Meta Gene 2 (2014) 782–798

Contents lists available at ScienceDirect

Meta Gene

Comparative in-silico genome analysis of *Leishmania* (*Leishmania*) *donovani*: A step towards its species specificity

Satheesh Kumar S., Gokulasuriyan R.K., Monidipa Ghosh *

Department of Biotechnology, National Institute of Technology, Mahatma Gandhi Avenue, Durgapur 713209, West Bengal, India

ARTICLE INFO

Article history:
Received 27 June 2014
Revised 12 September 2014
Accepted 4 October 2014
Available online 24 October 2014

Keywords:
Visceral leishmaniasis
Apical Membrane Antigen 1
A2 gene family
Amastin
Species specific genes

ABSTRACT

Comparative genome analysis of recently sequenced *Leishmania* (*L.*) *donovani* was unexplored so far. The present study deals with the complete scanning of *L.* (*L.*) *donovani* genome revealing its interspecies variations. 60 distinctly present genes in *L.* (*L.*) *donovani* were identified when the whole genome was compared with *Leishmania* (*L.*) *infantum*. Similarly 72, 159, and 265 species specific genes were identified in *L.* (*L.*) *donovani* when compared to *Leishmania* (*L.*) *major*, *Leishmania* (*L.*) *mexicana* and *Leishmania* (*Viannia*) *braziliensis* respectively. The cross comparison of *L.* (*L.*) *donovani* in parallel with the other sequenced species of leishmanial led to the identification of 55 genes which are highly specific and expressed exclusively in *L.* (*L.*) *donovani*. We found mainly the discrepancies of surface proteins such as amastins, proteases, and peptidases. Also 415 repeat containing proteins in *L.* (*L.*) *donovani* and their differential distribution in other leishmanial species were identified which might have a potential role during pathogenesis. The genes identified can be evaluated as drug targets for anti-leishmanial treatment, exploring the scope for extensive future investigations.

Introduction

Leishmaniasis is a vector-borne parasitic disease caused by obligate intracellular protozoa of the genus *Leishmania*. Leishmaniasis, an endemic disease of tropical and subtropical regions is the second-largest parasitic killer in the world (after malaria), responsible for an estimated 2 million cases each year and 350 million people at risk worldwide clearly imposing a major health problem globally except Australia and Antarctica

* Corresponding author. Tel.: +91 9433919252, +91 343 2755209 (Office); fax: +91 343 2547375.
E-mail addresses: sathishh18@yahoo.com (S.K. S.), suryah777@gmail.com (G. R.K.), gmonidipanitd@gmail.com, gmonidipa@yahoo.com (M. Ghosh)

http://dx.doi.org/10.1016/j.mgene.2014.10.003

(Desjeux, 2001). Approximately 88 countries are named to be infected with this sand fly borne disease. Although leishmaniasis is represented by at least 20 leishmanial species, the disease spectrum is generally categorized into 1) cutaneous leishmaniasis, 2) mucocutaneous leishmaniasis and 3) visceral leishmaniasis, depending upon the tropism of the disease and the species causing the infection. Visceral leishmaniasis which is the most severe and deadly form is caused by the old world species *Leishmania* (*L.*) *donovani* and *Leishmania* (*L.*) *infantum* (Africa, Asia, Europe) and the new world species *L.* (*L.*) *infantum* (South America). India, Bangladesh, Nepal, Sudan, and Brazil have been reported to have more than 90% of the world's cases of visceral leishmaniasis with an estimated incidence of 500,000 new cases and 60,000 deaths each year (News, 2006). The severity of the disease is further augmented by drug resistance and drug failure, particularly in *L.* (*L.*) *donovani* strains of India and Nepal, which has been recently documented (Downing et al., 2011). So the raising future risks by this neglected tropical disease among the neglected populations of the world made WHO to include leishmaniasis among the six major diseases targeted for intense research and control.

