

# Phylogenetic analysis of *Leishmania* RNA virus and *Leishmania* suggests ancient virus–parasite association

Giovanni Widmer\* and Sheryl Dooley

Division of Infectious Diseases, Department of Comparative Medicine, Tufts University School of Veterinary Medicine, North Grafton, MA 01536, USA

Received January 17, 1995; Revised and Accepted April 27, 1995

GenBank accession nos L39069, L39070, L41164, U23810

## ABSTRACT

Some strains of the protozoan parasite *Leishmania* belonging to the new world species *guyanensis* and *braziliensis* are infected with persistent, single-segmented, non-enveloped dsRNA viruses termed LRV1. A single old world strain classified as *L. major* was recently found to harbor a similar virus, designated LRV2-1. The genomic nucleotide sequences of two LRV1 types (1-1 and 1-4) isolated from two *L. guyanensis* strains have been determined and found to be highly conserved. In contrast, LRV1-specific cDNA probes derived from the conserved genomic 5' region failed to recognize LRV2 RNA on Northern blots, suggesting a greater degree of divergence between LRV1 and LRV2 than among LRV1 types. This observation suggests a long-term association and co-evolution of LRV within each parasite strain. We tested this concept by comparing nucleotide sequences of seven LRV types and PCR fingerprints of the parasite strains from which these viruses were derived. In support of the idea of virus–parasite co-evolution, we find that genetic distances between LRV types mirror the heterogeneity between parasite fingerprints and are clustered according to the geographical origin of the strains. In agreement with the postulated common origin of persistent dsRNA viruses of protozoa and fungi, we conclude that the infection of *Leishmania* with LRV pre-dates the divergence of *Leishmania* into different lineages.

## INTRODUCTION

Viruses found in some strains of the protozoan parasite *Leishmania* consist of spherical capsids containing a dsRNA genome of ~5280 nucleotides (nt) (1–3). The two overlapping open reading frames, ORF2 and ORF3, encode the 82 kDa capsid protein (4) and the RNA-dependent RNA polymerase (RDRP) (5), respectively. A comparison of the LRV1 RDRP sequence with homologous sequences from other viruses (5) showed that the yeast L-A/ScV virus RDRP gene is the sequence most similar to LRV.

This is consistent with the common ancestral origin of four dsRNA viruses which persistently infect protozoa and fungi, as suggested by the presence of conserved sequence motifs in the RDRP gene of L-A/ScV in yeast, UmVH in *Ustilago*, TvV in *Trichomonas* and LRV (6).

With the exception of one LRV-infected *L. major* strain, originally isolated from a human cutaneous lesion in the former Soviet Union, all LRV-infected strains described so far originate from South America. Guilbride *et al.* (7) described 12 such LRV1 types from a sample of 71 *Leishmania* strains; all infected strains were classified as *L. braziliensis* or *L. guyanensis* and most were isolated in the Amazon basin. The comparison of the entire genomic sequence of two LRV1 types showed 77% overall nucleotide similarity, and 92 and 82% amino acid sequence identity within ORF2 and ORF3, respectively (8). This high degree of amino acid sequence conservation is reflected in the frequent occurrence of third base substitutions within the ORFs. In contrast, the 5' untranslated region (UTR) spanning the first 450 nt of the (+) strand shows less divergence between LRV1-1 and 1-4 suggesting the existence of conserved regulatory sequence elements within this region. Experimental evidence for cap-independent translation mediated by this region has recently been obtained (Maga, Widmer and LeBowitz, in preparation).

Molecular and biochemical techniques have demonstrated significant genetic heterogeneity within the genus *Leishmania*. Based on such studies, *L. major* and *L. braziliensis/guyanensis* were assigned to different species complexes, namely *tropica* and *braziliensis* (9,10). The absence of cross-hybridization on Northern blots between LRV1 and LRV2 suggested that viruses from taxonomically distant *Leishmania* strains have diverged more than those originating from related strains. A similar conclusion is supported by Western blot analysis, where a LRV1-4-specific antiserum recognized other LRV1 types but failed to detect LRV2-1 (11). In spite of these differences, the *L. major* virus was assigned to LRV based on the apparent identical genome size and the profile of viral transcript generated during an *in vitro* polymerase assay (G.W., unpublished). The apparent correlation between parasite taxonomy and viral heterogeneity is consistent not only with the persistent nature of LRV, but also with a predominantly asexual mode of replication in *Leishmania* (12).

