Liver fatty acid binding protein (L-FABP) is critical for *Plasmodium* liver stage growth. Immuno-flourescent images of *Plasmodium berghei* liver stages grown in hepatoma cells with normal L-FABP expression levels (left panels) and in hepatoma cells with reduced L-FABP expression levels (right panels).

Available online at www.sciencedirect.com
ScienceDirect
International Journal for Parasitology 37 (2007) 565–576
www.elsevier.com/locate/ijpara

Glycoprotein 63 (*gp63*) genes show gene conversion and reveal the evolution of Old World *Leishmania*

Isabel L. Mauricio *, Michael W. Gaunt, J. Russell Stothard [1], Michael A. Miles

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Received 24 July 2006; received in revised form 14 November 2006; accepted 20 November 2006

Abstract

Species of the subgenus *Leishmania* (*Leishmania*) cause the debilitating disease leishmaniasis on four continents. Species grouped within the *Leishmania donovani* complex cause visceral leishmaniasis, a life-threatening disease, often associated with poverty, and affecting some 0.5 million people each year. The *Leishmania* glycoprotein GP63, or major surface protease, is a metalloprotease involved in parasite survival, infectivity and virulence. Here, we show that evolution of the *gp63* multigene family is influenced by mosaic or fragmental gene conversion. This is a major evolutionary force for both homogenisation and for generating diversity, even in the absence of sexual reproduction. We propose here that the high GC content at the third codon position in the *gp63* family of Old World *Leishmania* may be higher in multicopy regions, under the biased gene conversion model, because increased copy numbers may lead to increased rates of recombination. We confirm that one class of *gp63* genes with an extended 3′end signal, $gp63^{EXT}$, reveals genetic groups within the complex and gives insights into evolution and host associations. $Gp63^{EXT}$ genes can also provide the basis for rapid and reliable genotyping of strains in the *L. donovani* complex. Our results confirmed that a more stringent definition of *Leishmania infantum* is required and that the species *Leishmania archibaldi* should be suppressed.

Keywords: *Leishmania donovani* complex; Major surface protease; *gp63*; Taxonomy; Phylogeny; Gene conversion; Genotyping; GC3 bias

1. Introduction

Leishmaniasis is caused by the genus *Leishmania*, which comprises intracellular parasites of macrophages in the mammalian host, transmitted by the bite of sandflies. Most species of the subgenus *Leishmania (Leishmania)* cause cutaneous leishmaniasis (CL) in the Old World (OW; *Leishmania major*, *Leishmania tropica* and *Leishmania aethiopica*) and in the New World (NW; *Leishmania mexicana* and *Leishmania amazonensis*). Some OW species (*Leishmania arabica*, *Leishmania gerbilli* and *Leishmania turanica*) have never been found in humans. The remaining species cause visceral leishmaniasis (VL), with an estimated 0.5 million new cases annually worldwide (WHO, 1990), and belong to the *Leishmania donovani* complex: *L. donovani* (syn. *Leishmania archibaldi*) in Sudan (Ibrahim and Barker, 2001; Mauricio et al., 2001, 2006; Jamjoom et al., 2004; Zemanova et al., 2004; Kuhls et al., 2005), and *Leishmania infantum* (syn. *Leishmania chagasi*) in the NW (Mauricio et al., 2000).

Leishmania glycoprotein GP63 is a surface metalloprotease, also called leishmanolysin or major surface protease (MSP), encoded by a multigene family in chromosome 10 of *L. major* and its homologues in *L. infantum* and *Leishmania braziliensis*. GP63 is an important antigen with a role in internalization of the parasite into the macrophage (Chakrabarty et al., 1996; Brittingham et al., 1999; Puentes et al., 1999; Chen et al., 2000) and in avoidance of the host's complement mediated lysis (Joshi et al., 1998). The subgenus *L. (Leishmania)* has two main classes of *gp63* genes. One, $gp63^{SHORT}$, has a clear glycosylphosphatidylinositol

* Corresponding author. Tel.: +44 20 7927 2814; fax: +44 20 7636 8739.
E-mail address: isabel.mauricio@lshtm.ac.uk (I.L. Mauricio).
[1] Present address: Natural History Museum, Cromwell Road, London SW7 5BD, UK.

0020-7519/$30.00
doi:10.1016/j.ijpara.2006.11.020

(GPI) anchor signal (Medina-Acosta et al., 1993). The other, $gp63^{EXT}$, has a longer 3′end, which encodes a hydrophobic tail without a recognizable GPI anchor attachment signal. $Gp63^{EXT}$ seems to be single copy in *L. infantum*, *mspC* (Ramamoorthy et al., 1992) or *gp63-4* (in the *L. infantum* genome sequence) and in *L. major*, *gp63-6* (Voth et al., 1998) or *gp63-3* (in the *L. major* genome sequence). The *L. infantum* genome sequence (www.genedb.org) has, so far, only identified four different $gp63^{SHORT}$ genes in contrast with the initial description of the organization of *gp63* genes in South American strains of this species (Ramamoorthy et al., 1992), in which several copies had been identified of two classes of $gp63^{SHORT}$: *mspS* and *mspL*. In the *L. infantum* genome sequence, one of the $gp63^{SHORT}$ genes has a truncated 5′end (LinJ10.0800), whereas another has an extended 5′end and a truncated 3′end (LinJ10.0820). It is not clear whether these are accurate gene sequences, as the *L. infantum* genome is not yet complete and there are no equivalents in *L. major*. Similarly, seven genes have been described for *L. major* (Voth et al., 1998), whereas only five have been reported in the genome sequence. The genome organization of these genes differs between the two species. In *L. major*, all genes are located on the same DNA strand, whereas in *L. infantum gp63-1* to *3* are in opposite strands to *gp63-4* to *5*. Other putative *gp63* copies have been identified in the genomes of *L. infantum* and *L. major*, on chromosomes 28 (LinJ28.0580, LinJ28.3180 and LmjF28.0570) and 31 (LinJ31.2360 and LmjF31.2000). However, these genes have very low homology with the tandemly repeated genes on chromosome 10. It is thus not clear if they are true *gp63* genes and they are not analysed here. Mosaic gene conversion, or homologous recombination, in *gp63* genes has previously been proposed (Medina-Acosta et al., 1993; Victoir et al., 2003), based on the finding that the phylogeny of the carboxyl terminus differs from that of the amino terminus.

Here we exploit $gp63^{EXT}$ to investigate the genetic structure and evolution of the *L. donovani* complex and other OW *Leishmania*. We found that $gp63^{EXT}$ was an excellent target to analyse putative mosaic/fragmental gene conversion processes, and to explore host associations of $gp63^{EXT}$ diversity.

2. Materials and methods

Thirty-three strains of the *L. donovani* complex of diverse origins and zymodemes (multilocus enzyme electrophoretic variants) were used as representatives of the natural diversity inherent within the group, with reference strains of other OW *Leishmania* (Table 1) and published sequences for $gp63^{EXT}$ of *L. major* (AF039721) and *L. mexicana* (X64394). $Gp63^{SHORT}$ published sequences used here were: *L. amazonensis* (L46798), *L. donovani* complex (L19562-3, M60048, M80669, M80672, U48798, Z83677), *Leishmania guyanensis* (M85203, L16776-9), *L. major* (Y00647) and *Leishmania panamensis* (AF037165-6). These genes were compared with those obtained from the genome sequences of *L. major* ($gp63^{EXT}$ LmjF10.0470, $gp63^{SHORT}$ LmjF10.0460, LmjF10.0465 and LmjF10.0480), *L. infantum* ($gp63^{EXT}$ LinJ10.0810, $gp63^{SHORT}$ LinJ10.0780 and LinJ10.0790) and *L. braziliensis* (LbrM10_V2.0470-610, LbrM10_V2.01540-1590 and LbrM10_V2.01610-1720). In the analyses we used mainly the non-genome sequences available for *L. infantum* and *L. major* because they had been used in the initial comparisons and were equivalent to the genome sequences.

