Nucleic Acids Research, 1993, Vol. 21, No. 5 1067–1071

Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*

M.J.Gardner+, J.E.Feagin[1], D.J.Moore, K.Rangachari, D.H.Williamson and R.J.M.Wilson*
National Institute for Medical Research, Mill Hill, London NW7 2AB, UK and [1]Seattle Biomedical Research Institute, Nickerson Street, Seattle, WA 98109-1651, USA

Received December 24, 1992; Revised and Accepted January 28, 1993 EMBL accession no. X61660

ABSTRACT

The malaria parasite *Plasmodium falciparum* carries an extrachromosomal 35 kb circular DNA molecule of unknown provenance. A striking feature of the circle is a palindromic sequence of genes for subunit rRNAs and several tRNAs, spanning *ca.* 10.5 kb. The palindrome has an intriguing resemblance to the inverted repeat of plastid genomes, and the sequence and putative secondary structure of the malarial large subunit (LSU) rRNA described in this report were used as the basis of a phylogenetic study. The malarial rRNA was found to be highly divergent in comparison with a selected group of chloroplast LSU rRNAs but was more closely related to them than to mitochondrial LSU rRNA genes.

INTRODUCTION

Malaria parasites have two extrachromosomal DNA molecules (reviewed in (1)). The first comprises a 6 kb reiterated sequence (2) that specifies three mitochondrial proteins— cytochrome b and subunits I and III of cytochrome oxidase (3). It also encodes fragments of both small and large subunit (SSU and LSU) rRNA genes (4,5). Since, in addition, it is concentrated in a subcellular fraction enriched in mitochondria (6), it is believed to be of mitochondrial origin. The second extrachromosomal DNA is an (A+T)-rich (>80%) 35 kb circle whose organellar location and evolutionary origin remains to be established (7). The portion of the 35 kb circle that has been sequenced has two striking features, firstly, juxtaposed *rpo*B and *rpo*C genes with homology to those of *Escherichia coli* and chloroplasts (8), and secondly, a palindromic sequence comprising rRNA and tRNA genes that is reminiscent of the inverted repeat found in chloroplast genomes (9). These characteristics led us to propose that malaria parasites may have had a hitherto unsuspected photosynthetic ancestor (1), reviewed in (10). Here we report the sequence and predicted secondary structure of the LSU rRNA, as well as a phylogenetic analysis based on these data.

MATERIALS AND METHODS

DNA analysis

Isolation of the 35 kb circular DNA of *Plasmodium falciparum* and general procedures used in cloning and DNA sequencing have been described elsewhere (7,8). The IR_A end of the palindrome was obtained by conventional cloning (*Hind*III fragment 4 in Fig.2) and the IR_B end by PCR amplification from total *P.falciparum* DNA. The primers used for the latter purpose, 5′-CGTGAGACAGTTCGGTCC and 5′-CCCAAAATA-GATATGTTACC, were based on sequenced clones. The first primer sequence is located in *Hind*III fragment 4 near the 3′ end of the IR_A copy of the LSU rRNA gene and was expected to be repeated in the IR_B copy. The second, unique, primer was derived from a cloned fragment of *Hind*III fragment 2 (Fig.2), and was located about 1.5 kb downstream from the estimated 3′ end of the IR_B LSU rRNA gene. The resulting PCR-amplified product (*ca.* 2 kb) was cloned directly in the TA vector (Invitrogen) for sequencing. The sequences of the two 3′ ends of the LSU rRNA genes and adjacent downstream regions obtained in this way were determined using custom-made oligonucleotide primers. The 5′ and central portions of the LSU rRNA gene were determined by sequencing cloned copies of *Hind*III fragments 5 and 6a. The complete nucleotide sequence of the gene is deposited in the EMBL database under the accession number X61660.

RNA analyses

Procedures for RNA extraction, Northern blotting and transcript mapping have been described in detail elsewhere (7,8). Mapping of the 5′ end of the LSU transcript was carried out by primer extension using total RNA and an oligonucleotide sequence located 85 nucleotides down-stream from the presumptive 5′ end. The 3′ end was mapped by RNase protection, in which total RNA was hybridized with ^{32}P-labelled anti-sense transcripts made *in vitro* (5) from *Rsa*I or *Rsa*I/*Alu*I restriction fragments of cloned circular DNA containing the 3′ end of the LSU rRNA from IR_A. The hybrids were digested with RNase and electrophoresed on a 6% acrylamide, 7M urea sequencing gel.

