Curr Genet (1986) 10:405–410

Current Genetics

Structure and evolution of the small subunit ribosomal RNA gene of *Crithidia fasciculata*

Murray N. Schnare, James C. Collings, and Michael W. Gray

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Summary. We present the cloning and sequence analysis of the nuclear-encoded *Crithidia fasciculata* small subunit (SSU) rRNA gene, the longest (2,206 bp) such gene yet characterized by direct sequence analysis. Much of the sequence can be folded to fit a phylogenetically conserved secondary structure model, with the additional length of this gene being accommodated within discrete variable domains that are present in eukaryotic SSU rRNAs. On the basis of sequence comparisons, we conclude that *Crithidia* contains the most highly diverged SSU rRNA described to date among the eukaryotes, and therefore represents one of the earliest branchings within the eukaryotic primary kingdom.

Key words: *Crithidia fasciculata* – rRNA – Sequence – Evolution

Introduction

The ribosome of *Crithidia fasciculata*, a trypano somatid protozoan, is unusual in that its large subunit contains four novel small rRNAs in addition to the 5S and 5.8S rRNAs found in other eukaryotes (Gray 1979, 1981; Schnare et al. 1983). Decisions about whether the appearance of these novel small rRNAs in *Crithidia* was an ancient or relatively recent event in the evolution of eukaryotic rRNA genes will rest ultimately on any data we can obtain about the phylogenetic position of *Crithidia*.

Analysis of the *Crithidia* 5.8S rRNA sequence (Schnare and Gray 1982) and the 3′-terminal small subunit (SSU) rRNA sequence (Schnare and Gray 1981) previously led us to conclude that *Crithidia* contains the most highly diverged rRNAs of all the eukaryotes examined to date. However, it was subsequently concluded (McCarroll et al. 1983) that *Dictyostelium discoideum* 5.8S rRNA represents an earlier branching within the eukaryotic line of descent than does the 5.8S rRNA of *Crithidia*.

Comparisons of complete SSU rRNA sequences have been valuable in establishing secondary structure models (Woese et al. 1983; Nelles et al. 1984; Olsen et al. 1983; Atmadja et al. 1984; Hogan et al. 1984; Stiegler et al. 1981) and in determining phylogenetic relationships among organisms (McCarroll et al. 1983; Küntzel and Köchel 1981; Gray et al. 1984). We have therefore cloned and determined the sequence of a SSU rRNA gene from *Crithidia fasciculata* and have used the data to establish that *Crithidia* does indeed represent one of the earliest branchings within the eukaryotes, earlier in fact than that defined by the *Dictyostelium* SSU rRNA sequence (McCarroll et al. 1983).

Methods

Isolation of Crithidia DNA. Cells were harvested from 250 ml of a three day culture of *C. fasciculata* (Palmer 1973) and lysed by passage through a French pressure cell at 2,000 lb in^{-2}, as described (Gray 1981). The lysate was centrifuged at 15,000 × g for 15 min. The pellet was washed by resuspension in 0.44 M sucrose – 50 mM Tris · HCl (pH 8.0) – 1 mM EDTA followed by centrifugation at 1,000 × g for 10 min (2×). After final resuspension of the washed nuclear pellet in 10 ml of this same buffer, DNA was extracted by the phenol-detergent method (Parish and Kirby 1966), precipitated by addition of two volumes of 95% ethanol, and stored in ethanol at –20 °C.

Purification, end-labeling and 5′-terminal sequence analysis of Crithidia SSU rRNA. Crithidia SSU rRNA was prepared by electrophoresis of *Crithidia* 3 M NaCl-insoluble RNA (Gray 1979) in a 2.5% polyacrylamide gel for 5 h at 500 V (Schnare and Gray