\* To whom correspondence should be addressed

The idea of co-evolution of LRV and *Leishmania* was tested by comparing the sequences of two genomic regions from seven LRV types; similarly, genetic distances between the parasite strains harboring these viruses were estimated based on random amplified polymorphic DNA (RAPD) fingerprints (13). This technique was chosen since insufficient variation between homologous *Leishmania* sequences precludes direct sequence comparisons. The data presented here demonstrate a similar relationship between parasite strains on one side and viruses on the other, supporting the concept of a long-term association between LRV and *Leishmania*. These data are consistent with the co-evolution of *Leishmania* and LRV and support the view of a common origin of persistent dsRNA viruses of fungi and protozoa.

## MATERIALS AND METHODS

### Parasites

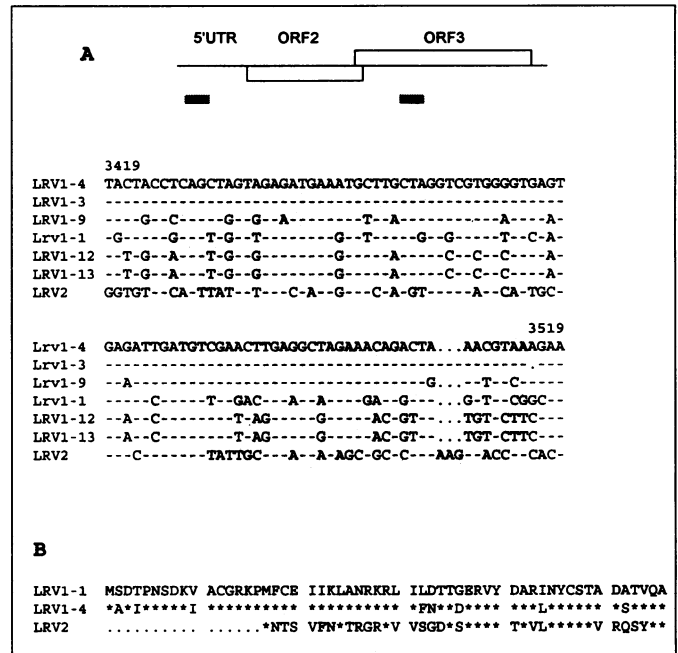
Table 1 shows the *Leishmania* strains used in this study and the corresponding LRV type. The taxonomic classification of strain 5-ASKH was confirmed at the species complex level by analysing the nucleotide sequence of the variable region of the small subunit rRNA gene identified by van Eys *et al.* (14). A fragment of ~100 nt of this region was PCR amplified and sequenced. A nucleotide deletion at position 1032, diagnostic of the tropica complex, was identified. The Bolivian isolates FY and NC (kindly provided by Dr Jean-Pierre Dedet, IBBA, La Paz) were isolated in 1990 from two patients with nasal lesions and cryopreserved shortly thereafter. Two mismatches within the 5' UTR of LRV1-13 and 1-14 confirmed that these viruses originated from different strains.

### Molecular techniques

**Synthesis of LRV2-1 cDNA.** Viral dsRNA, purified in low melt agarose gels from a 5-ASKH RNA preparation, was reverse transcribed with random hexanucleotides and PCR amplified with an arbitrary 9mer oligonucleotide. PCR products were cloned using Invitrogen's TA cloning system. To confirm the origin of the clones, inserts were <sup>32</sup>P-labelled by random priming (NEBlot kit, New England Biolabs) and used to probe Northern blots of LRV1 and LRV2-1 RNA. Clones which recognized only LRV2-1 were sequenced.

**Synthesis of LRV1-3, 1-9, 1-13 and 1-14 cDNA.** Promastigote RNA was reverse transcribed and PCR amplified with primers derived from consensus sequences flanking the regions of interest in LRV1-1 and 1-4.

