Volume 12 Number 10 1984 Nucleic Acids Research

Sequence analysis of the transcribed and 5' non-transcribed regions of the ribosomal RNA gene in *Dictyostelium discoideum*

Toshinori Ozaki, Yoshiko Hoshikawa, Yoichi Iida* and Masaki Iwabuchi+

Department of Botany and *Department of Chemistry II, Faculty of Science, Hokkaido University, Sapporo, Hokkaido 060, Japan

Received 9 March 1984; Revised and Accepted 19 April 1984

ABSTRACT

The nucleotide sequence of Dictyostelium discoideum rDNA extending over almost the entire transcribed region and a part of the 5' non-transcribed spacer region has been determined. Computer analysis revealed that there were several conserved sequences in the 17S, 5.8S and 26S coding regions when compared with the sequences at analogous positions in some eukaryotic rRNA genes. The data also showed that the D. discoideum rDNA contains several extra sequences, which have not been found in other eukaryotes' rDNAs, near the 3' terminus of the 17S coding region and the 5' terminus of the 26S coding region.

INTRODUCTION

Cytoplasmic ribosomes of eukaryotes contain 25-28S (large subunit), 17-18S (small subunit), 5.8S and 5S rRNAs. Over the past few years, data on the primary structures of several eukaryotic and prokaryotic rRNA genes have accumulated (1-17). Comparative analysis of the nucleotide sequences of these genes has suggested that there are some eukaryote-specific and evolutionally conserved sequences in the transcribed regions of the eukaryotic rDNAs, and that these conserved sequences may be important for the ribosome structure in connection with its function (9).

More recently, Olsen et al. (17) reported the secondary structure of D. discoideum 17S rRNA inferred from the nucleotide sequence of the cloned 17S rRNA gene. In the present study, we have determined almost all the nucleotide sequence of the transcribed and 5' non-transcribed regions of D. discoideum rDNA. Our data on the 17S coding sequence were somewhat different from those of Olsen et al. (16), suggesting that the D. discoideum rRNA genes of about 180 copies per haploid genome are not

homogeneous. The evolutionally conserved sequences were found by comparison of the D. discoideum and other eukaryotic rDNAs using a computer program.

MATERIALS AND METHODS

DNA preparation and DNA sequencing

Recombinant plasmid pDd 507 was provided by R. A. Firtel (18), which contains the 5' non-transcribed spacer and entire transcribed regions except the 3' terminal sequence in the 26S coding region.

Preparation of the recombinant plasmid and the DNA fragments from the restriction endonuclease-digested plasmid were the same as described previously (1). For DNA sequencing, the fragments were 5'-end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase after alkaline phosphatase treatment. The cleavage sites for restriction endonucleases on the rDNA fragment were determined as described by Smith & Birnstiel (19). DNA sequencing was carried out by the method of Maxam and Gilbert (20) with slight modifications (21) according to the strategy shown in Fig. 1.

Preparation of cytoplasmic rRNA

Cytoplasmic total rRNA and 17S rRNA were extracted from partially purified ribosomes prepared from D. discoideum A3 cells and purified by two cycles of sucrose gradient centrifugation.

S1 nuclease mapping

S1 nuclease protection mapping was carried out as described by Berk and Sharp (22). The ^{32}P-5'- or 3'- end labeled rDNA fragments were hybridized with cytoplasmic rRNA or 17S rRNA, followed by S1 nuclease treatment. The DNA-RNA hybrids protected from S1 nuclease digestion were extracted, denatured and electrophoresed on sequencing gels. 3'-end labeling of the DNA fragments was performed using [α-^{32}P]ddATP and terminal transferase.

Enzymes and radioisotopes

Restriction endonucleases were purchased from Takara Shuzo, Bethesda Research Laboratories Inc. and Boehringer Mannheim; bacterial alkaline phosphatase and T4 polynucleotide kinase from Takara Shuzo;and S1 nuclease from Boehringer Mannheim. [γ-^{32}P]ATP, [α-^{32}P]ddATP and the 3'-end labeling kit were obtained from Amersham.