Genome comparison of *Leishmania* (*L.*) *major*, *L.* (*L.*) *infantum*, *Leishmania* (*Viannia*) *braziliensis* showed great conservation of synteny and identifies only a small number of genes (approx 200) which are differentially distributed (Toledo et al., 2010; Peacock et al., 2007). These species specific genes may be a key factor for difference in pathogenesis between the species. Visceral leishmaniasis poses a fatality rate of greater than 100% within two years if untreated (Peacock et al., 2007). Since the genome information of *L.* (*L.*) *donovani* is very recent, so far, research on visceral leishmaniasis was mostly dependent on *L.* (*L.*) *infantum* genome information (Chappuis et al., 2007). Also most of the drugs commonly used to treat visceral leishmaniasis are toxic and exert unacceptable side effects (Khan et al., 2010). Though many control programs have been organized to control visceral leishmaniasis in Indian sub-continent, the success rate is severely compromised by the developing resistant strains of *L.* (*L.*) *donovani* at least in parts of India, particularly North Bihar and West Bengal. The diverse intra-strain genetic variability and drug resistance developed by the most severe *L.* (*L.*) *donovani* in Indian subcontinent and Nepal imposed the recent concern in the field of visceral leishmaniasis treatments (Pourshafie et al., 2004). So the treatment and control of leishmaniasis caused by *L.* (*L.*) *donovani* indeed require novel drug alternatives.

Although the outcome of infections and presentation of disease depend on many factors such as host immune response, host genetic variability, vector, and protozoan species (Guerin et al., 2002), it is obviously the specific genes of the species which determines the spectrum and severity of the disease. In view of the obvious clinical importance of this human pathogen, a genomic approach is highly desirable and may give insight into the complex mechanisms of pathogenesis. Here we report the comparative genome analysis of *L.* (*L.*) *donovani* with the other sequenced leishmanial species. This study therefore frameworks the experimental verification of few significant genes, consistent with independent existence, to set an avenue of genomic aspect of drug targeting to overcome the current problems in an effective way.

Results

Genome information of L. (L.) donovani

Despite many years of evolution, the genome content of *L.* (*L.*) *donovani* was greatly conserved with the other four leishmanial species except few specific genes. The lack of vast diversity among the leishmanial species over estimated 20–100 Ma of evolution may be due to lack of some machinery that causes diversity in eukaryotes and probably lack of transposable elements in the leishmanial species might be the cause (Khan et al., 2010). Though the presence of retrotransposons and RNAi machinery in *Trypanosoma brucei and L. braziliensis* was clearly reported, the evolutionary loss of these elements in leishmanial species preserves their genome content (Bringaud et al., 2006; Villanueva et al., 1991). Similarly the *L.* (*L.*) *donovani* genome lacks the retrotransposons and RNAi machinery. *L.* (*L.*) *donovani* contains 36 chromosomes and has a haploid genome size of 32.4 Mb. *L.* (*L.*) *donovani* genome encodes 8032 proteins, out of which 42.84% proteins were found to be functional homologs of other leishmanial species, 56.34% proteins lack functional assignment and the remaining ~0.8% proteins were exclusive to *L.* (*L.*) *donovani*. Approximately 5.6% genes were identified to code for repeat containing proteins which were conserved and probably might play a huge role in pathogenesis which was discussed elsewhere. The summary of *L.* (*L.*) *donovani* genome information was given in Supplementary Table S1.

Comparative analysis of L. (L.) donovani with the other sequenced leishmanial species

L. (L.) donovani species specific genes

Though *L.* (*L.*) *donovani* disease tropism differs greatly from other leishmanial species, till now no specific genes were reported for difference in disease presentation, except for the A2 gene family which was reported to be involved in the survival of the parasite in visceral organs (Zhang et al., 2003). Keeping this gap area in mind, we compared the proteome of *L.* (*L.*) *donovani* with the other four leishmanial species which identifies 55 gene coding proteins which were specific and expressed exclusively in *L.* (*L.*) *donovani*. The list of species specific genes of *L.* (*L.*) *donovani* which has been ascribed putative function was given in Table 1. A total list of *L.* (*L.*) *donovani* species specific genes were given in Supplementary Table S2. Out of the 55 *L.* (*L.*) *donovani* specific genes, only 36 genes were assigned putative function based on homology searches and Gene Ontology. The remaining genes encode for hypothetical proteins with conserved domain or unknown function which requires experimental documentation. Signal peptides were also detected for five specific *L.* (*L.*) *donovani* proteins which may have antigenic role in leishmanial pathogenesis. Among the 36 genes which encode proteins with putative function, few proteins were membrane related proteins, of which the important proteins being amastin like surface protein, lipophosphoglycan biosynthetic protein and phosphoglycan 1,3 galactosyltransferase which might have prime roles in pathogenesis, though the way or the mechanism it

Table 1
Species specific *L.* (*L.*) *donovani* proteins.

