

The divergent region of the *Leishmania tarentolae* kinetoplast maxicircle DNA contains a diverse set of repetitive sequences

Michael L. Muhich, Nicolas Neckelmann* and Larry Simpson+

Department of Biology and Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received 4 February 1985; Revised and Accepted 9 April 1985

ABSTRACT

A 2.76 kb segment of the 12 kb divergent region of the *Leishmania tarentolae* kinetoplast maxicircle DNA consists almost entirely of repeated sequences. The repeats can be grouped into six families, some of which are present throughout the remainder of the divergent region. The repeats are oriented in a head-to-tail fashion with the three simplest repeats clustered into large arrays. A 47 bp palindrome and two copies of a "supercluster" of three different types of repeats are also present in the sequenced region. A sequence change in the divergent region is described for a clonal strain of *L. tarentolae* which was passaged continuously for several years. The repetitive sequences found in the divergent region appear to be appropriate substrates for the presumed deletion/insertion/recombination events occurring in this rapidly evolving portion of the maxicircle.

INTRODUCTION

The mitochondrial DNA (mtDNA) of trypanosomatid protozoa consists of a catenated network of approximately 10⁴ minicircles and 20-50 maxicircles (see references 1-3 for review) termed the kinetoplast DNA (kDNA). The function of the minicircles is not known, although there is recent evidence for possible protein coding function of putative minicircle fragments in *Crithidia* (4). The maxicircle is thought to be the informational mtDNA in these cells. Cross-hybridization studies (5, 6) have shown that the conserved sequences of maxicircle DNAs from different species and genera are organized in a basically colinear fashion over a 15-17 kb region. The conserved region of the maxicircle is extensively transcribed (7-10), and in *Leishmania tarentolae* (11-13) and *Trypanosoma brucei* (14-16) contains genes for the 9S and 12S rRNAs and open reading frames (ORFs) homologous with cytochrome oxidase subunits (CO) I and II, cytochrome b and human unidentified reading frames 4 and 5. One significant difference,

however, is the presence of an ORF homologous with the COIII gene in the L. tarentolae maxicircle and its apparent absence from the T. brucei maxicircle (11, 16). This difference may account for one of the two small interruptions of homology between the conserved regions of the L. tarentolae and T. brucei maxicircles as noted by hybridization studies (5).

Maxicircle sequence evolution occurs predominantly in the divergent region. A lack of cross-species sequence homology (5,6) and reduced (10) or undetectable levels (5) of transcription characterize this portion of the molecule. Divergent region length variation is largely responsible for the size differences observed between maxicircles of different species (6) and genera (5, 6). Length variations of up to 1.5 kb have also been found in the divergent regions of maxicircle DNAs isolated from different strains of T. brucei (17).

Rapidly evolving repetitious sequences may be a common feature of the divergent region in maxicircle DNAs from all trypanosomatid species. Heteroduplex studies suggested that tandemly arranged imperfect repeats are located within the divergent region of the T. brucei maxicircle (18). Partial denaturation studies performed on the L. tarentolae maxicircle (8) have mapped six A+T-rich segments to the divergent region. In the L. tarentolae as well as the C. oncopelti maxicircle, pairs of identically spaced restriction sites are found repeated in the divergent region (19, 20).

The rapid rate of sequence evolution and the absence of transcriptional activity suggests a function other than encoding structural genes for the maxicircle divergent region. In the C. fasciculata maxicircle, the putative origin of replication of the leading strand has been mapped to the divergent region (21). A less direct approach has also indicated that possible maxicircle origin of replication function may be encoded by the divergent regions of the L. tarentolae (22) and C. oncopelti (cited in 6) maxicircles. In both of these cases divergent region fragments were found to exhibit autonomous replicating sequence (ars) activity in yeast.

In this report we show by direct nucleotide sequence analysis that a 2.76 kb segment of the L. tarentolae maxicircle

divergent region is composed almost entirely of repetitive sequences. The repeats can be grouped into at least six different families, the complexity and organization of which varies depending on the repeat. We also present evidence of the existence of these repeats throughout the remainder of the 12 kb untranscribed divergent region.

MATERIALS AND METHODS

Cell culture and kDNA isolation

Cultures of *L. tarentolae* were grown as previously described (23). Cells were harvested at stationary phase and the kDNA recovered (24). kDNA networks were purified by banding twice in CsCl/ethidium bromide gradients. Full length, linear maxicircle DNA was released from the networks by EcoRI digestion and isolated by the CsCl/Hoechst 33258 method (25). EcoRI linearized maxicircle DNA was digested with HindIII and the products electrophoresed through 1.0% agarose. The 1846 bp and the 913 bp HindIII divergent region fragments (Lt30 and Lt54, respectively) were recovered by electroelution.

Restriction endonuclease digestions and DNA hybridizations

Restriction endonucleases were purchased from New England Biolabs and Bethesda Research Labs. Digestion conditions were those recommended by the supplier. Agarose gel electrophoresis and Southern transfer methods were as described (5). Hybridizations were carried out in 0.75 M NaCl, 0.075 M Na citrate, pH 7.2, 0.2% NaDodSO₄, 0.5mg/ml sonicated denatured salmon sperm DNA, 0.5 mg/ml poly(rA), 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin and 50% formamide (8) at 37°C for 18-36 hours. Hybridized filters were washed with multiple changes of 0.015 M NaCl, 0.0015 M Na citrate, pH 7.2 at 45°C. Probe DNAs were labeled by nick translation (26) with ³²P-dATP and ³²P-dCTP to specific activities of approximately 10⁸ cpm/ug.