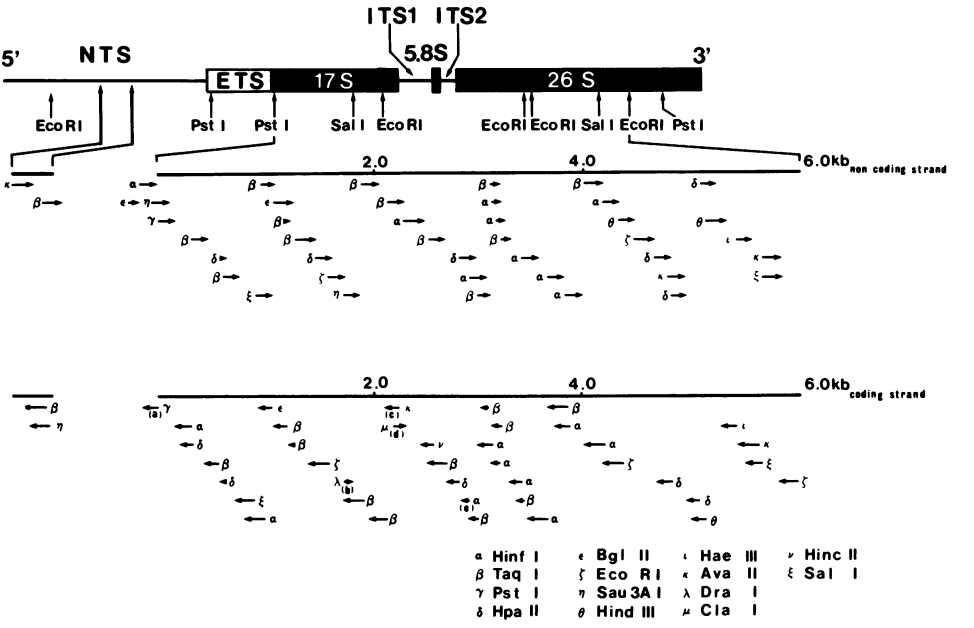
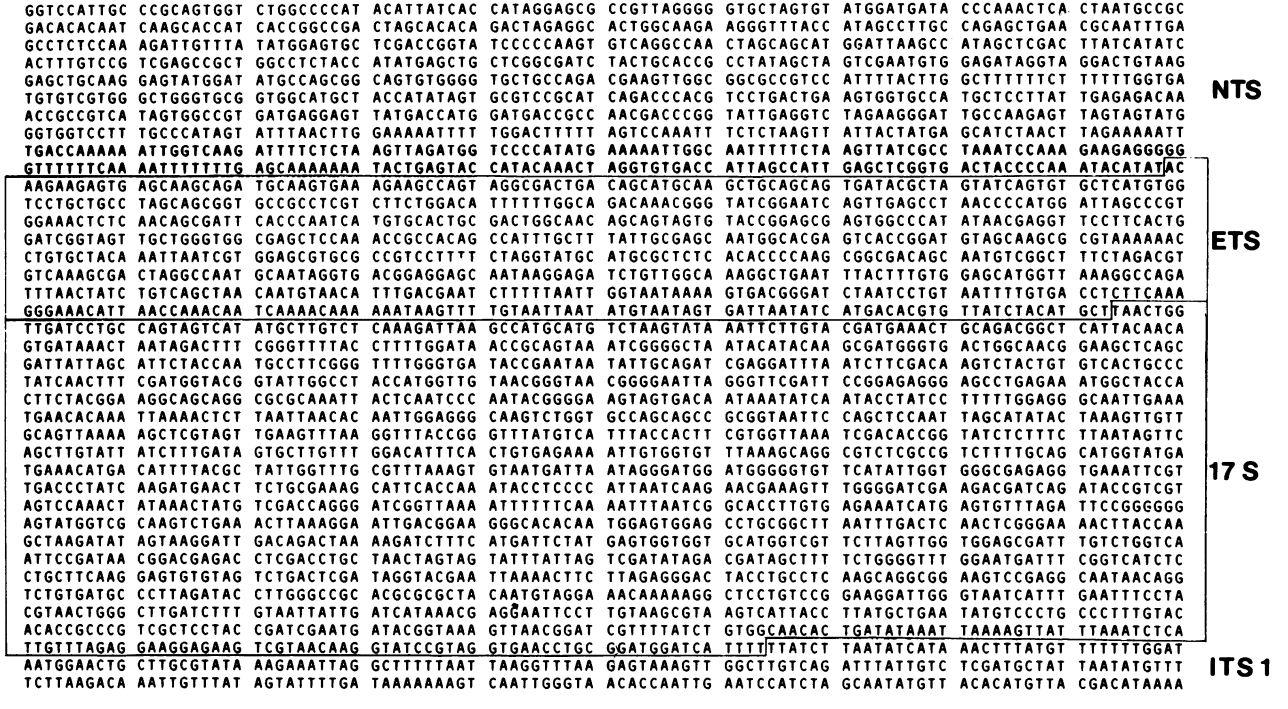