Functions	*L.* (*L.*) *donovani*[a]	Protein ID of *L.* (*L.*) *donovani*	*L.* (*L.*) *infantum*	*L.* (*L.*) *major*	*L.* (*Viannia*) *braziliensis*	*L.* (*L.*) *mexicana*
60s ribosomal, putative	LdBPK_150220	CBZ32839.1	–	–	–	–
Histone H3, putative	LdBPK_160600	CBZ33039.1	–	–	–	–
Paraflagellar rod protein 2C	LdBPK_161520	CBZ33130.1	–		–	–
Elongation factor 1-alpha	LdBPK_170200	CBZ33167.1	–	–	–	–
ATG8/AUT7/APG8/PAZ2, putative	LdBPK_190850	CBZ33555.1	–	–	–	–
Glycerol uptake protein, putative	LdBPK_191310	CBZ33599.1	–	–	–	–
Aminoacylase, putative	LdBPK_201730	CBZ33807.1	–	–	–	–
Cornichon homolog (*Drosophila*), isoform	LdBPK_240080	CBZ34396.1	–	–	–	–
CMP-sialic acid transporter, putative	LdBPK_240350	CBZ34423.1	–	–	–	–
Eukaryotic initiation factor 5a, putative	LdBPK_250760	CBZ34706.1	–	–	–	–
Succinyl-CoA synthetase α subunit	LdBPK_252230	CBZ34853.1	–		–	–
Aminopeptidase P1, putative	LdBPK_020010	CBZ34885.1	–	–	–	–
10 kDa heat shock protein, putative	LdBPK_260590	CBZ34948.1	–	–	–	–
Phosphoenolpyruvate carboxykinase	LdBPK_271710	CBZ35328.1	–	–	–	–
Heat-shock protein hsp70, putative	LdBPK_282960	CBZ35699.1	–	–	–	–
AMA1 protein, putative	LdBPK_301490	CBZ36170.1	–	–	–	–
Polyubiquitin	LdBPK_090950	CBZ36500.1	–	–	–	–
Glutaminyl-peptide cyclotransferase	LdBPK_312030	CBZ36603.1	–	–	–	–
Phosphoglycan β 1,3 galactosyltransferase	LdBPK_210010	CBZ36730.1	–	–	–	–
Ribosomal protein L3, putative	LdBPK_323330	CBZ37061.1	–	–		–
40S ribosomal protein S3, putative	LdBPK_151010	CBZ37261.1	–	–	–	–
Amastin-like surface protein, putative	LdBPK_342650	CBZ37742.1	–	–	–	–
Lipophosphoglycan biosynthetic protein	LdBPK_044281	CBZ37906.1	–	–	–	–
40S ribosomal protein S6, putative	LdBPK_212150	CBZ38085.1	–	–	–	–
60S ribosomal protein L22, putative	LdBPK_364640	CBZ38901.1	–	–	–	–
Beta-fructofuranosidase, putative	LdBPK_040310	CBZ31370.1	–		–	–
Metalloendopeptidase OMA1	LdBPK_041090	CBZ31447.1	–	–	–	–
ATPase alpha subunit	LdBPK_050500	CBZ31513.1	–	–	–	–
ATP-binding cassette protein subfamily G	LdBPK_060100	CBZ31594.1	–	–	–	–
U box domain-containing protein	LdBPK_070110	CBZ31730.1	–	–	–	–
Amastin-like protein	LdBPK_080760	CBZ31918.1	–		–	–
Translation initiation factor EIF-2B gamma	LdBPK_091140	CBZ32073.1	–	–	–	–
Cathepsin L-like protease	LdBPK_080950	CBZ31936.1	–	–	–	–
Histone H3	LdBPK_101050	CBZ32222.1	–	–	–	–
60S ribosomal protein L6, putative	LdBPK_151060	CBZ37264.1	–	–	–	–
Flagellar radial spoke protein-like	LdBPK_290690	CBZ35793.1	–	–	–	–

