ELSEVIER

Molecular & Biochemical Parasitology 113 (2001) 315–321

MOLECULAR & BIOCHEMICAL PARASITOLOGY

www.parasitology-online.com.

Short Communication

Conservation of the LD1 region in *Leishmania* includes DNA implicated in LD1 amplification☆

Susan M. Sunkin [a], Paul McDonagh [b], Mark L. Cunningham [c], Stephen M. Beverley [c], Kenneth Stuart [a,d], Peter J. Myler [a,d,*]

[a] *Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109-1651, USA*
[b] *Rosetta Inpharmatics, 115th Ave NE, Kirkland, WA 98034, USA*
[c] *Department of Molecular Microbiology, 760 McDonnell Science Building Box 8230, Washington University School of Medicine, 660 Euclid Avenue, St. Louis, MO 63110-1093, USA*
[d] *Department of Pathobiology, University of Washington, Seattle, WA 98195, USA*

Received 1 November 2000; received in revised form 11 January 2001

Keywords: *Leishmania major* Friedlin; Amplified DNA; *Leishmania* DNA 1 (LD1); Genomic organization; Sequence comparison

DNA amplification is ubiquitous in the biological world [1]: programmed amplification of genes occurs during development [2], and selection for DNA amplification can underlie cancerous growth [3], as well as drug resistance [4,5]. In *Leishmania*, protozoan parasites of humans, the frequency of DNA amplification events gave rise to the concept of genome plasticity [6–9]. Selection, for instance, enriches for *Leishmania* with amplifications of genes that encode drug targets, transporters or metabolic enzymes, such as R DNA (*DHFR-TS*), G DNA (*NAGT*), and H DNA (*PTR1* and *PGPA*) [6,8,10,11]. However, DNA amplifications have been found even in *Leishmania* strains that do not appear to have undergone selection [12–14]. The products of amplification in *Leishmania* are generally small episomal linear or circular molecules, often organized as inverted repeat dimers [6–9,11], however, the mechanism by which they are formed remains unclear.

One genomic region in *Leishmania* that is prone to numerous types of amplification is *Leishmania* DNA 1 (LD1). LD1 has been defined as the 27.2-kb DNA sequence derived from Chr35 which occurs as an inverted repeat on the multicopy 54.5-kb circular episomal element in *L. infantum* MHOM/BL/67/ITMAP263clone10 (LSB-7.1) [15] and other strains [14,16,17]. Comparative studies have identified stable linear minichromosomes ranging in size from 180 to 550 kb in *L. major*, *L. donovani*, *L. braziliensis*, and *L. mexicana*, that contain inverted repeats of LD1 [9,14,16,18–21].

LD1 amplification in the absence of intentional selection is quite frequent: circular or linear LD1 amplicons have been detected and stably maintained in ~15% of all *Leishmania* isolates tested [9,14]. The prevalence of LD1 amplification in *Leishmania* suggests that strains containing these amplified sequences have a selective advantage. Indeed, it has been recently demonstrated that the amplification and consequent overexpression of one of the genes encoded by LD1, *BT1*, confers a significant growth advantage in both naturally isolated and recombinant cell lines [22,23]. Also consistent with selection for LD1 amplification is the observation that in *L. tarentolae*, minichromosomes containing LD1 can appear after nutrient stress or subcloning cultures [19].

Abbreviations: bp, base pair; Chr1, chromosome 1; Chr3, chromosome 3; Chr27, chromosome 27; Chr35, chromosome 35; kb, kilobase pair; LdoK, *L. donovani* MHOM/SD/00/Khartoum; LinI, *L. infantum* MHOM/BL/67/ITMAP263clone10; LmjF, *Leishmania major* MHOM/IL/81/Friedlin; *NIMA*, never-in-mitosis gene A.

☆ *Note*: Nucleotide sequence data reported in this paper are available in GenBank™ under the accession numbers AC005801 (LmjF cosmid 511), AC005804 (LmjF cosmid 611), L27052 (LinI), U35460 (LinI), L25643 (LinI), U35461 (LdoK), U02459 (LdoK), and L38571 (LdoK)

* Corresponding author. Tel.: +1-206-2848846; fax: +1-206-2840313.

E-mail address: mylerpj@sbri.org (P.J. Myler).

PII: S0166-6851(01)00215-8

As part of our effort to understand the biological significance of LD1 amplification, we have undertaken a detailed comparison of this locus in *L. donovani* LSB-51.1 (LdoK), *L. infantum* LSB-7.1 (LinI), and *L. major* MHOM/IL/81/Friedlin (LSB-132.1) (LmjF), the strain chosen for the *Leishmania* genome sequencing project. We previously reported the cloning and sequence analysis of ~16 kb of LD1 from the circular amplified element in LinI [15,24–26]. We show in this study that there is a high degree of sequence conservation in the protein-coding (91–96%) and non-coding (79–85%) regions of LD1 and that sequences previously implicated in LD1 amplification [9,27] are highly conserved.

