reiter *et al.* (27).

Computer Analyses. Sequence alignment was done with the programs GAP, BESTFIT, LINEUP, PROFILE, and PROFILEGAP of the UWGCG package (28). Alignment was hand-corrected for mismatches of obvious short signatures.

The phylogenetic dendrograms were constructed with the help of the programs FITCHPRO, PROTPARSPRO, and DNAML of the PHYLIP, version 3.0, program package (13) and the algorithm of Lake (29) as described (30).

RESULTS

Cloning. RNA polymerase component C gene of *Su. acidocaldarius* and a component C gene fragment for *H. halobium* were identified after shotgun cloning of genomic DNA fragments into expression vectors using polyclonal anti-C component antibodies. Adjacent fragments were selected from genomic libraries and sequenced by the chain-termination method (26). The component C and B (or B′+B″) genes of both organisms were assigned using N-terminal protein sequence analyses. The A subunits of both enzymes had a blocked N terminus and their genes were identified by sequence homology with other A (or β′) components.

Organization of Genes. RNA polymerase component A and C genes were found to be closely linked in both organisms (Fig. 1). The A genes followed the B (*Su. acidocaldarius*) or B′ (*H. halobium*) genes directly, but in a different frame, and were followed directly by the C genes, also in a different frame. In *Su. acidocaldarius* an overlap of 5 nucleotides occurred between the B and A genes and in *H. halobium* the A gene overlapped the C gene by 5 nucleotides. Shine–Dalgarno sequences, complementary to the 3′ termini of 16S rRNA (32, 33), were a few nucleotides upstream of initiation codons for the A and C genes in both organisms.

Alignment. Derived amino acid sequences of the A and C components from *Su. acidocaldarius* and *H. halobium* were aligned with each other and with other available polymerase sequences—for the A subunit from *Methanobacterium thermoautotrophicum* (34), for the A components of pol I, II, and III from *Saccharomyces cerevisiae* (*Sa. cerevisiae*) (3, 35), for the A component of pol II from *Mus musculus* (ref. 36 and data not shown), for the A components of pol II and III from *Trypanosoma brucei* (37, 38), and for the β′ subunits from *E. coli* and *Marchantia polymorpha* chloroplasts (39, 40)—(Fig. 2) by using the UWGCG package (28). Several errors, caused by a tendency to rate an average fit higher than the conservation of signatures (short consensus sequences), were hand-corrected after all sections of the alignments were inspected.

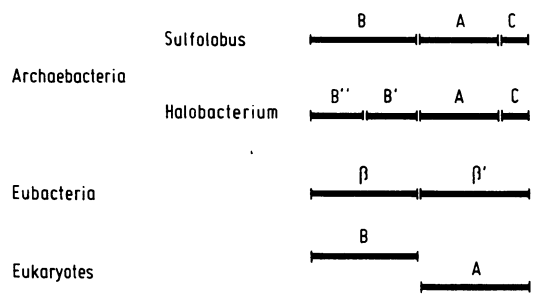


FIG. 1. Gene linkage for the large components of DNA-dependent RNA polymerase of two archaebacteria and eubacteria (31) and eukaryotes. The B and A subunit genes are unlinked in eukaryotes.

Nine domains that are highly conserved in all sequences are separated by stretches in which either the archaebacterial plus the eukaryotic pol II and III A components or the eubacterial components show exact gaps whereas the other sequences are well aligned. The pol I A sequence is ambivalent in that it corresponds to the eubacterial β′ components or to the other eukaryotic A components in such regions. In addition, there are some regions of questionable alignment.