Figure 1. Schematic representation of the structure of D. discoideum rDNA cloned in recombinant plasmid pDd 507 and the sequencing strategy. NTS, ETS and ITS denote the non-transcribed, external transcribed and internal transcribed spacer regions, respectively. Arrows indicate the direction of sequencing and the size of the sequenced DNA fragment. Greek letters denote the restriction site at which the fragment was ^{32}P-5'- or 3'- end labeled. The fragments denoted as (a)-(e) were also used as probes in the S1 mapping analysis to determine the nucleotides at the 5' and 3' termini of the coding region: (a) and (b) were the probes to map the 5' and 3' ends of the 17S rRNA gene, respectively; (c) and (d), those to map the 5' and 3' ends of the 5.8S rRNA gene, respectively; (e), that to map the 5' end of the 26S rRNA gene.

RESULTS AND DISCUSSION

The primary structure of D. discoideum rDNA.

We previously reported the nucleotide sequence around the transcriptional initiation site of D. discoideum rDNA (1). Here, we determined the nucleotide sequence of the remaining part, the transcribed and 5' non-transcribed spacer regions, of the same cloned rDNA (pDd 507). DNA sequencing was performed according to the strategy shown in Fig. 1. Fig. 2 shows the compiled data on the nucleotide sequences determined here (ca. 6.6 kb) and


NTS
ETS
17 S
ITS 1

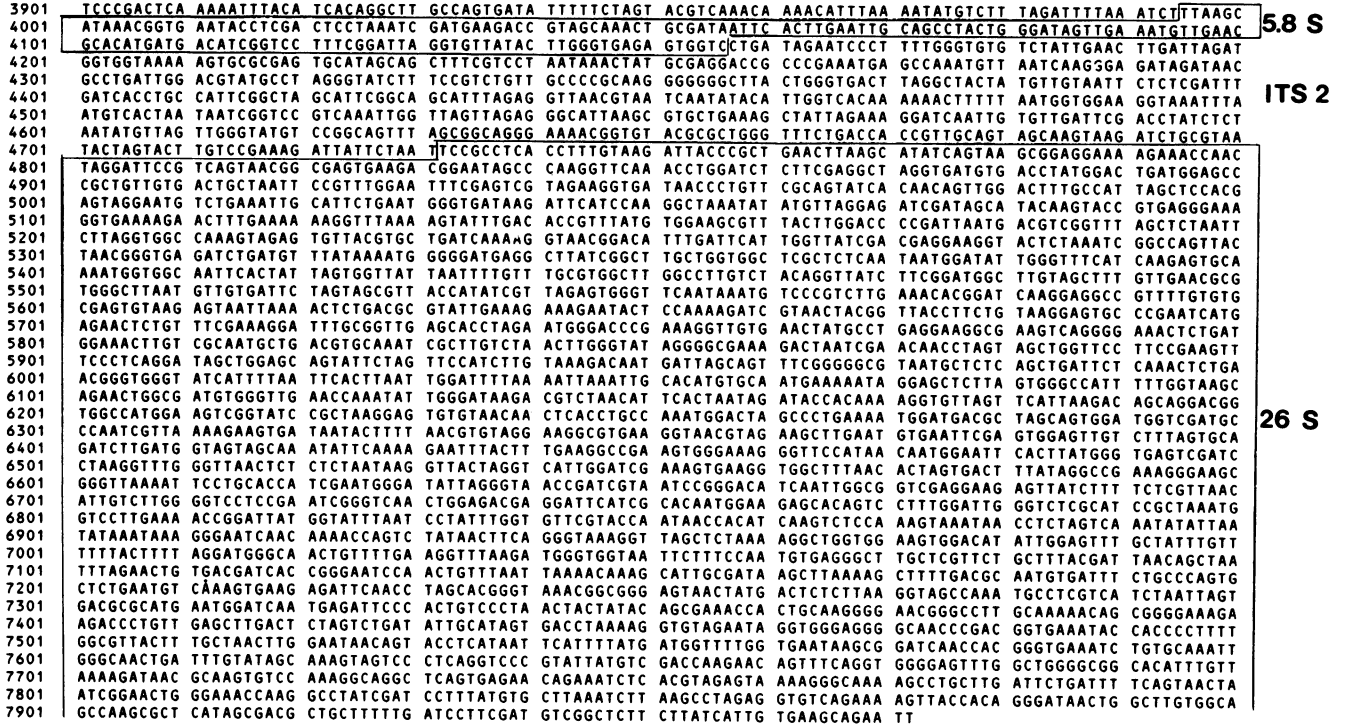


Figure 2. The nucleotide sequence of D. discoideum rDNA cloned in recombinant plasmid pDd 507. The sequencing data are compiled together with our previously reported ones (1). The ETS, 17S, 5.8S and 26S (partial) regions are boxed.