For this study, we completed the sequence of the circular LD1 molecule from LinI and sequenced selected portions of LD1 and flanking regions from LdoK (Fig. 1). DNA sequence analysis revealed that the LD1 region (both coding and non-coding) from LdoK and LinI are >99% identical and that the majority of nucleotide differences in the coding regions occur as silent changes. This is similar to the nucleotide identity seen between genes in *L. donovani* and *L. infantum* outside the LD1 locus, including genes encoding *S*-adenosylmethionine decarboxylase (*ADOMETDC*), putative mitogen-activated protein MAP kinase (*MKK*), DNA polymerase α (*POLA*) [28], *S*-adenosylmethionine synthase (*METK*), and heat shock protein 70 (*HSP70*) [29].

As part of the *Leishmania* Genome Network, we are currently sequencing LmjF Chr35, from one end of which LD1 is derived (Fig. 1) [8]. From our analyses of Chr35, we were able to obtain the sequence of the LD1 region by shotgun sequencing two cosmids (511 and 611) identified from a *L. major* MHOM/IL/81/FriedlinVI cLHYG library [30] by hybridization with the *BT1* gene [25,31]. The sequences obtained from both cosmids were assembled into a consensus sequence of 67,038-bp containing LD1 and its flanking region (Fig. 1), which was validated by Southern analysis (data not shown). Analysis of the sequence from cosmid 511 and 611 with the GCG program TESTCODE [32] and the gene-finding program Glimmer [33] revealed 21 putative protein-coding genes. All of the 21 putative