**RAPD amplifications.** Seven *Leishmania* strains (Table 1) were analysed with five arbitrary primers. To reduce the possibility of introducing artifactual differences, DNA was prepared simultaneously from all strains. All manipulations from cell harvest to PCR amplification were performed in parallel. Between  $1.8 \times 10^7$  and  $3 \times 10^7$  promastigotes were harvested from cultures in M199 media (15), washed once in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and digested overnight in TE, 0.1% SDS, 200 µg/ml proteinase K at 45°C. Nucleic acid was extracted, ethanol precipitated and resuspended in 50 µl TE supplemented with RNaseA (20 µg/ml). In each reaction, 2 µl of parasite DNA was amplified with a single arbitrary primer. The following primers were used: G, GCGGCCGCG; EM13, GGAAACAGTTATGAC-CATG; 33, GCTCAATATTAGTTTCATCGTC; 113, CGGTGTT-



**Figure 1.** (A) Alignment of seven LRV ORF3 sequences. Uppermost is a schematic representation of the LRV 5.2 kb genome showing both ORFs and UTRs. Shaded blocks indicate the approximate location of the sequences analysed. A portion of the sequence alignment from the ORF3 region is shown below. -, identity; ., deletion. (B) Alignment of the ORF2 amino terminal from LRV1-1, 1-4 and 2. Compared with LRV1, the LRV2 ORF2 initiates 17 residues further downstream. \*, identity.

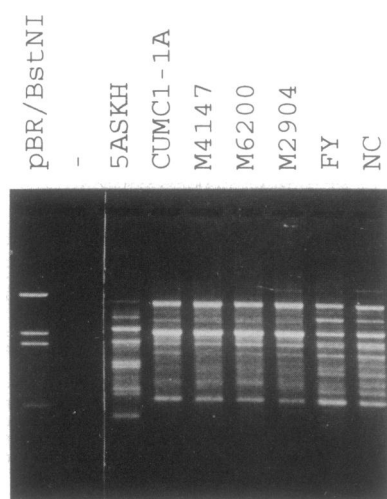
GTAAACCAATTATTC; 43, GTTTCATCGTGTCTGCGTC; 162, GCCGTGTAAGACTCTTTCAG. PCR was performed in a Robocycler (Stratagene) using the following program: 94°C, 1 min; 40°C, 1 min; 72°C, 2 min followed by 40 cycles of 94°C, 1 min; 55°C, 1 min; 72°C, 2 min.

### Data analysis

Nucleotide sequence alignments and LRV dendrograms based on the unweighted pair group clustering method (UPGMA) (16) were obtained with the Pileup program of the GCG sequence analysis package. Maximum likelihood trees were constructed using DNAm1 (Phylip version 3.5) (17) using the random order input option and four repetitions. Where included in the analysis, LRV2 was selected as outgroup. Distance indices (D) between RADP profiles (Table 2) were calculated using Jaccard's method (18) as  $D = 1 - C/(2N - C)$ , where  $C$  is the number of bands in common between two strains and  $N$  is the total number of different bands in each pairwise comparison. One  $N$  and one  $C$  value was determined for each of 21 pairwise comparisons using the entire set of RAPD fingerprints.

## RESULTS

In order to examine the relationship between LRV1 and 2, cDNA from LRV2 was cloned and sequenced. The GAP program (GCG) was used to determine the best alignment between the LRV2 clones and the complete LRV1-1 and 1-4 sequence. Only cDNA clones within the ORF3 could be aligned unambiguously (Fig. 1A), whereas other regions of the LRV2 genome were not



**Figure 2.** RAPD fingerprints from seven *Leishmania* strains obtained with an arbitrary 20mer primer. RAPD fingerprints (13) generated with arbitrary primer 43 segregate into three subgroups: the Bolivian strains FY and NC, the four *L. braziliensis/guyanensis* strains (CUMC1-1A, M4147, M6200, M2904) and the *L. major* strain 5-ASKH. pBR/BstNI digest was used as a size marker; band sizes: 1857 nt, 1058 nt, 929 nt, 383 nt, 121 nt. (-), negative control amplification without DNA.

sufficiently conserved. LRV2 fragment 5a-4 located within the upstream portion of ORF3 and corresponding to position 3419–3633 of LRV1 was included in a multiple sequence alignment with six LRV1 types. Within this fragment, the similarity at the nucleotide level between LRV2 and LRV1-4 was 50%, while similarities among LRV1 types ranged from 71% (LRV1-13 × LRV1-4) to 100% (LRV1-3 × LRV1-4). The characteristic trend for third base substitutions as previously seen between LRV1-1 and 1-4 (8) was also apparent between LRV2 and LRV1 (Fig. 1A). At the amino acid level 60% similarity between LRV2 and LRV1-4 was found within this ORF3 region. The two dendrograms generated from this data set are shown in Figure 3 (left side).