* To whom correspondence should be addressed

+ Present address: Division of Experimental Therapeutics, Walter Reed Army Institute of Research, Washington, DC 20307-5100, USA

Phylogenetic analysis

The malarial LSU rRNA sequence was aligned visually with the predicted secondary structure of *E.coli* LSU rRNA (11), allowing us to identify stretches of conserved sequence (a total of 608 nt) suitable for phylogenetic analysis. The GCG sequence programme PILEUP (12) was used to align these fragments from the malarial gene with corresponding regions from the specified organisms. With this alignment the programmes DNADIST, NEIGHBOR and DRAWTREE from J.Felsenstien's PHYLIP 3.4 package (13) were used to compute distance matrices and construct an unrooted tree.

RESULTS

The palindrome

Purified 35 kb circular DNA digested with *Hind*III yielded seven bands whose stoichiometric proportions are illustrated in Fig.1. Densitometric and restriction analysis (not shown) revealed that band 5 comprised two copies of the same fragment. Likewise, band 6 comprised a doublet, resolved in some gels, each of the two constituent copies of which, labelled 6a and 6b, being present in duplicate molar amounts. Further restriction mapping and cloning has now allowed each palindromic arm to be separately identified and they have been dubbed IR_A and IR_B by analogy with those of chloroplast genomes (14). As shown schematically in Fig.2, each arm of the palindrome contains one SSU and one LSU rRNA gene and a cluster of intervening tRNA genes (indicated in Fig.2, but to be presented in detail elsewhere). This overall arrangement is consistent with the occurrence on the circles of a cruciform structure (15), and with the detection of rapidly renaturing (snap-back) fragments following denaturation of molecules cleaved at pairs of restriction sites located symmetrically across the region (16, and Gardner *et al.*, in preparation). We have shown that the sequence of the SSU rRNA is distinct from the nucleus-encoded SSU rRNA genes and that its inferred secondary structure is consistent with an organellar origin (7).

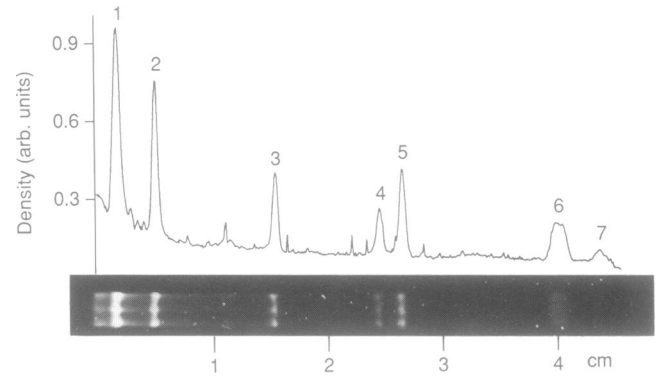

Figure 1. Densitometric analysis of a *Hind*III restriction digest of the 35 kb circular DNA from *P.falciparum*. The numbers were used in labelling the map shown in Fig.2. As discussed in the text, band 5 carries two molar copies of fragment 5. Band 6 is a doublet, resolved in some gels into two components, 6a and 6b, each of which is present in two molar copies.

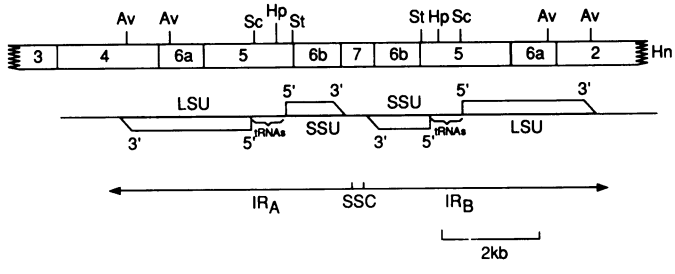

Figure 2. Schematic of the arrangement of the rRNA genes and restriction map in the palindromic segment of the malarial 35 kb circular DNA. The positions of two clusters of tRNAs are indicated. IR_A and IR_B denote the arms of the palindrome. *Av, Ava*II; *Sc, Sca*I; *Hp, Hpa*I; *St, Sty*I; *Hn, Hind*III. LSU, SSU, large and small rRNA subunits. SSC, small single copy sequence. Numbers in the bar are labels for the *Hind*III fragments (see Fig.1).