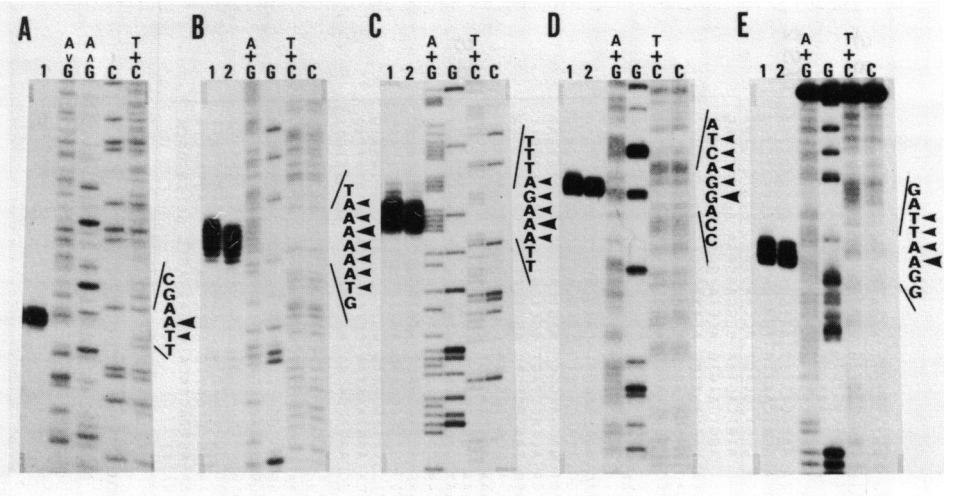


Figure 3. S1 nuclease mapping of the 5' and 3' ends of D. discoideum 17S, 5.8S and 26S rRNAs on the cloned rDNA. The ^{32}P-5'- or 3'-end labeled coding strand of the restricted DNA fragments, (a)-(e) in Fig. 1, was hybridized with whole cytoplasmic rRNA or partially purified 17S rRNA in 0.1 ml of a solution of 80% formamide and 2 x SSC at 50° C for 36 hr. The reaction mixture was diluted 1 : 10 with cold 50 mM sodium acetate (pH 4.5) containing 250 mM NaCl and 0.1 mM ZnSO and then digested with 3000 units (lane 1) or 4500 units (lane 2) of S1 nuclease at 37° C for 30 min. The S1-treated DNA-RNA hybrids were denatured and electrophoresed on 8 or 10% sequencing gels in parallel with the same 5'- or 3'-end labeled coding strand cleaved by nucleoside-specific chemical reactions. The large arrow head indicates the presumed 5' or 3' terminal nucleotide on the coding strand of the rRNA gene (for details, see RESULTS AND DISCUSSION). A and B, the 5' and 3' ends of the 17S rRNA gene, respectively; C and D, the 5' and 3' ends of the 5.8S rRNA gene, respectively; E, the 5' end of the 26S rRNA gene.

previously (ca. 1.3 kb). The nucleotides at the 5' and 3' termini of the 17S and 5.8S coding sequences and that at the 5' terminus of the 26S coding one were determined in the S1 mapping experiment. As the autoradiographic patterns in Fig. 3 show, multiple S1 protection bands were observed. The appearance of these multiple bands, which were probably due to nibbling and under-digestion of the DNA-RNA hybrids in the S1 nuclease treatment, made the results of end determination of the rRNA genes ambiguous. Therefore, referring to the data reported for the nucleotide sequences at the 5' and 3' terminal positions of the 17S, 5.8S and 26S rRNAs (15, 16, 23), we determined

Table 1. Comparison of the coding and internal transcribed spacer regions of eukaryotic rDNA as to nucleotide length and A-T content (represented as % in parentheses).

