Volume 16 Number 10 1988 Nucleic Acids Research

The genome of *Plasmodium cynomolgi* is partitioned into separable domains which appear to differ in sequence stability

Thomas F.McCutchan*, John B.Dame[1], Robert W.Gwadz and Kenneth D.Vernick

Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892 and [1]Department of Infectious Diseases, University of Florida, Gainesville, FL 32610, USA

Received January 14, 1988; Revised and Accepted April 19, 1988

ABSTRACT

The genome of Plasmodium cynomolgi is partitioned into at least 7 distinct genetic domains. Each domain is apparently uniform in DNA density and is separable from the others by CsCl density centrifugation in the presence of Hoechst dye. The protein-encoding genes that were tested are localized in the two heaviest density domains (isochores). The ribosomal genes are in two lighter isochores as well as in one of the isochores that contains protein encoding genes. Telomeric sequences are mainly, if not exclusively, in the lightest isochores, indicating that position with regard to chromosome ends may correlate with density. Blocks of a tandemly-repeating sequence which mark genetically hypervariable chromosome regions in malaria parasites are located in all isochores. However, the rate of change associated with the blocks of sequence is much slower in some isochores than in others. This indicates that the rate of genetic change in these parasites may differ with isochore and chromosomal position. These results may also have more general biological implications since they suggest that the genetic instability often noted for tandem repeat sequences in the eukaryotic genome may be limited to only a distinct subset of the genomic complement of such sequence blocks.

INTRODUCTION

We have recently described Plasmodium-derived DNA sequences which are associated with genetic hypervariability at meiosis but are essentially stable during mitotic replication (1; Vernick et al., manuscript submitted). Upon examining restriction fragment length polymorphism in the parents and progeny of a genetic cross, it was found that almost every progeny clone contained fragments which did not occur in either parent. The rearranged sequences were found to be located both in telomeric locations and in locations internal to the chromosome end, although the actual site of genetic variability was frequently localized in the internal fragment. It was not determined whether all or only a subset of fragments containing the hypervariable sequence were meiotically unstable. This question is interesting because if all such sequences are unstable it would indicate that meiotic hypervariability involved a sequence-related mechanism. Conversely, variability only in a subset of the sequence copies would tend to suggest that location on the chromosome was involved in hypervariability. Further, since the hypervariable sequence occurs as blocks of tandem repeats, the restriction of its genetic variability to a particular component of the genome could have general implications for our understanding of the mechanism of tandem repeat instability. In order to investigate the differential stability of sequences in the Plasmodium

genome, as well as to gain new insight into the overall organization of the genome, we have separated genomic DNA into distinctive and reproducible domains which appear to correspond to biologically-defined genetic partitions.

The reproducible separation of different DNAs typically requires prior treatment with a restriction endonuclease to make fragments containing a given segment of DNA uniform in size and composition. Large pieces of DNA can be physically separated from each other without restriction cleavage if the pieces being separated are characteristically different in some other way. The characteristic may be as simple as a uniform size difference. For example, the size of a chromosome is biologically defined rather than being defined by post-isolation treatment with restriction endonucleases. Chromosomes from some organisms may be separated from each other by pulsed field gel electrophoresis (2). Unrestricted DNA can also be separated on the basis of density difference. The separation of satellite DNA from other components of the genome during density ultracentrifugation in CsCl is a well-known example.

An example of a biologically-defined characteristic which correlates with density differences is the timing of DNA replication during the cell cycle. It has been shown for warmblooded vertebrates that replication units that are copied early are dG•dC rich, while units that are copied later are dA•dT rich and hence less dense than the early replicating DNA (3, 4). The difference in dG•dC content between the genomic regions has allowed the early and late replicating components of the mammalian genome to be separated by density gradient centrifugation (5).

We have studied the organization of genomic DNA from parasitic malaria protozoans of the genus Plasmodium in a similar manner. Some species like P. falciparum have genomes that are uniformly light (6) while others like P. cynomolgi contain distinct density components (isochores) like warm blooded vertebrates. Here we report on some of the characteristics of compartmentalization of the P. cynomolgi genome, and the relationship between the density organization of the genome and the rate of genetic change.