LSU rRNA sequence analysis

The predicted secondary structure of the LSU rRNA encoded by the *P.falciparum* 35 kb circular DNA is shown in Fig.3. It was derived from sequence assembled from cloned *Hind*III fragments 4, 5 and 6a (Fig.2). About 300 nt at the 3′ terminus of the IR_B copy (the limit of our cloned PCR-amplified sequence) was found to be identical in the two copies of the gene.

The format in Fig.3 is based on that of *E.coli* (11), capital letters being used to indicate bases conserved between the *P.falciparum* and *E.coli* sequences. The overall level of conservation is *ca.*40%, but in the core regions it averages 75% (5). Some portions of the molecule are extremely rich in A/U and may be idiosyncratic in their secondary structure; the putative helices indicated in these regions have not been substantiated structurally. As in bacterial and chloroplast LSU rRNAs, a possible helical structure exists between the 5′ and 3′ termini; this is indicated on both halves of Fig.3.

The complete DNA sequence of the LSU gene is deposited in the EMBL database under the accession number X61660.

Transcript analysis

Northern blots of total RNA extracted from parasitized erythrocytes probed with an oligonucleotide complementary to sequence near the 5′ end of the LSU rRNA revealed a single transcript of *ca.* 2.9 kb (data not shown). This transcript was somewhat larger than the size predicted from the LSU rRNA sequence alone (2694 nt). Its 5′ end was examined by primer extension (Fig.4A), and although it was not possible to read the sequence to the final nucleotide of the 85 nt extension product (see caret), the size of the transcript corresponded closely with that predicted from the secondary structure analysis. The 3′ end was determined from RNase protection studies using as probes ^{32}P-labelled *in vitro* transcripts complementary to the expected 3′ end of the LSU rRNA. Several protected products were detected using a probe corresponding to an 885 nt *Rsa*I fragment (Fig.4B); the largest product (244 nt) mapped to the LSU 3′ end predicted from the secondary structure. This fragment was the only one protected when the probe was shortened to include minimal sequence (*ca.*30 nt) 3′ of the predicted LSU end (Fig.4C), indicating that the fragments with estimated sizes of 110 nt, 76/78 nt and 63 nt were derived from transcripts of sequences downstream from the LSU rRNA gene, including the tRNA shown in Fig.4C.

Phylogenetic analysis

The LSU rRNA sequence was examined for phylogenetic information that might throw light on the provenance of the

