Volume 13 Number 6 1985 Nucleic Acids Research

Organization of *Plasmodium falciparum* genome: 1. Evidence for a highly repeated DNA sequence

Ramareddy V.Guntaka[1]*, Siddarame Gowda[1], Arepalli S.Rao[1] and Theodore J.Green[2]

Departments of [1]Microbiology and [2]Veterinary Microbiology, University of Missouri, Columbia, MO 65212, USA

Received 30 November 1984; Revised and Accepted 22 February 1985

ABSTRACT

Plasmodium falciparum DNA, isolated from the merozoite stage, was cleaved with HindIII and cloned in pBR322 and λL47.1 vectors. Plasmid clones containing 13.4, 7.0, 4.3, 4.1 and 1.5 kb inserts were characterized in some detail. The inserts contain several repeating units of smaller size. Nucleic acid hybridization studies showed that the repeat element is present in the Plasmodium DNA at a very high copy number and appears to be distributed widely throughout the genome.

INTRODUCTION

Malaria, a major infectious disease in tropical countries, is caused by species of Plasmodium, a genus that belongs to protozoa. The life cycle of plasmodia is very complex in that the asexual reproduction or schizogony occurs in a human host and the sexual reproduction or sporogony takes place in mosquito. At least four distinctly different stages - sporozoite, exoerythrocytic forms, merozoite and gametocytes -exist, during which quite different antigens are expressed. Thus, a different set of genes must be expressed in each of these stages. However, very little is known about the organization and developmental regulation of the genome.

The genome of Plasmodium species is also unusual in that it has extremely low G+C content (17-19%) and is only four to seven times larger than the E. coli genome (1, 2). Eukaryotic genomes contain highly repeated, moderately repeated and unique sequences organized in such a way that the unique coding sequences are interspersed by the repeat families of DNA (3). In Plasmodium berghei, about 18% of the DNA appears to be repetitive as evidenced by C_0t analysis (2), and the infectivity of the parasite has been shown to be directly related to the proportion of repetitive DNA (4). Some repeat sequences have been cloned and one of them is used to demonstrate the presence of multiple bands in P. falciparum DNA (4a). Other studies indicate the presence of tandem repeats within the coding sequences of some genes (5-9). However, the structure of repeat sequences and their arrangement in relation to the structural genes are not known. This is very important especially because of differential expression of genes during various stages of development in two diverse host species.

In order to elucidate the structural organization of the plasmodium genome, we have established genomic libraries in pBR322 and λL47.1 vectors and characterized in some detail a repeat element that is present at a high copy number in the genome of Plasmodium falciparum. This report describes the identification and characterization of this repetitive DNA sequence.

MATERIALS AND METHODS

Source of Plasmodium falciparum - FCR-3/FMG strain of Plasmodium falciparum was propagated by the method of Trager and Jensen (10) as described (11).

Isolation of DNA and RNA - All solutions were treated with diethylpyrocarbonate and autoclaved. The CsCl method for the simultaneous isolation of DNA and RNA was described (12). Briefly, human red blood cells infected by fully mature segmenters containing well developed merozoites, were pelleted by low speed centrifugation and the pellet was resuspended in 1 to 2 volumes of 5M guanidinium thiocyanate. The resulting lysate was mixed with solid CsCl (1 g/ml), layered on a cushion of CsCl solution (1.35 g/ml) and centrifuged at 25,000 rpm for 18 hours at 17°C in a SW41 rotor. A pellicle containing hemin and hemozoin pigment and membranes were at the top of the gradient. The RNA formed a translucent pellet and the DNA remained in the gradient. After removing the pellicle, the DNA was collected from the gradient, dialyzed against 10 mM Tris-HCL, pH 8.1, 1 mM EDTA (TE), extracted with phenol-chloroform and precipitated with ethanol at -20°C. It was resuspended in TE and stored at 4°C. The pelleted RNA was resuspended in TE containing 0.2% sarkosyl, made the solution to 0.2 M sodium acetate and precipitated with ethanol at -20°C.