Archaebacterial A sequences show colinearity with, and homology to, roughly the first two-thirds of eukaryotic A and eubacterial β′ sequences, and archaebacterial C sequences show colinearity with about the last third of the eukaryotic A and the eubacterial β′ sequences (see Fig. 1). Thus, two adjacent archaebacterial genes correspond to one contiguous gene in eukaryotes and eubacteria. However, the C-terminal heptapeptide repeat, characteristic for eukaryotic pol II A subunits (41), is absent from the archaebacterial C components. Two adjacent genes, the first for B″ and the second for B′, in *H. halobium* and *M. thermoautotrophicum*, correspond to one contiguous B gene in *Su. acidocaldarius*.

Homology. Two sequence distance matrices are shown in Fig. 3; the upper right was calculated from identities in the occupation of positions and the lower left evaluated each pair of amino acids as proposed by Dayhoff *et al.* (42). As expected, high homology (lowest sequence distance) is observed between the archaebacterial sequences. However, homologies between archaebacterial and eukaryotic pol II and III sequences are not much smaller. They are as high as, or even higher than, those between eukaryotic pol II and III sequences. The lowest homologies are seen between archaebacterial sequences and eukaryotic pol II and III sequences, respectively, and eubacterial sequences.

Phylogenetic Trees. Unrooted phylogenetic dendrograms calculated by the distance matrix method of Fitch and Margoliash (12, 13) from either matrix in Fig. 3 without specific correction showed that archaebacteria are a coherent group in the immediate neighborhood of eukaryotic pol II and III A lineages, whereas the pol I A lineage appears to be separate, showing a bifurcation with the eubacterial β′ lineage (Fig. 4). Stepwise restriction of the data basis to positions with increasing conservation (increasing identical occupation) did not change this unexpected result. Trees obtained both from the identity and the similarity (42) matrix were subjected to treatment by two algorithms designed to correct for artifacts possibly introduced by multiple exchanges in long branches. A correction employing the Poisson distribution (43) did not change the identity tree but moved the pol I lineage onto a common stem with pol II and III in the similarity tree. Treatment as described by Dayhoff *et al.* (42) did not change the topology of either tree, but led to a strong increase of the mean standard deviation. The standard deviations of the selected trees were generally only slightly lower than those calculated for alternative topologies, indicating that the choice of the most probable distance matrix dendrogram might not be significant.

The parsimony method (13, 14) yielded the same tree as the distance matrix method without correction. An evaluation of different dendrograms by comparing the numbers of nucle-