DNA cloning and sequencing

The HindIII divergent region fragments Lt54 and Lt30 were cloned into the HindIII sites of the M13 vectors mp8 and mp10, respectively. BAL-31 (Bethesda Research Labs) generated M13 mp9 deletion subclones were constructed for both orientations of the Lt30 clone by the method of Poncz *et al.* (27). In the case of

the Lt54 insert, BAL-31 digestion was carried out on both termini simultaneously and the orientations of the resultant deleted inserts deduced following cloning into M13 mp10. Portions of the initial nucleotide sequence accumulated for the Lt30 fragment was obtained from subclones generated by the method of Hong (28).

DNA sequence analysis was carried out by the dideoxy chain-termination method of Sanger *et al.* (29) using the technical improvements described by Garoff and Ansorge (30) and de la Cruz *et al.* (11). Due to the highly repetitious nature of the sequence, alignments were determined manually. Computer analysis of DNA sequences was performed using the programs of Staden (31) and the Los Alamos Sequence Analysis System (32) running on the VAX 11/780, and the Pustell and Kafatos programs (33) running on the IBM PC.

RESULTS

Hybridization evidence for repetitive sequences throughout the maxicircle divergent region

To test for the presence of repetitive sequences within the *L. tarentolae* maxicircle divergent region, two ³²P-labeled cloned divergent region fragments were hybridized to blots of digested total maxicircle DNA. The blots in Fig. 1 show that both of the HindIII divergent region fragments, Lt30 and Lt54, hybridize with the same set of contiguous fragments representing approximately 10 kb of the maxicircle divergent region. The 1.5 kb HaeIII-HindIII fragment (number 8, Fig. 1) does not appear to contain sufficient sequence homology with either Lt30 or Lt54 to form stable hybrids under these stringent hybridization conditions; there is however, preliminary nucleotide sequence data (Neckelmann, de la Cruz and Simpson, unpublished results) which reveals the presence of a small number of repetitive sequences homologous with those present in Lt30 and Lt54. In addition, this fragment is also part of the untranscribed portion of the maxicircle (5).

Nucleotide sequence and analysis of a portion of the divergent region

In order to define the nature of the repetitive elements giving rise to the hybridization pattern observed in Fig. 1, the

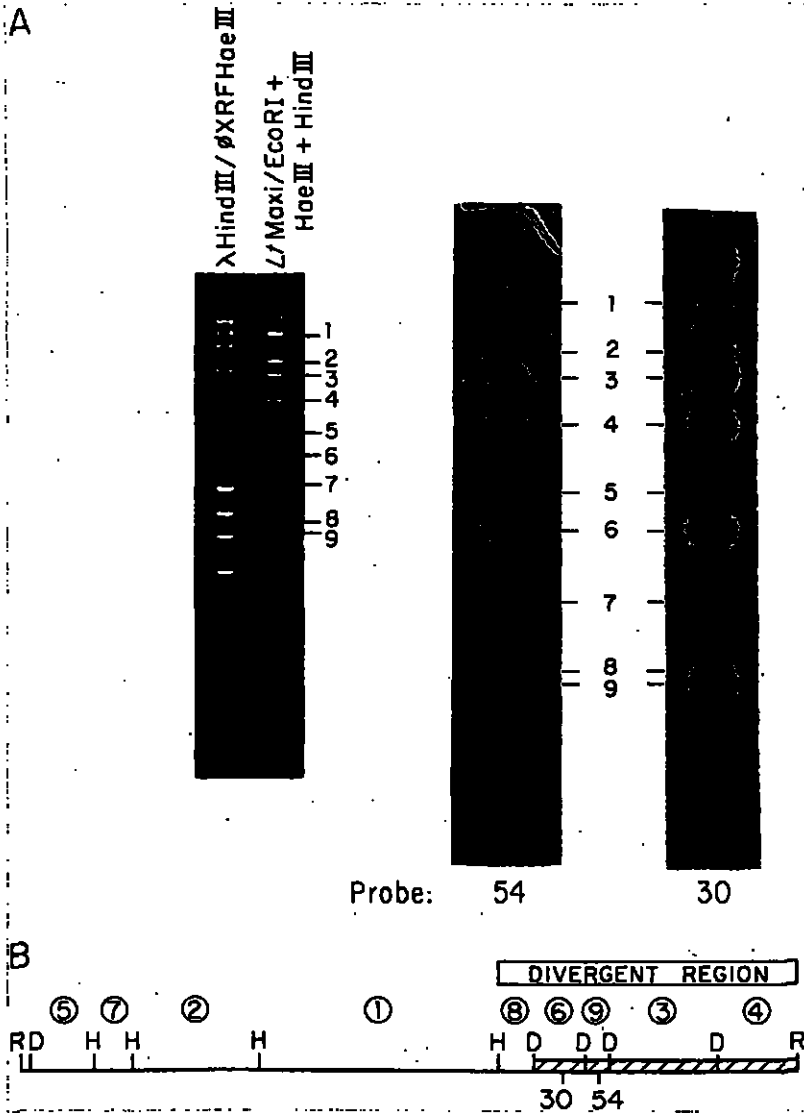


Figure 1. Evidence of a repetitive sequence element(s) dispersed throughout the divergent region of the *L. tarentolae* maxicircle. (A) Hybridization of the indicated 32 P-labeled HindIII divergent region fragments to blots of digested *L. tarentolae* maxicircle. The faint hybridization to fragments 1 and 2 is not considered significant. (B) The results are diagrammed with the hybridizing fragments indicated by hatched marks on the EcoRI linearized (30kb) maxicircle. D, HindIII; H, HaeIII; R, EcoRI. See Figs. 1 and 2 of reference 11 for the genomic organization of the *L. tarentolae* maxicircle.