Cloning in pBR322 - pBR322 DNA was cleaved with HindIII endonuclease and then dephosphorylated by bacterial alkaline phosphatase (BRL, Gaithesberg, MD). Approximately 1 µg of HindIII-digested Plasmodium falciparum DNA was ligated to 0.5 µg of dephosphorylated pBR322 DNA at 15°C with 4 units of T4 DNA ligase and the ligated DNA was used to transform E. coli HB101 as described previously (13). Transformants were scored on LB-agar plates containing ampicillin (amp) and then screened for tetracycline (tet) sensitivity. DNA was isolated from $amp^R tet^S$ colonies and characterized as described (13).

Cloning in λL47.1 vector - λL47.1 phage (14) was propagated in NZCYM medium supplemented with 0.2% maltose. Purification of phage and isolation of DNA were as described by Maniatis et al. (15). To purify λL47.1 HindIII or EcoRI arms, phage DNA was first annealed at 42°C, the cohesive ends were ligated by T4 ligase and the DNA was then digested by HindIII or EcoRI. Ligated arms were separated from the stuffer fragment on velocity sucrose gradients. All these steps were carried out essentially as described by Maniatis et al. (15). One µg HindIII-digested Plasmodium DNA was ligated

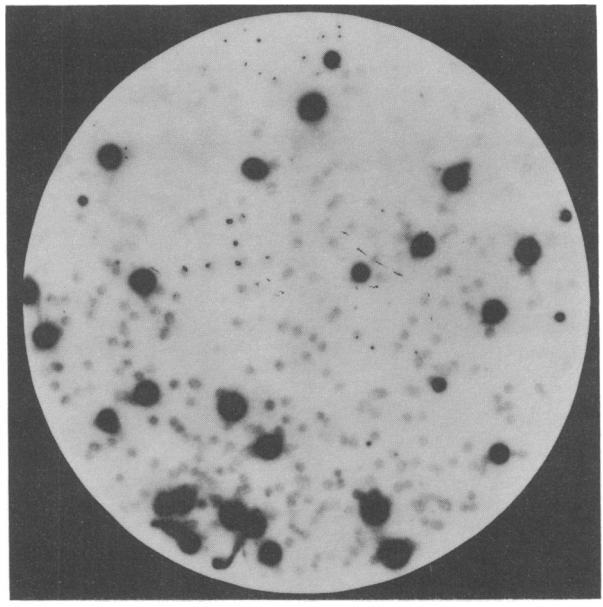

Figure 1. **Plaque hybridization.**
About 300 plaques from amplified λPL-HindIII library were plated on NZCYM plates, transferred to cellulose nitrate filters and hybridized with pPL-7 probe. The filters were washed and exposed against X-ray film with an intensifying screen for about 4 hours.

to 3 µg λ L47.1 arms at 15C in 20 µl. Packaging was carried out using Promega Biotech (Madison, Wisconsin) extracts as instructed by the suppliers.

Screening for recombinant plaques - Plaques were transferred to nitrocellulose filter discs, denatured and hybridized as described by Maniatis et al. (15). About 1 x 10^6 cpm/ml (specific activities 50-200 x 10^6 cpm/µg DNA) hybridization buffer (6 x SSC, 2 x Denhardt buffer, 3 mM EDTA, 100 µg/ml denatured calf thymus or herring sperm DNA) was used at 68°C for 18 to 20 hours. The filters were washed, dried and exposed to Kodak X.O.Mat AR film with intensifying screen as described previously (16).

RESULTS

Identification of the repeat element A HindIII library of plasmodium DNA was established in λL47.1 vector by standard methods (14). Approximately 100,000 plaque-forming units were obtained and more than 90% of these plaques contained inserts of various sizes. From this library about 300 plaques were screened for DNA sequences related to pPL-7, using the radiolabeled pPL-7 probe. pPL-7 is a recombinant pBR322 clone and contains a 13.4 kb plasmodium DNA insert. This clone was chosen for further