Genomic DNA was prepared as previously described (Kelly, 1993). Specific amplification of the entire $gp63^{EXT}$ coding sequence was done with primers C1F (5′gcg cct gca gag cca tg3′) and C8R, as previously described (Mauricio et al., 1999) but with an annealing temperature of 65 °C and 2.5 min extension. Direct sequencing used the PCR primers and internal primers (information available from the authors) with BigDye™ Terminator Cycle Sequencing V2 or V3.1 (ABI PRISM® Applied Biosystems) in ABI PRISM™ 377 or 3730 DNA sequencers (Applied Biosystems). Consensus sequences were obtained, from both strands, with Bioedit (Hall, 1999) and Clustal W (Thompson et al., 1994). The sequences were aligned so that gaps (all multiples of three) corresponded to amino acids. Sites with such gaps were manually removed from all alignments prior to phylogenetic analyses.

Phylogenetic trees were generated from the entire DNA sequence. In addition, and in order to test the robustness of the phylogeny, trees were built from the protein sequence, from non-degenerate and degenerate nucleotide sites and from all codon positions in the DNA sequence. The trees were built with Neighbor-joining and maximum parsimony in the MEGA version 2.1 (Kumar et al., 2001) and with maximum likelihood (DNAML and PROTML) in the PHYLIP package [Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package). Seattle, Distributed by the author. Department of Genetics, University of Washington]. In PHYLIP, all tests randomised the input order of sequences (10 replicates), with global rearrangements and the most thorough search option. Bootstrap tests were done for 1,000 replicates (Neighbor-joining) or 100 for maximum parsimony and maximum likelihood. Two maximum likelihood (ML) analyses were performed, one using an optimization of conditions, with empirical transition/transversion (ts/tv) ratios and base frequencies, and constant rate of variation among sites. The second ML method estimated ts/tv and the alpha parameter of the gamma distribution using a reiterative heuristic search (PAUP*4.0), these being ts/tv = 1.29 and $\propto = 0.31$, respectively. The gamma distribution models mutation rate heterogeneity across a locus. Taking into account that *gp63* are highly GC rich, we also produced a Neighbor-joining tree with LogDet transformed distances.

Recombination domains identified visually were tested statistically by measuring the phylogenetic correlation between different gene sequences with the programme PhylPro (Phylogenetic Profiles for Win32 PhylPro (beta) Version 0.90) (Weiller, 1998). This programme determines

Table 1
Leishmania donovani complex strains

WHO code	Zymodeme MON[a]	*mspC* group	Sequence code
Leishmania aethiopica			
MHOM/ET/1972/L100[b]	14	–	AJ532831
Leishmania arabica			
MPSM/SA/1983/JISH220[b]	(64)	–	AJ532830
Leishmania tropica			AJ495008
MHOM/SU/1974/K27[b]	60	–	AJ495009
Leishmania turanica			
MRHO/SU/1983/MARZ-051	(81)	–	AJ495010
L. donovani			
IMAR/KE/1962/LRC-L57	37	A	AJ290774
MCAN/KE/0000/D2	36	A	AJ290776
MHOM/ET/0000/Ayele8	(56)	A	AJ290777
MHOM/KE/1967/MRC(L)3	37	A	AJ010238[e]
MHOM/KE/1973/MRC74	2	A	AJ010239[e]
MHOM/KE/1980/Ndandu 4A	(44)	A	AJ290770
MHOM/KE/1975/Mutinga H9	32	A	AJ290780
MHOM/KE/0000/Neal-R1	(56)	A	AJ495002
MHOM/KE/1954/LRC-L53	36	A	AJ495003
MHOM/IN/1980/DD8[b]	2	B	AJ010235[e]
MHOM/IN/1977/Chowdhury-X	(41)	B	AJ290783
MHOM/IN/1982/Patna 1	(41)	B	AJ010236[e]
MHOM/IN/1979/STL1-79	(41)	B	AJ290784
MHOM/ET/1967/HU3 (LV9)[b]	18	C	AJ010237[e]
MHOM/ET/1972/Gebre 1[c]	82	C	AJ290778
MHOM/SD/1982/Gilani[d]	30	C	AJ290781
MHOM/LB/1984/Salti 4	(46)	C	AJ290771
MHOM/SD/0000/Khartoum	18	C	AJ290772
MHOM/SD/1985/A22	(48)	C	AJ290773
MCAN/IT/1976/Dora	(50)	C	AJ290775
MHOM/PT/1992/IMT 180	18	C	AJ290782
MHOM/ET/0000/Hussen	31	E	AJ290769
MHOM/ET/1984/Addis 164	83	E	AJ290779
MHOM/ET/0000/Ayele 5	(52)	E	AJ290768
MHOM/SA/1987/VL29	(42/52)	E	AF267730
MHOM/CN/0000/Wangjie 1	35	F	AJ495004
MCAN/IQ/1981/SUKKAR2	3	G	AJ495005
MHOM/IQ/1977/BUMM3	(43)	G	AJ495006
MHOM/SU/1984/MARZ-KRIM	73	H	AJ495007
Leishmania infantum ss			
MHOM/TU/1980/IPT-1[b]	1	D	AJ010234[e]
MHOM/MT/1985/Buck	78	D	AJ290785
IARI/PT/1989/IMT 171	24	D	AJ010240[e]
MHOM/CN/1980/Strain A	24	D	AJ010241[e]

In italics are partial sequences. Key to WHO codes: host/country/year/strain name. Hosts: IARI – *Phlebotomus ariasi*, IMAR – *Phlebotomus martini*, MCAN – *Canis familiaris*, MHOM – *Homo sapiens*, MPSA – *Psammomys obesus*, MRHO – *Rhombomys opimus*. Countries: CN, China; ET, Ethiopia; IN, India; IQ, Iraq; IT, Italy; KE, Kenya; LB, Lebanon; MT, Malta; PT, Portugal; SA, Saudi Arabia; SD, Sudan; SU, former Soviet Union; TU, Tunisia. 0000 is unknown year.

[a] Zymodemes are MLEE profiles; in parenthesis are LON zymodemes shown where MON zymodemes are not available.

[b] WHO reference strains.

[c] A strain previously classified as *L. archibaldi*.

[d] A strain previously classified as *L. infantum*.

[e] Completed or longer versions of partial sequences obtained previously (Mauricio et al., 1999).

the pairwise distances of all sequences in a series of sliding-windows along the sequence and evaluates, for each sequence, the degree to which the patterns of distances agree. Windows with low agreement, or low phylogenetic correlation, are indicative of boundaries between recombinant and non-recombinant regions. $Gp63^{EXT}$ sequences were compared with different classes of $gp63^{SHORT}$. For further confirmation, analyses were also performed with Recombinant Detection Program (RDP2; Martin et al., 2005), with the following recombination detection methods

used in an automated search: RDP, GENECONV (Padidam et al., 1999), Bootscan (Salminen et al., 1995), MaxChi and Chimaera (Maynard-Smith, 1992; Posada and Crandall, 2001) and SiScan (Gibbs et al., 2000).

Given that GP63 is an antigenic protein, it could be expected to be under selective pressure by the host. We thus tested neutrality of evolution in this gene using a codon-based *Z*-test (Nei and Kumar, 2000) in MEGA (Kumar et al., 2001). We chose this test because it is based on the relative frequencies of synonymous and non-synonymous mutations and our sample was not population based. It tests the null hypothesis (H_0) that H_0: $dN = dS$, where the number of non-synonymous substitutions per non-synonymous site (dN) is equal to the number of synonymous substitutions per synonymous site (dS), using a *Z*-test where $Z = (dN - dS)/\sqrt{(\sigma^2(dS) + \sigma^2(dN))}$ (σ^2 = variance). Neutrality was rejected if $P < 0.05$. In addition, refined ML analysis of positive selection was performed by splitting the data set into two separate alignments for non-recombining domains and the conserved, recombining domains. The phylogeny for each alignment was performed using a LogDet Neighbor-joining method. Each region was then subject to ML positive selection analysis using a codon model (PAML) and two estimates of site specific selection, firstly, by modelling variability of dN/dS per site, hereafter referred to as ω, by comparing a one-ratio model (M0 – PAML) against a site heterogeneity model (M3 – PAML) and secondly, by examining directly if $\omega > 1$ for each specific site incorporating the beta distribution, viz. beta versus beta ω. The significance between the models was determined using a likelihood ratio test (Yang, 2001).