EXPERIMENTAL PROCEDURES

P. cynomolgi Parasite Strains

The NIH and London strain were isolated from different geographical regions of Malaysia (7). The RO strain was isolated from a Rhesus monkey imported from Burma. All three strains are maintained at the National Institutes of Health

DNA Treatment

Genomic DNA was isolated from P. cynomolgi (NIH strain) as described (8). Twenty-five μg of this DNA was mixed with 25 μg of Hoechst dye 33258 and then centrifuged at 45K RPMs in a CsCl solution (refractive index of 1.3950) as described (6, 9, 10). A photograph of this type of gradient resulting in the separation of either P. cynomolgi or Plasmodium vivax DNA

has been published (10). Although a number of bands are visible upon examination of gradients, DNA preparations can contain varying amounts of Rhesus DNA which itself has highly repetitive elements like alpha satellites. We therefore have relied totally on Southern blot analysis for the identification of parasite material. The CsCl solution was then fractionated into thirty-three 1.2 ml aliquots after puncturing the bottom of the centrifuge tube. Each fraction was extracted with CsCl-saturated isopropanol, dialyzed against 0.01 M Tris-HCl and 0.001 M EDTA, pH 7.5, and then precipitated in 0.3 M NaAcetate and two volumes of ethanol. As shown below, several distinct isochores can be separated in this fashion. DNA from each fraction was digested in the presence of the restriction endonuclease HindIII. One-fifth of each digestion mix was electrophoresed in 0.8% agarose. The DNA was transferred to nitrocellulose filters and subjected to Southern blot analysis. The fraction in which a particular gene is most concentrated therefore defines its isochore. We use the term isochore to describe different density regions as Bernardi and his colleagues have done for other eukaryotic genomes (5).

RESULTS

The P. cynomolgi Genome is Partitioned into a Small Number of Density Domains

We have separated P. cynomolgi genomic DNA into several distinct density fractions by Hoechst dye-CsCl ultracentrifugation. By preferentially binding to dA•dT base pairs, Hoechst dye decreases the density of dA•dT-rich sequences relative to dG•dC-rich ones on an isopycnic gradient (11). Dye remains in contact with the DNA for the 18 hour period of centrifugation and hence the amount of dye bound to DNA probably reaches near saturation. Naturally occurring density differences within the genome are intensified by the dye in order to increase the magnitude of separation during CsCl ultracentrifugation. The DNA preparation used was isolated from NIH-strain parasites and was essentially free of host DNA contamination (data not shown).

The data in Figure 1 show that the density gradient isolated distinct isochores, or density fractions, away from one another. For this figure, Southern blots comparing HindIII-digested DNA of each of the 33 fractions from the CsCl gradient were prepared. The same percentage of each fraction was loaded on the gel because that approach gives information on the actual distribution of material. The lower-numbered fractions on the left side were heavier fractions from the bottom of the gradient while the higher-numbered fractions were lighter. Lanes 13 and 26 contain a HindIII restriction digest of bacteriophage lambda. The weak hybridization occasionally observed in lanes 13 and 26 resulted from homology between lambda DNA size markers and pBR322, which was part of some probes. In Figure 1A, whole DNA from fraction 24 was radiolabelled by nick translation and used to probe the Southern blots. Although this procedure would not detect low copy number homology, the fragments detected were mostly contained in three fractions of the gradient, 23, 24 and 25. Therefore, the DNA