Offprint requests to: M. N. Schnare

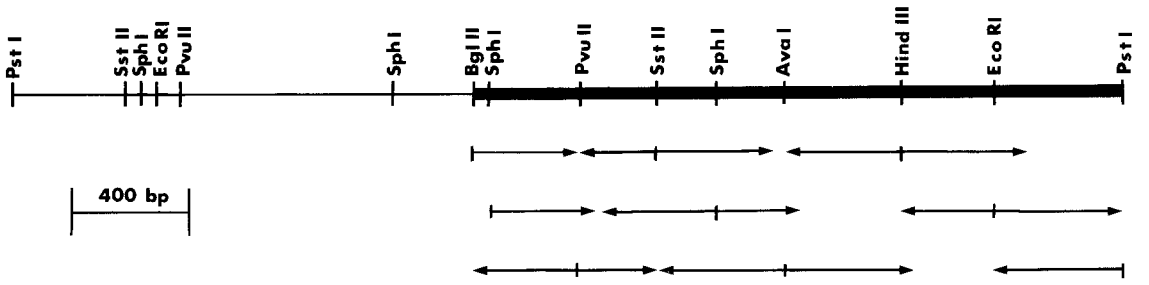

Fig. 1. Restriction map of the pCf1 insert and sequencing strategy for the *C. fasciculata* SSU rRNA gene. The entire sequence was determined from both strands and overlapping sequence data were obtained for each restriction site within the gene

1981). The SSU rRNA was visualized by U.V. shadowing (Hassur and Whitlock 1974) and extracted (Rubin 1973). Alternatively, *Crithidia* SSU rRNA was isolated by phenol extraction of purified small ribosomal subunits (Gray 1981).

C. fasciculata SSU rRNA was 5′-end-labeled (Schnare et al. 1985), re-purified by gel electrophoresis, and subjected to enzymatic sequence analysis (Donis-Keller et al. 1977; Simoncsits et al. 1977; Donis-Keller 1980) and terminal nucleotide analysis (MacKay et al. 1980). For use in hybridization experiments, SSU rRNA and 3 M NaCl-insoluble RNA were partially hydrolyzed with alkali (0.15 M NH_4OH, 90 °C, 2 min, followed by lyophilization) prior to 5′-end-labeling.

Cloning and sequence analysis of the C. fasciculata SSU rRNA gene. *Pst*I-digested pUC9 and *C. fasciculata* genomic DNA in a 4:1 mass ratio were ligated with T4 DNA ligase (Weiss et al. 1968). Competent *E. coli* JM83 cells (Messing 1979) were transformed with 10 ng of ligated DNA as described (Hanahan 1983). White colonies were chosen for in situ colony hybridization (Grunstein and Hogness 1975) with 5′-end-labeled RNA. Minipreparations of plasmid DNA (Holmes and Quigley 1981) from positive colonies were further characterized by digestion with *Pst*I, agarose gel electrophoresis, and Southern hybridization (Southern 1975) with end-labeled RNA as described (Bonen and Gray 1980). A large-scale plasmid preparation (Schlief and Wensink 1981) of clone pCf1 (whose 3.75 kbp insert contains the 18S rRNA gene) was used for detailed restriction site mapping (Falconet et al. 1984) and chemical sequence analysis (Spencer et al. 1984).

Results and discussion

A single 3.75 kbp DNA fragment was detected upon hybridization of *Crithidia* SSU rRNA to *Pst*I-digested *C. fasciculata* genomic DNA, while primary sequence analysis of the 3′-terminus of the SSU rRNA (Schnare and Gray 1981) revealed the presence of a *Pst*I site 13 nucleotides from the 3′-end of the molecule. We therefore concluded that the bulk of the SSU rRNA gene (minus the 3′-terminal 13 nucleotides) could be isolated as a single *Pst*I clone and that the 5′-terminus of the gene would be located toward the center of the clone. The restriction map of the corresponding clone (pCf1) is presented in Fig. 1. Southern hybridization experiments allowed positioning of the 5′-end of the SSU rRNA gene between the single *Bgl*II site and a *Pvu*II site located 360 bp to the right of this *Bgl*II site.