3. Results

3.1. Subgenus L. (Leishmania) $gp63^{EXT}$ genes

$Gp63^{EXT}$ was fully sequenced for 25 *Leishmania* strains and for reference *L. (Leishmania)* strains for which the gene could be amplified using primers C1F and C8R (Table 1). No specific band was amplified for *L. amazonensis* (MHOM/BR/1973/M2269), *L. gerbilli* (MRHO/CN/1960/Gerbilli), *L. major* (MHOM/SU/1973/5-ASKH), *L. mexicana* (MHOM/BZ/1982/BEL21) and *Leishmania venezuelensis* (MHOM/VE/1974/PM-H3) reference strains.

The $gp63^{EXT}$ open reading frame varied in size between groups of species (Fig. 1a). Two Stop codons were identified, the first having been lost in both *L. tropica* and *L. aethiopica*, thus extending the open reading frame in these species. Putative in-frame deletions and insertions were found in all OW *Leishmania* sequences (Fig. 1a), particularly in the 5′end, but including one allele in *L. tropica* with a deletion of 16 codons in the 3′end. Average DNA sequence diversity varied from 1.4% (2.3% for amino acids) within *L. donovani* to 12.4% (21% for amino acids) between *L. mexicana* and OW *Leishmania* (Table 1, Supplementary data), which was associated with high genetic diversity at the first and second codon positions.

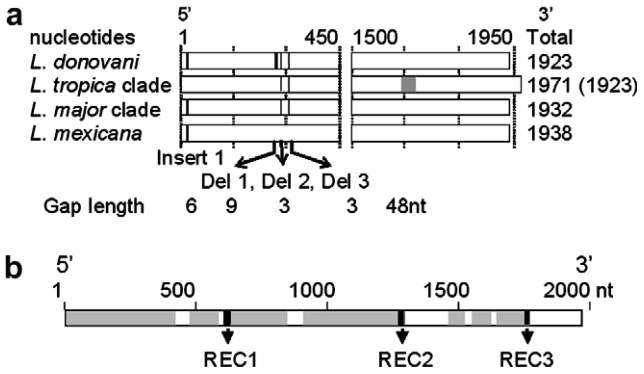

Fig. 1. Schematic representations of subgenus *Leishmania* (*Leishmania*) $gp63^{EXT}$ DNA sequences. (a) Distribution and size of gaps in the nucleotide alignment in the different species. The grey area in the 3′end of the *Leishmania tropica* clade sequence is a deletion present in only one allele. Vertical lines are gaps, which are all multiples of three and correspond to codons. Insert 1 is an extension from three to five serines in the *L. tropica* and *Leishmania aethiopica* protein sequences. Del 1 is a probable deletion in the *Leishmania donovani* complex, as it is present in all other species. Del 2-3 may have been deletions in the ancestor of all Old World *Leishmania* or insertions in the ancestor of *Leishmania mexicana*. (b) Distribution of putative *L. donovani* $gp63^{EXT}$ gene conversion domains (in grey) and of group specific gene conversion domains, REC1-3 (in black). Most of the 5′end is putatively recombinant and most of the 3′end is non-recombinant.

3.2. Gene conversion in gp63 genes

Visual inspection of $gp63^{EXT}$ alignments revealed clusters of single nucleotide polymorphisms (SNPs) in some *L. donovani* complex sequences (Fig. 2). Some of these clusters were found in a number of strains to be virtually identical to the corresponding regions in the published sequences of $gp63^{SHORT}$, which, in the remaining strains, were otherwise clearly distinct from those of $gp63^{EXT}$. All further analyses of the data suggested that these regions have probably been acquired by $gp63^{EXT}$ in some strains by gene conversion from $gp63^{SHORT}$ (see below), although the reverse might also be possible. The regions involved are here named as REC1 at 583–606 nucleotides, REC2 at 1204–1230 nucleotides and REC3 at 1693–1716 nucleotides in $gp63^{EXT}$ (Fig. 1b and 2, and full alignment in Fig. 3 of Supplementary data). As a consequence, we hypothesised that gene conversion had a role in maintaining identity of the regions that are similar across classes of genes in the *L. donovani* complex (LdConv). This is in contrast with the regions that are different across gene classes (LdNon-Conv) (Fig. 1b). We investigated if these regions had different characteristics that would support the hypothesis.

First, a phylogenetic correlation analysis (PhylPro) showed valleys of low correlation (indicative of recombination) flanking REC1-3 and, to some extent, also LdConv (Fig. 3). In further analyses, RDP2 GENECONV and Bootscan supported these findings and independently confirmed the $gp63^{SHORT}$ origin of REC regions (Table 2, Supplementary data). Second, the phylogenies of LdConv and LdNon-Conv domains for all *gp63* gene classes were distinct in an analysis with all available *Leishmania* species.

```
                                 REC1                         REC2                  REC3
                           5'    590         600                     1220                 1700          3'
                                 ..|....|....|....|....|..    ....|....|....|....    ..|....|....|....
Lc-mspC M80671                   CCGGCGCACATCACTGAAGGCTTCA    ATGAGAACGAGGTGACTAT    GCCAACGTCAAGGGAGC
Lc-mspS4 L19562                  ..............C..T...C.G.    ..............G....    .G......GC...CT..
Lc-mspS2 L19563                  ..............C..T...C.G.    .C---.G...T.ACG.CC.    .G......GC...CT..
Lc-mspL/S M80672/M80669          ..............C..T...C.G.    ...................    .G......GC...CT..
Kenya (MRC(L)3)                  .A.......................    ...T..GT.T..ACGT.G.    .................
MRC(L)3 SHORT                    ..............C..T...C.G.    .Y..SRMYKMSKYR.RN.K    ND
India                            ..............C..T...C.G.    ...T..GT.T..ACGT.G.    ........G........
SUKKAR2|BUMM3                    ..............C..T...C.G.    ...T..GT.T..ACGTCG.    ........G........
MARZ-KRIM                        ..............C..T...C.G.    ......GT.T..ACGTCG.    ........G........
Sudan (LV9)                      ..............C..T...C.G.    ...T..GT.T..ACGTCG.    ........G........
LV9 SHORT                        ..............C..T...C.G.    .YR.S..Y..S.Y..YK.K    ND
Ethiopia/SA                      .A.......................    ...................    .G......GC...CT..
Wangjie 1                        .A.......................    ......G............    .G......GC...CT..
L. infantum ss (IPT-1)           .........................    ...................    .................
IPT-1 SHORT                      ..............C..T...C.G.    ...................    ND
L. tropica                       .A............C..G.......    ......G.....ACGTC..    .G...............
L. aethiopica                    .A............C..G.......    ......G.....ACG.C..    .G............T..
L. turanica                      .A............CA.G.......    ......GA....ACG.C..    .G...............
L. arabica                       .A............C..N.......    ......G.....ACG.C..    .G...T...........
L. major AF039721                .A............C..G.......    ......GA....ACG.C..    .G......GC...CT..
L. major SHORT Y00647            .A............C..G.......    ......G.....ACG.C..    .G......GC...CT..
L. mexicana X64394               ....A.........G......G.G.    ......GT.C..CC..C..    .G.........A.CT..
L. amazonensis L46798            .........G....C.G....G.G.    ......GT.C..ACG.C..    .G....CGC....CTAT
```

Fig. 2. Alignments of *gp63* sequences showing REC regions. In the boxes are comparisons of sequences for *gp63*^{EXT} and *gp63*^{SHORT} for the same strain. The latter includes preliminary DNA sequence data for MRC(L)3, LV9 and IPT-1, obtained by direct sequencing of *gp63*^{SHORT} PCR products. These sequences show that *gp63*^{SHORT} in the same strain have REC sequences for the relevant strains. Standard codes are used for ambiguous bases.