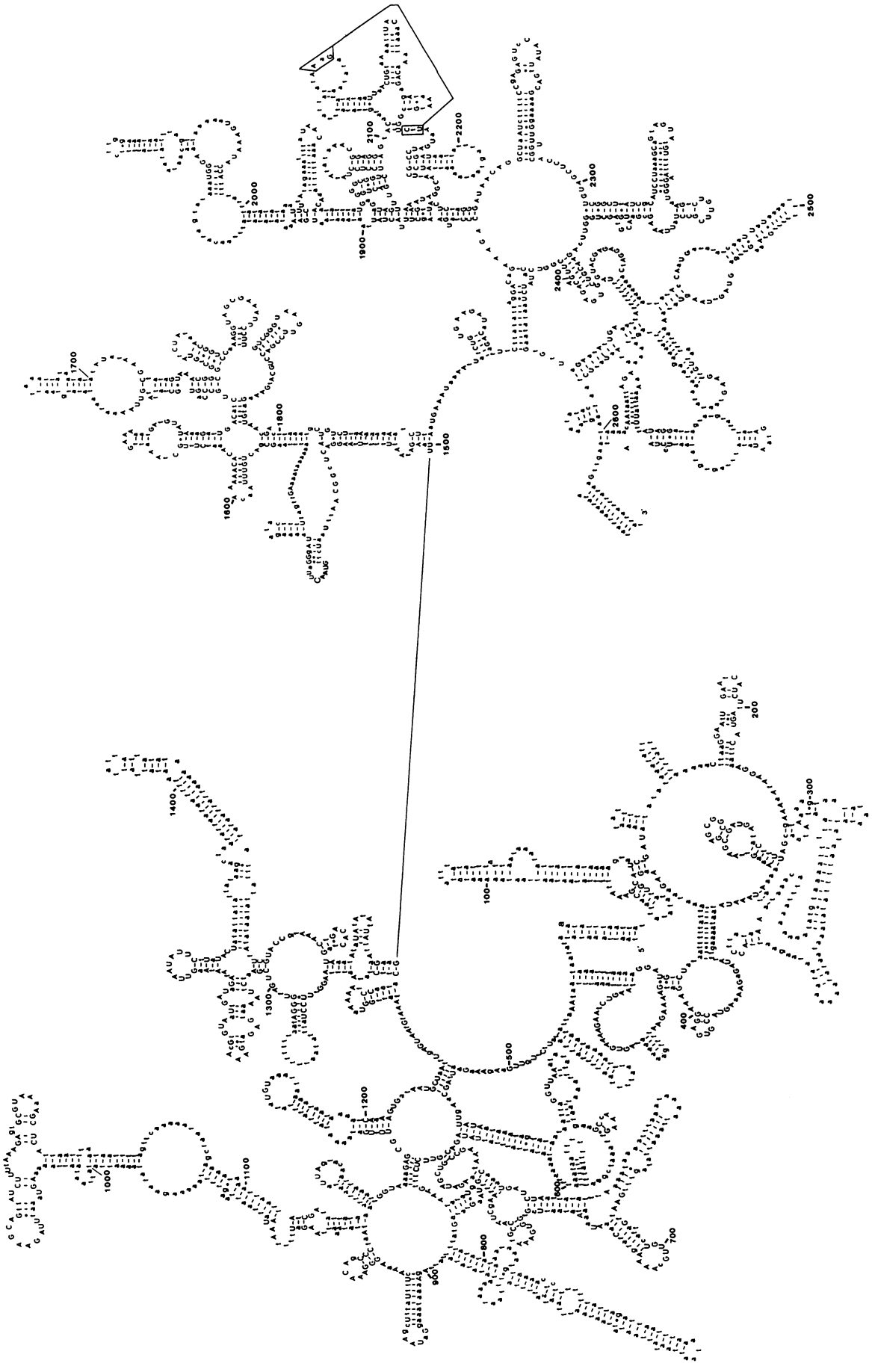

Figure 3. Predicted secondary structure of extrachromosomal, 35 kb circle, LSU rRNA of *P.falciparum* based on that of *E.coli* (11). Conserved bases are indicated in capital letters, parasite-specific positions are indicated with small letters for clarity, using t for uridine.