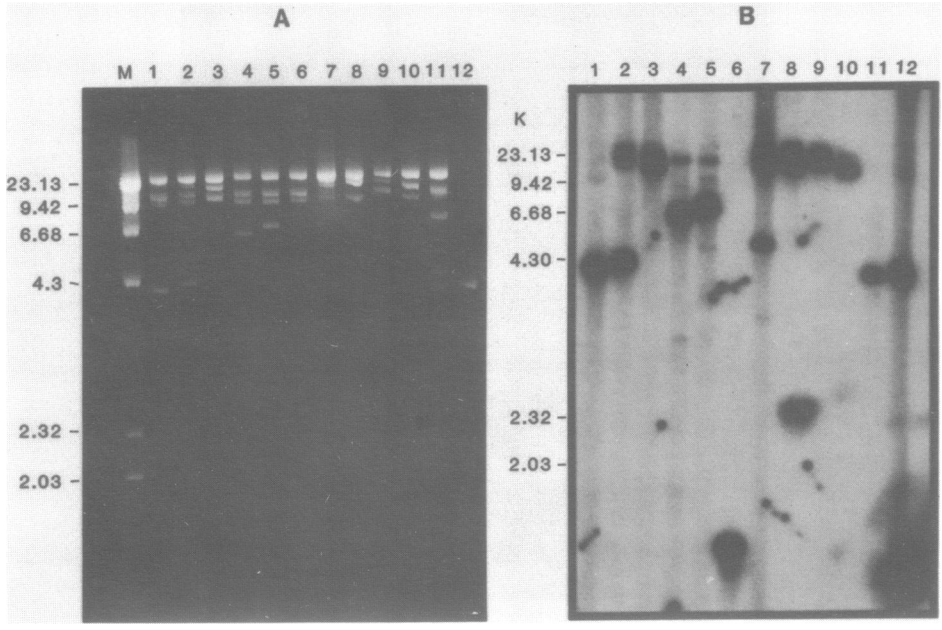

Figure 2. **Characterization of λPL-7 clones.**
The phages from individual positive plaques (Fig. 2) were amplified and banded in CsCl gradient (Maniatis et al., 1982). The DNA was isolated, restricted with HindIII and analyzed on a 0.8% agarose gel. The gel was blotted and hybridized to pPL-7 probe. Lane M contained HindIII-digested λ DNA marker. Lanes 1-11 contain DNAs from clones λPL-7-1, 7-3, 7-6, 7-8, 7-9, 7-10, 7-14, 7-22, 7-23, 7-24, and 7-25, respectively. Lane 12 contains the purified 4.1 kb fragment from pPL-15. Panel A: Ethidium bromide staining of this gel. Panel B: autogradiogram of the Southern blot.

analysis because of its unique structure and composition as evidenced by its resistance to cleavage by several restriction enzymes such as HpaII, EcoRI and HinfI. An unusually high proportion (25 out of 300) of them showed hybridization (Fig. 1) indicating the presence of sequences homologous to pPL-7. Recombinant phage from each positive plaque was amplified and DNA was extracted. The DNAs were restricted with HindIII and fractionated on agarose gel. It is clear from the data (Fig. 2, panel A) that each clone contains an insert of different size. The size of the inserts ranged from 1.5 to 20 kb. Most of the clones contain more than one insert. This is probably due to the minimum size requirement for packaging. In order to identify the fragments that specifically contain the repeat element, the gel was blotted and hybridized to pPL-7 probe. The results (Fig. 2, panel B) indicate that at least one fragment from each recombinant hybridized and that the fragment is different from clone to clone. For example, clones λPL-7-10 (lane 6), λPL-7-1 (lane 2), λPL-7-9 (lane 5) and λPL-7-23