[a] The IDs represent the GeneDb IDs of *L.* (*L.*) *donovani*.

differs from other species was not clearly understood. Also 6 ribosomal genes specific to *L.* (*L.*) *donovani* were reported, though the exact involvement of the product of these genes in pathogenesis was unknown. In addition 2 specific genes (gene IDs: LdBPK_020010, LdBPK_260590) of *L.* (*L.*) *donovani* encode peptidases and heat shock proteins which were well known to have been involved directly or indirectly in pathogenesis. Also 3 other genes (gene IDs: LdBPK_252230, LdBPK_271710, LdBPK_040310) involved in sugar metabolism were identified which might possibly be involved in virulence and pathogenesis (Loughman and Caparon, 2006; Moyrand et al., 2007). Though the relationship between sugar metabolism and virulence remains greatly undefined, the novel proteins encoded by these genes might be involved in sugar metabolism of *L.* (*L.*) *donovani* and may influence the virulence in an unknown manner.

Interestingly two novel and specific genes of *L.* (*L.*) *donovani* that encode 1) Apical Membrane Antigen 1 (AMA1) (gene ID: LdBPK_301490) and 2) cathepsin L like protease (gene ID: LdBPK_080950) were identified and its involvement in pathogenesis was discussed as follows.

Apical Membrane Antigen 1 (AMA1), highly suspected protein in parasite interaction and invasion

Apical Membrane Antigen 1 was documented as a protein directly involved in invasion of apicomplexon parasites into the host (Tonkin et al., 2011). In plasmodium and toxoplasma, the AMA1 proteins were secreted from microneme of rhoptries and it was targeted to the apical membrane where it gets integrated with the parasite plasma membrane. Earlier studies with plasmodium clearly showed that the integrated AMA1 forms a complex with RON2 which in turn helps the parasite to attach with the host cell to promote invasion. The involvement of AMA1 in signaling and parasite replication was also documented in toxoplasma (Santos et al., 2011). In conjunction, three AMA1 genes were identified in *L.* (*L.*) *donovani* out of which two were conserved in all leishmanial species and the third one was specific to *L.* (*L.*) *donovani*. The homology search showed very less sequence similarity between AMA1 from plasmodium and leishmanial species. The absence of microneme and RON2 in *Leishmania* species clearly indicates the absence of this mechanism in leishmanial species. Surprisingly the Gene Ontology (GO) studies of specific AMA1 (gene ID: LdBPK_301490) reported in *L.* (*L.*) *donovani* shows cholesterol binding ability. The importance of host membrane cholesterol in *L.* (*L.*) *donovani* infection was already well documented (Pucadyil and Chattopadhyay, 2006). In *L.* (*L.*) *donovani*, cholesterol is required for binding and internalization of the parasite inside a host cell (Pucadyil et al., 2004; Tewary et al., 2006). Similar phenomenon was also documented in *L.* (*L.*) *infantum* (Rodriguez et al., 2006). Also the *L.* (*L.*) *donovani* specific AMA1 is leucine rich protein and the significance of leucine residues in interacting with host cell membrane was already documented in many organisms including leishmanial species (Kedzierski et al., 2004). Also signal peptide was detected which possibly confers that the protein was secretory. All these clues together made us to hypothesize that in *L.* (*L.*) *donovani*, AMA1 is secreted from an unknown organelle into the apical membrane of *L.* (*L.*) *donovani* where it interacts with the cholesterol of host membrane through leucine residues and helps in parasite internalization. The species specific expression of this protein in *L.* (*L.*) *donovani* indicates the probable role of this protein in visceralization, which needs future experimentations. The ongoing experimental studies in our laboratory regarding AMA1 will be a huge hope in the future for anti-leishmanial therapy.