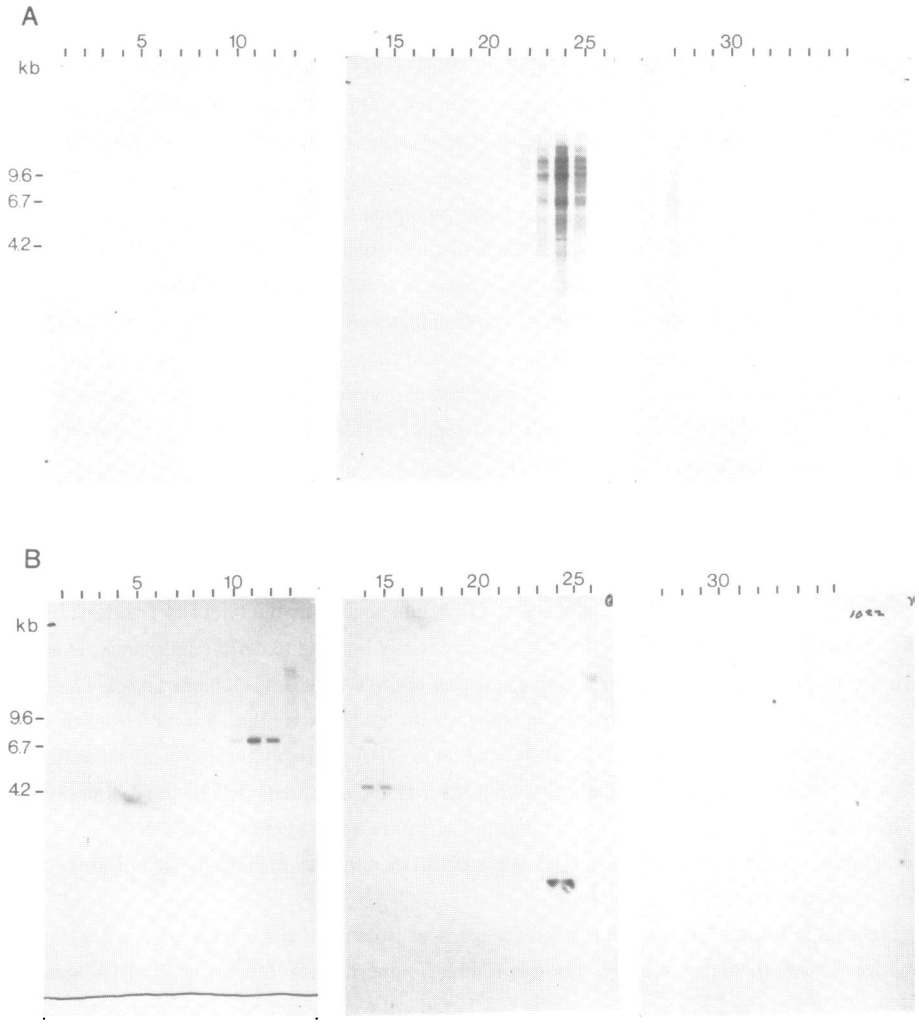

Figure 1. Southern blot analysis of DNA from fractions collected from a Hoechst dye-CsCl density gradient of genomic DNA from P. cynomolgi, NIH strain. Thirty-six fractions were collected. The lowest numbered fractions contain the denser material. The DNA from each was digested with HindIII, electrophoresed, and blotted to nitrocellulose.
A. Autoradiogram of Southern blots that were hybridized with radiolabelled DNA from fraction 24.
B. Autoradiogram of Southern blots that were hybridized with a ß-actin probe from chicken.

sequences from fraction 24, although clearly complex in makeup, was not spread over the gradient. Figure 1B shows that sequences which hybridize to specific DNA probes were found in distinctly different areas of the gradient. In this experiment, an actin-specific probe

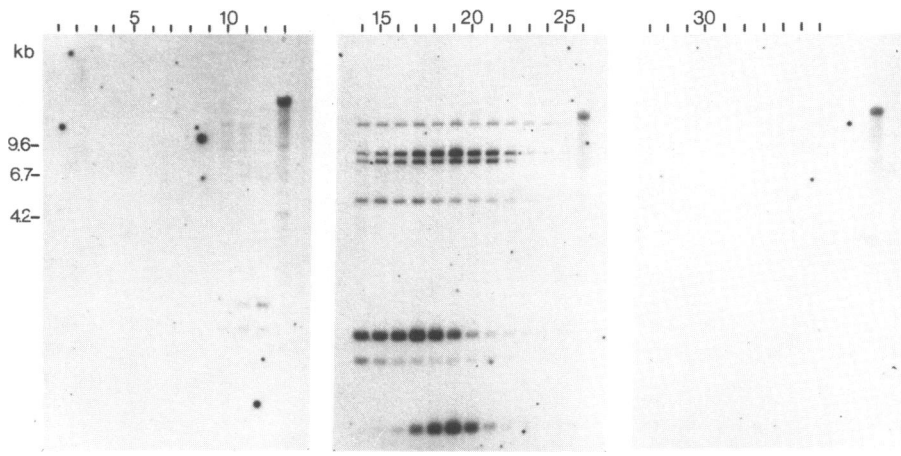

Figure 2. Gradient blots hybridized with a ribosomal RNA gene probe. Treatment of DNAs were as in Figure 1.

isolated from the chicken genome (12) was used to detect homologous sequences in P. cynomolgi DNA. Two fragments were detected. One fragment was most enriched in fraction 11 of the gradient while the other fragment was most enriched in fractions 14 and 15. The actin probe therefore identified two P. cynomolgi isochores.