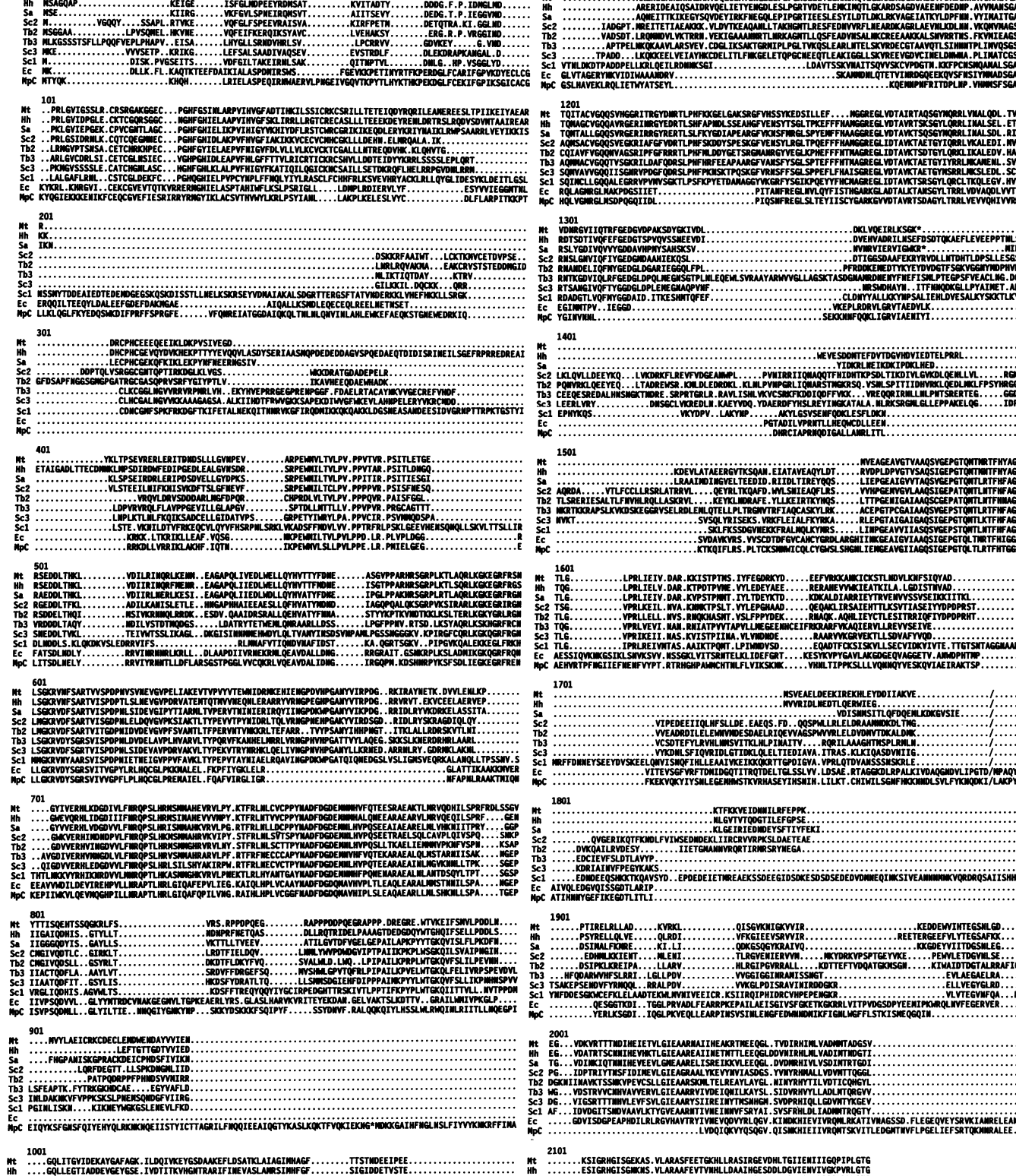


FIG. 2. Aligned sequences of the A plus C components of the DNA-dependent RNA polymerases of *H. halobium* (Hh), *Su. acidocaldarius* (Sa), and *Methanobacterium thermoautotrophicum* (Mt), the A components of pol II from *T. brucei* (Tb2) and *Sa. cerevisiae* (Sc2), of pol III from *Sa. cerevisiae* (Sc3) and *T. brucei* (Tb3), and of pol I from *Sa. cerevisiae* (Sc1), of the β' subunits from *E. coli* (Ec), and of the β' and β'' components of *M. polymorpha* chloroplasts (MpC). The numbers refer to positions in the alignment including gaps. The tilted bar at position 1735 indicates an insert present only in the chloroplast β'' sequences.

	Mt	Hh	Sa	Mm2	Sc2	Tb2	Tb3	Sc3	Sc1	Ec	MpC	SoC
Mt	-	53.8	50.1	39.2	40.0	33.6	35.3	37.4	30.4	26.9	24.5	24.2
Hh	62.1	-	47.4	36.8	38.5	32.8	33.5	35.6	28.4	28.9	25.0	24.8
Sa	52.6	53.6	-	40.2	40.0	35.5	36.3	40.0	31.7	28.5	26.5	25.9
Mm2	41.6	39.2	42.5	-	53.7	41.4	33.1	34.8	30.2	25.0	22.2	21.8
Sc2	41.8	40.6	40.8	60.3	-	41.8	32.7	35.1	30.5	24.3	23.7	23.0
Tb2	34.6	34.0	36.7	43.7	44.2	-	32.5	32.4	27.9	22.4	22.4	22.1
Tb3	37.2	35.5	38.0	30.8	31.3	32.0	-	40.0	29.3	25.1	24.2	23.3
Sc3	39.4	38.2	42.5	34.9	35.7	31.7	41.8	-	30.6	24.5	23.1	22.7
Sc1	27.8	25.3	29.8	28.4	29.3	25.5	26.9	29.0	-	23.2	20.4	19.3
Ec	25.8	26.3	25.8	21.7	19.6	18.3	19.3	20.5	17.7	-	40.7	37.8
MpC	19.6	19.6	21.2	15.4	15.7	14.7	16.5	15.7	14.0	40.4	-	58.8
SoC	18.8	19.6	20.2	13.9	14.6	13.1	15.5	14.4	11.0	37.2	63.6	-