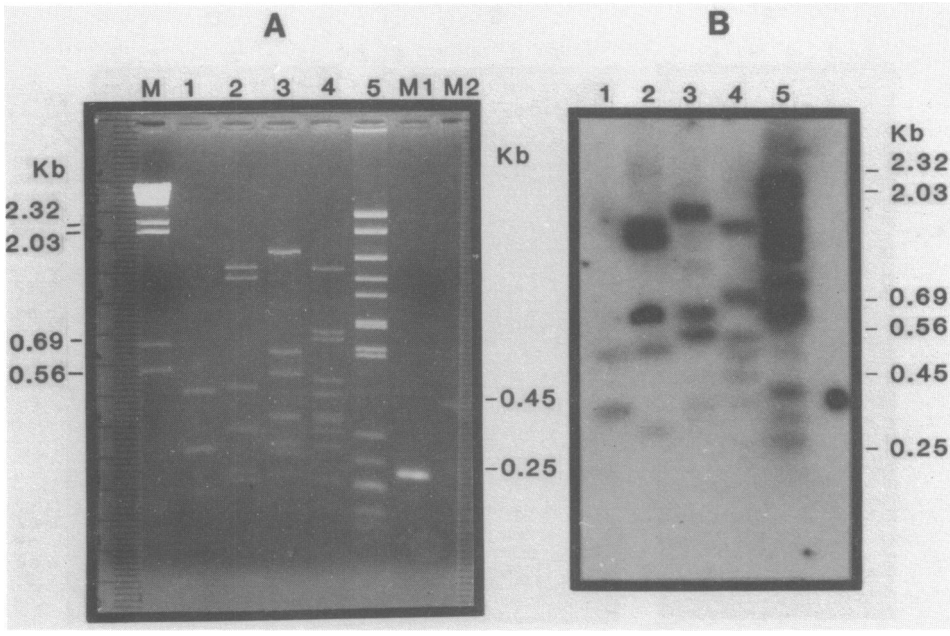

Figure 3. **Sequence organization of various clones.**
The 1.5, 4.1, 4.3, 7.0, and 13.4 kb inserts from various λ and pBR322 clones were gel purified and subjected to RsaI digestion. Following fractionation on a 1.5% agarose gel and staining with ethidium bromide to visualize the fragments (panel A), they were blotted onto two filters and hybridized with pPL-7 (panel B). Lanes 1-5 contain the RsaI fragment of the 1.5, 4.1, 4.3, 7.0 and 13.4 kb fragments, respectively. Lanes M, M1 and M2 refer to λ-HindIII, 0.25 and 0.45 kb markers, respectively.

(lane 9) carry fragments of 1.5, 4.1, 7.0 and 18 kb, respectively, that hybridized to pPL-7. Some clones (lane 1 and lane 7) contain two different size fragments both of which carry the repeat element as evidence by the hybridization to pPL-7 probe. These results suggest that a DNA sequence which is common to all the clones is present in each fragment. Additional evidence that each insert is different not only in their size but in their sequence organization was obtained by purifying the 1.5, 4.1, 4.3, 7.0 and 13.4 kb fragments, digesting them with RsaI and analyzing these on a 1.5% agarose gel. The restriction enzyme pattern (Fig. 3) is quite different for these inserts. This enzyme, although cleaves each fragment at many sites, does not generate common fragments among these cloned inserts (Fig. 3, panel A). In order to identify the sequence that is homologous, the gel was blotted and hybridized to pPL-7 probe (Fig. 3, panel B). The results show hybridization only to certain fragments and that each fragment that hybridized is different from clone to clone. For example in the 1.5 kb fragment, RsaI fragments of 0.5 and 0.32 (Panel B and lane 1) show hybridization