Other single and low copy number genes were used as probes for the same panel of DNA fractions used in Figure 1. The data obtained indicates that there are a limited number of isochores in the genome, and that both stage-specific and housekeeping genes can be contained in a single dG•dC rich isochore. Bands that were detected by Chlamydomonas ß-tubulin (13) and a gene of P. knowlesi that is transcribed during late schizogeny (14) were both in fraction 11 (not shown), like one of the two bands detected by ß-actin. An oligonucleotide that is common to the circumsporozoite genes of P. knowlesi, P. vivax, and P. falciparum has been shown to hybridize to two bands in P. cynomolgi DNA (15-17). We found that one of these bands was contained in fraction 11 and the other in fraction 14 (not shown). This is the same pattern that was seen with a ß-actin probe. This data indicates that a relatively small number of isochores exist, each of which contain multiple genes. Genes from the same isochore may or may not be found on the same chromosome. In other eukaryotic genomes, a single isochore can contain genes from different chromosomes (5).

Figure 2 shows data resulting from hybridization of the density gradient Southern blots with a probe specific to the 3' end of the gene for the Plasmodium large ribosomal subunit (9). There are approximately seven copies of the ribosomal DNA gene in P. cynomolgi, which are interspersed among other DNA sequences (McCutchan, unpublished observation.), as is the

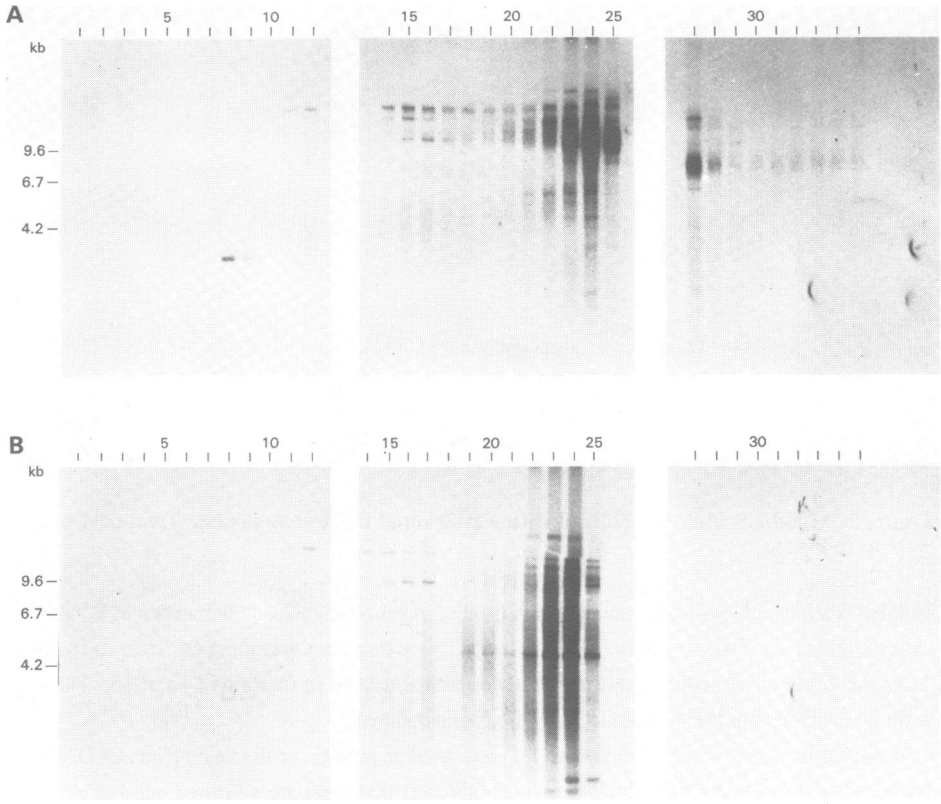

Figure 3. Gradient blots hybridized with the Plasmodium hypervariable sequence $(CCCT(A/G)AA)_3$. The source of DNA in panel A was the P. cynomolgi NIH strain and for panel B was the P. cynomolgi London strain.