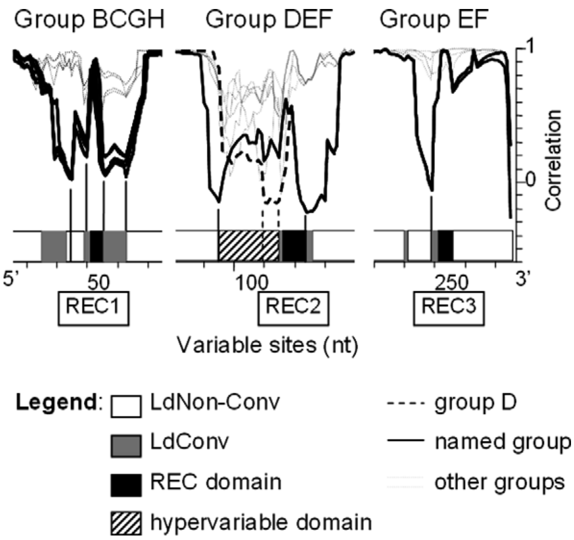

Fig. 3. Phylogenetic correlation profile generated by PhylPro with a sliding window of six differences for all *Leishmania donovani* *gp63*^{EXT}. Only variable sites are shown for clarity and space-saving purposes. Low correlation valleys (<0) indicate putative recombination sites, marked by vertical lines, and flank putative gene conversion domains. *Leishmania* *gp63*^{SHORT} sequences (M60048, M80669, M80672, U48798, and Z83677) used as references are not shown. LdConv have the strongest recombination signals at REC domains. Phylogenetic correlation is very close to 1 in the regions not shown in the figure.

Specifically, sequences clustered according to species in LdConv, showing concerted evolution of the different gene classes, and to gene class in LdNon-Conv (Fig. 4). Third, a preliminary sequence analysis of the sequences of *gp63*^{SHORT} in strains selected to represent REC regions 1 and 2 (LV9 and IPT-1) or their absence (MRC(L)3) showed that the large majority of SNPs in LdConv regions across the *L. donovani* complex are present in both classes of *gp63* in the same strain (approximately 90%). As expected, all other regions had much lower agreement: namely polymorphic regions, REC or otherwise, had only about 30% shared SNPs, whereas the two double nucleotide polymorphisms (GC) had about 60% (data not shown). Fourth, this preliminary analysis has shown that REC sequences are present in all studied *gp63*^{SHORT} genes. Fifth, in *L. donovani* LdConv had much higher GC3 content than LdNon-Conv (Fig. 5), as expected for recombinant regions under the biased gene conversion model (Birdsell, 2002). Furthermore, GC3 and GC1 content in LdNon-Conv were higher in *gp63*^{SHORT} than in *gp63*^{EXT}, which has a lower copy number and has thus lower chances to recombine. Sixth, we found different putative regions of gene conversion within the *L. donovani* complex and within other OW *Leishmania* species (Fig. 3, Supplementary data). Seventh, we found that neutrality could not be statistically ruled out against purifying selection in LdConv, showing that selection does not explain sequence conservation across these regions. Instead, we found some evidence for positive selection (see Section 3.4). Finally, most deletions and insertions were found in repeat regions and are thus consistent with alignment errors during gene conversion.

3.3. Phylogenies and taxonomy

The *gp63*^{EXT} gene produced robust phylogenies of *L. donovani* and *L. (Leishmania)*, which were congruent

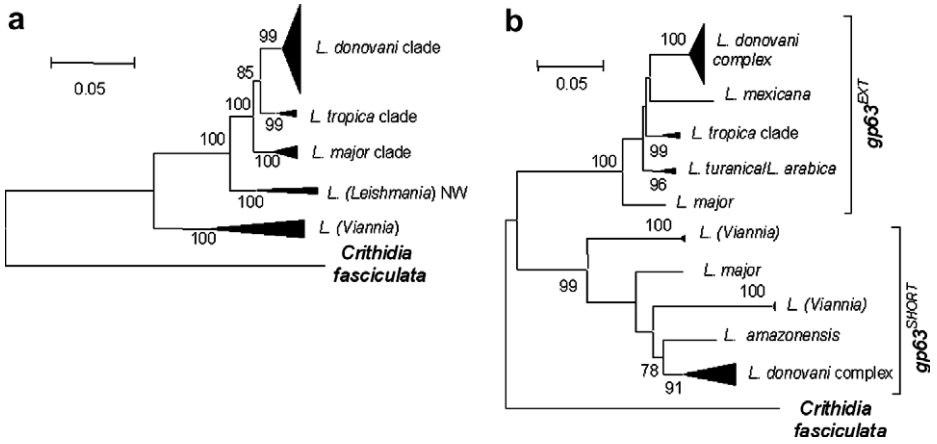

Fig. 4. Phylogenetic trees for putative (a) LdConv and (b) LdNon-Conv domains (and their equivalent in other species and gene classes), built with Neighbor-joining from Kimura-2 parameter distances with 1,000 bootstrap replicates (values above 50% shown) and complete deletion option. The trees were rooted by *Crithidia fasciculata* and some branches were collapsed for clarity of presentation. The LdConv tree was built from an alignment of 1,329 sites, of which 1,214 were used to build the tree and 306 were parsimony-informative sites. The LdNon-Conv tree was built from 447 sites, of which 424 were used in the tree and 272 were parsimony-informative. The LdConv tree shows the phylogeny of *Leishmania*, whilst the LdNon-Conv tree shows the phylogeny of the genes.

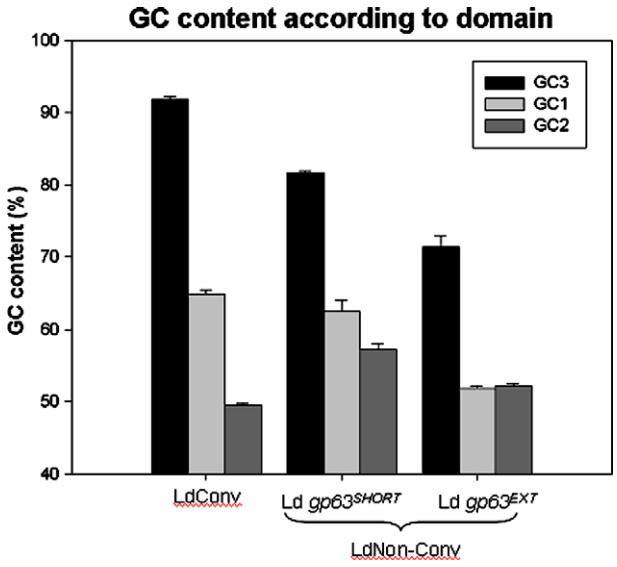

Fig. 5. Comparison of GC content per codon position in *Leishmania donovani gp63* genes according to class of gene and type of domain. GC3 and GC1 are higher in LdConv than in LdNon-Conv domains, and GC1 are higher in $gp63^{SHORT}$ than in $gp63^{EXT}$ genes. SD bars are shown.

even using different methods and data types, including the use of GC corrected distances and accommodating for different rates amongst sites (see Section 2) (bootstrap support >75%) (Fig. 6; and data not shown).

Three main clades were identified in OW *Leishmania*: the *L. donovani* clade, a second clade with *L. tropica* and *L. aethiopica* (*L. tropica* clade) and a third with *L. major*, *L. turanica* and *L. arabica* (*L. major* clade; Fig. 6). Deletion and insertion events were diagnostic for each of these clades (Fig. 6).

The *L. donovani* clade had a total of eight robust sub-clades (>75% bootstrap support with at least one phylogenetic method), which were here named as groups A–H (Table 1, Fig. 6). LdConv REC1-3 in $gp63^{EXT}$ were diagnostic, respectively, of clades BCGH, DEF and EF (Fig. 6). Eighteen candidate group-specific SNPs were identified within the *L. donovani* complex, covering all groups but F (Table 2).