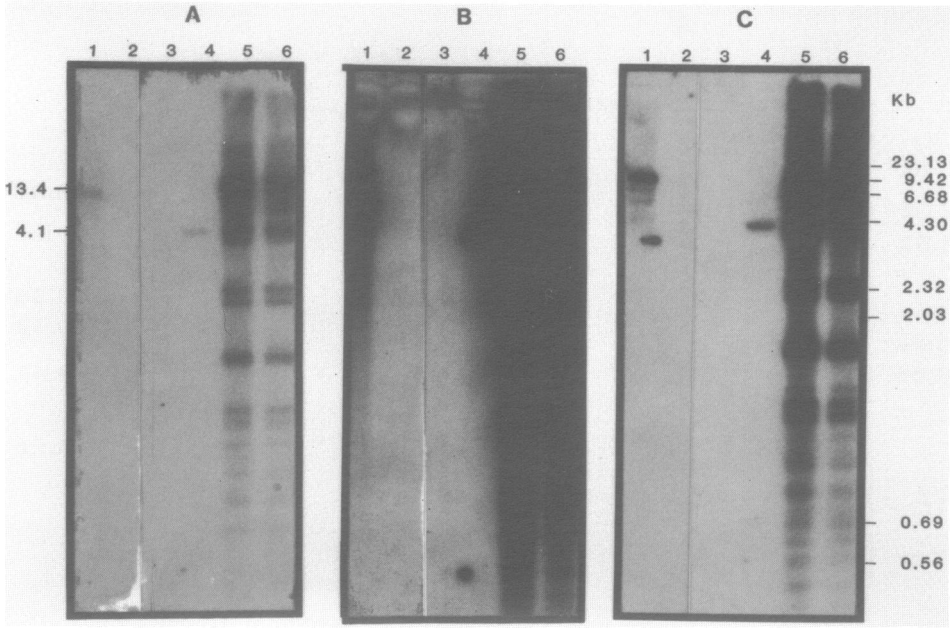

Figure 4. **Evidence that the repeat element is specific for Plasmodium genome.** About 2 to 3 µg DNA from two different batches of P. falciparum merozoites were digested with HindIII and fractionated on a 1% agarose gel. Following Southern blotting, they were hybridized with pPL-7 probe. Subsequently, the pPL-7 probe was removed and the filter was rehybridized with pPL-15 probe. Hybridization with pPL-7 probe followed by 2 hour (panel A) or 18 hour (panel B) exposure against X-ray film. Panel C contains results of rehybridization with pPL-15 probe. Lanes 2 and 3 in each panel contains 20 µg of HindIII-digested human KB cell or mouse 3T3 DNAs. Lanes 1 and 4, respectively, contain purified 13.4 and 4.1 kb insert DNAs used as markers. Lanes 5 and 6 contain two different batches of plasmodium DNA.

whereas in the 4.1 kb insert two RsaI fragments of 1.45 and 1.25 kb in addition to the 0.52 kb and 0.25 kb fragments (lane 2), show hybridization. Three RsaI fragments of 1.7, 0.6 and 0.55 kb of the 4.3 kb insert and at least six fragments of the 7.0 kb insert hybridized to pPL-7 probe. These results indicate that the repeat element is present at specific regions of each insert and RsaI is probably cleaving at least once in the repeat element. In support of this, we have shown that in the 4.1 kb insert, the repeat sequence is present at two distinct regions, one at one end and the other in the middle (data not shown). Further experiments, including sequencing, are required to understand the detailed characteristics and organization of these elements.

Specificity of the repeat element for Plasmodium genome - The following experiment shows that this novel repeat sequence is specific for plasmodium DNA and is present at a very high copy number. Different batches of P. falciparum DNA as well as human

and mouse DNA were digested with HindIII, fractionated on a 1% agarose gel, blotted and hybridized to radiolabeled pPL-7 probe. Both human and mouse DNAs gave negative results (Fig. 4, lanes 2 and 3 in all three panels) even after 18 hours (panel B) exposure where as several distinctly different bands (Fig. 4, lanes 4 and 5) ranging in size from 0.5 to 25 kb could be readily observed in plasmodium DNA, strongly suggesting that the repeat sequence is Plasmodium-specific. It should be pointed out that only 2 to 3 µg Plasmodium DNA was used in these experiments as opposed to 15-20 µg of human and mouse DNAs. Even with these smaller amounts of plasmodium DNA, repeat sequence-specific bands could be detected only after 15 minute exposure; an example of these results after a 2 hour exposure is given in Fig. 4 (Panel A, lanes 4 and 5). Fig. 4, panel B shows results of an autoradiogram after 18 hour exposure against X-ray film. Since a single copy gene in human and mouse DNAs could be detected in less than 7 to 8 hours of exposure under these conditions, we conclude that the repeat sequence we have identified is specific for plasmodium DNA. Employing a single copy gene clone (λPL-Ag2) as a standard, we have obtained evidence to indicate that up to 2 to 3 percent of the plasmodium DNA is represented by this repeat sequence.