case with the four P. berghei ribosomal genes (9). The individual ribosomal genes were concentrated in at least 3 different density fractions. The three bands below the 4.2 kb marker appear to be concentrated in fractions 17, 14-15, and 19, in size order from larger to smaller bands, respectively. The upper 4 bands above the 4.2 kb marker can also be localized in one of these three density fractions. Thus, the hybridization identified at least two new isochores centered in fractions 17 and 19. Hybridization signal to fractions 14-15 identified an isochore seen previously. Combined with the data in Figure 1, this identified at least 4 different isochores in the P. cynomolgi genome.

Chromosome-Internal Tandem Blocks of a Genetically Hypervariable Sequence Occur in All Isochores

The rate of evolutionary sequence change in each isochore was tested by determining the degree of restriction site polymorphism in the DNA of different parasite strains, using as a

probe a repetitive Plasmodium-specific sequence known to be unstable during meiosis (Vernick et al., manuscript submitted). We reasoned that the amount of divergence between two strains of a species for a given DNA sequence should be a function of the relative genetic stability of the particular sequences involved, as well as of the amount of time the two strains have been reproductively isolated from each other. We wanted to know if sequence environment or chromosome location was a factor in the stability of chromosomal sequences. Thus, if some gradient fractions displayed hybridization patterns conserved between P. cynomolgi strains for a sequence known to be unstable, then such density fractions could be considered to be essentially stable density domains. Significantly, it would then follow that the between-strain variability previously observed for the unstable sequence (1) had arisen by an evolutionary mechanism which restricted the relevant sequence rearrangement to certain density domains and excluded it from others.

Figure 3 shows data resulting from hybridization to gradient Southern blots by an oligonucleotide probe, $CCCT(A/G)AA_3$, consisting of three copies of the simple telomeric repeat unit from P. falciparum (1) which is conserved in P. cynomolgi (Vernick and McCutchan, unpublished observation). The stringency of hybridization and washing was such that the probe would detect no less than three tandem copies of the simple repeat. In P. falciparum, this sequence was found both at telomeres as well as internal to chromosome ends (1). Moreover, it was known that P. falciparum genomic sequences which hybridized to the probe were genetically hypervariable, rearranging at high frequency during meiosis particularly after outcrossing. The sequences were essentially stable during mitotic expansion (Vernick et al., manuscript submitted).

Hybridizations of the hypervariable probe were done to density gradient Southern blots made from DNA of three different strains of P. cynomolgi. The NIH, London, and RO strains used were from geographically isolated parasite populations which display small morphological differences and differ in the sequence of their circumsporozoite protein. The entire genomic density gradients for the NIH and London parasite strains are shown in Fig. 3A and 3B, respectively. The heavier isochores of the London and RO strains are shown in Fig. 4B and 4C, respectively. The NIH-strain hybridization shown in Figure 3A used the same density gradient Southern blots as in Figures 1 and 2, but the lambda/HindIII marker lanes 13 and 26 are not included in Figure 3A. Hybridization centered in lanes 8, 24, and 27 represented new isochores not described above, although the isochore centered in fraction 24 seems to be the same as the fractions identified by whole fraction 24 labelled DNA in Figure 1A. Hybridization signal in fractions 14-15 identified an isochore seen previously. This raised the total number of isochores to seven.

Two types of bands were seen in the hybridizations. Diffuse bands were located at ends of the chromosome and thus had a HindIII site at only one of the two termini. The diffuse appearance of the terminal telomeric bands results from variability in the number of terminal