FIG. 3. Percent sequence homologies between pairs of archaebacterial A plus C, eukaryotic nuclear A, and eubacterial β' components of the DNA-dependent RNA polymerases, calculated from the alignment presented in Fig. 2. In the upper right triangle only pairwise identities were considered; in the lower left triangle values were calculated from scores obtained with the aid of the mutational matrix of Dayhoff *et al.* (42). Mm2, pol II from *Mus musculus*; SoC, *Spinacea oleracea* chloroplast. Other abbreviations are as in Fig. 2.

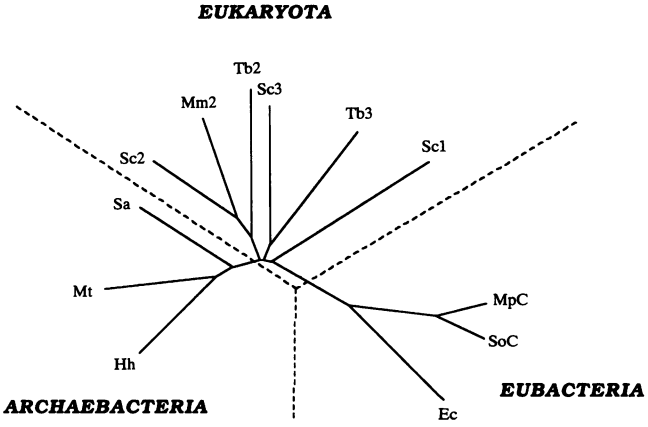


FIG. 4. Unrooted phylogenetic tree of the A plus C, A, β', and β' plus β'' components of the DNA-dependent RNA polymerases, listed in Fig. 3, calculated from the sequence homologies according to Fitch and Margoliash (12) and Felsenstein (13) from the 715 positions in the alignment in which all 12 sequences are represented, considering identities only with no further correction. Abbreviations are as in Figs. 2 and 3.

otide exchanges required for their construction showed that trees in which the pol II and III lineages changed positions and that trees in which pol II and III originated from a common stem seem to be insignificantly less probable. Trees in which pol I originated from one stem with pol II and III are, however, clearly less significant. The tree in which the "eocyte" *Su. acidocaldarius* (44) bifurcates with the eukaryotes and the "photocyte" *H. halobium* (45) with the eubacteria has the lowest significance. This result remained essentially unchanged when instead of all positions only the highly conserved positions without a gap were considered (Fig. 5).

The archaebacterial and the pol III and II lineages are of about equal length. Their relations are thus probably correctly described by the discussed trees. But the pol I and especially the β' lineages are longer. The maximum likelihood program of Felsenstein (46), a DNA parsimony algorithm considering possible branch length (multiple exchange) effects on tree topology, also yielded a dendrogram in which pol II bifurcated with the archaebacteria and pol I bifurcated with the eubacteria.

The analysis of the positions of the alignment on the DNA level, employing the proposed rate-invariant evolutionary parsimony algorithm of Lake (29), yielded a strong preponderance of the topologies of four-membered dendrograms, indicating the unity of the archaebacteria rather than suggesting the eocyte–photocyte tree. Determination of distance values yielded a tree in which pol II and III share a stem that bifurcates with the archaebacteria (in which *Su. acidocaldarius* and *H. halobium* have changed positions) and pol I, again, bifurcates with the eubacteria.