Evidence presented above indicated that all the λ clones contained an identical or similar repeat sequence. For example, we have demonstrated that pPL-7 and pPL-15 contain a homologous repeat element although they contain quite different size inserts (13.4 and 4.1 kb, respectively). To further demonstrate that these sequences are similar in the plasmodium genome, the radioactivity of the blot used in Fig. 4 was removed and the filter was rehybridized with pPL-15 probe. Since the same set of bands could be detected (Fig. 4, panel C), the results suggest that the repeat element is very similar.

Plasmodium DNA, digested with HindIII, was fractionated on a 1% agarose gel to identify most of the fragments that have a sequence common to pPL-7 probe. Since the resolution of fragments larger than 4 or 5 kb is not good in 1% agarose gel (Fig. 5, panel A), and in order to resolve the larger fragments, plasmodium DNA was digested with HindIII and analyzed on a 0.5% gel. The results (Fig. 5, panel B) indicate that several unique bands of different size could be readily detected under these electrophoresis conditions as opposed to the poor resolution in 1% agarose (Fig. 4). This experiment indicated that multiple bands (more than 15-20) of size greater than 4 kb are present. Moreover, these results demonstrate that some bands of larger size are less intense than some bands of smaller size fragments. This might be due to the presence of multiple copies of the same repeat sequence in a given fragment or a single copy of repeat sequence in multiple fragments of similar size. A similar banding pattern was observed with HaeIII (Panel C, lane 1) or with NdeI (Panel C, lane 2) which cleave the 13.4 kb DNA at five and one site, respectively. Recently we have cloned a 250 bp sequence, which is part of the repeat element in M13mp7 and used this as a probe in

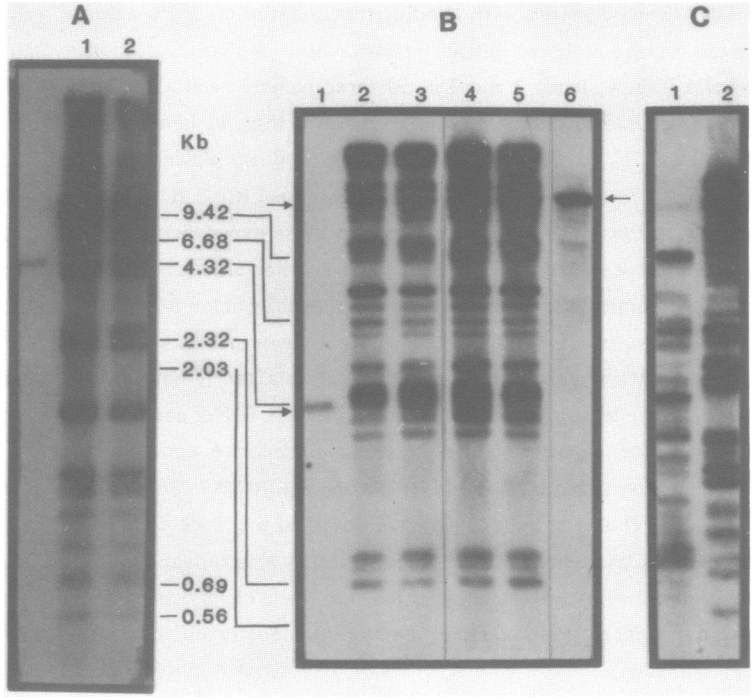

Figure 5. As in Fig. 4 except a 0.5% agarose gel was used to demonstrate multiple discrete bands ranging in size from 4 to 25 kb. Panel A. Two different batches (lanes 1 and 2) of Plasmodium DNA digested with HindIII and ran as in Fig. 5. Panel B. As in panel A except the DNA was run on a 0.5% gel, blotted, hybridized and exposed for 3 hours (lanes 2 and 3) or 6 hours (lanes 4 and 5). Lanes 1 and 6 contain 50 pg of isolated 4.1 and 13.4 kb fragments. Panel C. Plasmodium DNA digested with HaeIII (lane 1) or NdeI (lane 2).

similar genome blots. The repeat pattern is identical to those shown above (data not shown). These results suggest that the repeat element is localized at several discrete sites in the genome. Further experiments are necessary to deal with the organization of the repeat element.