Cathepsin L like protease

Cathepsin L like protease, a type of lysosomal endopeptidases, has been already identified as class of drug targets which participate in many essential biological processes of the parasite such as embryogenesis, molting, and immune evasion (Lustigman et al., 2004; Dalton et al., 2003). Cathepsin L like protease is expressed in both promastigote and amastigote of leishmanial species (Sakanari et al., 1997), and probably the novel gene (gene ID: LdBPK_080950) from *L.* (*L.*) *donovani* which encodes cathepsin L like protease may assist in evading the host immune system.

Differential gene distribution in leishmanial species

Further the comparative analysis establishes few genes specific between the *L.* (*L.*) *donovani* and other leishmanial species. *L.* (*L.*) *donovani* was found to be closest with *L.* (*L.*) *infantum* and there were only 60 genes which were distinctly observed between *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum*; 72 *L.* (*L.*) *donovani* specific genes were distinguished while comparing with *L.* (*L.*) *major*; 159 specific genes were found distinct while comparing with *Leishmania* (*L.*) *mexicana*; 265 specific genes were found distinct while comparing with *L.* (*Viannia*)

braziliensis. The comparison was depicted in Fig. 1. The lists of few important genes and its protein product from *L.* (*L.*) *donovani* which share homology with some leishmanial species but not conserved with the other leishmanial species were given in Table 2. The full list of such genes including those which encode hypothetical proteins was given in Supplementary Table S3. The Gene Ontology (GO) for the *L.* (*L.*) *donovani* genes which were observed in few leishmanial species (~83 genes which encode functional proteins) showed that these genes mainly encode lipid containing membrane proteins, possibly involved in binding. Also some genes encode proteins which contain the enzymatic activity, like kinase proteins were expected to be involved in signal transduction or sugar metabolism. The Gene Ontology (GO) results were depicted in Fig. 2.

A specific gene (gene ID: LdBPK_290650) found to be conserved between *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum* but was completely absent in other leishmanial species encodes a putative BET1-like protein. Though the exact function of this protein is not known, its close similarity with BET1 shows that the protein might participate in vesicular transport and may function as SNARE in docking of ER-derived vesicles with the cis-golgi membrane. The possible involvement of this protein in pathogenesis needs future experimental investigations. The other gene (gene ID: LdBPK_161550) from *L.* (*L.*) *donovani* which was found homologous to *L.* (*L.*) *infantum* and *L.* (*Viannia*) *braziliensis* but absent in *L.* (*L.*) *major* and *L.* (*L.*) *mexicana* encodes a kinesin protein which may be involved in flagellar movement inside the host cell. Though leishmanial genome encodes many conserved kinesin molecules involved in flagellar locomotion, the importance of this specific kinesin needs further studies. Another gene (gene ID: LdBPK_367030) which was found in *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum* but has become a pseudogene or absent in other leishmanial species encodes a putative mitochondrial carrier protein (Agcp2438-like protein) which is a mitochondrial transmembrane protein involved in transport, probably playing a key role inside the oxidative environment of macrophages (Dolezal et al., 2012).

The important surface proteins of leishmanial species amastin, amastin-like surface protein, cysteine protease B (CPB), lipophosphoglycan LPG3 and the leishmanolysin GP63, were clearly reported for their potential role in parasite virulence (Rochette et al., 2005; Mottram et al., 2004; Vinet et al., 2009; Joshi et al., 2002). Though we found no obvious difference in the gene structure of lipophosphoglycan LPG3 (gene IDs: LdBPK_044280 and LdBPK_044281) in *L.* (*L.*) *donovani* in comparison with the other leishmanial species, we do found the discrete gene pattern of amastin, amastin like surface protein, leishmolysin and cysteine