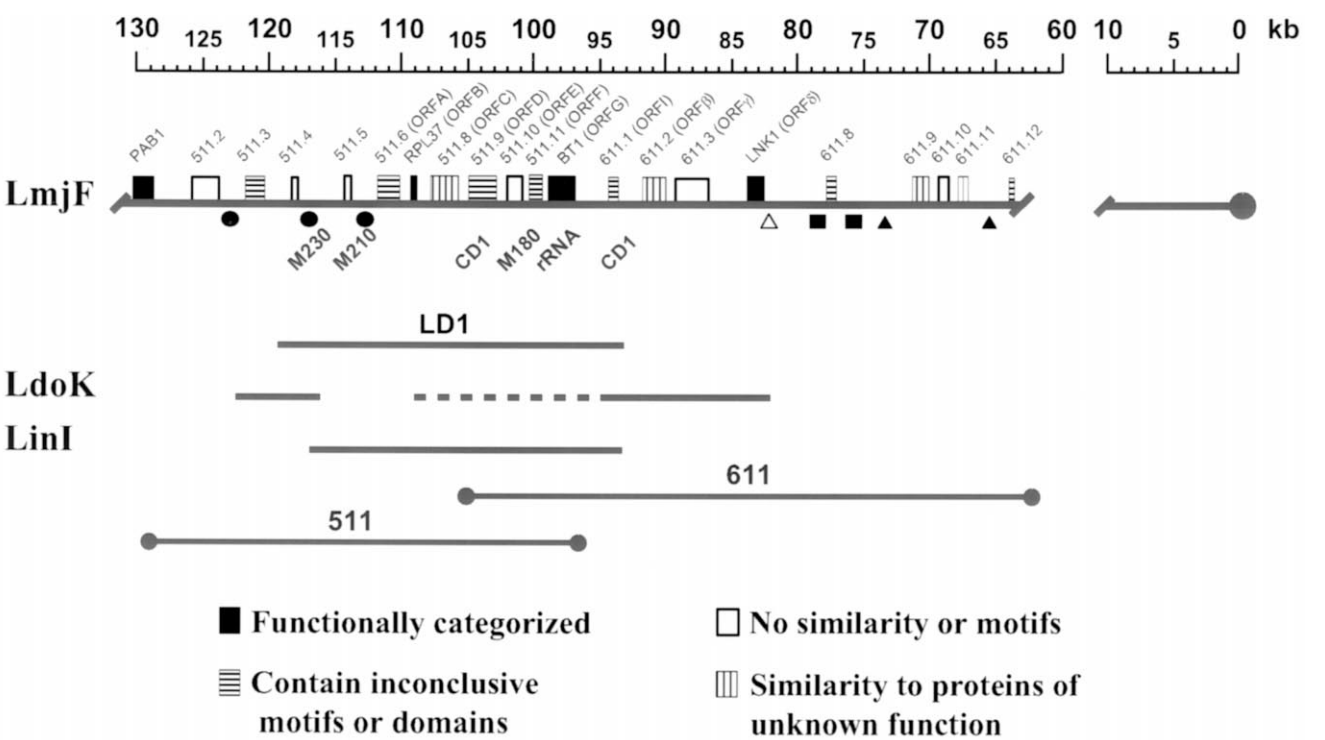

Fig. 1. Gene organization of the LD1 region in *Leishmania*. The location and coding strand of the 21 putative protein-coding ORFs in *L. major* Friedlin cosmid 511 and 611 are indicated by boxes. All are predicted to be encoded by the same DNA strand and to be transcribed from left to right. A legend for the gene classification is shown below the figure. The telomere is indicated by a circle on the right end [8]. The regions previously implicated in the amplification of LD1 are marked (M230, M210, CD1, M180, and rRNA) [9,27,46,47]. The area originally described as LD1 (*511.4–611.1*) is shown below the boxes. The regions sequenced from LdoK and LinI are indicated (accession numbers: L27052, U35460, L25643, U35461, U02459, and L38571). For LdoK, low-pass sequenced regions are shown by dashed lines. The extent of DNA contained in cosmid 511 and 611 is indicated in the bottom two lines. Intact cosmid DNA was prepared, ligated to M13mp18 RF DNA, and sequenced as previously described [37]. Generally, 600–1000 sequencing reactions were performed during the shotgun phase before assembly using the PHRED/PHRAP/CONSED software suite [49]. Repeats within this region are indicated by the following symbols: 559-bp (●); 6-bp (△); 515-bp (■); 306-bp (▲). Scale is in 10-kb increments and indicates the putative distance from the telomere.

genes are flanked by the pyrimidine-rich tracts common to kinetoplastid intergenic regions, which provide part of the processing signals for trans-splicing of the 39-bp 5′ spliced-leader sequence and polyadenylation [34–36].

Overall, the gene density (one gene per 3.2 kb) within cosmid 511 and 611 is nearly identical to that of LmjF Chr1 (one gene per 3.3 kb) [37], and higher than that reported in *Plasmodium falciparum* (one gene per 4.8 kb) [38] and *Caenorhabditis elegans* (one gene per 5.0 kb) [39]. Within these cosmids, however, DNA that encodes proteins is unevenly distributed: ~49% of LD1 is protein-coding (similar to the amount (~55%) of protein-coding information on Chr1 [37]), while the region flanking LD1 is only 35% coding. This lower-than-usual gene density may be due, in part, to the presence of repetitive DNA sequences flanking LD1 (as described below). In addition, the frequent amplification of LD1 may have influenced the evolution of its DNA sequence. Similar to gene clusters found on Chr1, Chr3, and Chr4 [37,40], all of the 21 genes in this region are on the same DNA strand (Fig. 1). This organization is consistent with polycistronic transcription and indeed, it has been shown previously that numerous transcripts containing genes in the LD1 region include RNA from adjacent genes [25–27].

A combination of database searches and protein-sequence analyses were used to assign the 21 genes identified from cosmid 511 and 611 into categories based on their putative biological function (Table 1). Four genes encode proteins which could be assigned specific functional roles: PolyA Binding protein 1 (*PAB1*) [41], Ribosomal Protein L37 (*RPL37*) [24], Biopterin Transporter 1 (*BT1)* [25,31], and *Leishmania NIMA*-related Kinase 1 (*LNK1*) [42]. *511.3* (*ORFα*), *511.8* (*ORFC*) [26], *611.1* (*ORFI*) [25], *611.9*, and *611.11* encode predicted proteins with sequence similarity to prokaryotic and eukaryotic proteins without ascribed functions; while the proteins predicted from *511.3* (*ORFα*), *511.6* (*ORFA*), *511.8* (*ORFC*), *511.9* (*ORFD*), *511.11* (*ORFF*), *611.1* (*ORFI*), *611.2* (*ORFβ*) [21], *611.8*, and *611.12* contained motifs or domains that were not informative enough to indicate a precise function. The remaining six genes (*511.2*, *511.4*, *511.5*, *511.10*, *611.3*, and *611.10*) have no identifying features or similarities. A search of the Kinetoplastida sequence databases revealed orthologues in *Trypanosoma brucei* for 11 of the 21 genes, of which five also showed orthologues in *T. cruzi* (Table 1). The amino acid identity for the *T. brucei* orthologues ranges from 35% (*BT1*) to 97% (*RPL37*) and from 40% (*LNK1*) to 94% (*RPL37*) in *T. cruzi* (Table 2), although in most cases the *T. brucei* and *T. cruzi* sequences did not represent the entire gene, being predicted from GSS and EST sequences. The lack of orthologues in the current Kinetoplastida databases for the remaining genes is difficult to interpret and may simply reflect the incomplete nature of these databases.