	17-18S	ITS-1	5.8S	ITS-2	26-28S
D.discoideum	1871(57)	331(74)	162(57)	575(57)	3241(57)**
S.cerevisiae	1789(55)	355(64)*	158(54)	234(62)*	3392(52)
P.polycephalum			155(45)	492(50)	3788(46)
X.laevis	1825(46)	557(16)	162(40)	262(12)	4110(34)
Rat	1869(44)	1067(25)	156(42)	765(20)	4718(33)

Data are cited from ; S. cerevisiae (4-6), X. laevis (7-9), P. polycephalum (13,14) and rat (10-12).
*, Data on S. carlsbergensis (25,26).
**, The about 0.7kb sequence 5' to the 3' terminus is not included because it was not sequenced.

the putative 5' and 3' nucleotides in the coding region of rDNA and their positions are indicated by the large arrow-heads in Fig.3. When the results obtained here were put in order together with our previously reported ones, it became possible to assign the NTS (non-transcribed spacer), ETS (external transcribed spacer), 17S, ITS (internal transcribed spacer)-1, 5.8S, ITS-2 and 26S regions in the rDNA sequence of about 7.9 kb long.

The lengths and A-T contents of the coding, ITS-1 and ITS-2 regions of D. discoideum rDNA were compared with those of the corresponding regions of several eukaryotic rDNAs. As can be seen from the data summarized in Table 1, in D. discoideum the 17S and 5.8S coding regions tended to be somewhat larger compared to those of other eukaryotes, and the A-T content was similar to that of Saccharomyces cerevisiae but quite different from those of Xenopus laevis, Physarum polycephalum and rat. In general, there was the tendency that the A-T content of rDNA of the lower eukaryotes such as yeast, Dictyostelium and Physarum was higher than that of higher eukaryotes, and the ITS-1 region was extremely A-T rich.

More recently, McCarroll et al. (16) who had already reported the 5.8S rRNA sequence of D. discoideum (15), determined the DNA sequence of the D. discoideum 17S rRNA gene. Our sequencing data on the 17S and 5.8S rRNA genes are almost the same as those of Olsen et al. (15) and McCarroll et al. (16) except for some minor differences (Table 2). Since it is known

Table 2. Differences of the nucleotide sequences of the 17S and 5.8S coding regions between two different D. discoideum rDNA clones.

17S rDNA			5.8S rDNA		
Position*	Our data	Other data**	Position*	Our data	Other data***
277	A	-	31	T	C
543	G	N (G)	36	C	T
564	A	R (A)			
587	T	-			
786-787	-	T			
787-788	-	A			
952	C	T			
1258	G	A			
1573-1574	-	G			

*, Nucleotide positions numbered according to our sequencing data.
**, Data of McCarroll et al. (16).
***, Data inferred from the nucleotide sequence of 5.8S rRNA determined by Olsen et al. (15).

that in D. discoideum there are rRNA genes of about 180 copies per haploid genome, these differences may reflect the heterogeneity of the rRNA genes, if no point mutation had occurred on keeping the two recombinant plasmids containing the rDNA insert, and if there had been no misreading of the sequence ladder on the autoradiogram.

Comparison of the primary structure of the small subunit rRNA genes.

Computer analysis was performed to compare the nucleotide sequence of the D. discoideum 17S rRNA gene with those of the S. cerevisiae (4), X. laevis (7) and rat (10) 18S rRNA genes. The nucleotide sequences of the 17-18S rRNA genes in D. discoideum and the above species were aligned with insertion of some deletions at appropriate positions according to the computer program designed to give the maximum homology by Iida (unpublished). The sequence homology was calculated every 50 nucleotides from the 5' end using the formula proposed by Iida (unpublished) and expressed as the percentage of the conserved nucleotides.

The overall sequence homology of the 17-18S rRNA genes between D. discoideum and S. cerevisiae, X. laevis and rat was 70, 67 and 66%, respectively. However, in both the 5' and 3'