The nucleotide sequence from the *Bgl*II site to the *Pst*I site at the 3′-end of the gene was determined according to the sequencing strategy shown in Fig. 1. Comparison of the 5′-terminal SSU rRNA sequence (see Methods) and the DNA sequence presented in Fig. 2 revealed that the 5′-terminus of the gene coincides precisely with the position of the *Bgl*II site. The sequence at the 3′-terminus of the gene is in good agreement with our previously published RNA sequence (Schnare and Gray 1981). However, T_{2148} in the DNA was reported as a C in the RNA sequence. This discrepancy is most likely the result of an error in interpreting the RNA sequencing gels, probably due to the presence of a modified nucleotide at this position in the mature rRNA.

The *Crithidia* SSU rRNA gene has a G + C content of 49.7% and is 2,206 bp long. This length is consistent with the unusually large size of *Crithidia* SSU rRNA previously measured by polyacrylamide gel electrophoresis (Gray 1979). Other sequenced eukaryotic (nuclear-encoded) SSU rRNA genes range in size from 1,753 bp in *Tetrahymena thermophila* (Spangler and Blackburn 1985) to 1,872 bp in *D. discoideum* (McCarroll et al. 1983). The *Crithidia* SSU rRNA gene is therefore 334 bp longer than the longest eukaryotic SSU rRNA gene sequence previously reported.

In order to define precisely the positions of the extra nucleotides in *Crithidia* SSU rRNA relative to other eukaryotic SSU rRNAs, we have attempted to fit the *Crithidia* sequence to the most recently proposed secondary structure model for eukaryotic SSU rRNAs (Nelles et al. 1984). In regions where this highly base-paired structure is not applicable to the *Crithidia* sequence (e.g., positions 352–403), we have used the more open structure of the *D. discoideum* model (Olsen et al. 1983). In each of these secondary structure models two alternative helical structures have been proposed for the central region of the molecule. The *Crithidia* sequence has about 105 extra nucleotides in this region (V8 of Fig. 3) relative to the other known sequences and therefore the eukaryote-specific structure proposed here (Nelles et al. 1984; Ol-