Table 1
Classification of gene products from LmjF Chr35 cosmids 511 and 611[a]

Category
Protein synthesis
PolyA binding protein 1 (*PAB1*)[Tb (AQ648923)]
Ribosomal protein L37 (*RPL37/ORFB*)*[Tb (AQ660139), Tc (AA874771)]
Transport
Biopterin Transporter 1 (*BT1/ORFG*)*[Tb (X73565), Tc (AI080876)]
Signal transduction
Leishmania NIMA-related kinase 1 (*LNK1/ORFδ*)[Tb (L03778), Tc (AI110486)]
Similarity to proteins of unknown function
Hypothetical proteins in Sc[b], Ce[c], and Sp[d] (*511.3/ORFα*)†
Hypothetical protein in several organisms (*511.8/ORFC*)*[g Tb (AQ643840)]
Hypothetical protein in Ph[e] and Mj[f] (*611.1/ORFI*)*[g]
Hypothetical proteins in several organisms (*611.9*)[g Tb (AQ657512), Tc (AI717789)]
Hypothetical protein in Ce[2] (*611.11*)[g Tb (AQ640674)]
Contain inconclusive motifs or domains
Bipartite nuclear localization signal (*511.6/ORFA*)*[Tb (AQ660388), Tc (AA676106)]
Proline-rich motif (*511.9/ORFD*)*[Tb (AQ650956)]
SAM-dependent methyltransferase motif (*511.11/ORFF*)*
ATP/GTP-binding site motif A *(611.2/ORFβ)*
Zinc finger domain (*611.8*)
Aldehyde dehydrogenase cysteine active site (*611.12*)
No similarity or motifs
511.2, *511.4* (*ORF2*)*, *511.5* (*ORF10*)*, *511.10* (*ORFE*)*[Tb (AQ658937)], *611.3* (*ORFγ*)[Tb (AQ657513)], *611.10*

[a] [Tb], Orthologue identified in *T. brucei* with GenBank™ accession number; [Tc], orthologue identified in *T. cruzi* with GenBank™ accession number; *, denotes gene located in LD1.
[b] *Saccharomyces cerevisiae*.
[c] *Caenorhabditis elegans*.
[d] *Schizosaccharomyces pombe*.
[e] *Pyrococcus horikoshii*.
[f] *Methanococcus jannaschii*.
[g] Also contain inconclusive motifs or domains.

In addition to protein-coding genes, cosmid 511 and 611 contain repetitive DNA (Fig. 1). Thirty-nine copies of an imperfect tandem repeat of 6-bp are located between *LNK1* and *611.8*, forming part of the polypyrimidine tracts. Two copies of an interspersed repeat of 515-bp are also found in this region, flanking *611.8*. Located between *611.8* and *611.12* are two copies of a 306-bp repeat. Three copies of an ~559-bp repeat are positioned between *511.2* and *511.6*, with an additional copy located upstream of *PAB1* (data not shown). In addition to direct repeats, a 14-bp inverted repeat separated by eight nucleotides (AGACCGCAACGCG-CacacacgcGCGCGTTGCGGTCT) is located upstream of *511.3*.

With the complete sequencing of the LD1 region from LmjF, this part of Chr35 can now be compared to the corresponding DNA in LdoK and LinI. Since the