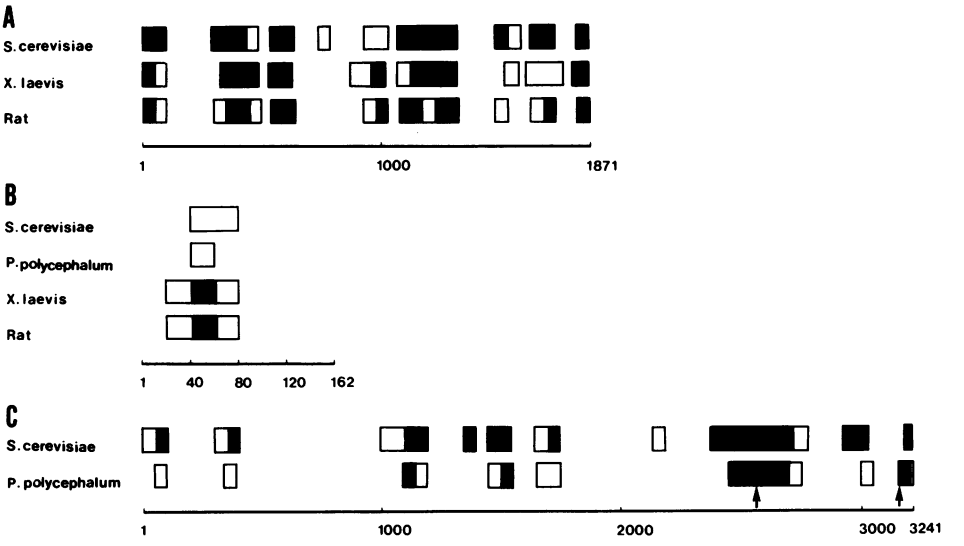


Figure 4. Schematic representation for comparison of the sequence homology of the rRNA genes between D. discoideum and other species. The open and filled boxes represent regions showing 70-80 and 80-100% homology, respectively, when compared with the corresponding region of the D. discoideum rRNA gene. A, small subunit rRNA gene; B, 5.8S rRNA gene; C, large subunit rRNA gene. The vertical arrow indicates the insertion site of the introns of the P. polycephalum 26S rRNA gene. The 3' regions of the large subunit rRNA genes are not compared because the about 0.7 kb sequence 5' to the 3' terminus of the D. discoideum 26S rRNA gene could not be sequenced.

terminal regions the sequence homology was more than 85%, suggesting that the sequence of the terminal region of the small subunit rRNA gene may be highly conserved in eukaryotes. Several sequences showing relatively high sequence homology (70% or more) were dispersedly located at comparable regions throughout the lengths of the 17-18S rRNA genes in these four eukaryotes. The results are schematically summarized in Fig. 4A. Therefore, it is conceivable that these conserved regions play certain important roles in connection with the structure and function of ribosomes, although we have no direct evidence of this.

On comparison of the nucleotide sequences of the 17-18S coding regions we observed several extra sequences in the non-conserved regions, which were more than 10 nucleotides in length and species-specific. These sequences are listed in Table 3. In

Table 3. The extra sequences of the 17S coding region of D. discoideum rDNA.

Comparison with S.cerevisiae		Comparison with X.laevis		Comparison with Rat	
Position*	Extra sequence	Position*	Extra sequence	Position*	Extra sequence
982-991	GATCGAAGAC	1362-1375	TATAGACGATAG CT	180-181	(CCCCCTTCCCGT GG)**
1349-1368	ATTTATTAGTCG ATATAGAC	1383-1397	GGTTTGGAATGA	211-212	(AAACCAACCCGG)**
1385-1398	TTGGAATGATTT C	1403-1416	ATCTCCTGCTTC AA	281-282	(CGCCCTCCGTG)**
1406-1451	TCCTGCTTCAAG GAGTGTGTAGTC TGACTCGATAGG TACGAATTAA	1421-1448	TGTGTAGTCTGA CTCGATAGGTAC GAAT	1383-1396	GGTTTGGAATGA TT
				1409-1450	TGCTTCAAGGAG TGTGTAGTCTGA CTCGATAGGTAC GAATTA

*, Nucleotide positions numbered according to our sequencing data.
**, The extra nucleotide sequences in rat 18S rDNA (10) which is not seen in the corresponding region of D. discoideum 17S rDNA.

D. discoideum, such extra sequences were localized near the 3' terminal in the 17S coding region. Thus, it is suggested that the existence of these extra sequences may be due to the minor differences in nucleotide length of the 17-18S coding regions among the four species as mentioned above.

Comparison of the primary structure in the 5.8S coding region.

We have compared the nucleotide sequences of the 5.8S rRNA genes in five species in the same way as in the case of the 17-18S rRNA genes, except that the sequence homology was computed every 20 nucleotides from the 5' end. The overall sequence homology of the D. discoideum 5.8S rRNA gene was 60% for S. cerevisiae (6), 59% for X. laevis (8), 51% for P. polycephalum (14) and 61% for rat (11). As indicated in Fig. 4B, however, the sequences showing sequence homology of more than 70% were distributed near the 5' terminal position in the 5.8S coding regions of these 5 species. The comparative analysis also revealed that in D. discoideum the sequence of the 5.8S rRNA gene was less conserved relative to those of the 17S and 26S rRNA genes. It has been pointed out (14, 24) that in some eukaryotes, the entire nucleotide sequence of the 5.8S rRNA gene is very similar to the 5' terminal sequence of the Escherichia coli 23S rRNA gene (3). This was the case for the D. discoideum

5.8S rRNA gene as well. Thus, it can be considered that the eukaryotic 5.8S rRNA gene might be derived from the 5' terminal region of the E. coli 23S rRNA gene, as suggested by some workers (14, 24), and its function is probably analogous to that of the bacterial 5' terminal sequence.