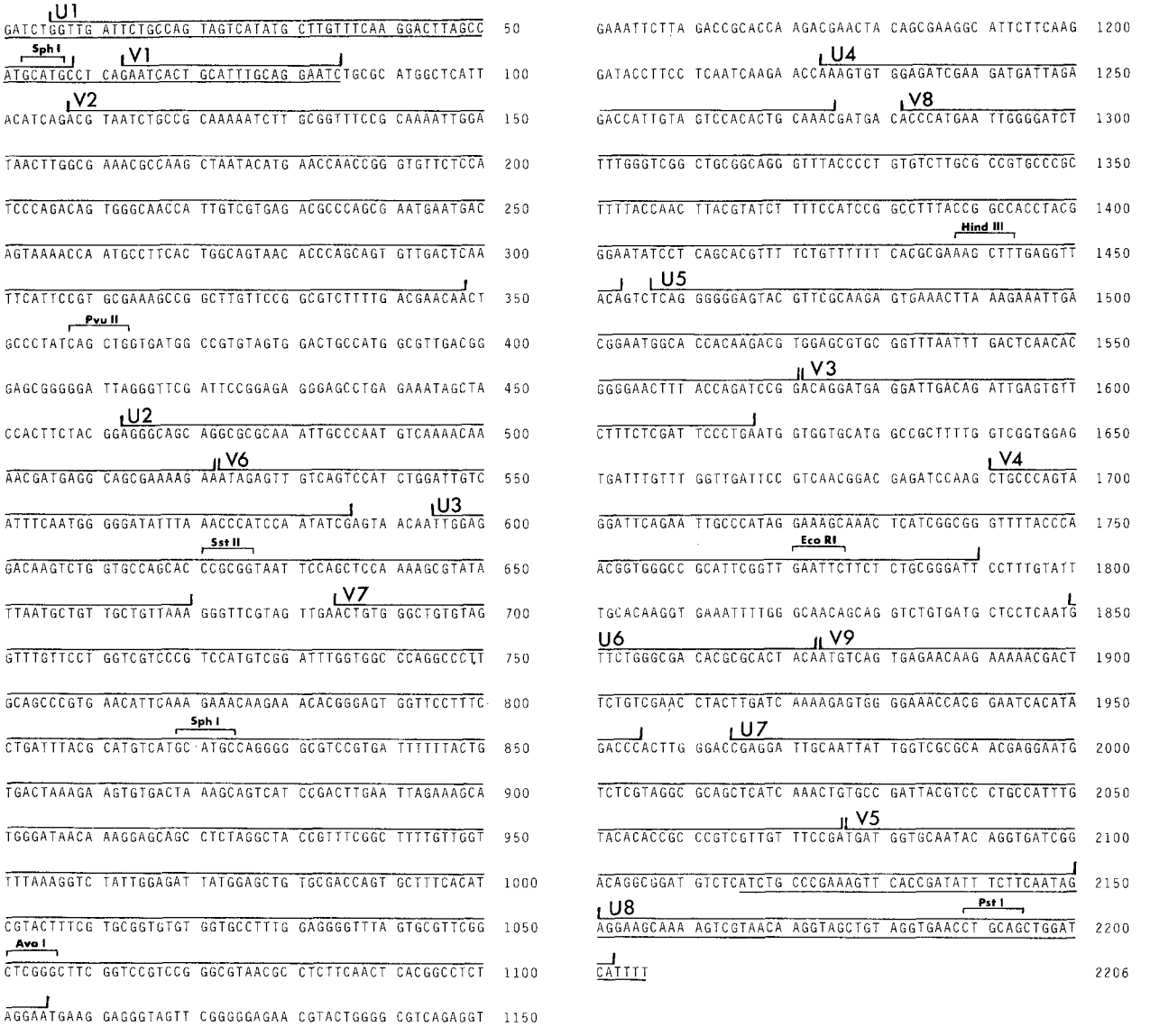

Fig. 2. Nucleotide sequence (RNA-like strand) of the SSU rRNA gene of *C. fasciculata*. The sequence of the 3′-terminal 91 nucleotides and the 5′-terminal 85 nucleotides (underlined regions) have also been determined by direct sequence analysis of end-labeled RNA (Schnare and Gray 1981; this paper). Restriction sites shown in Fig. 1 are indicated, as are the universal (U) and variable (V) regions

sen et al. 1983; Atmadja et al. 1984) cannot be drawn with confidence for the *Crithidia* SSU rRNA sequence. We have used the prokaryotic-like structure (Hogan et al. 1984) in this region (*Crithidia* positions 660–665 and 1,458–1,463), which represents part of a universally conserved secondary structural core (Stiegler et al. 1981; Spencer et al. 1984; Gray et al. 1984) in SSU rRNA. In fact, much of the rest of the central core of the secondary structure (encompassing residues 666–1,281, 1,454–1,523 and 1,956–2,054) fits a eubacterial model (Woese et al. 1983) of this region somewhat better than it fits proposed eukaryotic models (Nelles et al. 1984; Olsen et al. 1983). In this region we have generally followed the yeast secondary structure model (Hogan et al. 1984) which is more closely patterned after the eubacterial model.

Our analysis indicates the presence of eight regions of major structural variation within eukaryotic SSU rRNAs. These variable regions (V1, V2, V4–V9 of Fig. 3) are dispersed throughout the length of the molecule and are separated by regions of highly conserved primary and/or secondary structure. The variable regions correspond closely in position to regions of pronounced structural variation previously identified in a comparison of eubacterial and organellar (plastid and mitochondrial) SSU rRNA sequences (Spencer et al. 1984; Gray et al. 1984), and they have been numbered accordingly. An additional region, V3, differs from its eubacterial counterpart, but is

