

Nucleotide sequence analysis of the rRNA transcription unit of a pathogenic *Entamoeba histolytica* strain HM-1:IMSS

S.Ramachandran, A.Bhattacharya¹ and S.Bhattacharya*

School of Environmental Sciences and ¹School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Received February 22, 1993; Accepted February 26, 1993

EMBL accession nos X65163 and X64142

The protozoan parasite, *E.histolytica* causes amoebiasis, a widespread disease in developing countries. The rRNA genes of this parasite are present on an extrachromosomal, circular plasmid which is 25 kb in size and present in about 200 copies per cell (1). Each plasmid molecule contains two rRNA transcription units arranged as inverted repeats (2). We have determined the nucleotide sequence of one of the transcription units which was almost entirely contained in a 6.3 kb HindIII fragment. Overlapping deletions of this fragment, cloned in pTZ18U, were generated with exonuclease III (3) and double strand sequencing was done by the dideoxy chain termination method using appropriate primers (4). Northern blots of total *E.histolytica* RNA were probed with EcoRI subfragments of the 6.3 kb fragment to confirm the location and size of ssu- and lsu-rRNA genes. The Northern blot data together with sequence alignment with known rRNA sequences from a large number of diverse eukaryotes using the computer programs BESTFIT and LINEUP of UWGCG ver 7.1 (5) led to the deduction of 5'- and 3'-ends of genes for ssu-, 5.8S-, and lsu rRNAs. The organisation of the transcription unit is shown in Figure 1. The inferred length of ssu rRNA gene was 1946 nt. This agreed well with the data from Northern blot, where a 2 kb band hybridized with the coding region for ssu rRNA. The 28S rRNA equivalent appeared to be fragmented into two equal fragments as shown previously (6). The 3'-end of this gene was just outside the 6.3 kb HindIII fragment and is therefore not included in the sequence reported here. In general, the coding regions showed a higher percentage of G+C than spacer regions which had a G+C content comparable with the reported values for *E.histolytica* genome (22–27%; ref 7). Compared with lengths of ITS1 and ITS2 in other eukaryotes, the *E.histolytica* spacers are very short, although the length of its rRNAs conforms with the general values for most eukaryotes. Amongst eukaryotes, including protozoa, separated by large evolutionary distances, the length of the two spacers is not the same but they are nearly identical in G+C content (8). In *E.histolytica*, the reverse situation is encountered. The two spacers are almost identical in length, which is a rare feature found only in a few plant species (9). On the other hand they vary significantly in G+C content (13% and 29% G+C) which amongst eukaryotes is found only in *Dictyostelium discoideum*, a species phylogenetically very close to *E.histolytica* (10).

The secondary structure model of ssu rRNA of *E. histolytica*

inferred from the sequence data was in large part very similar to the structures proposed for other eukaryotic ssu rRNAs. The lengths of helix-loop conformations in the conserved regions were in close agreement with most other eukaryotes. Detailed phylogenetic analysis using the ssu rRNA sequence will be published elsewhere (manuscript in preparation).

ACKNOWLEDGEMENTS

S.R. thanks the UGC for a fellowship. This work was supported in part by funds from the Department of Science and Technology, Govt. of India.

REFERENCES

- Bhattacharya, S., Bhattacharya, A., Diamond, L.S. and Soldo, A.T. (1989) *J. Protozool.* **36**, 365–368.
- Huber, M., Koller, B., Gitler, C., Mirelman, D., Revel, M., Rozenblatt, S. and Garfinkel, L. (1989) *Mol. Biochem. Parasitol.* **32**, 285–296.
- Henikoff, S. (1984) *Gene* **28**, 351–359.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Deveureux, J., Haeblerli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Albach, R.A. (1989) *J. Protozool.* **36**, 197–205.
- Gelderman, A.H., Bartgis, I.L., Keister, D.B. and Diamond, L.S. (1971) *J. Parasitol.* **57**, 912–916.
- Schnare, M.N., Cook, J.R. and Gray, M.W. (1990) *J. Mol. Biol.* **215**, 85–91.
- Torres, R.A., Ganai, M. and Hemleben, V. (1990) *J. Mol. Evol.* **30**, 170–181.
- Sogin, M.L. (1989) *Amer. Zool.* **29**, 487–499.

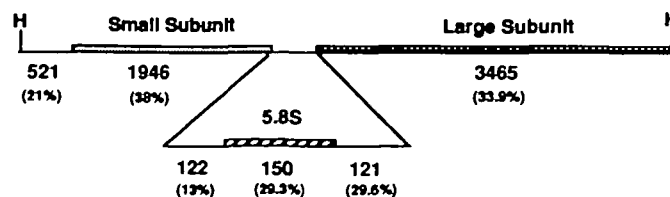


Figure 1. Organisation of the rRNA transcription unit of *E.histolytica*. The number of nucleotides in each segment of the transcription unit is given below the line. Numbers in parentheses give percentage G+C content of each segment. H, HindIII.

* To whom correspondence should be addressed