The 5' terminal sequence of a 4 kb contig consisting of three overlapping LRV2 cDNA clones was aligned to the LRV1-1 and 1-4 sequence. Using the relatively conserved ORF3 sequence (Fig. 1A) as an 'anchor', the 5' end of the contig corresponded to position 132 on the LRV1-1 genome. The 5' most fragment of the LRV2 sequence was obtained from tailed cDNA and is thought to be the 5' end of the genome, although additional cDNA clones will need to be generated to verify this assumption. The accuracy

of the alignment between LRV1 and 2 was confirmed by the presence in LRV2 of an ORF initiating at approximately the same position as the LRV1 ORF2. The homology between these ORFs is supported by the presence of 17 conserved positions within the 40 N-terminal residues (Fig. 1B). In contrast, within the 5'-UTR the similarity between LRV1 and 2 was too low to obtain a meaningful alignment. This finding confirmed previous Northern blot analyses which showed that cDNA probes derived from the LRV1-4 5'-UTR did not hybridize to LRV2 RNA. The six remaining LRV1 types showed a high degree of conservation within a 211 nt fragment spanning positions 42–252 on the LRV1-4 genome. Similarity values within this region ranged from 88% (LRV1-13 × LRV1-4) to 100% (LRV1-4 × LRV1-3). The left dendrograms in Figure 4 illustrate the relationship between sequences for this region. Due to the inability to obtain a meaningful alignment between the 5'-UTR of LRV1 and 2, LRV2 was excluded from 5'-UTR trees.

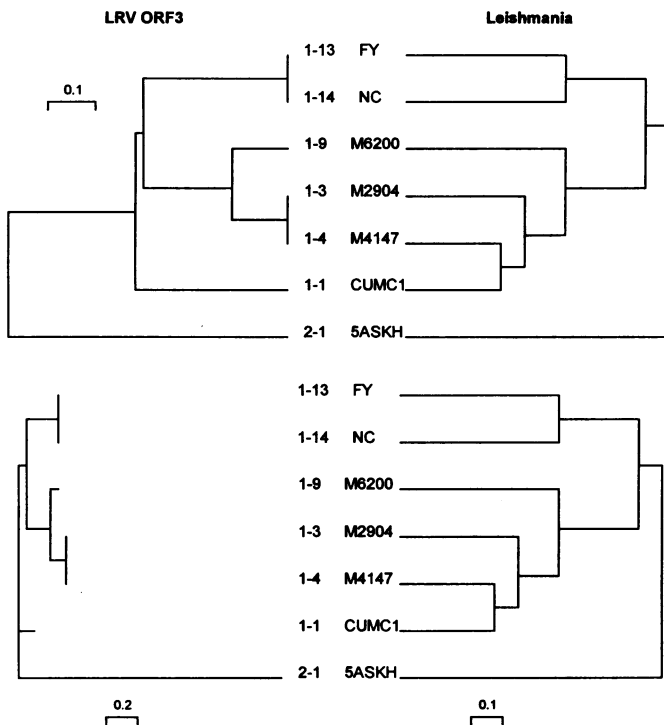
As a means of determining genetic distances between seven LRV-infected *Leishmania* strains (Table 1), DNA was obtained from each strain and subject to RAPD analysis. To avoid any possible bias caused by the choice of primer, multiple RAPD analyses were performed, each with a different primer. Figure 2 shows RAPD fingerprints obtained from seven strains with primer 43, an arbitrary 20mer. Fingerprints clearly segregated into three groups: the *L. major* strain (5-ASKH), the two Bolivian strains (FY, NC), and the remaining four *L. guyanensis/braziliensis* strains (CUMC1-1A, M4147, M2904, M6200). Although these three subgroups were apparent in most RAPDs, the degree of heterogeneity seen among the *braziliensis* group was strongly dependent on the primer. For example, primer 162 generated numerous differences among the *L. guyanensis/braziliensis* subgroup (not shown), while other primers, such as 43 (Fig. 2) gave similar profiles. In order to avoid any bias in the interpretation of the RAPDs, the first six primers giving clear, multibanded patterns were included in the numerical evaluation, regardless of the extent or type of heterogeneity produced. As expected from the small number of shared bands, distances between 5-ASKH and all other strains were close to 1 (Table 2). Distances among the *braziliensis* complex, excluding the two Bolivian strains, were between 0.34 and 0.58 and between the Bolivian strains and the remaining *L. guyanensis/braziliensis* subgroup ranged between 0.78 and 0.82. To reduce the possibility that fingerprints were artifactually affected by the DNA preparations, RAPD analyses were repeated with five of the seven strains (excluding FY and NC) using a different set of DNA samples. A similar clustering was obtained (not shown).