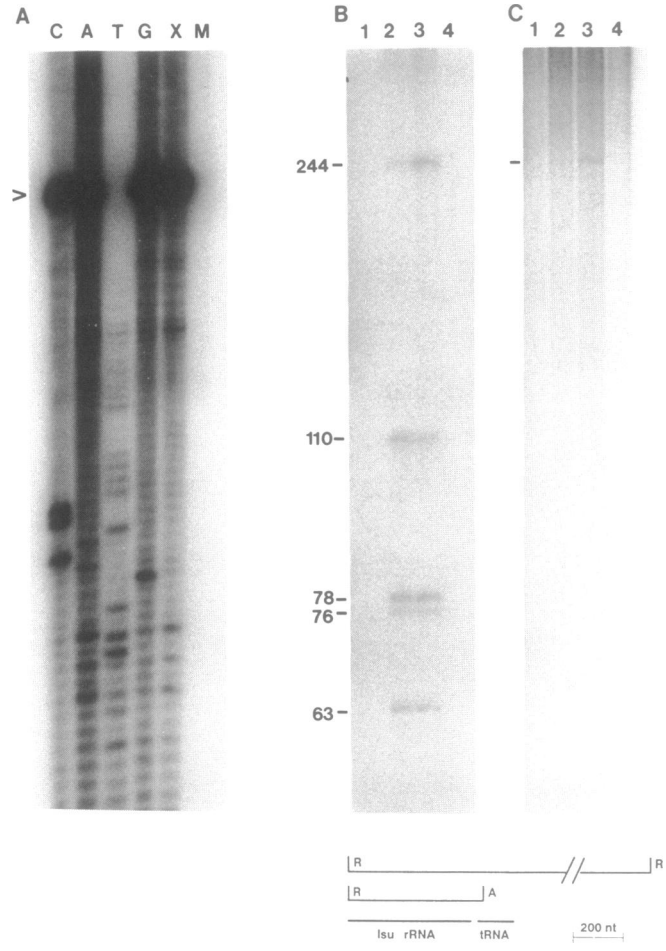

Figure 4. A. Determination of the 5′ end of the LSU rRNA. Total *P.falciparum* RNA was hybridized to an oligonucleotide complementary to sequences near the expected 5′ end of the LSU rRNA. Sequencing was by the dideoxy chain-termination method. No dideoxynucleotides were added to lane X and primer was omitted in lane M. The size of the major extension product (caret = 85 nt) was estimated from adjacent sequencing ladders. **B.** RNase protection study to determine the 3′ end of the LSU rRNA. Total *P.falciparum* RNA (5 and 10 μg in lanes 2 and 3, respectively) was hybridized with ^{32}P-labelled *in vitro* RNA transcripts complementary to the expected 3′ end of the LSU rRNA (*RsaI* fragment R/R) shown in the schematic map below panels B and C. The hybrids were digested with RNase and electrophoresed on a 6% acrylamide, 7M urea sequencing gel. Lanes 1 and 4 are controls with no RNA and 10μg of *Trypanosoma brucei* total RNA respectively, hybridized and digested as above. Approximate sizes of the protected fragments were estimated from end-labelled *HpaII* fragments of pBR322 and are indicated in nt. **C.** As in B, except that the ^{32}P-labelled *in vitro* transcript was produced from the *RsaI-AluI* fragment (R/A) shown in the schematic map.

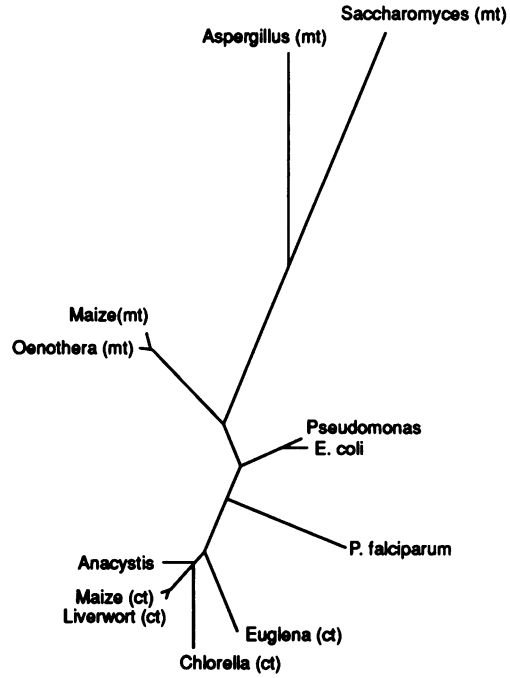

Figure 5. Phylogenetic analysis of the malarial 35 kb LSU rRNA (*Plasmodium falciparum*). An unrooted tree was computed using the PHYLIP 3.4 programmes DNA DIST, NEIGHBOR and DRAWTREE. The sequences are as follows; Eubacterial: *Pseudomonas aeruginosa* and *E.coli*. Cyanobacterial: *Anacystis nidulans*. Mitochondrial sequences of Maize (*Zea mays*), *Oenothera berteriana*, *Aspergillus nidulans* and *Saccharomyces cerevisiae* are designated 'mt'. Chloroplast sequences of Maize, Liverwort (*Marchantia polymorpha*), *Chlorella ellipsoidea* and *Euglena gracilis* are designated 'ct'.