Comparison of the primary structure of the large subunit rRNA genes.

As mentioned above, we could not determine the about 0.7 kb sequence 5' to the 3' end in the 26S coding region since the 3' terminal sequence was not included in the recombinant plasmid used here. So, we tried to compare the nucleotide sequence of about 3.2 kb at the 5' side in the 26S coding region with that in the corresponding regions of S. cerevisiae (5) and P. polycephalum (13), using the same computer program as employed in the analysis of the small subunit rRNA gene. The overall sequence homology in the corresponding region of D. discoideum for S. cerevisiae and P. polycephalum was 65 and 58%, respectively. As Fig. 4C shows, the sequences showing sequence homology of more than 70% were located at comparable portions throughout the lengths of the 26S rRNA genes in the three species. These conserved sequences may be situated at structurally and functionally important regions of ribosomes. In addition, we found several extra sequences consisting of 10 nucleotides or more in D. discoideum which were not seen in the other two species. As can be seen from these sequences listed in Table 4, they tended to be distributed at the 5' side in the coding region. As can be seen on comparison of D. discoideum and P. polycephalum, there were three extra sequences at the 3' side in P. polycephalum which were not present in D. discoideum. Furthermore, the comparative analysis revealed that two introns present in the P. polycephalum 26S rRNA gene (13) are located in highly conserved regions existing commonly in the two species (see Fig.4C). Our unpublished data on the secondary structure of D. discoideum 26S rRNA inferred from the DNA sequence suggested that the sequence at the 5' terminal region of the 26S rRNA can interact with the 5.8S rRNA to construct a stable secondary structure, as has already been pointed out in other eukaryotes (13, 26).

Table 4. The extra sequences in a part of the 26S coding region of D. discoideum rDNA.

Comparison with S.cerevisiae		Comparison with P.polycephalum	
Position*	Extra sequence	Position*	Extra sequence
456-468	GTTTAGCTCTAAT	402-423	TATTTGACACCGTTTATGTG GA
531-553	GGTTATCGACGAGGAAGGTA CTC	485-495	TAGAGTGTTAC
589-598	TTTATAAAAT	619-630	TTGCTGGTGGCT
687-709	TATTAGTGGTTATTAATTTT GTT	654-668	TTTCATCAAGATGC
725-743	TGTCTACAGGTTATCTTCG	1304-1318	TTTTAAAATTAAATT
910-924	AAAGAATACTCCAAA	1586-1587	(CGGGCTTCGGCTCGCA)**
1302-1313	GATTTTAAAATT	2805-2806	(CCGTAAAAGGTGGGGGAAGG GATAGG)**
1844-1856	GTGACTTTATAGG	2881-2882	(CCGGCGAGTGC)**
2854-2863	GAAATCTGTG		
2881-2907	TTGTATAGCAAAGTAGTCCC TCAGGTC		

*, Nucleotide positions numbered according to our sequencing data.
**, The extra nucleotide sequences in P. polycephalum 26S rDNA (13) which are not present in the corresponding regions of D. discoideum 26S rDNA.

It is known that the eukaryotic nuclear rRNA gene is relatively well conserved among organisms and also that it has some nucleotide sequences partially the same as those of the bacterial (e.g., E. coli) rRNA gene. The results in the present study suggest that the coding regions of eukaryotic rRNA genes are roughly divided into two kinds of portions with highly conserved and relatively less conserved sequences, and the former sequences particularly tend to be located at comparable regions throughout the length of rDNA. In order to discuss the biological significance of the highly conserved sequences, it is necessary to determine the secondary structure of rRNA on the basis of the sequencing data on rDNA. Since the recent results of Olsen et al. (17) have suggested that the highly conserved nucleotide sequences present in D. discoideum 17S rRNA tend to occupy important positions in the secondary structure of the rRNA, the same thing can be considered for the highly conserved sequences in the 26S coding region.