DISCUSSION

As described in detail (ref. 15 and G.P., unpublished results), the genes for the large components of the archaebacterial RNA polymerases are organized in transcription units in an arrangement corresponding to that of the *E. coli* β and β' genes in the *rpoBC* operon. The neighboring genes, however, do not correspond to those found in *E. coli* at the same positions. These transcription units thus add to increasing evidence for a similarity, though not identity, of the organization of archaebacterial and eubacterial genomes.

The high homology between the archaebacterial A plus C genes and the corresponding eukaryotic pol II and III A genes is in contrast to the situation found for rRNA (1). It is, however, in line with the discovery that archaebacterial translation factor and ribosomal protein genes, which are also organized in operons resembling those found in *E. coli*, show considerably more sequence homology to their eukaryotic than to their eubacterial counterparts (47, 48).

In accord with the immunochemical analysis (4, 8) and with the nature of putative archaebacterial transcription promoters (9, 10) and terminators (49–51), the sequence data thus provide strong evidence that the structure and function of

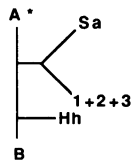
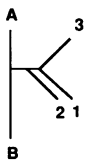
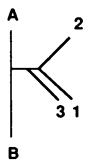
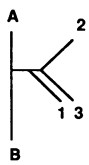
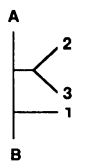
Positions of entire alignment	Tree topology (schematic)						
							
2172	18522	18259	18299	17605	16907	16893	16909
(all)	+75.0	+62.9	+64.7	+32.8	+0.6	0.0	+0.7
715	3828	3792	3789	3801	3768	3758	3749
(without gap)	+11.0	+6.0	+5.6	+7.3	+2.7	+1.3	0.0

FIG. 5. Nucleotide exchanges (upper values) required for construction of parsimony trees with various branching topologies. A, archaebacteria; 2, 3, 1, pol II, III, and I, respectively; B, eubacteria. The tree in column 1 (A*, methanogens, Sa, *Sulfolobus;* Hh, *Halobacterium*) was proposed by Lake (45). The number of exchanges per 100 amino acid residues required in addition to that needed for the optimal topology is given (lower values). The seven columns give the values for different fractions of the alignment from 2172 positions (first line), down to only the 715 sites, in which all 12 sequences are represented (second line).

archaebacterial RNA polymerases and eukaryotic pol II and III are of the same type.

Even in the most probable phylogenetic dendrograms, the relative placement of the eukaryotic pol II and III lineages, including their possible origin from a common ancestral enzyme, remains ambiguous. However, these dendrograms show the pol I lineage arising separately from a bifurcation with the eubacterial β' lineage. The three specialized eukaryotic nuclear RNA polymerases have thus not necessarily arisen by diversification from one eurkaryotic polymerase.

The bifurcations of pol II/III with the archaebacteria and pol I with the eubacteria could be explained by assuming that these lineages arose separately—pol II/III from the early archaebacterial and pol I from the early eubacterial lineage(s)—and became components of the eukaryotic genome through a fusion incident or by acquisition through endosymbionts or by horizontal gene transfer, events discussed for the genesis and the development of the eukaryotic chimera. The alternative that the archaebacterial and/or the eubacterial enzymes arose by reduction events out of the eukaryotic lineage from the already diversified pol II and I, respectively, appears far less probable. The insertion of further especially deep branching and short eubacterial and pol I lineages could increase the significance of this hypothesis. But the uncertainty regarding the origins of the eukaryotic polymerases accentuates their early separation. Archaebacteria thus testify to an ancestral archaebacterial nature of a fraction of the eukaryotic nuclear genome. If archaebacteria are primitive (i.e., closer to the root than the apparently highly streamlined eubacteria), the "archaebacterial" genes of eukaryotes are primitive in the same sense. The factors responsible for retarding evolution in the two kingdoms might, however, be fundamentally different—e.g., seclusion in extreme niches for the archaebacteria and the existence of a complex network of intermolecular interactions for the eukaryotes.