Leishmania donovani groups correlated with geographical distribution and isoenzyme type, or zymodeme (Table 1). Group D (*L. infantum*) did not form a lineage separated from *L. donovani*. Zymodemes MON-30 (“*L. infantum*”), MON-82 (“*L. archibaldi*”) and MON-18 (*L. donovani*) formed a single clade.

3.4. $Gp63^{EXT}$ and selection pressure

GP63 are immunogenic proteins with a role in cell invasion and complement avoidance in the mammalian host, so *gp63* genes would be expected to be under selective pressure. However, neutrality was not rejected against positive selection by the codon based test of selection. Instead, there was evidence for purifying selection only for regions unique to $gp63^{EXT}$ (LdNon-Conv) and, for *L. donovani*, only in group A. Therefore, we analysed selection at individual sites. In both data sets comprising LdNon-Conv and LdConv, highly significant between-site heterogeneity was observed for ω (dN/dS per site), viz. $2\Delta L = 27.8$, $P \ll 0.0001$ (χ^2) at 4 df (degrees of freedom) LdNon-Conv and $2\Delta L = 89.1$, $P \ll 0.0001$ (χ^2) at 4 df for LdConv. Using the beta models of identifying positive selection at given sites, however, LdNon-Conv was equivocal, viz. $2\Delta L = 7.8$ $P = 0.02$ (χ^2) at 2 df, $\omega = 2.9$, and whilst 11 sites showed $\omega > 1$, only 1 site was marginally significant ($P = 0.96$). In contrast, LdConv showed strong, highly significant site specific selection, viz. $2\Delta L = 34.3$, $P \lll 0.0001$, $\omega = 4.0$, with 44 nucleotide sites showing $\omega > 1$ with two sites showing highly significant selection

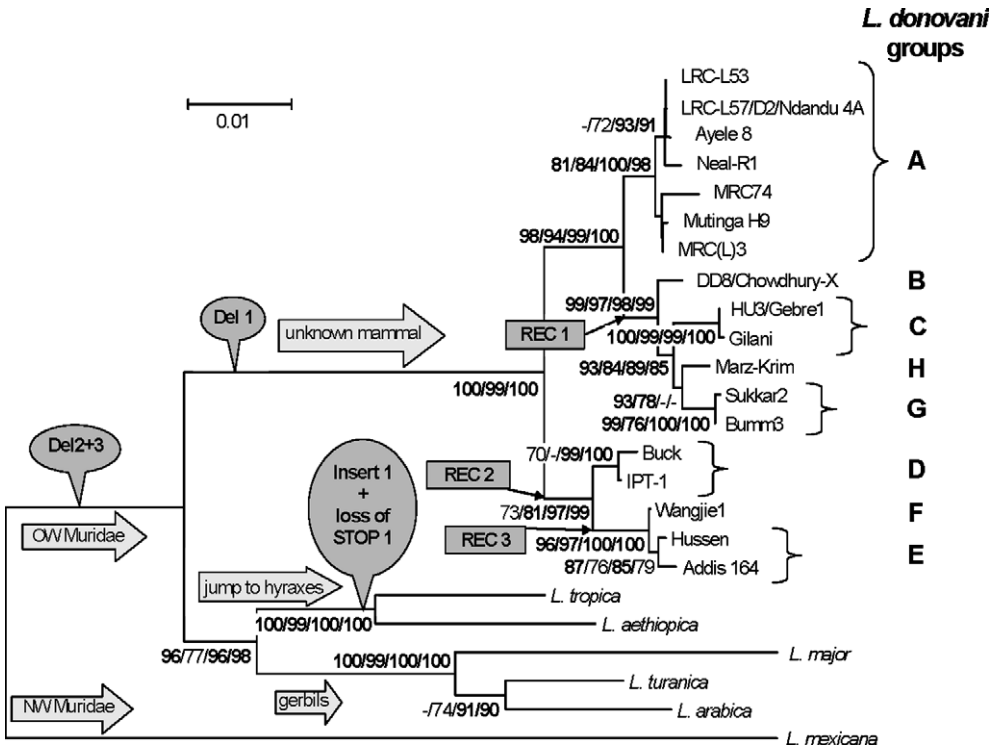

Fig. 6. Rooted Neighbor-joining tree of $gp63^{EXT}$ DNA coding sequences with MEGA, Kimura-2-parameter, for all $gp63^{EXT}$ of the subgenus *Leishmania* (*Leishmania*). Alignment of 1,911 bp with 227 informative sites. Bootstrap values >70% from, respectively, 100 (maximum likelihood and parsimony – MEGA) or 1,000 replicates (Neighbor-joining with Kimura-2 parameter and LogDet distances) are shown. Significant evolutionary events are depicted: insertion, deletions and acquisition of REC domains, as in Fig. 1, and loss of a stop codon in the first position. Shaded arrows show hypothesised parasite-host co-evolution. *Leishmania mexicana* was used here as an outgroup to Old World *Leishmania*.

($P > 0.99$) and three additional sites showing significant selection ($P > 0.95$). There was no apparent association between any of these sites and function or location, except for one located in the active site (see below). Overall, the average global dN/dS was also higher for LdConv; for example, using the beta ω models, this was 0.54 for LdNon-Conv compared with 0.94 for LdConv. In summary there appeared to be far greater selection pressure on LdConv than LdNon-Conv.

In addition, we investigated regions of known functional significance, starting with those involved in GP63 protease activity. The zinc binding domain (HEMAH) was conserved throughout all *L. (Leishmania) gp63*^{EXT} genes. However, the proposed active site (Morales et al., 1997) was polymorphic. Its putative ancestral sequence was CNESE and only the CN motif was conserved throughout the subgenus (CNE**N**E in group ED, CNES**V** in group H, CN**V**S**V** in groups ABCG, CNE**R**E in *L. major* and *L. turanica*). Only two amino acid changes were not charge conservative (S to R and E to V). In fact, the fourth site (with variations of S/N/R) had a highly significant score ($P > 0.99$) for positive selection and another site, three amino acids downstream scored significantly ($P > 0.95$). These are located within REC1.

We further investigated regions of *gp63* genes that are known to be involved in interaction with the mammalian host. Internalization domains identified in *L. guyanensis* (subgenus *Viannia*) (Puentes et al., 1999) and the macrophage internalization domain SRYD (Soteriadou et al., 1992) were conserved in $gp63^{EXT}$. However, the putative fibronectin binding site was EYLEV in the *L. major* clade (Brittingham et al., 1999) but EYLE**I** for *L. donovani* and *L. tropica*, and **D**YLE**I** for *L. aethiopica*, without charge alterations. Immunodominant domains located between positions 459 amino acids and 517 amino acids of *L. donovani* $gp63^{EXT}$ were polymorphic but the number of polymorphic amino acid sites (8%) was similar to the gene average (8.2%); these domains were conserved within clade DEF and groups A and B (Fig. 1, Supplementary data).

4. Discussion

4.1. Comparison of $gp63^{EXT}$ genes

As far as we are aware, this study is amongst the most comprehensive analyses of *L. donovani* and of OW *Leishmania* using a coding sequence, and of the intra-specific and intra-subgeneric diversity of a class of *gp63* genes. This complements the ribosomal internal transcribed spacer study of OW *Leishmania* (Berzunza-Cruz et al., 2002), the studies on the intergenic regions of *gp63* (Mauricio et al., 2001), ITS (Kuhls et al., 2005), ITS with mini-exon (Mauricio et al., 2004), cytochrome oxidase II (Ibrahim

Table 2
Alignment of clade-specific regions of *Leishmania donovani* complex $gp63^{EXT}$ coding sequences

Sequence position[a]		**162**	**173**	**243**	**404**	**529**	**545**	**768**	**834**	**835**	**841**	873	874	**975**	**1341**	**1356**	**1382**	**1455**	1477	1694	1702	1706	1704	**1775**	**1854**	**1863**
Codon position		**3**	**2**	**3**	**2**	**1**	**2**	**3**	**3**	**1**	**1**	3	1	**3**	**3**	**3**	**2**	**3**	1	2	1	2	3	**2**	**3**	**3**
Specific from OW[b]		+	+	+	+	+	–	+	+	+	–	–	–	+	+	+	+	+	+	–	–	–	–	+	–	–
Other Ld base		G	G	C	C	A	A	G	C	C	A	T	A	C	G	T	A	G	A	C	A	G	A	T	G	T
Group	**A**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.
	B	.	A	.	T	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	C	.	.	C	.	.	.	.	T	.	G	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.
	G	C	.	.	.	.	.	A	.	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
	H	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G
	D	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	E	.	.	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	C	G	C	C	T	C	.	.
	F	.	.	.	.	.	.	.	.	.	.	C	G	.	.	.	.	.	C	G	C	C	T	.	.	.