sequences from LdoK and LinI are essentially identical, when a percentage identity is noted, it refers to that between LmjF and either of the other two species. The nucleotide identity in the LD1 protein-coding ORFs between these *Leishmania* species ranges from 93% (*BT1*) to 96% (*511.4* and *RPL37*) (Table 2). Protein-coding ORFs outside this genomic region, including those for *MKK*, *POLA* [28], *HSP70* [29], *HSP100* [43], and β-tubulin, between *L. major* and *L. donovani* or *L. infantum* share a similar level of nucleotide identity, ~93%. *ORFH*, one of the previously identified putative genes in LD1 [25], is only 84% conserved between LmjF and LinI, contains a frame-shift within the LmjF sequence, and both TESTCODE and Glimmer fail to predict this ORF as protein-coding. Thus, we feel that it is unlikely that this ORF has protein-coding function, despite the apparent presence of a stable transcript from this region in LinI [25]. Contributing to the nucleotide identity in the LD1 region is the lack of size variation between the genes. Only *511.3* (*ORFα*), *BT1*, and *LNK1* are predicted to be 54, 12, and 30 nucleotides larger respectively in LmjF, as compared to the same genes in the other two species (Table 2). As expected, the deduced amino acid identity is quite high for the 14 gene products listed in Table 2, ranging from 90 (*511.8* and *BT1*) to 100% (*RPL37*). It is interesting to note that in many instances, the amino acid identity was lower than the nucleotide identity, suggesting that constraints on amino acid sequence conservation may not be particularly severe. *BT1* has also been sequenced from *L. mexicana* [44], and *611.2* and *611.3* from *L. braziliensis* [21]. The nucleotide (~87%) and amino acid (~82%) identities between these genes and their respective gene products in LmjF are somewhat lower than the identity seen between LmjF and LdoK/LinI. This is consistent with the greater evolutionary distance between *L. major* and *L. mexicana* or *L. braziliensis*, as compared to that for *L. major* and *L. donovani* or *L. infantum*. Surprisingly, the DNA sequence from a *PAB1* gene from LmjF reported by others [41] is different at several nucleotide positions from that of *PAB1* from cosmid 511. These differences may reflect sequence variability between the Chr35 homologues, sequencing errors, or the mechanism(s) responsible for the frequent amplification of this genomic region.

There is less conservation in the non-coding regions compared to the protein-coding ORFs, with nucleotide identities ranging from 79% to 85% (Table 3) with the greater identity in the larger non-coding regions (85%) than in the smaller ones (79%). In addition, the sizes of the non-coding regions are more variable than those of the coding regions, with four and nine non-coding regions smaller and larger respectively in LmjF as compared to non-coding regions from LdoK/LinI (Table 3). The greatest size variability, 119 nucleotides, was found for the *611.3–LNK1* non-coding region. From limited analyses of non-coding regions outside of the LD1 region, (~500-bp of sequence immediately flanking *HSP100* and *HSP70*), the nucleotide identity is 87% and 95% between *L. major* and *L. donovani*, and *L. major* and *L. infantum* respectively.

In addition to analyzing the overall conservation of the coding and non-coding regions of LD1, we examined the specific sequences previously implicated in the generation of various LD1 amplicons (Fig. 1). The 8-bp sequence CCACCTCC, for instance, has been linked to the formation of the circular molecule CD1 from *L. mexicana* [9,45]. This sequence is located in identical positions in *511.9* (*ORFD*) and *611.1* (*ORFI*) in LmjF, LdoK/LinI, and *L. mexicana*. The two 13-bp sequences

Table 2
Nucleotide and amino acid identity of the LD1 region between *Leishmania* and *Trypanosoma* species[a]

Gene Range	Predicted amino acids	Nucleotide identity (%) 91–96%	*Leishmania* amino acid identity (%) 90–100%	*T. brucei* amino acid identity (%) 35–97%	*T. cruzi* amino acid identity (%) 40–94%
511.3 (*ORFα*)	522	91	91	–	–
511.4 (*ORF2*)*	168	96	99	–	–
511.5 (*ORF10*)*	171	95	95	–	–
511.6 (*ORFA*)*	584	95	94	46	48
RPL37 (*ORFB*)*	83	96	100	97	94
511.8 (*ORFC*)*	741	94	90	61	–
511.9 (*ORFD*)*	720	95	94	42	–
511.10 (*ORFE*)*	370	94	92	37	–
511.11 (*ORFF*)*	327	95	91	–	–
BT1 (*ORFG*)*	631	93	90	35	46
611.1 (*ORFI*)*	306	94	93	–	–
611.2 (*ORFβ*)	416	95	91	–	–
611.3 (*ORFγ*)	637	95	92	43	–
LNK1 (*ORFδ*)	477	93	98	39	40

[a] The columns indicate the gene, the number of amino acid residues predicted from the LmjF ORF, and the nucleotide and amino acid identity of the ORF between LmjF and LdoK or LinI, and LmjF and *T. brucei* and *T. cruzi*. *, Denotes gene located in LD1.