Table 1. Sequence positions (*E.coli*) of LSU rRNA genes used for alignment (based on Gutell and Fox, reference 11)

026–034	0971–0984	1883–1891	2381–2394
188–205	1057–1073	1897–1905	2418–2435
240–259	1082–1098	1906–1918	2447–2455
447–460	1193–1202	1927–1940	2466–2478
497–514	1248–1256	1945–1955	2490–2505
560–579	1307–1316	1962–1971	2552–2582
580–588	1349–1367	1989–1995	2583–2589
664–677	1599–1613	2009–2023	2607–2617
745–753	1662–1677	2057–2065	2653–2667
773–794	1814–1826	2068–2081	2719–2728
799–812	1833–1843	2259–2275	2877–2883

circular DNA molecule. The 608 nt we used were selected from 44 segments (Table 1) from positions throughout the molecule including the most conserved portions (17). Using PILEUP, they were aligned with corresponding sequences from the eubacteria *Escherichia coli* and *Pseudomonas aeruginosa*, the cyanobacterium *Anacystis nidulans*, the chloroplast 23S ribosomal RNAs of *Zea mays*, *Marchantia polymorpha*, *Chlorella ellipsoidea* and *Euglena gracilis*, together with the corresponding mitochondrial genes from two plants *Zea mays* and *Oenothera berteriana* and two fungae, *Apergillus nidulans* and *Saccharomyces cerevisiae* (11). The PHYLIP programmes DNADIST and NEIGHBOR (18) were used to compute an unrooted tree (Fig.5). To infer a root, *E.coli* was taken as the outgroup, based on the considerable body of evidence indicating that mitochondria are of eubacterial origin (19). Our analysis thus indicates that although the malarial LSU rRNA was distantly related to all of the selected chloroplast sequences, it was more closely related to plastids than to mitochondria.

DISCUSSION

Although we have not completely sequenced both LSU rRNA genes within the palindrome, we have shown that the 3′ end is identical in both copies for at least 300 nt, and both restriction mapping (Fig.2) and denaturation studies (16 and Gardner *et al.*, unpublished observations) indicate that most probably the two

genes resemble each other. We therefore feel justified in commenting on various aspects of the LSU rRNA sequence presented here.

LSU rRNAs are frequently fragmented, with short regions at the 5′ and 3′ends separately encoded as 5.8S and 4.5S RNAs, respectively (11). We have found no evidence for the occurence of either of these two fragments, and the sequence reported here is continuous. This is in striking contrast to the LSU rRNA gene from *P.falciparum's* so-called '6 kb element', which is of mitochondrial origin. This gene is extensively fragmented and at the sequence level is essentially unrelated to its counterparts on the circle (5).

We have previously pointed out that the malarial 35 kb circular DNA resembles the residual plastid genome of non-photosynthetic plants (1,10), and the presence of the inverted repeat containing ribosomal RNA and tRNA genes was one feature leading us to this view. However, certain features of the repeat distinguish it from that found on the plastid genomes of higher plants and most algae (9): thus the LSU genes on the malarial circle are distal rather than proximal to the small single copy (SSC) region, the direction of transcription of the malarial LSU rRNA is away from the SSC rather than towards it, and the SSC in the malarial palindrome has been reduced substantially. In this last feature, however, the malarial circle is like the residual plastid genome of the non-photosynthetic plant *Epiphagus* in which most of the SSC region has also been deleted, leaving only two genes (20).

It was with this background that we conducted the phylogenetic analysis of the LSU gene reported here. The unrooted tree of Fig.5, based on the nearest neighbour model of Saitou and Nei (18) and LSU rRNAs from representative eubacteria, chloroplasts, and mitochondria, indicates that those from mitochondria are the most distantly related to the malarial gene. However, the malarial gene is also highly diverged from the chloroplast and cynaobacterial genes used in the analysis. Similar results have been obtained with phylogenetic analyses of the 35 kb circle *rpo*B gene (Gardner *et al.*,in preparation) and a portion of the *rpo*C gene (21), except that in both these instances the malarial gene was more firmly placed with those of chloroplasts than has been the case in the present analysis. Phylogenetic analyses by other workers of nucleus-encoded SSU rRNA genes confirmed earlier speculations that malaria parasites and related organisms (Phylum Apicomplexa) may have evolved from dinoflagellates (22–24). However, to the best of our knowledge, no DNA sequence information is available about the plastid genomes of photosynthetic dinoflagellates, and additional phylogenetic analyses with genes from this source and from other algae will probably be required before a definitive conclusion can be reached about the provenance of the malarial 35 kb circle.