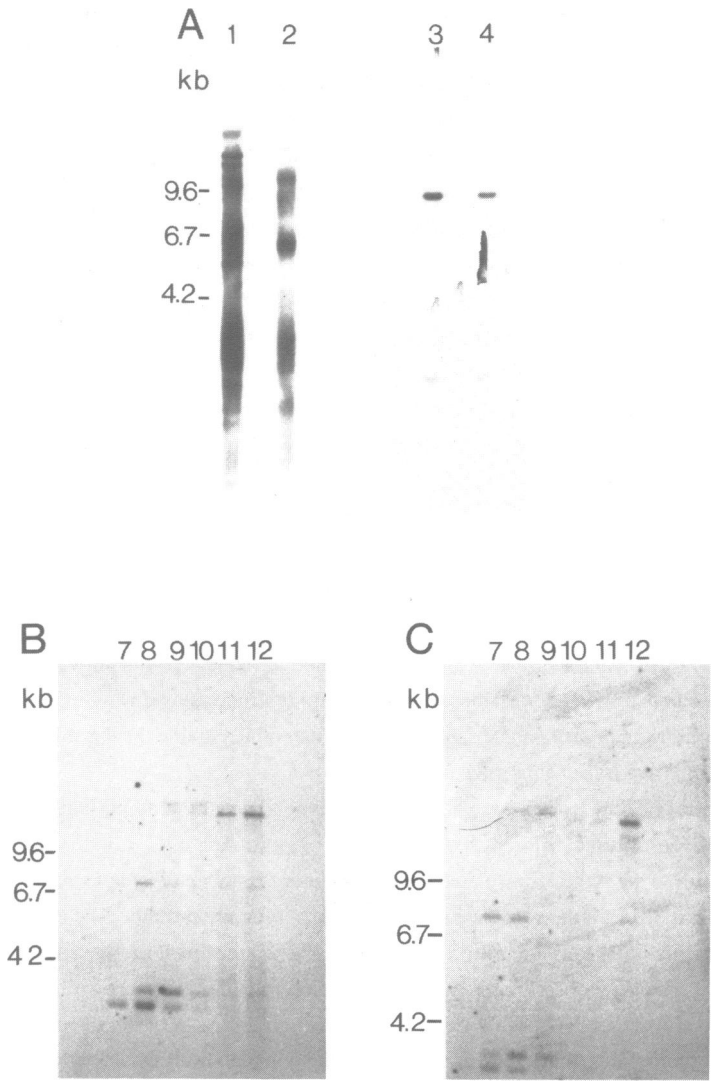

Figure 4. A. Southern blots of HindIII restricted total P. cynomolgi DNA. A single blot was hybridized to the hypervariable probe $(CCCT(A/G)AA)_3$ (lanes 1 and 2). The blot was then stripped of label and subsequently hybridized with an oligonucleotide, CCATG(T/C)ACTGAAC(T/C)TGTGGAAAGGT, which is homologous to the circumsporozoite protein of P. cynomolgi, as a control for the completeness of restriction endonuclease reactions. Lanes 1 and 3 contain DNA from the London strain while Lanes 2 and 4 contain DNA from the RO strain.
B. Longer autoradiograph exposure of gradient fractions 7-12 of the London strain DNA shown in Fig. 3.
C. A Southern blot corresponding to the DNA in gradient fractions 7-12 of a separation of RO strain DNA.

repeat units added to the ends of chromosomes (18). Concise, sharply defined bands possessed a HindIII site at both termini and thus were internal to the ends of chromosomes. We verified that the diffuse bands were Bal 31 sensitive while the sharply defined ones were Bal 31 insensitive (data not shown). Since all of the diffuse bands which hybridized to the telomeric probe were located in the lightest isochores (Fig. 3), we conclude that chromosomal telomeres were predominantly confined to these isochores.

Blocks of Genetically Hypervariable Sequence Are Evolutionarily Stable in the Heavier Isochores

It is striking that the sharply defined bands representing internal blocks of tandemly repeated sequence in the heavier isochores were evolutionarily quite stable between strains both in size and number of bands, while the internal blocks in lighter isochores were variable. For example, in DNA from the NIH strain (Fig. 3A), London strain (Fig. 3B and 4B), and RO strain (Fig. 4C), the hybridization patterns of the two heaviest isochores in fractions 8 and 11 appeared to be nearly identical. Bands in the lighter isochores (fractions 20-27) seemed to be much less related in size and number (Fig. 3A and 3B; RO strain not shown). This point is emphasized by comparing the pattern of hypervariable probe hybridization to HindIII-restricted total DNA from the RO and London strains (Fig. 4A, left panel), which show little similarity, with the pattern obtained with the same DNAs after gradient fractionation (Fig. 4B). The hybridization patterns of both strain DNAs with a control probe for the circumsporozoite protein gene were identical (Fig. 4A, right panel. Thus, there is evolutionary conservation of a portion of the blocks of a genetically hypervariable sequence in the genome, which only became apparent following density gradient separation of isochores in genomic DNA.