Table 3
Length and sequence variation in non-coding regions[a]

Non-coding region Range	Nucleotide identity (%) 79–85%	LmjF size	LdoK/LinI size	Size difference −82 to +119
511.3–511.4	82	1955	1950	5
511.4–511.5	84	3653	3599	54
511.5–511.6	81	2082	2054	28
511.6–RPL37	84	748	722	26
RPL37–511.8	85	916	921	−5
511.8–511.9	82	612	586	26
511.9–511.10	81	977	923	54
511.10–511.11	82	1090	1058	32
511.11–BT1	82	346	362	−16
BT1–611.1	84	2764	2846	−82
611.1–611.2	85	1725	1712	13
611.2–611.3	79	878	884	−6
611.3–LNK1	82	2546	2427	119

[a] The non-coding region is defined as the region between the putative stop codon of the upstream gene and the putative start codon of the downstream gene. The size difference column shows the difference in size relative to LmjF. For example, 5 and −5 for *511.3−511.4* and *511.7−511.8* indicates that the LmjF non-coding region is five nucleotides larger and five nucleotides shorter, respectively, than the non-coding region in LdoK or LinI. Length of LmjF and LdoK/LinI non-coding regions is in bp.

associated with the translocation of *511.11* (*ORFF*) and *BT1* to Chr27 in LdoK are conserved in LmjF with 12 and 11 nucleotides, respectively, matching in the 5′ boundary and internal deletion sites [27]. The sequences thought to mediate the generation of *L. major* LV-561 minichromosomes (M180, M210, and M230) that contain LD1 are also conserved [9,17,46,47]. For M180, a 26-bp A/T-rich region and 62-bp putative stem loop structure are conserved in LdoK/LinI with 96% and 92% nucleotide identity respectively. For M210, a 21-bp A/T-rich region is also present in LinI. For M230, the 9-bp sequence GGATGGCGC is located in similar positions (downstream of *511.4*, *511.5*, and *611.2*) in LdoK/LinI.

In summary, every DNA sequence implicated in the generation of specific LD1 amplicons is well conserved in the *Leishmania* species examined. This suggests that any of the various observed LD1 amplicons can be generated in any *Leishmania* species. For example, if the conserved sequence CCACCTCC is specifically involved in the generation of the circular CD1 molecule in *L. mexicana*, then this amplicon also could be generated in LmjF and LdoK/LinI. To date, however, only other types of LD1 amplicons have been found in these latter species. As a further complication, different LD1 amplicons are found in subclones derived from the same strain (*L. mexicana* M379 can have either a 210-kb minichromosome or a 53-kb circular element [45], and *L. major* LV-561 has three different mini-chromosomes: M180, M210, and M230 [9,17,46,47]). These observations suggest that the DNA sequence itself is not sufficient to determine which LD1 amplicon will be found in a particular *Leishmania* species or strain.

Individual LD1 amplicons may be generated when environmental stress leads to aberrant DNA replication, as has been shown to occur in the SV40 in vitro system [48]. In cells under such stress, the DNA polymerase could stall at various positions in the LD1 region, then switch DNA strands, utilizing the newly synthesized strand as its template, resulting in the generation of an inverted repeat. Individual LD1 amplicon variation could be dependent on the region where the DNA polymerase pauses. Consistent with this mechanism, nearly all of the LD1 amplicons contain an inverted repeat of LD1 [9,17,21].

The reason that particular amplicons are found in *Leishmania* strains may be explained by selection. Because of the function of the genes in LD1, as yet largely unknown, strains in which this region is amplified may be selected, with different amplicons favored based on environmental conditions as well as strain genetic background. Variability in gene copy number, gene content, and nucleotide polymorphisms, which may effect protein function and expression, may contribute to strain differences in pathogenicity. In the case of the LD1 region, there are only minor differences in the protein-coding genes in *Leishmania*, but extensive variability in gene copy number depending on the presence of a particular LD1 amplicon in a given strain, but it is unclear whether these differences have functional significance. Additional analyses may lead to clues about which genetic factor(s) contribute to the variability in *Leishmania* pathogenesis.

Acknowledgements

We thank the members of the SBRI sequencing team (Ellen Sisk, Greg Hixson, Patti Kiser, Erika Rickel, Melissa Hassebrock, Jason Cawthra, and Felicia Hand-

ley) for their excellent technical assistance. This study was supported by PHS grants R01 AI 40599 and R01 AI 17375 from the National Institutes of Health to K. Stuart and NIH R37 AI 21903 to S.M. Beverley.