These data support the view that extant life should be classified into three (1, 2, 52) rather than two or five distinct kingdoms (44, 45). Like many other features [e.g., rRNA primary structures or the nature of the lipids (for a more complete listing, see ref. 53)], RNA polymerase structure and gene organization indicate the coherence of the archaebacterial kingdom for the following reasons: (*i*) The component sequences form a coherent cluster in the most probable phylogenetic trees. (*ii*) The open reading frames flanking the polymerase genes in the genomes of the two archaebacteria correspond to each other but not to those found in corresponding positions in eubacteria. (*iii*) The unique A–C gene separation, in contrast to the contiguous A and β' genes in eukaryotes and *E. coli*, respectively, appears to be the rule among archaebacteria, although the apparent subunit sizes and immunochemical data suggest that *Thermoplasma* (54) and *Archaeoglobus* (55) may constitute exceptions. (*iv*) *H. halobium* and *Methanobacterium thermoautotrophicum* show the same B gene separation.

1. Woese, C. R. & Olsen, G. J. (1986) *Syst. Appl. Microbiol.* **7**, 161–177.
2. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
3. Memet, S., Gouy, M., Marck, C., Sentenac, A. & Buhler, J.-M. (1988) *J. Biol. Chem.* **263**, 2830–2839.
4. Schnabel, R., Thomm, M., Gerardy-Schahn, R., Zillig, W., Stetter, K. O. & Huet, J. (1983) *EMBO J.* **2**, 751–755.
5. Gropp, F., Reiter, W. D., Sentenac, A., Zillig, W., Schnabel, R., Thomm, M. & Stetter, K. O. (1986) *Syst. Appl. Microbiol.* **7**, 95–101.
6. Stetter, K. O. & Zillig, W. (1985) in *The Bacteria*, eds. Gunsalus, I. C., Sokatch, J. R., Ornston, L. N., Woese, C. R. & Wolfe, R. S. (Academic, Orlando, FL), Vol. 8, pp. 85–170.
7. Zillig, W., Holz, I., Klenk, H.-P., Trent, J., Wunderl, S., Janekovic, D., Imsel, E. & Haas, B. (1987) *Syst. Appl. Microbiol.* **9**, 62–70.
8. Huet, J., Schnabel, R., Sentenac, A. & Zillig, W. (1983) *EMBO J.* **2**, 1291–1294.
9. Reiter, W.-D., Palm, P. & Zillig, W. (1988) *Nucleic Acids Res.* **16**, 1–19.
10. Thomm, M. & Wich, G. (1988) *Nucleic Acids Res.* **16**, 151–163.
11. Bucher, P. & Trifonov, E. N. (1986) *Nucleic Acids Res.* **14**, 10009–10027.
12. Fitch, W. M. & Margoliash, E. (1967) *Science* **155**, 279–284.
13. Felsenstein, J. (1984) in *Cladistics: Perspectives in the Reconstruction of Evolutionary History*, eds. Ducan, T. & Stuessy, T. F. (Columbia Univ. Press, New York), pp. 169–191.
14. Hennig, W. (1966) *Phylogenetic Systematics* (Univ. of Illinois Press, Urbana).
15. Leffers, H., Gropp, F. & Garrett, R. (1989) *J. Mol. Biol.* **206**, 1–17.
16. Oesterhelt, D. & Stoeckenius, W. (1974) *Methods Enzymol.* **31**, 667–678.
17. Madon, J. & Zillig, W. (1983) *Eur. J. Biochem.* **133**, 471–474.
18. Inouye, M. (1983) *Experimental Manipulation of Gene Expression*, ed. Inouye, M. (Academic, New York).
19. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Konheiser, U., Pasti, G., Bollschweiler, C. & Klein, A. (1984) *Mol. Gen. Genet.* **198**, 146–152.
22. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