**Table 1.** Parasite strains used in this study

Code	Species	Geographical origin	LRV type	Reference
MHOM/SU/73/ <u>5-ASKH</u>	<i>major</i>	Soviet Union <sup>a</sup>	LRV2-1	this study
MHOM/SR/81/ <u>CUMC1-1A</u>	<i>guyanensis</i>	Surinam	LRV1-1	1
MHOM/BR/75/ <u>M4147</u>	<i>guyanensis</i>	Amazon basin	LRV1-4	2
MHOM/BR/75/ <u>M2904</u>	<i>braziliensis</i>	Amazon basin	LRV1-3	7
MCHO/BR/80/ <u>M6200</u>	<i>guyanensis</i>	Amazon basin	LRV1-9	7
FY	<i>braziliensis</i>	Bolivia	LRV1-13	this study
NC	<i>braziliensis</i>	Bolivia	LRV1-14	this study

Where available, the complete WHO code is given. Short code used in the text is underlined.

<sup>a</sup>Origin within current political boundaries is not known.

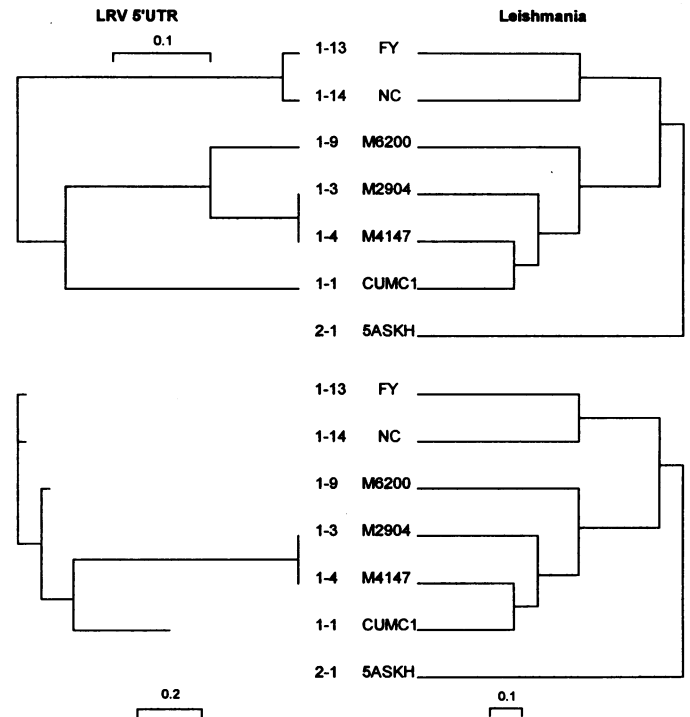


**Figure 3.** Comparison of trees obtained from the viral ORF3 sequences to a *Leishmania* tree based on RAPD fingerprint analysis. Uppermost: LRV tree obtained from the alignment of 215 nt of the ORF3 (RDRP gene) by the UPGMA method (left) is compared with a parasite tree derived from the RAPD data by the same method (right). Lowermost: viral phylogenetic tree obtained by the maximum likelihood method (17) is compared with the same parasite tree. The same subgroups (FY/NC; M6200/M2904/M4147; 5-ASKH) observed with the RAPD analysis were also apparent in both viral trees. The similar clustering of strains and LRV types supports the notion that *Leishmania* and LRV have co-evolved. Scalebars: maximum likelihood trees, units of expected nt substitution per site (17); UPGMA trees, genetic distance, where 0 is equal to identity and 1 is absence of shared characters (RAPD bands).