References

[1] Wintersberger E. DNA amplification: new insights into its mechanism. Chromosoma 1994;103:73–81.
[2] Wellauer PK, Reeder RH, Dawid IB, Brown DD. The arrangement of length heterogeneity in repeating units and chromosomal rRNA from *Xenopus laevis*. J Mol Biol 1976;105:487–98.
[3] Stark GR, Debatisse M, Giulotto E, Wahl GM. Recent progress in understanding mechanisms of mammalian DNA amplification. Cell 1989;57:901–8.
[4] Hyrien O, Debatisse M, Buttin G, de Saint Vincent BR. The multicopy appearance of a large inverted duplication and the sequence at the inversion point suggest a new model for gene amplification. EMBO J 1988;7:407–17.
[5] Ma C, Martin S, Trask B, Hamlin JL. Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. Genes Dev 1993;7:605–20.
[6] Beverley SM. Gene amplification in *Leishmania*. Annu Rev Microbiol 1991;45:417–44.
[7] Stuart KD. Circular and linear multicopy DNAs in *Leishmania*. Parasitol Today 1991;7:158–9.
[8] Segovia M. *Leishmania* gene amplification: a mechanism of drug resistance. Ann Trop Med Parasitol 1994;88:123–30.
[9] Segovia M, Ortiz G. LD1 amplification in *Leishmania*. Parasitol Today 1997;13:342–8.
[10] Papadopoulou B, Roy G, Ouellette M. A novel antifolate resistance gene on the amplified H circle of *Leishmania*. EMBO J 1992;11:3601–8.
[11] Ouellette M, Papadopoulou B. Mechanisms of drug resistance in *Leishmania*. Parasitol Today 1993;9:150–3.
[12] Hightower RC, Ruiz-Perez LM, Wong ML, Santi DV. Extrachromosomal elements in the lower eukaryote *Leishmania*. J Biol Chem 1988;263:16970–6.
[13] Petrillo-Peixoto ML, Beverley SM. Amplification of a new region of DNA in an unselected laboratory stock of *L. tarentolae*: The T region. J Protozool 1989;36:257–61.
[14] Tripp CA, Myler PJ, Stuart K. A DNA sequence (LD1) which occurs in several genomic organizations in *Leishmania*. Mol Biochem Parasitol 1991;47:151–60.
[15] Tripp CA, Wisdom WA, Myler PJ, Stuart K. A multicopy, extrachromosomal DNA in *Leishmania infantum* contains two inverted repeats of the 27.5 kb LD1 sequence and encodes numerous transcripts. Mol Biochem Parasitol 1992;55:39–50.
[16] Liu J, Gajendran N, Muthui D, Muyldermans S, Dujardin J-C, De Doncker S, Jacquet D, Le Ray D, Mathieu-Daudé F, Hamers R. Chromosome rearrangement in *Leishmania mexicana* M379. Mol Biochem Parasitol 1991;46:53–60.
[17] Navarro M, Liu J, Muthui D, Ortiz G, Segovia M, Hamers R. Inverted repeat structure and homologous sequences in the LD1 amplicons of *Leishmania* spp. Mol Biochem Parasitol 1994;68:69–80.
[18] Beverley SM, Coburn CM. Recurrent de novo appearance of small linear DNAs in *Leishmania major* and relationship to extra-chromosomal DNAs in other species. Mol Biochem Parasitol 1990;42:133–42.
[19] Rovai L, Tripp C, Stuart K, Simpson L. Recurrent polymorphisms in small chromosomes of *Leishmania tarentolae* after nutrient stress or subcloning. Mol Biochem Parasitol 1992;50:115–26.
[20] Navarro M, Maingon R, Hamers R, Segovia M. Dynamics and size polymorphisms of minichromosomes in *Leishmania major* LV-561 cloned lines. Mol Biochem Parasitol 1992;55:65–74.
[21] Fu G, Melville S, Brewster S, Warner J, Barker DC. Analysis of the genomic organisation of a small chromosome of *Leishmania braziliensis* M2903 reveals two genes encoding GTP-binding proteins, one of which belongs to a new G-protein family and is an antigen. Gene 1998;210:325–33.
[22] Moore, Beverley. Abstract 107, Woods Hole Molecular Parasitology Meeting, 1996.
[23] Cunningham et al. Abstract 236, Woods Hole Molecular Parasitology Meeting, 1997.
[24] Myler PJ, Tripp CA, Thomas L, Venkataraman GM, Merlin G, Stuart KD. The LD1 amplified element from *Leishmania infantum* encodes a homolog of ribosomal protein L37. Mol Biochem Parasitol 1993;62:147–52.
[25] Myler PJ, Lodes MJ, Merlin G, deVos T, Stuart KD. An amplified DNA element in *Leishmania* encodes potential integral membrane and nucleotide binding proteins. Mol Biochem Parasitol 1994;66:11–20.
[26] Myler PJ, Venkataraman GM, Lodes MJ, Stuart KD. A frequently amplified region in *Leishmania* contains a gene that is conserved in prokaryotes and eukaryotes. Gene 1994;148:187–93.
[27] Lodes MJ, Merlin G, deVos T, Ghosh A, Madhubala R, Myler PJ, Stuart K. Increased expression of *LD1* genes transcribed by RNA polymerase I in *Leishmania* donovani as a result of duplication into the rRNA gene locus. Mol Cell Biol 1995;15:6845–53.
[28] Croan DG, Morrison DA, Ellis JT. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. Mol Biochem Parasitol 1997;89:149–59.
[29] MacFalane J, Blaxter ML, Bishop RP, Miles MA, Kelly J. Identification and characterisation of a *Leishmania donovani* antigen belonging to the 70-kDa heat-shock protein family. Eur J Biochem 1990;190:377–84.
[30] Ivens AC, Lewis SM, Bagherzadeh A, Zhang L, Chang HM, Smith DF. A physical map of the *Leishmania major* Friedlin genome. Genome Res 1998;8:135–45.
[31] Lemley C, Yan S, Cunningham MW, Dole V, Raj VS, Ghosh A, Madhubala R, Beverley SM, Myler PJ, Stuart KD. The *Leishmania donovani* LD1 locus gene *ORFG* encodes a biopterin transporter (BT1). Mol Biochem Parasitol 1999;104:93–105.
[32] Fickett JW. Recognition of protein coding regions in DNA sequences. Nucl Acids Res 1982;10:5303–18.
[33] Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. Nucl Acids Res 1998;26:544–8.
[34] Perry K, Agabian N. mRNA processing in the Trypanosomatidae. Experientia 1991;47:118–27.
[35] LeBowitz JH, Smith HQ, Rusche L, Beverley SM. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. Genes Dev 1993;7:996–1007.
[36] Matthews KR, Tschudi C, Ullu E. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev 1994;8:491–501.
[37] Myler PJ, Audleman L, deVos T, Hixson G, Kiser P, Lemley C, Magness C, Rickell E, Sisk E, Sunkin S, Swartzell S, Westlake T, Bastien P, Fu G, Ivens A, Stuart K. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. Proc Natl Acad Sci USA 1999;96:2902–6.
[38] Bowman S, Lawson D, Basham D, Brown D, Chillingworth T, Churcher CM, Craig A, Davies RM, Devlin K, Feltwell T, Gentles S, Gwilliam R, Hamlin N, Harris D, Holroyd S, Hornsby T, Horrocks P, Jagels K, Jassal B, Kyes S, McLean J, Moule S, Mungall K, Murphy L, Oliver K, Quail MA, Rajandr-