Reproducibility of the Gradient Separation

The reproducibility of this procedure is attested to by the data in Figures 3 and 4. Here DNA from three different members of a species complex all separate in a similar fashion. The evolutionarily conserved restriction fragments found to hybridize to the hypervariable probe are consistently found in the same part of the gradient regardless of the source of the DNA. DNA from the Gombak and Mulligan strain of P. cynomolgi and the SalI strain of P. vivax also separate on these gradients (data not shown). The question also arises as to whether the size of the genomic DNA is critical for the separation to occur. Standard handling procedures as described (8) produces DNA in the size range that is reproducibly separable. We find that this size range is between 50 and 100 kb. We have analyzed DNA as small as an average of 25 kb with no effect on the distribution of fragments hybridizing to the telomere related hypervariable probe that is described above (data not shown). The size of each fraction after separation was also determined. In some gradients the size was reduced from 100 kb to around 50 kb but this had no apparent effect on restriction analysis of the product. No fraction contained significantly smaller DNA than any other fraction.

Any reduction in size is thought to result from shear during the retrieval of DNA from the gradient but this has not been investigated.

Variability of dye binding might also affect separation profiles. The fact, however, that the concentrated dye is in contact with the DNA sample for the entire 18 hours that the gradient forms suggests that saturation of binding has occurred in each sample.

DISCUSSION

We have shown that the genome of P. cynomolgi is partitioned into regions that are each defined by a distinctive density. The density domains, or isochores, can be separated on a CsCl density gradient in the presence of the dA•dT-binding Hoechst dye. The exact size range of the different isochore has not been determined but certain information on their sizes can be gathered from this work. Isochores are likely to average 100 kb or greater since we were able to cleanly separate DNA that ranged up to this size. A uniform pattern of separation was attained with DNA that was as small as 25 kb in length indicating some homogeneity in density across an isochore. As the 14 chromosomes of P. falciparum range in size from 800 to 3500 kb in length (20), a given isochore may represent only a small portion of a single chromosome. It is likely, therefore, that members of a single isochore may be represented on a number of different chromosomes or even multiply on a single chromosome. For example, the telomeric ends of a number of different chromosomes all appeared in the lightest isochores.

We also found that the density domains differ markedly in their rates of sequence change. The pattern of restriction fragment length polymorphisms (RFLPs) identifying blocks of a tandemly repeating genetically hypervariable sequence in the heaviest isochores was highly conserved among three different members of the P. cynomolgi species complex, while the corresponding patterns for the same probe in the lightest isochores was not stable at all. Interestingly, this indicates that a genetically hypervariable sequence was not innately unstable, but instead was conditionally unstable in a way which was dependent upon its isochore location. We do not know whether the RFLP variability we saw in lighter isochores resulted from the loss of repeat units between stable flanking restriction sites, or whether there was a generalized sequence plasticity which included the restriction sites. We also do not know whether there is a physical correlation between the density of isochores and their genetic stability. However, our data clearly suggest that certain sequences may be more prone to sequence change than others simply by virtue of their isochore location in the genome, and hence their chromosomal position. An alternative explanation that could be proposed is that the evolutionary progenitor of the NIH, Ro and London strains of P. cynomolgi had a number of short tandem repeats fortuitously associated with the heaviest isochores. As these repeats would have been much shorter than the average repeat length and hence less likely to undergo unequal crossing over their appearance of stability would be

the direct result of structural arrangement as opposed to their general genomic location. This of course depends on the chance association of the heaviest isochore with short tandem repeats rather than particular genomic environments being less stable.

We envisage the Plasmodium chromosome to be composed of a patchwork of heavy and light isochores, like the vertebrate chromosome. An apparently similar form of partitioning of the DNA into different density domains occurs in the genome of warm blooded vertebrates (5). One presumes that the genome partitioning among these vertebrates has a biological meaning, since it is conserved through their evolution. Biological significance is also suggested by observations which indicate that biochemical processes involved in DNA metabolism seem to recognize isochore boundaries. For example, DNA replication in eukaryotic cells is biphasic in nature, with replication regions which are categorized as either early or late. In mammalian genomes, the Giemsa banding pattern coincides with the arrangement of early and late replicons along the chromosome (3). The early replicating DNAs are Giemsa light and dG•dC rich, while late replicating DNAs are Giemsa dark and dA•dT rich (4). Additionally, mammalian housekeeping genes reside in the dG•dC rich, early replicating DNA, while tissue specific genes and heterochromatin are in the dA•dT rich, late replicating DNA. The rate of DNA sequence change is also faster in the lighter isochores (5). Independently of isochore separation, it has also been seen that mammalian late-replicating DNA contains fragile sites which are evolutionarily quite labile, as well as being associated with breakpoints of some oncogenic chromosomal rearrangements (19).