DISCUSSION

Earlier recombinant DNA work by Scaife and his colleagues (4a) indicated the presence of repeat sequences in P. falciparum but they have not characterized the sequences in their clones. The results reported here demonstrate the presence of a highly repeated DNA sequence in the genome of P. falciparum. The repeat element appears to have been distributed over a large range of restriction fragments and the smallest fragment containing this sequence is about 450-500 bp. Whether an identical

repeat sequence is present in all these different fragments or whether similar but not identical repeats are present throughout the genome has to be determined. Isolation of the entire repeat element from different restriction fragments followed by nucleotide sequencing should reveal the characteristics of this sequence.

The clone pPL-7, containing the repeat sequence is also unusual in that very few restriction endonucleases cleave the insert. For example, HinfI, which cleaves at GANTC, has a single site in the 13.4 kb fragment. EcoRI (GAATTC) and AhaIII (TTTAAA) do not cleave at all, NdeI (CATATG) cleaves once, whereas HpaI (GTTAAC) cleaves at 6 sites suggesting that the sequence TTAA present more frequently than AATT or ATAT. HaeIII (GGCC) cleaves the fragment at five sites, whereas HpaII (CCGG) does not cleave this fragment. This might be due to low G + C content of P. falciparum DNA and therefore clustered G + C sequences are very rare. In agreement with this, other enzymes such as RsaI (GTAC) and AluI (AGCT) cleave the fragment more frequently than HaeIII but definitely less than the expected frequency, again reflecting high A + T content of this clone which carries the repeat element.

It is also remarkable in that the discrete bands of the repeat element-containing sequences are present in the P. falciparum genome. The smallest size HindIII fragment that contains the repeat sequence is about 450-500 bp with increments of about 50 to 100 bp that go up to about 1.6 kb. Then there are gaps in the gel before another set of bands of larger size appear. Although the precise organization is not known, it is tempting to speculate that small repeat clusters are present in a tandem array to give rise to large repeat elements. In other words, it is possible that several short repeats are arranged in tandem to give rise to a large repeat of 1 to 1.2 kb and several of these are in turn arranged in tandem in the larger DNA fragments such as the 13.4 kb of pPL-7. In support of this, we have observed that the 13.4 kb fragment, upon digestion with HaeIII, gives rise to fragments of 1.15, 2.3 and 3.5 kb all of which contain sequences homologous to the basic repeat element (data not shown).

The eukaryotic genome contains highly repeated, moderately repeated and unique sequences organized in such an order that the unique coding sequences are interspersed by these repetitive sequences (3). The most intensively characterized repeat unit is the Alu family, which is about 165-300 bp long and is represented by about 300,000 to 500,000 copies per human diploid genome (17). Recently several other repeat units such as Kpn, O and R families in human DNA, R and Bam5 families in mouse DNA in addition to the Alu family, have been described (17-19). Studies on the organization and expression of these repeat units indicate that these are localized adjacent to the unique coding sequences and are usually transcribed into RNA by RNA polymerase III (20) or by RNA polymerase II (21).

Experiments to examine the organization of this repeat unit in P. falciparum

indicate that other unique sequences are interspersed by these repeat sequences. Whether these unique sequences are transcribed into RNA remains to be determined. So far we were unable to detect RNA specific for the repeat sequences in P. falciparum merozoites suggesting that these sequences are probably not transcribed. However, it is possible that the repeat sequence might contain regulatory sites such as promoter in which case specific transcripts will not be found. Or, short leader sequences, like those present in the mRNAs of trypanosomes, might be provided by this repeat sequence (22-24). It is also possible that this repeat element might serve as an origin of DNA replication or function as a transposable element (3, 25). In this regard, it is noteworthy to point out that the amount of repetitive DNA in P. berghei appears to contribute to the infectivity of the parasite (4). Studies to determine the arrangement and expression of this repeat DNA sequence are in progress.