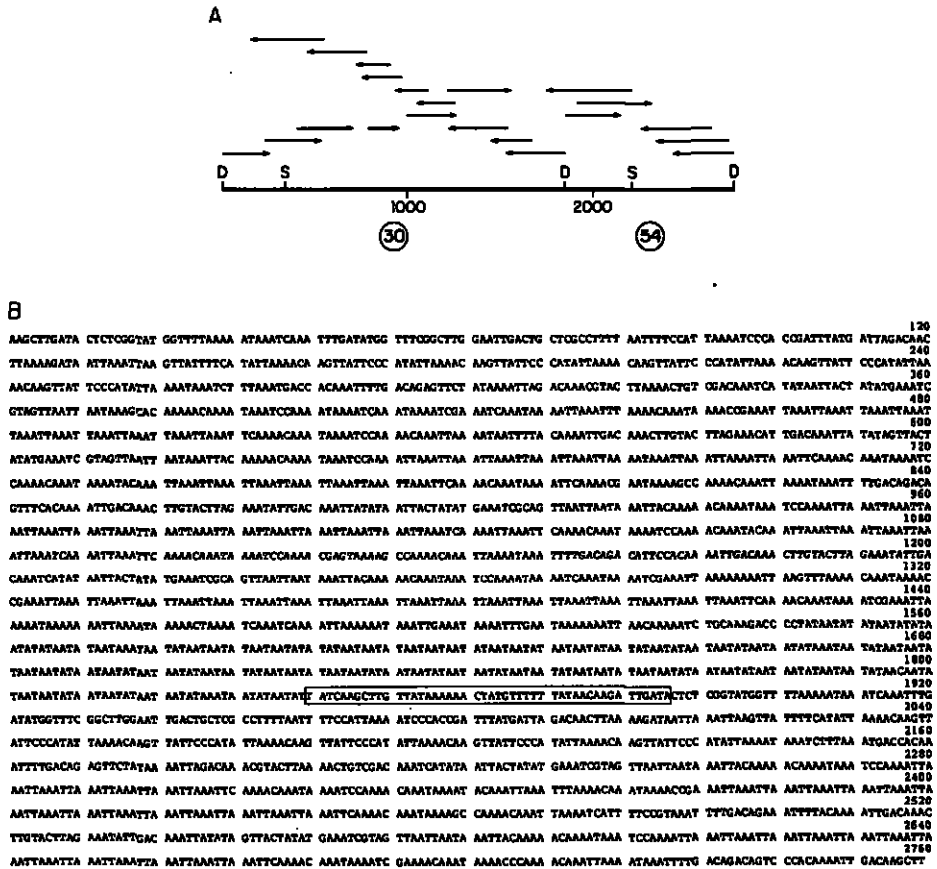


Figure 2. Nucleotide sequence of a portion of the *L. tarentolae* maxicircle divergent region. (A) the sequence was obtained from a series of overlapping BAL-31 subclones of the Lt30 and Lt54 HindIII fragments (Fig. 1) using the Sanger dideoxy-chain termination method. The relative orientation and contiguity of Lt30 and Lt54 was established by sequencing the termini of the internal 1875 bp SalI fragment. (B) nucleotide sequence of the 2759 bp which comprise the Lt30 and Lt54 fragments. The box indicates the position of a 47 nucleotide (21 bp stem) palindrome.

nucleotide sequence of the Lt30 and Lt54 HindIII divergent region fragments was determined. The complete sequence of the Lt30 and Lt54 fragments as well as the sequencing strategy employed are shown in Fig. 2. Nucleotide position number one is downstream of the HURF5 gene of the conserved, transcribed portion of the