In Plasmodium, we saw no partitioning of stage-specific genes into lighter isochores. A gene whose expression occurs in early parasite shizogony and one whose expression is limited to the sporozoite stage of the life cycle both hybridized to the heavy isochores. However, we feel that the difference in the rate of sequence change between isochores in itself strongly suggests a biological significance for the partitioning. For example, it seems likely from our data that the arrangement and environment of genes residing in the lighter isochores would be genetically variable in the population. Presumably, these would be genes for which there was some type of adaptive advantage in genetic variability, such as parasite genes which respond to the highly variable selective pressure of mammalian host immune response, or genes for which a population mosaic enhances group fitness. In fact, two nonessential genes which are thought to be involved with parasite growth rate (20) and sequestration (21) are known to be located in terminal areas of the chromosome. It is possible that the lighter isochores could serve to maintain a mosaic population with regard to a certain subset of genes while the parasite retains a stable genetic background of genes in the heavier isochores.

*To whom correspondence should be addressed

REFERENCES

1. Vernick, K.D. and McCutchan, T.F. (1987). Mol. Biochem. Parasitol. 28, 85-94.
2. Schwartz, D.C. and Cantor, C.R. (1984). Cell 37, 67-75.
3. Holmquist, G., Gray, M., Porter, T., and Jordan, J. (1982). Cell 31, 121-129.
4. Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A. and Nag, A. (1984). Science 224, 686-692.
5. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. (1985). Science 228, 953-958.
6. Dame, J.B. and McCutchan, T.F. (1987). Exp. Parasitology, in press.
7. Coatney, G.R. (1971). In: The Primate Malarias. U.S. Dept. of Health, Education and Welfare, Washington, D.C.
8. Dame, J.B. and McCutchan, T.F. (1983). Mol. Biochem. Parasitol. 8, 263-279.
9. Dame, J.B. and McCutchan, T.F. (1983). J. Biol. Chem. 258, 6984-6990.
10. McCutchan, T.F., Dame, J.B., Miller, L.H., and Barnwell, J.W. (1984). Science 225, 808-811.
11. Simpson, L. (1979). Proc. Natl. Acad. Sci. U.S.A. 76, 1585.
12. Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J., and Kirschner, M.W. (1980). Cell 20, 95-105.
13. Silflow, C.D. and Rosenbaum , J.L. (1981). Cell 24, 81-88.
14. Hansen, J.L. and McCutchan, T.F. (1986). Mol. Biochem. Parasitol. 19, 11-18.
15. Ozaki, L.S., Svec, P., Nussenzweig, R.S., Nussenzweig, V., and Godson, G.N. (1983). Cell 34, 815-822.
16. 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.S., Reddy, E.P., Diggs, C.L., and Miller, L.H. (1984). Science 225, 593-599.
17. McCutchan, T.F., Lal, A.A., de la Cruz, V.F., Miller, L.H., Maloy, W.L., Charoenvit, Y., Beaudoin, R.L., Guerry, P., Wistar, Jr., R., Hoffman, S.L., Hockmeyer, W.T., Collins, W.E., and Wirth, D. (1985). Science 230, 1381-1383.
18. Blackburn, E.H. and Szostak, J.W. (1984). Ann. Rev. Biochem. 53, 163-194.
19. Laird, C., Jaffe, E., Karpen, G., Lamb, M., and Nelson, R. (1987). Trends in Genetics 3, 274-281.
20. Wellems, T.E., Walliker, D., Smith, C.L., do Rosario, V.E., Maloy, W.L., Howard, R.J., Carter, R., and McCutchan, T.F. (1987). Cell 49, 633-642.
21. Pologe, L.H. and Ravetch, J.V. (1986). Nature 322, 474-477.