ACKNOWLEDGEMENTS

We thank Drs. Harold E. Varmus and Mark A. McIntosh for critical reading of the manuscript, Mr. Vincent Parsons for his assistance in preparing λ47.1 arms and Theresa Bean and Faye Bullard for growing Plasmodium falciparum. This work was supported by an U.S.A.I.D. contract No. AID-DPE-0453-COO-3058-00 and funds from the School of Medicine, University of Missouri.

*To whom correspondence should be addressed

LITERATURE CITED

1. Pollack, U., Katzen, A. L., Spira, D. T., and Golenser, J. (1982) Nucleic Acids Res. **10**, 539-546.
2. Dore, E., Birago, C., Frontali, C., and Battaglia, P. A. (1980) Mol. and Biochem. Parasitology **1**. 199-208.
3. Davidson, E., and Britten, R. (1979) Science **204**, 1052-1059.
4. Birago, C., Bucci, A., Dore, E., Frontali, C., and Zenobi, P. (1982) Mol. and Biochem. Parasitology **6**, 1-12.
4a. Goman, M., Langsley, G., Hyde, J., Yankovsky, N. K., Zolg, J. W., and Scaife, J. G. (1982) Mol. Biochem. Parasitology **5**, 391-400.
5. Kemp, D. J., Coppel, R. L., Cowman, A. F., Saint, R. B., Brown, G. V., and Anders, R. F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3787-3791.
6. Anders, R. F., Brown, G. V., and Edwards, A. (1983) Proc. Natl. Acad. Sci. **80**, 6652-6656.
7. Hyde, J. E., Zolg, J. W., and Scaife, J. G. (1981) Mol. and Biochem. Parasitology **4**, 283-290.
8. Ozaki, L. S., Svec, P., Nussenzweig, R. S., Nussenzweig, V., and Godson, G. N. (1983) Cell **34**, 815-822.
9. Dame, J. B., Williams, J. L., McCutchan, T. F., Weber, J. L., Wirtz, R. A., Hockmeyer, W. T., Maloy, W. L., Haynes, J. D., Schneider, I., Roberts, D., Sanders, G., Ready, E. P., Diggs, C. L., and Miller, L. H. (1984) Science **225**, 593-599.
10. Trager, W., and Jensen, J. B. (1976) Science **193**, 673-675.

11. Green, T. J., Morhardt, M., Brackett, R. G., and Jacobs, R. L. (1981) Infect. and Immun. **31**, 1203-1208.
12. Guntaka, R. V., and Weiner, A. J. (1978) Nature **276**, 276-278.
13. Mitsialis, S. A., Young, F., Palese, P., and Guntaka, R. V. (1981) Gene **16**, 217-225.
14. Loenen, W. A., and Brammer, W. J. (1980) Gene **10**, 249-262.
15. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning. Cold Spring Harbor Laboratory Manual. Cold Spring Harbor, New York.
16. Guntaka, R. V., Rao, P. Y., Mitsialis, S. A., and Katz, R. A. (1980) J. Virol. **34**, 569-572.
17. Schmid, C., and Jelinek, W. (1982) Science **216**, 1065-1070.
18. Shafit-Zagardo, B., Maio, J., and Brown, F. (1982) Nucleic Acids Res. **10**, 3175-3193.
19. Gebhard, W., Meitinger, T., Hoechtl, J., and Zachau, H. (1982) J. Mol. Biol. **157**, 453-471.
20. Haynes, S., and Jelinek, W. R. (1981) Proc. Natl. Acad. Sci. USA **78**, 6130-6134.
21. Heller, D., Jackson, M., and Leinwald, L. (1984) J. Mol. Biol. **173**, 419-436.
22. Parsons, M., Nelson, R. G., Watkins, K. P., and Agabian, N. (1984) Cell **38**, 309-316.
23. Campbell, D. A., Thorton, D. A., and Boothroyd, J. C. (1984) Nature **311**, 350-355.
24. Dorfman, D. M., and Donelson, J. E. (1984) Nucleic Acids Res. **12**, 4907-4920.
25. Haynes, S., Toomey, T., Leinwald, L., and Jelinek, W. (1981) Mol. Cell. Biol. **1**, 573-583.