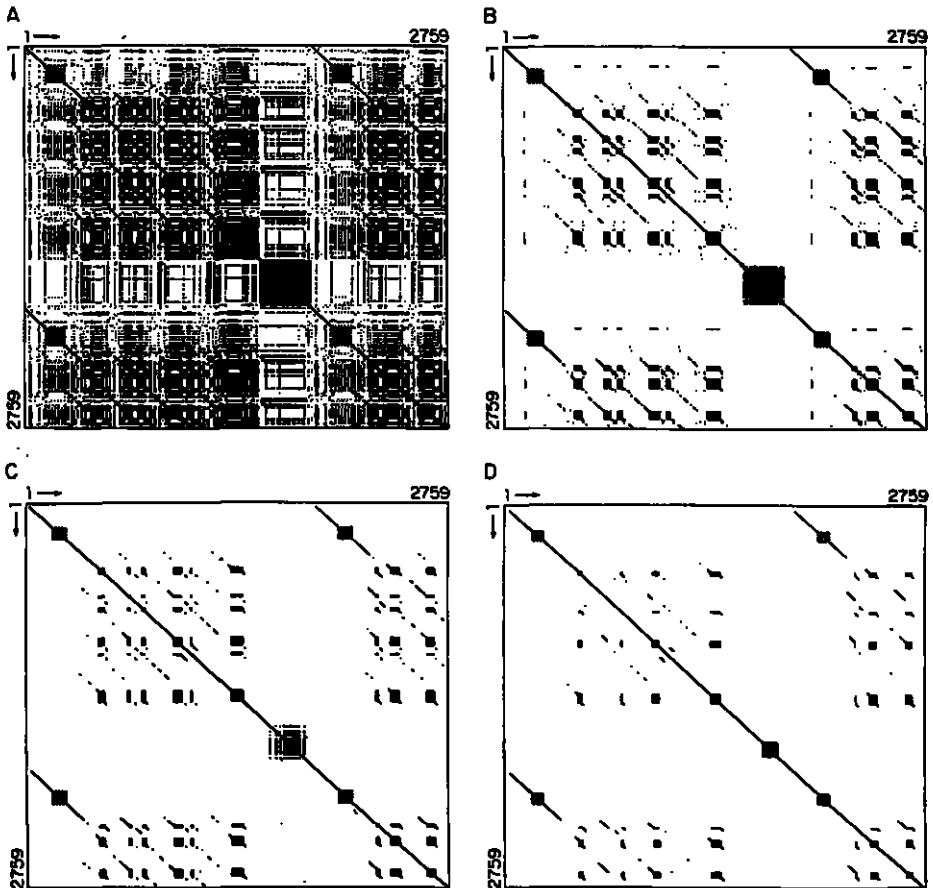


Figure 3. Evidence of the clustered organization of the simple repeats within the *L. tarentolae* maxicircle divergent region. Staden DIAGON analysis of the Lt30-54 divergent region sequence. The 2759 nucleotides was run versus itself using the perfect match algorithm with span lengths (L) and minimal scores (S) of (A) L=5, S=5; (B) L=11, S=11; (C) L=21, S=21; (D) L=35, S=35.

maxicircle, and nucleotide number 2759 is upstream of the 12S rRNA gene (11).

A DIAGON (31) dot matrix homology search of the Lt30-54 divergent region sequence run versus itself is shown in Fig. 3. The box structures residing on and distributed symmetrically about the line of identity are indicative of a head to tail clustered organization of repetitive sequences. As the

Table 1. Consensus Nucleotide Sequences and Frequency of Appearance of Repeating Elements in the Lt30-54 Segment of the *L. tarentolae* Maxicircle Divergent Region.

Repeat	Consensus Nucleotide Sequence	Reiteration Number
A	AATAATAT	29
B	AAATT	103 ^c
C	33-239 nt ^a	10
D	149 nt ^b	2
E	ATATTAACAAGTTATCCC	10
F	ACAAATTTGACAGAGTTCTATAAA ATTAGACAAACGTACTTAAACTGTC	2

- a. See Fig. 5 for nucleotide sequences
 b. nt 1-149 and 1877-2026, Fig. 2
 c. Number of repeating elements present in clusters of five or more contiguous units

stringency of the homology search is increased, the tandem head to tail arrangement of the type-A, -B and -E (Table 1 and Fig. 4) repeat family monomeric units becomes apparent. The clustered arrays of repeats account for approximately 38% of the Lt30-54 sequence. The remainder of the 2759 bp is composed almost entirely of repeated sequences of greater sequence complexity and dispersed organization. The distributions of the various repeat families are presented in Fig. 4. Table 1 shows the consensus nucleotide sequences and the frequency of appearance of the repetitive elements in this portion of the divergent region.

The monomeric unit of the type-B repeat, which is the simplest repetitive element identified, is AAATT. This sequence is found clustered in nine head to tail arrays ranging in size from five to nineteen monomeric units (Fig. 4B). The eleven unit cluster at nucleotide position (nt), 649-704 contains a small degree of degeneracy within two of the repetitive units. The nine type-B repeat clusters give rise to the box structures representing the majority of the pattern symmetry in the dot matrix in Fig. 3, as can be seen by comparing the positions of the type-B clusters on the line of identity in Fig. 3 with their positions in Fig. 4. Approximately 32% (49 of 152), of the total type-B monomeric units are found dispersed throughout the Lt30-54

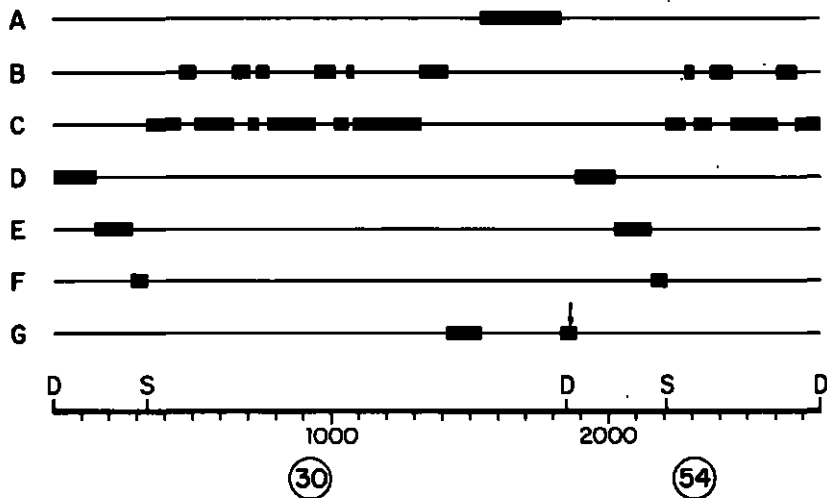


Figure 4. Diagrammatic representation of the distribution of the various types, -A, -B, -C, -D, -E, -F and -G, of repetitive sequences comprising the Lt30 and Lt54 fragments of the *L. tarentolae* maxicircle divergent region. The boxes in G represent unique sequences in that they are each present only once in the entire 2759 bp. The arrow indicates the two fold axis of symmetry of the 47 nucleotide palindrome (Fig. 2).

sequence as components of other repeat families. These dispersed type-B repeats are not included in Fig. 4 in order to simplify the presentation.