Transcribed spacer regions

It has been supposed that in the 5' and 3' transcribed

spacer regions adjacent to the rRNA coding regions there may be certain signal sequences involved in the processing of rRNA precursor. So, we searched such sequences but could not find any particular sequences in the ETS, ITS-1 and ITS-2 regions. In addition, in rough comparison of the sequences of these three regions of D. discoideum rDNA with those of the corresponding regions of other eukaryotes' ones, there was little sequence homology. The length of the sequence of the ITS-1 and ITS-2 regions varied from species to species (see Table 1). This was also the case for the ETS region. Although we did not further compare the sequence of the transcribed spacer regions, short meaningful conserved sequences may be found if a lot of the data has been accumlated and analyzed in more detail. At the moment, it seems certain that the transcribed spacer regions are variable compared with the coding regions.

ACKNOWLEDGMENTS

We thank Dr. R. A. Firtel for providing the recombinant plasmid pDd 507. This study was supported by Grants-in-Aid for Special Project Research from the Ministry of Education, Science and Culture of Japan (Project No. 58119002), and a Naito Foundation Research Grant for 1983.

+To whom correspondence should be addressed

REFERENCES

1. Hoshikawa,Y., Iida,Y. and Iwabuchi,M. (1983) Nucleic Acids Res. 11, 1725-1734.
2. Brosius,J., Palmer,M.L., Kennedy,P.J. and Noller,H.F. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 4801-4805.
3. Brosius,J., Dull,T.J. and Noller,H.F. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 201-204.
4. Rubtsov,P.M., Musakhanov,M.M., Zakharyev,V.M., Krayev,A.S., Skryabin,K.G. and Bayev,A.A. (1980) Nucleic Acids Res. 8, 5779-5794.
5. Georgiev,O.I., Nikolaev,N. and Hadjiolov,A.A. (1981) Nucleic Acids Res. 9, 6953-6958.
6. Rubin,G.M. (1973) J.Biol.Chem. 248, 3860-3875.
7. Salim,M. and Maden,B.E.H. (1981) Nature 291, 205-208.
8. Hall,L.M.C. and Maden,B.E.H. (1980) Nucleic Acids Res. 8, 5993-6005.
9. Ware,V.C., Tague,B.W., Clark,C.G., Gourse,R.L., Brand,R.C. and Gerbi,S.A. (1983) Nucleic Acids Res. 11, 7795-7817.

10. Torczynski,R., Bollon,A.P. and Fuke,M. (1983) Nucleic Acids Res. 11, 4879-4890.
11. Subrahmanyam,C.S., Cassidy,B., Busch,H. and Rothblum,L.I. (1982) Nucleic Acids Res. 10, 3667-3680.
12. Chan,Y.L., Olvera,J. and Wool,I.G. (1983) Nucleic Acids Res. 11, 7819-7831.
13. Otsuka,T., Nomiyama,H., Yoshida,H., Kukita,T., Kuhara,S. and Sakaki,Y. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3163-3167.
14. Otsuka,T., Nomiyama,H., Sakaki,Y. and Takagi,Y. (1982) Nucleic Acids Res. 10, 2379-2385.
15. Olsen,G.J. and Sogin,M.L. (1982) Biochemistry 21, 2335-2343.
16. McCarroll,R., Olsen,G.J., Stahl,Y.D., Woese,C.R. and Sogin,M.L. (1983) Biochemistry 22, 5858-5868.
17. Olsen,G.J., McCarroll,R. and Sogin,M.L. (1983) Nucleic Acids Res. 11, 8037-8049.
18. Cockburn,A.F., Newkirk,M.J. and Firtel,R.A. (1976) Cell 9, 605-613.
19. Smith,H.O. and Birnstiel,M.L. (1976) Nucleic Acids Res. 3, 2387-2399.
20. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
21. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory. Box 100. Cold Spring Harbor, New York.
22. Berk,A.J. and Sharp,P.A. (1977) Cell 12, 721-732.
23. Sakuma,K., Komiyama,R. and Muramatsu,M. (1976) Eur. J. Biochem. 63, 339-350.
24. Jacq,B. (1981) Nucleic Acids Res. 9, 2913-2932.
25. Veldman,G.M., Brand,R.C., Klootwijk,J. and Planta,R.J. (1980) Nucleic Acids Res. 8, 2907-2920.
26. Veldman,G.M., Klootwijk,J., Heerikhuizen,H.V. and Planta,R. J. (1981) Nucleic Acids Res. 9, 4847-4862.