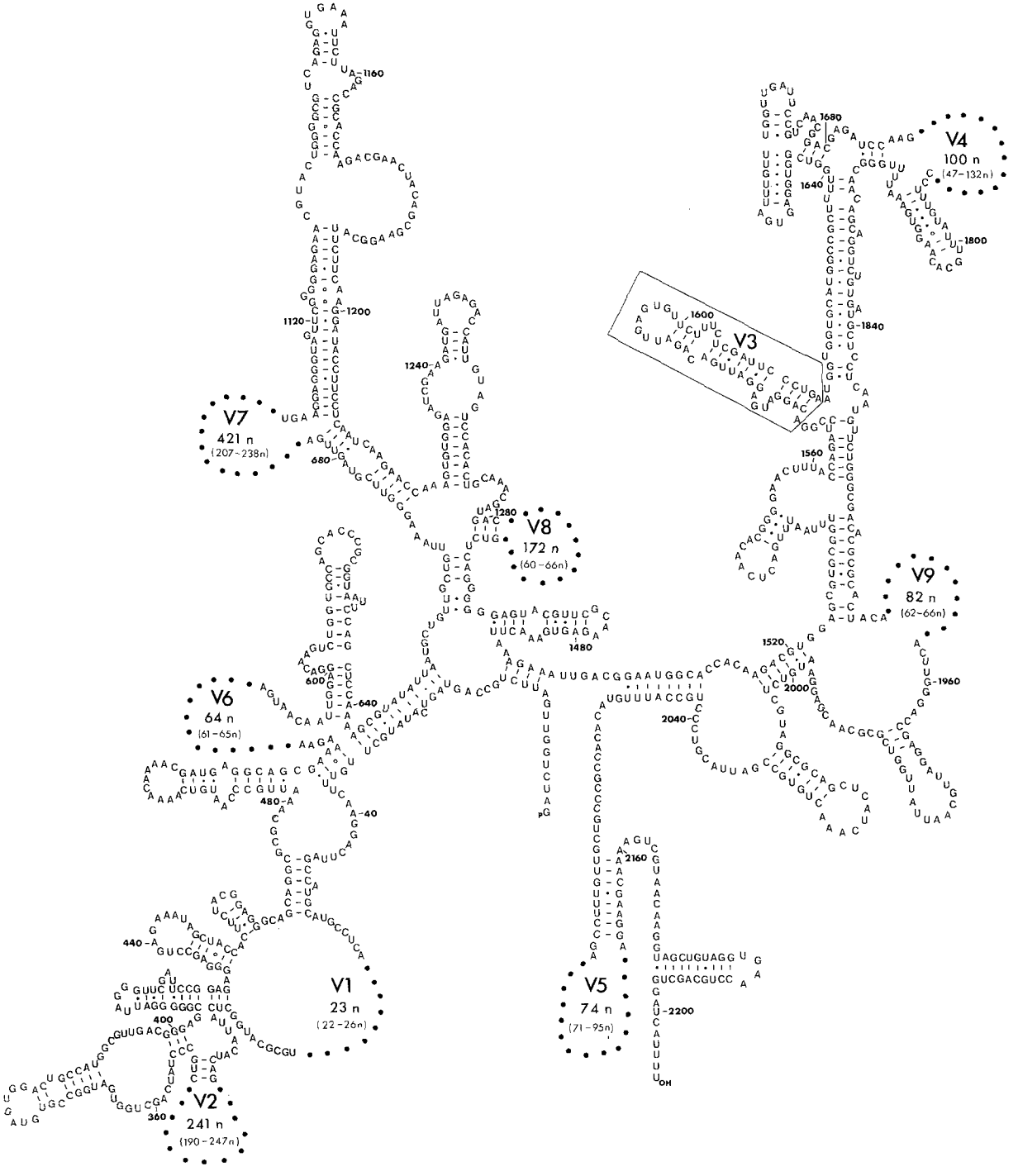

Fig. 3. Potential secondary structure of *C. fasciculata* SSU rRNA. The structure is adapted from the *Artemia* model (Nelles et al. 1984) with modifications as described in the text. Eight regions of primary and/or secondary structure variability among eukaryotes (V1, V2, V4–V9) have been identified. In each of these regions the length of the *Crithidia* sequence is given. The numbers in parentheses represent the range of lengths within the V regions of other eukaryotes. References for eukaryotic SSU rRNA sequences are given in the legend to Table 1