The remaining simple, pure A+T repeat, AATAATAT (Fig. 4A), is present in a large head to tail cluster at nt 1543-1825. The sequence integrity of the monomeric unit degenerates at the Lt54 distal terminus of the cluster and a single G:C base pair is found at position 1796. There are no additional large clusters of this type-A repeat present within this portion of the divergent region; however, at least one additional cluster is found outside of this region in the 1.5 kb *Hae*III-*Hind*III fragment (fragment 8 in Fig. 1) (Neckelman, de la Cruz and Simpson, unpublished results).

The type-E repeat is the most complex clustered repeat found. The 21 bp monomeric unit shown in Table 1E is organized into a head to tail cluster of five perfect units (nt 150-254 and 2026-2130) followed by a single degenerative repeat (nt 255-280

and 2131-2156). Surrounding the type-E clusters in the same relative positions are the 149 bp type-D repeats (nt 1-149 and 1877-2025) and the 51 bp type-F repeats (nt 281-331 and 2157-2207). These two repeats, types-D and -F, like the type-E cluster are present only once in each of the two HindIII fragments sequenced. Interestingly, these repeats and portions of two C family repeats appear to be organized into a "supercluster" which is itself repeated at nt 1-402 and 1877-2278.

The type-C sequences (Fig. 4C) represent the most complex family of repeats. Members of this family range in size from 33 to 239 bp (Fig. 5), and are found interspersed between the type-B family clusters and adjacent to the two type-F repeats. The organization of these type-C sequences consists of homologous sequence blocks of variable sizes which are found repeated in different arrangements in the 10 members of this family, Fig. 4C and Fig. 5.

There are two stretches of divergent region sequence (nt 1418-1542 and 1826-1876), which are unique in the Lt30-54 segment. For convenience they are listed as the type-G family in Fig. 4. These two sequences bear no significant homology with the various repeat families or with each other. The 51 bp sequence (nt 1826-1876) contributes to a 47 bp palindrome that can be drawn with its dyad axis at position 1863 (Fig. 2) and which includes 10 bp of the adjacent type-D repeat. Since the nucleotide sequence of the entire divergent region is not known, it is possible that both the 125 bp and the 51 bp sequences are not unique and that they are repeated elsewhere in the divergent region (Fig. 8).

Computer translation of the Lt30-54 sequence was carried out using a modified universal genetic code in which TGA codes for tryptophan (11). Numerous ORFs were found, of which the longest was 164 amino acids, located at nt 860-1354 on the reverse complement of the sequence presented in Fig. 2. This sequence encompasses the entire type-A repeat cluster and as a result the ORF is composed predominantly of isoleucine, leucine and tyrosine residues. Considering the apparent lack of transcription of the *L. tarentolae* divergent region (5) and the repetitious nature of the sequence presented here, it seems unlikely that the ORFs

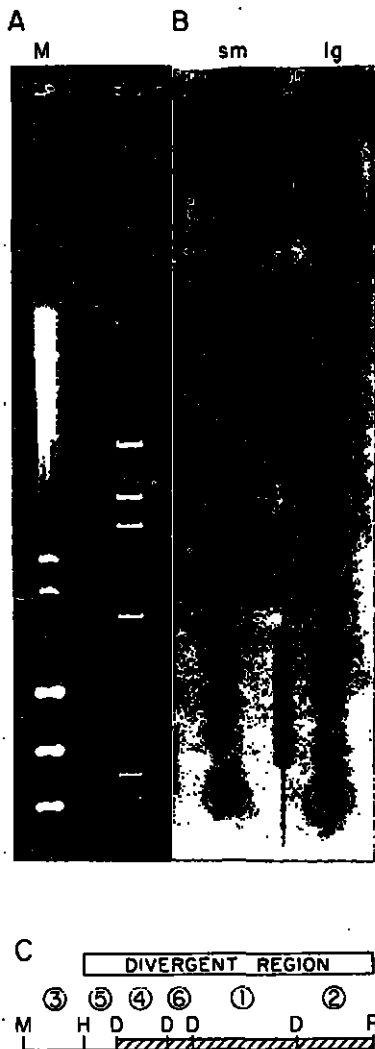


Figure 6. Distribution of repetitive sequences throughout the maxicircle divergent region. (A) The 14.5 kb EcoRI-MspI *L. tarentolae* maxicircle fragment was digested with HaeIII and HindIII and electrophoresed through 1.2% agarose. (B) Hybridization of the 361 bp (sm) and the 552 bp (lg) ³²P-labeled HindIII-SalI subfragments of Lt54 to blots of the digests in (A). (C) The hybridizing fragments in (B) are indicated by hatched marks on the diagram of the 14.5 kb EcoRI-MspI *L. tarentolae* maxicircle fragment. The apparent low level hybridization to fragment 3 is due to a slightly larger partial digestion product composed of fragments 4 and 5. D, HindIII; H, HaeIII; M, MspI; R, EcoRI. M, marker DNAs: HindIII digested lambda DNA and HaeIII digested ϕ X174 RF DNA.

encoded by the Lt30-54 segment specify proteins *in vivo*.