In the meantime, we aim to complete the sequence of the malarial circular DNA, to localize it within the cell, and to gain insight into its function.

ACKNOWLEDGMENTS

For helpful guidance on the phylogenetic analysis we thank Drs M.Gray and D.Spencer, (Dalhousie University), Nova Scotia, and Nick Goldman (National Institute for Medical Research, Mill Hill). We thank Peter Moore for technical assistance. This work was funded in part by the UNDP/WORLD BANK/WHO Special Programme for Research in Tropical Diseases (TDR)—R.J.M.W.; NIH Grant AI25513, and NATO Grant RG.0132/88—J.E.F.

REFERENCES

1. Wilson,R.J.M., Gardner,M.J., Feagin,J.E. and Williamson,D.H. (1991) Parasitol. Today, **7**, 134–136.
2. Vaidya,A.B., Akella,R. and Suplick,K. (1989) Mol. Biochem. Parasitol., **35**, 97–108.
3. Feagin,J.E. (1992) Mol. Biochem. Parasitol., **52**, 145–148.
4. Aldritt,S.M., Joseph,J.T. and Wirth,D.F. (1989) Mol. Cell. Biol., **9**, 3614–3620.
5. Feagin,J.E., Werner,E., Gardner,M.J., Williamson,D.H. and Wilson,R.J. (1992) Nucleic Acids Res., **20**, 879–887.
6. Wilson,R.J.M., Fry,M., Gardner,M.J., Feagin,J.E. and Williamson,D.H. (1992) Curr. Genet, **21**, 405–408.
7. Gardner,M.J., Feagin,J.E., Moore,D.J., Spencer,D.F., Gray,M.W., Williamson,D.H. and Wilson,R.J. (1991) Mol. Biochem. Parasitol., **48**, 77–88.
8. Gardner,M.J., Williamson,D.H. and Wilson,R.J. (1991) Mol. Biochem. Parasitol., **44**, 115–123.
9. Palmer,J.D. (1985) Ann. Rev. Genet., **19**, 325–354.
10. Palmer,J.D. (1992) Curr. Biol., **2**, 318–320.
11. Gutell,R.R. and Fox,G.E. (1988) Nucleic Acids Res., **16**, r175-r269.
12. Devereaux,J., Haeberli,P. and Smithies,O. (1984) Nucleic Acids Res., **12**, 387–395.
13. Felsenstein,J. (1991) PHYLIP (PHYLogenetic Inference Package), version 3.4, documentation.
14. Kohchi,T., Shirai,H., Fukuzawa,H., Sano,T., Komano,T., Umesono,K., Inokuchi,H., Ozeki,H. and Ohyama,K. (1988) J. Mol. Biol., **203**, 353–372.
15. Williamson,D.H., Wilson,R.J.M., Bates,P.A., McCready,S., Perler,F. and Qiang,B. (1985) Mol. Biochem. Parasitol., **14**, 199–209.
16. Wilson,R.J.M., Gardner,M.J., Rangachari,K. and Williamson,D.H. (1993) NATO ASI series, *in press*.
17. Noller,H.F. (1984) Annu. Rev. Biochem., **53**, 119–162.
18. Saitou,N. and Nei,M. (1987) Mol. Biol. Evol., **4**, 406–425.
19. Gray,M.W. (1988) Biochem. Cell. Biol., **66**, 325–348.
20. Wolfe,K.H., Morden,C.W. and Palmer,J.D. (1992) J. Mol. Biol., **223**, 95–104.
21. Howe,C.J. (1992) J. Theor. Biol., **158**, 199–205.
22. Levine,N.D. (1987) Phylum II. Apicomplexa. In: An illustrated guide to the protozoa. Lee,J.J., Hutner,S.H. and Bovee,E.C. (eds), pp. 322–357, Society of Protozoologists, Lawrence, Kansas.
23. Barta,J.R., Jenkins,M.C. and Danforth,H.D. (1991) Mol. Biol. Evol., **8**, 345–355.
24. Gajadhar,A.A., Marquardt,W.C., Hall,R., Gunderson,J., Ariztia-Carmona,E.V. and Sogin,M.L. (1991) Mol. Biochem. Parasitol., **45**, 147–154.