very well conserved among eukaryotes. It is evident from Fig. 3 that most of the extra length of the *C. fasciculata* SSU rRNA gene relative to other eukaryotic SSU rRNA genes can be accounted for by the extra ≈200 nucleotides in V7 and the extra ≈100 nucleotides in V8. The length of V4 is expanded in both *Crithidia* (100 n) and *Dictyostelium* (132 n) SSU rRNAs relative to those of other eukaryotes (47–56 n). We previously encountered a similar situation with the exceptionally long *Crithidia* 5.8S rRNA, in which the extra nucleotides are located

Table 1. Sequence conservation in pairwise comparisons of the universal regions of *Crithidia fasciculata* and other SSU rRNAs. The universal regions have been defined previously (Stiegler et al. 1981; Gray et al. 1984; Spencer et al. 1984) and correspond to *C. fasciculata* positions 6–58 (U1), 463–522 (U2), 595–670 (U3), 1,224–1,275 (U4), 1,457–1,571 (U5), 1,850–1,873 (U6), 1,965–2,076 (U7), and 2,051–2,202 (U8) (see also Fig. 2). The sequences are from *Escherichia coli* (Brosius et al. 1978; Carbon et al. 1979), ECO; *Xenopus laevis* (Salim and Maden 1981), XEN; mouse (Raynal et al. 1984), MOU; rabbit (Connaughton et al. 1984), RAB; *Artemia salina* (Nelles et al. 1984), ASA; *Saccharomyces cerevisiae* (Rubtsov et al. 1980; see also Gray et al. 1984), SCE; *Zea mays* (Messing et al. 1984), MAI; *Dictyostelium discoideum* (McCarroll et al. 1983), DIC; *Tetrahymena thermophila* (Spangler and Blackburn 1985), TET; *Crithidia fasciculata,* CRI; *Halobacterium volcanii* (Gupta et al. 1983), HVO. Within the universal regions, the rat sequence (Chan et al. 1984; Torczynski et al. 1983) is identical to that of mouse (Raynal et al. 1984), while the rice sequence (Takaiwa et al. 1984) has only one single base substitution and one single base deletion relative to the maize sequence (Messing et al. 1984)

	Identity (%)									
	XEN	MOU	RAB	ASA	SCE	MAI	DIC	TET	CRI	HVO
ECO	63.7	63.6	62.6	63.2	61.5	63.7	62.9	60.3	61.9	71.3
XEN		99.2	97.7	93.8	90.9	91.2	87.2	82.8	79.8	66.3
MOU			97.9	94.2	91.1	91.1	87.4	83.1	80.1	66.5
RAB				92.7	89.8	89.8	86.1	81.9	79.5	65.8
ASA					90.6	90.2	87.2	82.7	79.7	69.3
SCE						95.0	90.3	88.6	82.2	67.0
MAI							90.7	86.7	82.4	67.4
DIC								85.0	80.6	66.9
TET									77.5	66.4
CRI										65.4

in a region whose primary sequence and length are highly variable among 5.8S rRNAs (Schnare and Gray 1982). It should be noted that without additonal information on *Crithidia* SSU rRNA (e.g., T1 oligonucleotide catalogue, S1 analysis), we cannot rigorously exclude the possibility that some of the variable regions we identify at the gene level are, in fact, introns, and therefore do not appear in the mature rRNA. However, since the estimated size of *Crithidia* SSU rRNA (≈2,400 nucleotides; Gray 1979) closely matches the size of the gene encoding it (2,206 nucleotides; this paper), we consider this possibility highly unlikely.