Distribution of repeat families throughout the entire divergent region

At least five types of repeats, type-B, -C and the supercluster of repeat types-D,-E and-F, exist within both the Lt30 and Lt54 HindIII fragments of the divergent region. Furthermore, the hybridization results in Fig. 1 indicate that the two HindIII+EcoRI fragments adjacent to Lt54 in the divergent region also contain at least one of these five repeat families. The distribution of divergent region sequences homologous with the type-D-E-F supercluster sequences and the type-B and type-C repeats is presented in Fig. 6. The 361 bp HindIII-SalI restriction fragment of the cloned Lt54 insert (Fig. 4), which contains the type-D-E-F supercluster, and the 552 bp HindIII-SalI Lt54 fragment, which contains type-B and type-C repeat sequences, were used to probe digests of the 14.5 kb EcoRI-MspI maxicircle fragment which encompasses almost the entire divergent region. Each probe hybridized to the same four bands, representing approximately 10 kb of the divergent region, indicating that the two largest divergent region digest products (fragments 1 and 2, Fig. 6C) contain at least some portion of the D-E-F supercluster and at least some portion of either the type-B and/or type-C repeat sequences.

Sequence changes within the divergent region can be seen over an 11 year period

kDNA network preparations from an initially cloned strain of *L. tarentolae* were available from cultures passaged continuously over an 11 year period (1973-1984). The cells, which divide every 9-12 hr, were subcultured every 4-7 days, and viable samples were frozen in liquid nitrogen periodically for later retrieval. The hybridization of a ³²P-labeled Lt30 fragment with a blot of SalI digested kDNA networks which had been isolated from cells of the indicated year is shown in Fig. 7. The 1973 profile is comprised of the four expected SalI fragments (19), all of which contain sequences homologous with the divergent region probe. The products of the 1984 digest include in addition to the four expected SalI fragments, a 1.7 kb fragment (arrow, Fig.7B) which also hybridizes with the divergent

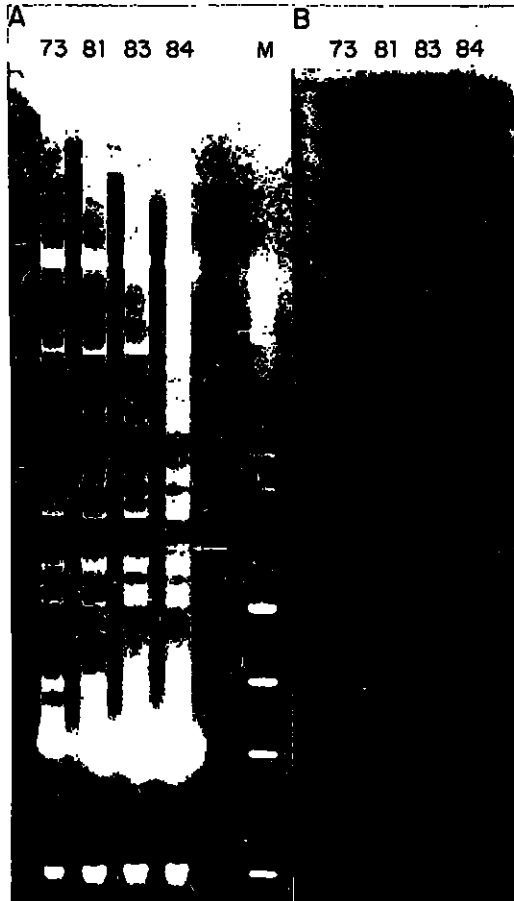


Figure 7. Sequence changes within the maxicircle divergent region occurring within an eleven year culture period. (A) Approximately 4.5 ug of kDNA networks from *L. tarentolae* cells of the indicated year were digested to completion with SaliI and electrophoresed through 1.2% agarose. (B) Hybridization of ³²p-labeled Lt30 probe with the blot of the digests in (A). The arrows indicate the position of the recently appearing SaliI restriction fragment. M, marker DNAs: HindIII digested lambda DNA and HaeIII digested ϕ X174 RF DNA.

region probe. The low relative autoradiographic signal intensity of the 1.7 kb band suggests that a reduced number of divergent region homologous sequences are available for hybridization with the probe. Examination of the ethidium bromide-stained kDNA digests indicates that this fragment is at a significantly lower concentration in the kDNA network isolates relative to the other

```

***** C-----B-----
GAATTC AAAA CAAATAAAAC CCAAACAAA TAAATCCAA ATTA AATTAA ATTA AATTAA 60
***** C-----
ATTA AATTAA ATTA AAAACA AATT TGGTGTG ACACAAC TTT TAACAATTG TTT CATATAT 120

AATCTAAAT TTTGTTGTG CTTGTTTTAT ACGAAAAGT TTGTAATATT ATTTTACTT 180

TTAAGTAAA GTTTAATTTA TGTTTTAATT TTATTTATAA ACTACAATCA ATAAAATATA 240
***** A-----
GAAANGCTTA TTTATGATAT AATACAATAC CACATGATTA TAAAAATATT AATTATAATA 300
***** G-----
ATATTTACAC TATTTATTA AAGTTAGTAA AAAGGAAATA GAAATAAATA ATCATACTCC 360

ATATTTATTA ATTATTTTTC TACACTTGTA AAAATATAGA ATATATTTAT TTTTATAGTG 420

TAATTATAGT ATTAAAA

```

Figure 8. Nucleotide sequence of the 437 bp portion of the *L. tarentolae* maxicircle extending from the unique EcoRI site to the 5' terminus of the 12S rRNA gene (13). Nucleotide identities with the Lt30-54 divergent region sequence are indicated by asterisks (*). One character pads were inserted at positions 285 and 292 to achieve maximum alignment of homologous sequences. A, B, C and G indicate the type of repeats (Table 1) with which the bracketed sequences are homologous.

four Sali cleavage products (arrow, Fig. 7A). Sali digestion of isolated EcoRI linearized maxicircle DNA indicates that this 1.7 kb fragment is a component of the maxicircle genome (data not shown) and not a minicircle catenane.