**Table 2.** Genetic distances between *Leishmania* strains were derived by the method of Jaccard (18) from five RAPD analyses

	CUMC1-1A	M4147	M6200	M2904	FY	NC
5-ASKH	0.94	0.93	0.90	0.96	0.91	0.92
CUMC1-1A		0.34	0.53	0.41	0.78	0.83
M4147			0.46	0.42	0.80	0.80
M6200				0.58	0.80	0.86
M2904					0.85	0.82
FY						0.55

A preliminary inspection of the distance matrix obtained from the fingerprints (Table 2) and the sequence similarities between LRV types revealed a good correlation between data sets: *Leishmania* strains harboring closely related LRV types generated similar RAPD fingerprints and vice versa. To visualize this trend, dendrograms were drawn from the RAPD data using the unweighted pair-group arithmetic average (UPGMA) clustering method (16) and juxtaposed to the LRV trees (Figs 3 and 4). Four trees were generated from the viral sequences; using the UPGMA



**Figure 4.** Comparison of viral trees from the 5'-UTR sequences to *Leishmania* tree. Analogous comparison of the relationship between LRV types and between *Leishmania* strains as shown in Figure 3 but based on 211 nt from the LRV 5'-UTR. Uppermost left, viral UPGMA tree and lowermost left, viral maximum likelihood tree are compared to the same *Leishmania* dendrogram (right). Due to insufficient sequences conservation between LRV1 and 2, the LRV2-1 sequence was excluded from this analysis. See Results section for details. Scalebars as in Figure 3.

and maximum likelihood methods two trees each were drawn from the ORF3 (Fig. 3) and 5'-UTR (Fig. 4) sequences. With the exception of the 5'-UTR maximum likelihood tree (Fig. 4), the same subgroups as seen with the parasite dendrogram were found in the viral trees, namely LRV1-13/14, LRV1-3/4/9 and LRV2-1. For unknown reasons, the CUMC1-1A/LRV1-1 sample did not follow this trend and the position of LRV1-1 was also inconsistent among the four LRV trees. The analogy between LRV trees obtained by both methods (UPGMA and maximum likelihood) was readily apparent with the ORF3 data, which is based on seven samples, whereas some inconsistencies were observed between the viral 5'-UTR trees, which are based on six sequences only. The high degree of sequence conservation among the LRV1 5'-UTRs could have contributed to this effect, since small differences between sequences can influence the topology of the tree. Significantly, in spite of the fact that two functionally different LRV regions were chosen, similar clustering of LRV types were obtained, in particular with the UPGMA method. The maximum likelihood trees were less conserved possibly due to the different number of sequences analysed.

In summary, the juxtaposition of *Leishmania* and LRV trees showed a good correlation between parasite and viral clusterings. The overall similarity between viral and *Leishmania* trees is consistent with the geographical origin of the parasite strains (Table 1) and with a long-term association of each LRV type and *Leishmania* strain.

## DISCUSSION

Genetic distances between LRV types and between *Leishmania* strains infected with these viruses are compatible with the parallel evolution of *Leishmania* and LRV. This model is supported by the observation that the same subgroups were obtained from the analysis of the parasite strains and viral types. Co-evolution is also consistent with the absence of infectious LRV virions and the predominantly clonal, asexual mode of reproduction of *Leishmania* (12,19).

In agreement with the high frequency of mutation in RNA genomes (20), our model predicts a lower degree of conservation between LRV types than between homologous *Leishmania* sequences, which is confirmed by this study; for instance, the dihydrofolate reductase gene from *L.major* and *L.amazonensis*, two species belonging to different complexes, are 91% similar, as compared with 50% similarity between the ORF3 of LRV1 and 2. Because of the absence of sequence variation within *Leishmania* species, RAPD analysis was chosen for assessing genetic distances between strains.