Table 1 shows the extent of sequence identity between the universal regions (Spencer et al. 1984; Gray et al. 1984) of *Crithidia* SSU rRNA and other eukaryotic SSU rRNAs, as well as a representative of the eubacterial (*E. coli*) and archaebacterial (*H. volcanii*) kingdoms. The data presented in Table 1 clearly show that *Crithidia* SSU rRNA is not as closely related to other eukaryotic SSU rRNAs as these are to each other. Within the "universal core" region, the *Crithidia* sequence is as distant from the *Dictyostelium* sequence as it is from each of the other eukaryotic sequences, and is significantly less closely related to other eukaryotic sequences (80.6 ± 1.3% sequence identity; mean ± SD) than is the *Dictyostelium* sequence (88.2 ± 1.9%) ($P < 0.0005$). This analysis establishes that *Crithidia* is a representative of one of the earliest branchings within the eukaryotic line of descent. The high degree of divergence of the *Crithidia* sequence relative to those of other eukaryotes cannot be explained by a "fast evolutionary clock" (McCarroll et al. 1983) in *Crithidia,* since the *Crithidia* sequence has not diverged significantly further away from the prokaryotic sequences than have those of other eukaryotes (Table 1).

A phylogenetic tree based on SSU rRNA sequences has been constructed by McCarroll et al. (1983), who concluded that the point of divergence of *Dictyostelium* from other eukaryotes was shortly after the point at which the mitochondrial line diverged from the rest of the eubacteria. They further concluded that "eukaryotes that represent substantially deeper branches than *D. discoideum* will not have mitochondria as we know them". With this in mind, it is interesting that *Crithidia,* which appears from these data (Table 1) to branch much earlier than *D. discoideum,* contains a kinetoplast instead of typical mitochondria (Wallace 1979).

After completion of this paper, the sequence of the SSU rRNA gene of *Tetrahymena thermophila* (1,753 bp) was published (Spangler and Blackburn 1985), and we have included in Table 1 the similarity values obtained from comparison of its "universal core" with those of the other eukaryotes considered here. The data indicate that the *Tetrahymena* sequence is intermediate between those of *Dictyostelium* and *Crithidia* in its degree of divergence from the mammalian, fungal, and plant sequences; moreover, although the *Tetrahymena* "universal core" is no more distant from the *Dictyostelium* core (85.0% sequence identity) than it is from the mammalian,

fungal, and plant sequences (84.3 ± 2.7%), it is much less closely related to the *Crithidia* "universal core" (77.5%). In view of evidence suggesting that *Tetrahymena* branched off extremely early in the evolution of the eukaryotic lineage (Tarr and Fitch 1976; Glover and Gorovsky 1979), and the recent demonstration of a deviation from the universal genetic code in nuclear genes of *Tetrahymena* (Horowitz and Gorovsky 1985), it is particularly noteworthy that the trypanosomatid flagellates, as exemplified by *Crithidia*, appear to represent an even earlier branch than the one leading to the ciliates.

Acknowledgements. We thank Jean D. Gray for statistical analysis of the data in Table 1 and Lisa Laskey for typing this manuscript. This work was supported by a grant (A8387) from the Natural Sciences and Engineering Research Council of Canada to M. W. G.