The 1.7 kb divergent region-homologous Sali fragment may have arisen through either the appearance of a new Sali recognition site or by a sequence change of a larger magnitude. The appearance of an additional Sali site within any of the three largest Sali cleavage products (22.5 kb, 4.2 kb, 1.9 kb and 1.4 kb) could yield a 1.7 kb fragment upon digestion. However, the cognate Sali fragment (20.8 kb, 2.5 kb or 0.2 kb) expected to be generated by cleavage at the newly appearing site has never been observed; hybridization of ^{32}P -labeled Lt30 probe with blots of Sali digested 1973 and 1984 kDNAs which had been electrophoresed through 2% agarose failed to detect any of the expected cognate bands (data not shown).

The autoradiograph in Fig. 7 reveals that both the 1981 and 1983 digests contain the 1.7 kb divergent region-homologous

band, but at significantly lower concentrations as compared with the 1984 results. Heterogeneity within the cell population or at the level of the kDNA networks is the most likely cause of the nonstoichiometric level of the 1.7 kb band and the increase in the relative copy number of this sequence as a function of time in culture.

Divergent and conserved region junction

The highly conserved sequence of the 12S rRNA gene defines the terminus of one end of the divergent region. The 437 bp extending upstream of the 12S rRNA gene to the unique EcoRI site hybridizes with divergent region probes only under non-stringent conditions (data not shown). This region lacks detectable transcriptional activity (8) and contains open reading frames no larger than 52 predicted amino acids. The results of a search of this sequence (previously published, 13) for homologies with the sequence of the Lt30 and Lt54 divergent region fragments are presented in Fig. 8. The most extensive homology, nt 2-83, is to sequences of the type-B and type-C repeat families. A cluster of seven type-B repeats is found abutting a type-C repeat sequence in a fashion similar to that observed in both the Lt30 and Lt54 fragments. At positions 253-304 a degenerative homology exists with the type-A repeat cluster of the Lt30 fragment. Interestingly, 18 of the 21 nucleotides at positions 328-348 are homologous with one of the "unique" regions (nt 1483-1503) of the 30 fragment, implying that this sequence may also be repeated in the divergent region.

DISCUSSION

Our results clearly demonstrate the existence of repeated sequences throughout the entire 12 kb *L. tarentolae* maxicircle divergent region. Nucleotide sequence analysis of the two divergent region fragments, Lt30 and Lt54, shows that a diverse group of repeated sequences occupy this portion of the divergent region. The complexity of these repeats ranges from 5 to 239 bp. All of the repeats were found oriented head to tail and to be organized in either a clustered (types-A, -B and -E) or a dispersive (types-C, -D, -F and -G) fashion. The approximately 360 bp D-E-F superclusters are bracketed at one end by a HindIII site

and at the other end by a Sali I site. The restriction map of the *L. tarentolae* maxicircle (19) shows two additional, similarly spaced and oriented, HindIII-Sali pairs located between the Lt30-54 segment and the 12S rRNA gene. Based on the results presented here, it seems likely that these two (12S gene proximal) HindIII-Sali fragments also contain D-E-F superclusters with the same sequence polarity as those in the Lt30-54 region. A further expectation from these data is the association of the D-E-F supercluster with the type-B and the complex type-C repeats throughout the maxicircle divergent region. Our results also suggest that the six denaturation bubbles (A+T-rich regions) previously identified in the *L. tarentolae* divergent region (8) probably correspond to large clusters of type-A or -B or some other high A+T repeat. Analysis of the published 437 bp nucleotide sequence directly upstream of the 12S rRNA gene (13) also revealed the presence of sequences with homologies to repeat types-A, -B, -C and -G. Furthermore, preliminary sequence data from the portion of the divergent region upstream of the Lt30 fragment (fragment no. 8, Fig. 1) indicates the presence of repeats of the types-A, -B and -G (Neckelmann, de la Cruz and Simpson, unpublished data).

The extensive variation among trypanosomatid protozoa of divergent region size and sequence and the lack of appreciable homology with discrete, stable RNAs makes it unlikely that the divergent region encodes structural genes. Borst *et al.* (17) have suggested, on the basis of studies of divergent region size variation in the maxicircle DNAs of different strains of *T. brucei*, that the divergent region may be analogous to the A+T-rich untranscribed region of *Drosophila* mtDNA that is known to contain an origin of replication. While the *T. brucei* divergent region is apparently no more A+T-rich than the rest of the maxicircle (18), the *L. tarentolae* divergent region does contain long stretches of pure, or nearly pure, A+T sequence. Functional proof of a divergent region localized origin of replication is lacking; however, the A+T-rich regions and direct repeats of the Lt30-54 segment are consistent with features common to many bacterial and phage replication origins (34). Of further interest is the presence of a 47 bp palindrome (nt 1840-1886) which

overlaps the 30 and 54 fragments in the divergent region. In addition we previously showed that the Lt30 fragment exhibits ars activity in yeast (22); the Lt54 fragment, however, has not yet been tested. Hajduk *et al.* (21) have also presented evidence suggesting that the origin of leading strand synthesis in the maxicircle of *C. fasciculata* is located in the divergent region.