23. Guo, L. H., Yang, R. C. A. & Wu, R. (1983) *Nucleic Acids Res.* **11**, 5521–5540.
24. Kieny, M. P., Lathe, R. & Lecocq, I. P. (1983) *Gene* **26**, 91–99.
25. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–114.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Reiter, W.-D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W. & Grampp, B. (1987) *Mol. Gen. Genet.* **206**, 144–153.
28. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
29. Lake, J. A. (1987) *J. Mol. Evol.* **26**, 59–73.
30. Holmquist, R., Miyamoto, M. M. & Goodman, M. (1988) *Mol. Biol. Evol.* **5**, 217–236.
31. Scaife, J. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 207–225.
32. Olsen, G. J., Pace, N. R., Nuell, M., Kaine, B. P., Gupta, R. & Woese, C. R. (1985) *J. Mol. Evol.* **22**, 301–307.
33. Mankin, A. S., Kagramanova, V. K., Teterina, N. L., Rubtsov, P. M., Belova, E. N., Kopylov, A. M., Baratova, L. A. & Bogdanov, A. A. (1985) *Gene* **37**, 181–189.
34. Berghöfer, B., Kröckel, L., Körtner, C., Truss, M., Schallenberg, J. & Klein, A. (1988) *Nucleic Acids Res.* **16**, 8113–8128.
35. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) *Cell* **42**, 599–610.
36. Corden, L. J., Cadena, D. L., Ahearn, J. M. & Dahmus, M. E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7934–7938.
37. Evers, R., Hammer, A., Köck, J., Jess, W., Borst, P., Memet, S. & Cornelissen, A. W. C. A. (1989) *Cell* **56**, 585–597.
38. Köck, J., Evers, R. & Cornelissen, A. W. C. A. (1988) *Nucleic Acids Res.* **16**, 8753–8772.
39. Ovchinnikov, Y. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Salomatina, I. S., Shuvaeva, T. M., Lipkin, V. M. & Sverdlov, E. D. (1982) *Nucleic Acids Res.* **10**, 4035–4044.
40. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1986) *Nature (London)* **322**, 572–574.
41. Nonet, M., Sweetser, D. & Young, R. A. (1987) *Cell* **50**, 909–915.
42. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), pp. 345–358.
43. Feng, D. F., Johnson, M. S. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21**, 112–125.
44. Lake, J. A., Anderson, E., Oakes, M. & Clark, M. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3786–3790.
45. Lake, J. A. (1986) *Nature (London)* **319**, 626.
46. Felsenstein, J. (1973) *Syst. Zool.* **22**, 240–249.
47. Lechner, K. & Böck, A. (1987) *Mol. Gen. Genet.* **208**, 523–528.
48. Auer, J., Lechner, K. & Böck, A. (1989) *Can. J. Microbiol.* **35**, 200–204.
49. Kjems, J. & Garrett, R. A. (1987) *EMBO J.* **6**, 3521–3530.
50. Reiter, W.-D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W. & Grampp, B. (1987) *Mol. Gen. Genet.* **206**, 144–153.
51. Murphy, S., Di Liegro, C. & Melli, M. (1987) *Cell* **51**, 81–87.
52. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magnum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. (1980) *Science* **209**, 457–463.
53. Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Pühler, G. & Klenk, H.-P. (1988) *Eur. J. Biochem.* **173**, 473–482.
54. Sturm, S., Schönefeld, U., Zillig, W., Janekovic, D. & Stetter, K. O. (1980) *Zentralbl. Bakteriol. Abt. 1 Orig. C* **1**, 12–25.
55. Stelter, K. O. (1988) *System. Appl. Microbiol.* **10**, 172–173.