Dots indicate identity to the reference base; bold positions are specific for only one group; sites also specific in relation to other Old World (OW) *Leishmania* species are indicated by +; group F does not have a single specific site, so sites specific for clade EF are shown.

[a] From Start codon.

[b] Yes (+) or No (–).

and Barker, 2001) and enzyme coding genes (Mauricio et al., 2006).

$Gp63^{EXT}$ genes varied in size due to a combination of different mechanisms: in frame deletions and insertions and a SNP that changed the first stop codon site in the *L. tropica* clade. Most gaps lay within the signal peptide and thus may have some effect on expression of this protein. The deletion and extension of the carboxyl terminus in *L. tropica* may alter protein attachment, if this region is a trans-membrane domain as proposed (Voth et al., 1998).

4.2. Gene conversion in gp63 genes

Mosaic gene conversion was identified between $gp63^{EXT}$ and $gp63^{SHORT}$, as suggested previously but not studied in detail (Medina-Acosta et al., 1993; Steinkraus et al., 1993; Alvarez-Valin et al., 2000; Victoir and Dujardin, 2002). Here, we identified putative recombinant domains in the *L. donovani* complex $gp63^{EXT}$ by comparison with *L. infantum* and *L. donovani* $gp63^{SHORT}$ genes. Despite the small number of *L. donovani* strains with $gp63^{SHORT}$ gene sequences available, the comparatively low number of SNPs suggested that they provide a good representation of the genetic diversity for this gene class in *L. donovani*. This was confirmed by analysis of the *L. infantum* genome and in a preliminary analysis of $gp63^{SHORT}$ in selected strains (MRC(L)3, HU3 and IPT-1; data not shown), which also showed that REC regions 1 and 2 are present throughout the *gp63* gene family and are thus more likely to have been acquired by gene conversion than by sexual recombination events, which are considered to be rarer in *Leishmania*. We have excluded the possibility that the gene conversion regions are artefacts produced by formation of chimeras during PCR, because we have directly sequenced the PCR products and have not cloned them. Furthermore, we have found the same inserts in strains shown to be related by other methods and targets (Ibrahim and Barker, 2001; Mauricio et al., 2001, 2006; Jamjoom et al., 2004; Zemanova et al., 2004; Kuhls et al., 2005).

Evidence for gene conversion was considerable. It appeared particularly strong in REC domains, as identified by several recombination detection methods. For example, with PhylPro, REC domains as well as some LdConv regions were flanked by regions of low phylogenetic correlation, indicating recombination sites. There was no evidence for negative selection in maintaining conservation at LdConv domains. In contrast, sites with statistically significant results for positive selection were more frequent in LdConv than in LdNon-Conv domains. Further investigation is required to minimise the recombination signal from the sites identified as being positively selected before a clear understanding is obtained. However, an association between gene conversion and positive selection may be creating genetic diversity at surface domains that are exposed to strong selection pressure.

In addition, we have shown that gene classes evolved in a concerted fashion in each lineage at LdConv domains, whilst evolving independently at LdNon-Conv. We have also uncovered preliminary evidence that in LdConv regions the different classes of genes are more similar in the same strain than between strains. Furthermore, the rates of synonymous and non-synonymous substitutions were of the same order of magnitude, only 20–50% higher depending on the method used. This is a more likely effect of gene conversion rather than a "birth-and-death" model of evolution, such as has been detected in human MHC and Ig loci (Rooney et al., 2002), in which mutations spread throughout a gene family by gene duplication and deletion.

As indirect evidence of gene conversion, the highest GC3 content was found in LdConv and in regions in $gp63^{SHORT}$ that are in the same relative locations as $gp63^{EXT}$ LdNon-Conv, as expected under the biased gene conversion (BGC) model (Birdsell, 2002). Regions in $gp63^{SHORT}$ that are equivalent to $gp63^{EXT}$ LdNon-Conv are conserved amongst $gp63^{SHORT}$, and are thus also likely to be affected by gene conversion. The BGC model proposes that allele recombination causes high synonymous GC content because mismatches produced by recombination are mostly corrected to C or G rather than to A or T, particularly in organisms with GC bias, such as *Leishmania*. The heterogeneity of GC3 content within the same gene cannot be explained by higher expression levels, as shown for differences between genes in *Trypanosoma sp.* (Alvarez et al., 1994). This was further supported by our observations that non-translated regions (NTR) of the *gp63* tandem repeat with the highest copy number, such as all 5′NTR and *mspL* 3′NTR, also had higher GC content than domains with lower copy number such as *mspC* 3′NTR (data not shown). We thus propose that a high GC3 content signature may also be generated by gene conversion in *Leishmania*, which is more frequent in regions with high copy number.

In the genome sequences of *L. major*, *L. infantum* and *L. braziliensis*, *gp63*-like genes are reported on two different chromosomes (31 and 28). These genes have markedly different sequences from the typical *gp63* genes studied here and so were not analysed, but they are conserved between species and are probably active genes. However, they are not likely to be involved in gene conversion with $gp63^{EXT}$ and $gp63^{SHORT}$ as there are no large regions of similarity and their GC3 content is relatively low (<77% in all three species). It is possible that these genes moved to different parts of the genome early in *Leishmania* evolution and due to lack of physical proximity had less chance to recombine with the clustered *gp63* genes. Alternatively, they may be cases of convergent evolution.

Alvarez-Valin et al. (2001) have found a pattern of strong co-variance of synonymous and non-synonymous substitutions, which can normally be attributed to gene conversion. However, they discarded this explanation of their data on the assumption that repeated gene conversion would be required in the same region of the gene, in distinct lineages. Here, gene conversion in *gp63* does seem to occur repeatedly in the same low divergence regions, and thus may be a factor in maintaining co-variance of synonymous and non-synonymous substitutions in regions where divergence is fuelled primarily by mutation. Indeed, as predicted by these authors, we found different patches of divergence (from putative rare gene conversion events) between distinct *L. donovani* complex lineages and other OW *Leishmania*.

We have searched for regions, eg. CCTCCCT and CCCCACCCC (Hellenthal and Stephens, 2006), involved in homologous recombination, but did not find any. Gene conversion in these genes is likely to be driven by large homology regions (Papadopoulou and Dumas, 1997), instead of specific signals. It is, however, possible that the short repeats found at the 5′end may facilitate recombination. We have also investigated LdConv and LdNon-Conv for correlation with function or location. As expected, functional regions, such as the Zinc binding site, the fibronectin binding site and the macrophage internalization site were located in LdConv regions. However, part of the active site was located in a REC region, thus creating differences within the same species. The immunodominant region (Fig. 1, Supplementary data) is located largely in a LdNon-Conv region, which may create a different immune signature for the two classes of gene. LdNon-Conv regions concentrate in three regions exposed to the surface, which may be important for exposure to the immune system. However, most of the protein surface is made up by LdConv regions (Fig. 2 and movies 1 and 2, Supplementary data). Similarly, REC regions were present in a mix of surface and interior domains.

Mosaic gene conversion in *Leishmania* has been proposed (Victoir and Dujardin, 2002) to have adaptive effects that are similar to sexual recombination, particularly in multigene families, by creating diversity and by spreading advantageous mutations across alleles. This study of $gp63^{EXT}$ confirms this idea, as errors in gene conversion are likely to have produced the observed deletions and insertions and to have introduced clusters of polymorphic sites in some lineages. However, here we have also found that gene conversion may have an additional or alternative effect in *gp63* as a cause of homogeneity between alleles and classes of genes. Gene conversion may cause the observed high homozygosity level of $gp63^{EXT}$ in *L. donovani* by spreading or eliminating mutations. This is further supported by the higher level of heterozygosity found in enzyme coding genes (Mauricio et al., 2006), which are single copy and thus less likely to suffer gene conversion.