The likelihood of a long-term, stable association between LRV and *Leishmania* raises the question of a possible biological function of LRV and the mechanism responsible for viral maintenance. Presently, no evidence has emerged pointing to any phenotypic properties of *Leishmania* associated with viral infection, although the lack of isogenic infected and uninfected strains has precluded a systematic study of this issue. The stability of the parasite-virus association suggested by the present study argues against selective neutrality of LRV, since in the absence of a mechanism favoring virus maintenance, LRV<sup>-</sup> lines would presumably arise sporadically and possibly overgrow the LRV<sup>+</sup> population. Clarifying the role of LRV and the mechanism of viral persistence might reveal whether LRV affects the virulence of the parasite in the vertebrate host or its transmission by the sandfly vector.

Finally, it remains to be explained why only a few strains belonging to three species have been found infected with LRV. Based on the data presented here and the evolution of the RDRP gene in dsRNA viruses (6), it appears that LRV has been acquired prior to the divergence of *Leishmania* into multiple lineages. Consequently, one has to assume that LRV was eliminated from selected species during the evolution of the genus.

## ACKNOWLEDGEMENTS

We thank Dr John Kelly, London School of Hygiene and Tropical Medicine, for providing the *L.major* strain and pointing out the possible existence of a virus in this strain, Robert Means, Harvard Medical School, for sequencing the 5-ASKH ribosomal gene fragment, Dr Greta Carraway for help with RAPD fingerprinting, and Drs Saul Tzipori and Ramaswamy Balakrishnan for reviewing the manuscript. Supported by PHS grant AI29390 from the NIAID.

## REFERENCES

- 1 Tarr, P.I., Aline, R.F., Smiley, B.L., Keithly, J. and Stuart K (1989) *Proc. Natl. Acad. Sci. USA* **85**, 9572-9575.
- 2 Widmer, G., Comeau, A.M., Furlong, D.B., Wirth, D.F. and Patterson J.L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5979-5982.
- 3 Widmer G. and Patterson, J.L. (1991) *J. Virol.* **65**, 4211-4215.
- 4 Cadd, T.L., and Patterson, J.L. (1994) *J. Virol.* **68**, 358-365.
- 5 Stuart, K.D., Weeks, R., Guilbride, L. and Myler, P.J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8596-8600.
- 6 Bruenn, J.A. (1993) *Nucleic Acids Res.* **21**, 5667-5669.
- 7 Guilbride, L., Myler, P.J. and Stuart, K. (1992) *Mol. Biochem. Parasitol.* **54**, 101-104.
- 8 Scheffter, S., Widmer, G. and Patterson, J.L. (1994) *Virology* **199**, 479-483.
- 9 Beverley, S.M., Ismach, R.B. and McMahon-Pratt, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 484-488.
- 10 Kreutzer, R.D., Souraty, N. and Semko, M.E. (1987) *Am. J. Trop. Med. Hyg.* **36**, 22-32.
- 11 Cadd, T.L., Keenan, M.C. and Patterson J.L. (1993) *J. Virol.* **67**, 5647-5650.
- 12 Panton, L.J., Tesh, R.B., Nadeau, K.C. and Beverley, S.M. (1991) *J. Protozool.* **38**, 224-228.
- 13 Welsh, J. and McClelland, M. (1990) *Nucleic Acids Res.* **18**, 7213-7218.
- 14 Van Eys, G.J.J.M., Schoone, G.J., Kroon, N.C.M. and Ebeling, S.B. (1992) *Mol. Biochem. Parasitol.* **51**, 133-142.
- 15 Coderre, J.A., Beverley, S.M., Schimke, R.T. and Santi, D.V. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2132-2136.
- 16 Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*, Freeman, San Francisco.
- 17 Felsenstein, J. (1993) PHYLIP, Phylogeny Interference Package version 3.5. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA, USA.
- 18 Jaccard, P. (1908) *Bull. Soc. Vaudoise Sci. Nat.* **44**, 223-270.
- 19 Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Breniere, S.F., Darde, M.L. and Ayala, F.J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5129-5133.
- 20 Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and VandePol, S. (1982) *Science* **215**, 1577-1585.
- 21 Wang, A.L., Wang, C.C. and Alderete, J.F. (1987) *J. Exp. Med.* **166**, 142-150.