The existence of rapidly evolving repetitive sequences within the divergent region could account for the lack of cross-species sequence homology as well as for the great variation in size of this portion of the maxicircle. The type-C repeats, which are the most complex family of repeats identified within the Lt30-54 segment, may represent an example of extensive sequence variation occurring within a *L. tarentolae* divergent region repeat family. Individual members contain blocks of sequence with varying degrees of homology to sequences of other members of the family. The different juxtapositions of these homologous sequence blocks among the various family members could be explained by deletion or duplication events occurring as the result of the misalignment of repetitive sequences, leading to an unequal recombinational exchange (35, 18). The divergent region sequence change we have identified in the *L. tarentolae* maxicircle (Fig. 7) may represent the product of such an unequal exchange.

In conclusion, although the details of the mechanism responsible for the rapid rate of maxicircle divergent region sequence evolution remain unclear, numerous data suggest that repetitive sequences are involved in the process.

ACKNOWLEDGEMENTS

We thank Dallas Hughes for the initial cloning of the Lt54 fragment and Vidal de la Cruz for helpful discussions. This work was supported by grant AI-09102 from the National Institutes of Health to L.S.. M.L.M. was the recipient of a U.S. Public Health Service National Research Service Award, GM-07104.

* Current address: Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322

+ To whom reprint requests should be addressed

REFERENCES

1. Simpson, L. (1972) *Int. Rev. Cytol.*, **32**, 139-207.
2. Borst, P. and Hoeijmakers, J. (1979) *Plasmid*, **2**, 20-40.
3. Englund, P., Hajduk, S. and Marini, J. (1982) *Ann. Rev. Biochem.*, **51**, 695-726.
4. Shlomai, J. and Zadok, A. (1984) *Nucl. Acids Res.*, **12**, 8017-8028.
5. Muhich, M., Simpson, L. and Simpson, A.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4060-4064.
6. Maslov, D., Kolesnikov, A. and Zaitseva, G. (1984) *Mol. Biochem. Parasitol.*, **12**, 351-364.
7. Hoeijmakers, J., Snijders, A., Janssen, J. and Borst, P. (1981) *Plasmid*, **5**, 329-350.
8. Simpson, L., Simpson, A. and Livingston, L. (1982) *Mol. Biochem. Parasitol.*, **6**, 237-257.
9. Hoeijmakers, J., Schoutsen, B. and Borst, P. (1982) *Plasmid*, **7**, 199-209.
10. Stuart, K. and Gelvin, S. (1982) *Mol. Cell. Biol.*, **2**, 845-852.
11. de la Cruz, V., Neckelmann, N. and Simpson, L. (1984) *J. Biol. Chem.*, **259**, 15136-15147.
12. de la Cruz, V. F., Lake, J. A., Simpson, A.M. and Simpson, L. (1985a) *Proc. Natl. Acad. Sci. USA*, **82**, 1401-1405.
13. de la Cruz, V., Simpson, A., Lake, J. and Simpson, L. (1985) *Nucl. Acids Res.*, **in press**.
14. Eperon, I., Janssen, J., Hoeijmakers, J. and Borst, P. (1983) *Nucl. Acids Res.*, **11**, 105-125.
15. Benne, R., DeVries, B., Van den Burg, J. and Klaver, B. (1983) *Nucl. Acids Res.*, **11**, 6925-6941.
16. Hensgens, A., Brakenhoff, J., De Vries, B., Sloof, P., Tromp, M., Van Boom, J. and Benne, R. (1984) *Nucl. Acids Res.*, **12**, 7327-7344.
17. Borst, P., Fase-Fowler, F., Hoeijmakers, J. and Frasch, A. (1980) *Biochim. Biophys. Acta.*, **610**, 197-210.
18. Borst, P., Weijers, P. and Brakenhoff, B. (1982) *Biochim. Biophys. Acta*, **699**, 272-280.
19. Masuda, H., Simpson, L., Rosenblatt, H. and Simpson, A.M. (1979) *Gene*, **6**, 51-73.
20. Maslov, D. A., Entelis, N. S., Kolesnikov, A. A., Zaitseva, G. N. c Kouzmin, N. P. and Fodor, I. (1981) *Dokl. Akad. Nauk S. S. R.*, **261**, 1271-1273.
21. Hajduk, S., Klein, V. and Englund, P. (1984) *Cell*, **36**, 463-492.
22. Hughes, D., Simpson, L., Kayne, P. and Neckelmann, N. (1984) *Mol. Biochem. Parasitol.*, **13**, 263-275.
23. Simpson, L. and da Silva, A. (1971) *J. Mol. Biol.*, **56**, 443-473.
24. Simpson, A.M. and Simpson, L. (1980) *Mol. Biochem. Parasitol.*, **2**, 93-108.
25. Simpson, L. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1585-1588.
26. Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.*, **113**, 234-251.
27. Poncz, M., Solowiejczyk, D., Ballatine, M., Schwartz, E. and Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4298-4302.
28. Hong, G.F. (1982) *J. Mol. Biol.*, **158**, 539-549.

29. Sanger, F. Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
30. Garoff, H. and Ansorge, W. (1981) Anal. Biochem., 115, 450-457.
31. Staden, R. (1982) Nucl. Acids Res., 10, 2951-2961.
32. Kanehisa, M.I. (1982) Nucl. Acids Res., 10, 183-196.
33. Pustell, J. and Kafatos, F.C. (1984) Nucl. Acids Res., 12, 643-655.
34. Vocke, C. and Bastia, D. (1983) Proc. Natl. Acad. Sci. USA, 80, 6557-6561.
35. Smith, G. (1973) Cold Spring Harbor Symp. Quant. Biol., 37, 507-513.