4.3. Phylogenies and taxonomy

$Gp63^{EXT}$ was an excellent target for phylogenetics of *L. donovani* and OW *Leishmania* because it is polymorphic within genetic groups and produced robust phylogenies. The recombinant regions did not alter the phylogenetic signal in $gp63^{EXT}$, as they evolve in concert within the gene family. This is in apparent contradiction with the observation that LdConv and LdNon-Conv domains produce

distinct phylogenies for all available *Leishmania gp63* sequences (respectively, by species and by gene class). However, in the LdNon-Conv tree the species phylogenies are evident within each gene class (Fig. 4).

Phylogenetic analyses of this gene confirmed and added to our previous independent findings on genetic diversity of the *L. donovani* complex (Mauricio et al., 2001, 2004, 2006). Specifically, clades A–F corresponded to groups 1–6 described from restriction analysis of *gp63* intergenic regions, ITS and mini-exon, and to the groups "Kenya", "India", "Sudan", *L. infantum sensu stricto*, "Saudi Arabia/other" and "China" of multilocus sequence typing (MLST). Here we further confirmed that zymodemes MON-18, -30 and -82 form a single clade ("Sudan"), that "*L. archibaldi*" is not a valid species and that Sudanese "*L. infantum*" is one of many genetic groups of the *L. donovani* complex (Ibrahim and Barker, 2001; Mauricio et al., 2001, 2006; Jamjoom et al., 2004; Zemanova et al., 2004; Kuhls et al., 2005). This is in contradiction to the published phylogenies based on MLEE (Rioux et al., 1990; Pratlong et al., 2001) that show Sudanese *L. donovani, "L. infantum"* and "*L. archibaldi*" as clearly separate species. Instead, the $gp63^{EXT}$ phylogeny is consistent with that produced by MLST of five genes for enzymes used in MLEE (Mauricio et al., 2006). We can thus conclude that the MLEE data and *gp63* data are compatible, but that the current MLEE based phylogenies are not reliable. The exception to the agreement between MLST and *gp63* phylogenies were strains here separated into groups E and G, which were not distinguished by MLST. It is possible that the MLST genes were not sufficiently discriminatory, or that the $gp63^{EXT}$ phylogeny reflects recombination between the ancestors of genetic groups E and D or G and BCH; a similar phenomenon was found for some genes used in MLST.

The phylogeny of OW *Leishmania* also differed somewhat from that inferred from isoenzymes (Rioux et al., 1990), which had a *L. donovani* complex/*L. tropica* clade. However, it was similar to the ribosomal internal transcribed spacer based phylogeny (Berzunza-Cruz et al., 2002): *L. aethiopica* and *L. tropica* were in the same clade, whereas the *L. donovani* complex was an independent lineage from other OW *Leishmania*.

Due to the high number of polymorphic sites and robustness of the genetic groups, the $gp63^{EXT}$ gene is particularly suitable for genetic typing of *L. donovani* complex groups and of OW *Leishmania*. We have successfully used this gene for identification of the *L. donovani* complex strains (Sharma et al., 2005). The identified putative group-specific diagnostic sites could be applied to rapid identification of strains.

4.4. $Gp63^{EXT}$ and selection pressure

The enzyme GP63 is important in the relationship between *Leishmania* and the host. Function related domains, such as macrophage internalisation domains and the zinc cation binding domain (HEMAH), were conserved throughout the subgenus *Leishmania*. The SNPs in the putative fibronectin binding site (Brittingham et al., 1999) did not alter the protein charge and did not correlate with host preference. However, we have found an apparent association between recombination and positive selection at LdConv domains, which may have a functional basis and should be investigated further. It is likely that selection acts to maintain GP63 epitopes that allow *Leishmania* to manipulate the immune system to its own advantage, rather than evading it, which is supported by experimental evidence of the activity of GP63 in the mammalian host (Joshi et al., 1998). Surprisingly, the catalytic site was polymorphic, involving charge changes that may affect substrate specificity or enzyme activity (e.g. at different pH) and thus possibly survival or virulence within the host. It is remarkable that a REC region affects part of the active site, with one of its most polymorphic amino acids scoring as highly significant for positive selection.

GP63 may be either a marker or a factor for host preference. All main clades of OW *Leishmania* included human pathogens, and species or clades vary from zoonotic to anthroponotic. The *L. major* clade is known to be associated with gerbils (Rodentia, Muridae, Gerbillinae). Within the *L. tropica* clade, *L. aethiopica* is associated with hyraxes (Hyracoidea, Procaviidae), in which strains of *L. tropica* have also been found (Jacobson et al., 2003; Jaffe et al., 2004), despite being anthroponotic throughout most of its range. So, we propose an association between hyraxes and the initial differentiation of the *L. tropica*/*L. aethiopica* lineage from the *L. major* clade (Schonian et al., 2001; Berzunza-Cruz et al., 2002). The *L. donovani* group D (*L. infantum ss*) is associated with canids (Carnivora, Canidae), in which group C has also been found. However, if there is another main host associated with evolution of the *L. donovani* complex it is still a mystery, although rodents and a later association with canids cannot be excluded. Anthroponotic behaviour must be a derived character in OW *Leishmania*, as it is present only in a few lineages, most *L. tropica* and "Indian" *L. donovani*.

We have presented here the $gp63^{EXT}$ gene as an excellent marker for phylogenetics and genotyping of OW *Leishmania*, in particular for the *L. donovani* complex. Our evidence supports a role of GP63 in using the host immune system, rather than evasion through antigenic variability. We have also shown that gene conversion, between different alleles of the same *gp63* gene class and also between different members of this gene family, is a determinant factor in generating genetic diversity, but also in maintaining homozygosity and homogeneity between genes in the *gp63* family.

Acknowledgements

We thank Praxis XXI for personal support to I.L.M. and the European Commission for additional funding

(Contracts No. ERBTS3*CT920113 and QLK2-CT-2001-01810).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpara.2006.11.020.

References

Alvarez-Valin, F., Tort, J.F., Bernardi, G., 2000. Nonrandom spatial distribution of synonymous substitutions in the gp63 gene from *Leishmania*. Genetics 155, 1683–1692.

Alvarez, F., Robello, C., Vignali, M., 1994. Evolution of codon usage and base contents in kinetoplastid protozoans. Mol. Biol. Evol. 11, 790–802.

Berzunza-Cruz, M., Cabrera, N., Crippa-Rossi, M., Sosa Cabrera, T., Perez-Montfort, R., Becker, I., 2002. Polymorphism analysis of the internal transcribed spacer and small subunit of ribosomal RNA genes of *Leishmania mexicana*. Parasitol. Res. 88, 918–925.

Birdsell, J.A., 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. 19, 1181–1197.

Brittingham, A., Chen, G., McGwire, B.S., Chang, K.-P., Mosser, D.M., 1999. Interaction of *Leishmania* gp63 with cellular receptors for fibronectin. Infect. Immun. 67, 4477–4484.

Chakrabarty, R., Mukherjee, S., Lu, H.G., McGwire, B.S., Chang, K.P., Basu, M.K., 1996. Kinetics of entry of virulent and avirulent strains of *Leishmania donovani* into macrophages: a possible role of virulence molecules (GP63 and LPG). J. Parasitol. 82, 632–635.

Chen, D.-Q., Kolli, B.K., Yadava, N., Lu, H.G., Gilman-Sachs, A., Peterson, D.A., Chang, K.-P., 2000. Episomal expression of specific sense and antisense mRNAs in *Leishmania amazonensis*: modulation of GP63 level in promastigotes and their infection of macrophages in vitro. Infect. Immun. 68, 80–86.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hellenthal, G., Stephens, M., 2006. Insights into recombination from population genetic variation. Curr. Opin. Genet. Dev. 16, 565–572.