References

Atmadja J, Brimacombe R, Maden BEH (1984) Nucleic Acids Res 12:2649–2667
Bonen L, Gray MW (1980) Nucleic Acids Res 8:319–335
Brosius J, Palmer ML, Kennedy PJ, Noller HF (1978) Proc Natl Acad Sci USA 75:4801–4805
Carbon P, Ehresmann C, Ehresmann B, Ebel J-P (1979) Eur J Biochem 100:399–410
Chan Y-L, Gutell R, Noller HF, Wool IG (1984) J Biol Chem 259:224–230
Connaughton JF, Rairkar A, Lockard RE, Kumar A (1984) Nucleic Acids Res 12:4731–4745
Donis-Keller H (1980) Nucleic Acids Res 8:3133–3142
Donis-Keller H, Maxam AM, Gilbert W (1977) Nucleic Acids Res 4:2527–2538
Falconet D, Lejeune B, Quetier F, Gray MW (1984) EMBO J 3: 297–302
Glover CVC, Gorovsky MA (1979) Proc Natl Acad Sci USA 76: 585–589
Gray MW (1979) Can J Biochem 57:914–926
Gray MW (1981) Mol Cell Biol 1:347–357
Gray MW, Sankoff D, Cedergren RJ (1984) Nucleic Acids Res 12:5837–5852
Grunstein M, Hogness DS (1975) Proc Natl Acad Sci USA 72: 3961–3965
Gupta R, Lanter JM, Woese CR (1983) Science 221:656–659
Hanahan D (1983) J Mol Biol 166:557–580
Hassur SM, Whitlock HW Jr (1974) Anal Biochem 59:162–164
Hogan JJ, Gutell RR, Noller HF (1984) Biochemistry 23:3322–3330
Holmes DS, Quigley M (1981) Anal Biochem 114:193–197
Horowitz S, Gorovsky MA (1985) Proc Natl Acad Sci USA 82: 2452–2455
Küntzel H, Köchel HG (1981) Nature (London) 293:751–755
MacKay RM, Spencer DF, Doolittle WF, Gray MW (1980) Eur J Biochem 112:561–576
McCarroll R, Olsen GJ, Stahl YD, Woese CR, Sogin ML (1983) Biochemistry 22:5858–5868
Messing J (1979) Recombinant DNA Technical Bulletin, NIH Publication no 79–99:2, no 2:43–48
Messing J, Carlson J, Hagen G, Rubenstein I, Oleson A (1984) DNA 3:31–40
Nelles L, Fang B-L, Volckaert G, Vandenberghe A, De Wachter R (1984) Nucleic Acids Res 12:8749–8768
Olsen GJ, McCarroll R, Sogin ML (1983) Nucleic Acids Res 11: 8037–8949
Palmer FBStC (1973) Biochim Biophys Acta 316:296–304
Parish JH, Kirby KS (1966) Biochim Biophys Acta 129:554–562
Raynal F, Michot B, Bachellerie J-P (1984) FEBS Lett 167:263–268
Rubin GM (1973) J Biol Chem 248:3860–3875
Rubtsov PM, Musakhanov MM, Zakharyev VM, Krayev AS, Skryabin KG, Bayev AA (1980) Nucleic Acids Res 8:5779–5794
Salim M, Maden BEH (1981) Nature (London) 291:205–208
Schlief RF, Wensink PC (1981) Practical methods in molecular biology. Springer, New York Heidelberg Berlin, pp 106–108
Schnare MN, Gray MW (1981) FEBS Lett 128:298–304
Schnare MN, Gray MW (1982) Nucleic Acids Res 10:2085–2092
Schnare MN, Spencer DF, Gray MW (1983) Can J Biochem Cell Biol 61:38–45
Schnare MN, Heinonen TYK, Young PG, Gray MW (1985) Curr Genet 9:389–393
Simoncsits A, Brownlee GG, Brown RS, Rubin JR, Guilley H (1977) Nature (London) 269:833–836
Southern EM (1975) J Mol Biol 98:503–517
Spangler EA, Blackburn EH (1985) J Biol Chem 260:6334–6340
Spencer DF, Schnare MN, Gray MW (1984) Proc Natl Acad Sci USA 81:493–497
Stiegler P, Carbon P, Ebel J-P, Ehresmann C (1981) Eur J Biochem 120:487–495
Takaiwa F, Oono K, Sugiura M (1984) Nucleic Acids Res 12: 5441–5448
Tarr GE, Fitch WM (1976) Biochem J 159:193–199
Torczynski R, Bollon AP, Fuke M (1983) Nucleic Acids Res 11: 4879–4890
Wallace FG (1979) In: Lumsden WHR, Evans DA (eds) Biology of the kinetoplastida, vol. 2. Academic Press, New York London, pp 213–240
Weiss B, Jacquemin-Sablon A, Live TR, Fareed GC, Richardson CC (1968) J Biol Chem 243:4543–4555
Woese CR, Gutell R, Gupta R, Noller HF (1983) Microbiol Rev 47:621–669

Communicated by C. W. Birky

Received August 12/September 5, 1985