ream M-A, Rutter S, Skelton J, Squares R, Squares S, Sulston JE, Whitehead S, Woodward JR, Newbold C, Barrell BG. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. Nature 1999;400:532–8.

[39] The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 1998;282:2012–2018.

[40] Myler PJ, Stuart KD. Recent developments from the *Leishmania* genome project. Curr Opin Microbiol 2000;3:412–6.

[41] Bates EJ, Knuepfer E, Smith DF. Poly(A) binding protein I of *Leishmania*: functional analysis and localization in trypanosomatid parasites. Nucl Acids Res 2000;28:1211–20.

[42] Osmani SA, May GS, Morris NR. Regulation of the mRNA levels of nimA, a gene required for the G2-M transition in *Aspergillus nidulans*. J Cell Biol 1987;104:1495–504.

[43] Hübel A, Brandau S, Dresel A, Clos J. A member of the clpb family of stress proteins is expressed during heat shock in *Leishmania* spp. Mol Biochem Parasitol 1995;70:107–18.

[44] Kündig C, Haimeur A, Légaré D, Papadopoulou B, Ouellette M. Increased transport of pteridines compensates for mutations in the high affinity folate transporter and contributes to methotrexate resistance in the protozoan parasite *Leishmania tarentolae*. EMBO J 1999;18:2342–51.

[45] Liu J, Salinas G, Gajendran N, Muthui D, Muyldermans S, Hamers R. DNA recombination associated with short direct repeats in *Leishmania mexicana* M379. Mol Biochem Parasitol 1992;50:351–4.

[46] Ortiz G, Navarro M, Segovia M. Location in the source chromosome of the 180-kb minichromosome of *Leishmania major* and characterization of the novel junction. Mol Biochem Parasitol 1995;71:153–61.

[47] Ortiz G, Segovia M. Characterisation of the novel junctions of two minichromosomes of *Leishmania major*. Mol Biochem Parasitol 1996;82:137–44.

[48] Aladjem MI, Lavi S. The mechanism of carcinogen-induced DNA amplification: in vivo and in vitro studies. Mutat Res 1992;276:339–44.

[49] Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Bonfield J, Burton J, Connell M, Copsey T, Cooper J, Coulson A, Craxton M, Dear S, Du Z, Durbin R, Favello A, Fraser A, Fulton L, Gardner A, Green P, Hawkins T, Hillier L, Jier M, Johnston L. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 1994;368:32–8.