Ibrahim, M.E., Barker, D.C., 2001. The origin and evolution of the *Leishmania donovani* complex as inferred from a mitochondrial cytochrome oxidase II gene sequence. Infect. Genet. Evol. 1, 61–68.

Jacobson, R., Eisenberger, C.L., Svobodova, M., Baneth, G., Sztern, J., Carvalho, J., Nasereddin, A., El Fari, M., Shalom, U., Volf, P., Votypka, J., Dedet, J.-P., Pratlong, F., Schonian, G., Schnur, L.F., Jaffe, C.L., Warburg, A., 2003. Outbreak of cutaneous leishmaniasis in northern Israel. J. Infect. Dis. 188, 1065–1073.

Jaffe, C.L., Baneth, G., Abdeen, Z.A., Schlein, Y., Warburg, A., 2004. Leishmaniasis in Israel and the Palestinian Authority. Trends Parasitology. 20, 328–332.

Jamjoom, M.B., Ashford, R.W., Bates, P.A., Chance, M.L., Kemp, S.J., Watts, P.C., Noyes, H.A., 2004. *Leishmania donovani* is the only cause of visceral leishmaniasis in East Africa; previous descriptions of *L. infantum* and *L. archibaldi* from this region are a consequence of convergent evolution in the isoenzyme data. Parasitology 129, 399–409.

Joshi, P.B., Sacks, D.L., Modi, G., McMaster, W.R., 1998. Targeted gene deletion of *Leishmania major* genes encoding developmental stage-specific leishmanolysin (GP63). Mol. Microbiol. 27, 519–530.

Kelly, J.M., 1993. Isolation of RNA and DNA from *Leishmania*. In: Hyde, J.E. (Ed.), Methods in Molecular Biology, Vol. 21. New Jersey, Humana Press, pp. 123–131.

Kuhls, K., Mauricio, I.L., Pratlong, F., Presber, W., Schonian, G., 2005. Analysis of ribosomal DNA internal transcribed spacer sequences of the *Leishmania donovani* complex. Microbes Infect. 7, 1224–1234.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Mauricio, I.L., Gaunt, M.W., Stothard, J.R., Miles, M.A., 2001. Genetic typing and phylogenetic analysis of the *Leishmania donovani* complex by restriction analysis of PCR amplified gp63 intergenic regions. Parasitology 122, 393–403.

Mauricio, I.L., Howard, M.K., Stothard, J.R., Miles, M.A., 1999. Genetic diversity in the *Leishmania donovani* complex. Parasitology 119, 237–246.

Mauricio, I.L., Stothard, J.R., Miles, M.A., 2000. The strange case of *Leishmania chagasi*. Parasitol. Today 16, 188–199.

Mauricio, I.L., Stothard, J.R., Miles, M.A., 2004. *Leishmania donovani* complex: genotyping with the ribosomal internal transcribed spacer and the mini-exon. Parasitology 128, 1–5.

Mauricio, I.L., Yeo, M., Baghaei, M., Doto, D., Pratlong, F., Zemanova, E., Dedet, J.-P., Lukes, J., Miles, M.A., 2006. Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int. J. Parasitol. 36, 757–769.

Maynard-Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Medina-Acosta, E., Beverley, S.M., Russell, D.G., 1993. Evolution and expression of the *Leishmania* surface proteinase (gp63) gene locus. Infect. Agents Dis. 2, 25–34.

Morales, G., Carrillo, G., Requena, J.M., Guzman, F., Gomez, L.C., Patarroyo, M.E., Alonso, C., 1997. Mapping of the antigenic determinants of the *Leishmania infantum* gp63 protein recognized by antibodies elicited during canine visceral leishmaniasis. Parasitology 114, 507–516.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Papadopoulou, B., Dumas, C., 1997. Parameters controlling the rate of gene targeting frequency in the protozoan parasite *Leishmania*. Nucleic Acids Res. 25, 4278–4286.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762.

Pratlong, F., Dereure, J., Bucheton, B., El-Safi, S., Dessein, A., Lanotte, G., Dedet, J.P., 2001. Sudan: the possible original focus of visceral leishmaniasis. Parasitology 122, 599–605.

Puentes, F., Guzman, F., Marin, V., Alonso, C., Patarroyo, M.E., Moreno, A., 1999. *Leishmania*: fine mapping of the leishmanolysin molecule's conserved core domains involved in binding and internalization. Exp. Parasitol. 93, 7–22.

Ramamoorthy, R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C., Wilson, M.E., 1992. Three distinct RNAs for the surface protease gp63 are differentially expressed during development of *Leishmania donovani chagasi* promastigotes to an infectious form. J. Biol. Chem. 267, 1888–1895.

Rioux, J.A., Lanotte, G., Serres, E., Pratlong, F., Bastien, P., Perieres, J., 1990. Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. Ann. Parasitol. Hum. Comp. 65, 111–125.

Rooney, A.P., Piontkivska, H., Nei, M., 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol. Biol. Evol. 19, 68–75.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning. AIDS Res. Hum. Retroviruses 11, 1423–1425.

Schonian, G., Schnur, L., el Fari, M., Oskam, L., Kolesnikov, A.A., Sokolowska-Kohler, W., Presber, W., 2001. Genetic heterogeneity in the species *Leishmania tropica* revealed by different PCR-based methods. Trans. R. Soc. Trop. Med. Hyg. 95, 217–224.

Sharma, N.L., Mahajan, V.K., Kanga, A., Sood, A., Katoch, V.M., Mauricio, I., Singh, C.D., Parwan, U.C., Sharma, V.K., Sharma, R.C., 2005. Localized cutaneous leishmaniasis due to *Leishmania donovani* and *Leishmania tropica*: preliminary findings of the study of 161 new cases from a new endemic focus in Himachal Pradesh, India. Am. J. Trop. Med. Hyg. 72, 819–824.

Soteriadou, K.P., Remoundos, M.S., Katsikas, M.C., Tzinia, A.K., Tsikaris, V., Sakarellos, C., Tzartos, S.J., 1992. The Ser-Arg-Tyr-Asp region of the major surface glycoprotein of *Leishmania* mimics the Arg-Gly-Asp-Ser cell attachment region of fibronectin. J. Biol. Chem. 267, 13980–13985.

Steinkraus, H.B., Greer, J.M., Stephenson, D.C., Langer, P.J., 1993. Sequence heterogeneity and polymorphic gene arrangements of the *Leishmania guyanensis* gp63 genes. Mol. Biochem. Parasitol. 62, 173–185.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Victoir, K., De Doncker, S., Cabrera, L., Alvarez, E., Arevalo, J., Llanos-Cuentas, A., Le Ray, D., Dujardin, J.-C., 2003. Direct identification of *Leishmania* species in biopsies from patients with American tegumentary leishmaniasis. Trans. R. Soc. Trop. Med. Hyg. 97, 80–87.

Victoir, K., Dujardin, J.-C., 2002. How to succeed in parasitic life without sex? Asking *Leishmania*. Trends Parasitol. 18, 81–85.

Voth, B.R., Kelly, B.L., Joshi, P.B., Ivens, A.C., McMaster, W.R., 1998. Differentially expressed *Leishmania major* gp63 genes encode cell surface leishmanolysin with distinct signals for glycosylphosphatidylinositol attachment. Mol. Biochem. Parasitol. 93, 31–41.

Weiller, G., 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol. Biol. Evol. 15, 326–335.

WHO, E.C., 1990. Control of the leishmaniases. WHO Tech. Rep. Ser. 793, 1–158.

Yang, Z., 2001. Maximum likelihood analysis of adaptive evolution in HIV-1 GP120 ENV gene. Pacific Symp. Biocomput. 6, 226–237.

Zemanova, E., Jirku, M., Mauricio, I.L., Miles, M.A., Lukes, J., 2004. Genetic polymorphism within the *Leishmania donovani* complex: correlation with geographic origin. Am. J. Trop. Med. Hyg. 70, 613–617.