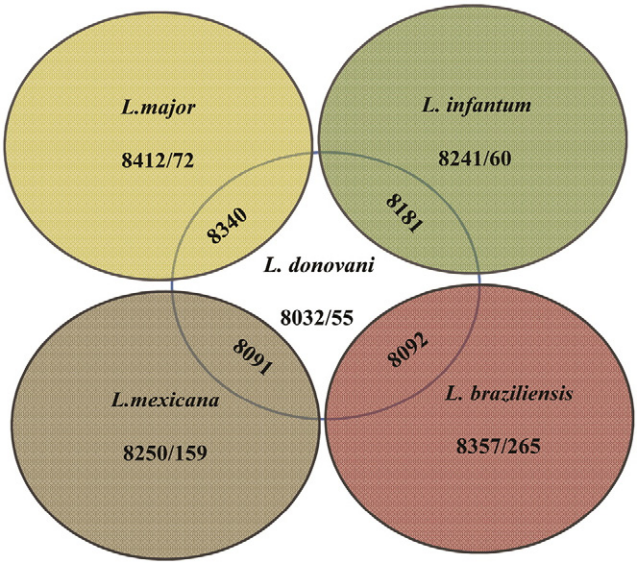

Fig. 1. Comparative analysis of *L.* (*L.*) *donovani* with four other leishmanial species. Individual circles represent the individual leishmanial species. The numbers in the main part of the individual circles represent the total number of protein coding genes of individual leishmanial species and the numbers after the slash represent the genes encoding the number of proteins which are distinct between *L.* (*L.*) *donovani* and any of the other four leishmanial species. The portion which was shared by two circles was used to represent the genes encoding the proteins which were homologous between any two leishmanial species. *L.* (*L.*) *donovani*, *L.* (*L.*) *infantum*, *L.* (*L.*) *major*, *L.* (*L.*) *mexicana*, and *L.* (*Viannia*) *braziliensis* are represented as *L. donovani*, *L. infantum*, *L. major*, *L. mexicana* and *L. braziliensis* respectively.

Table 2
Comparative analysis of *L.* (*L.*) *donovani* with the other four leishmanial species.

Functions	*L.* (*L.*) *donovani*[a]	Protein ID of *L.* (*L.*) *donovani*	*L.* (*L.*) *infantum*[a]	*L.* (*L.*) *major*[a]	*L.* (*Viannia*) *braziliensis*[a]	*L.* (*L.*) *mexicana*[a]
Protein kinase, putative	LdBPK_220370	CBZ34082.1	LinJ.22.0370	**LmjF.22.0130**	**–**	–
RNA-binding protein, putative, UPB1	LdBPK_250500	CBZ34680.1	LinJ.25.0500	**–**	**–**	–
Kinetoplast-associated protein-like protein	LdBPK_270240	CBZ35185.1	LinJ.27.0250	**LmjF.27.0745**	**LbrM.27.0030**	–
3′-Nucleotidase/nuclease, putative	LdBPK_312370	CBZ36636.1	LinJ.31.2370	LmjF.31.2300	**LbrM.31.3140**	–
Amastin-like surface protein, putative	LdBPK_342660	CBZ37743.1	LinJ.34.2660	–	**LbrM.34.3890**	–
Ribosomal protein S11 homolog	LdBPK_211790	CBZ33991.1		–	**–**	–
RNA-editing complex protein MP81, putative	LdBPK_020380	CBZ31208.1	LinJ.02.0380	**LmjF02.0100**	**–**	–
Oxidoreductase-like protein	LdBPK_020700	CBZ31240.1	LinJ.02.0700	LmjF.27.2650	**–**	–
GP63, leishmanolysin, metallo-peptidase	LdBPK_100520	CBZ32170.1	LinJ.10.0520	**LmjF.10.0425**	**–**	–
40S ribosomal protein S4, putative	LdBPK_131130	CBZ32612.1	LinJ.13.1130	–	**LbrM.13.1160**	–
U1 small nuclear ribonucleoprotein	LdBPK_161690	CBZ33146.1	LinJ.16.1690	–	**LbrM.16.1700**	–
ATPase subunit 9, putative	LdBPK_260450	CBZ34934.1	LinJ.26.0450	LmjF.26.0460	**–**	–
BET1-like protein, putative	LdBPK_290650	CBZ35789.1	LinJ.29.0650	–	**–**	–
Amastin, putative	LdBPK_093030	CBZ35869.1	LinJ.29.3030	LmjF.34.1080	**–**	–
Amastin, putative	LdBPK_291450	CBZ35870.1	LinJ.29.3030	LmjF.34.1080	**–**	–
Calcium-binding protein, putative	LdBPK_301300	CBZ36151.1	LinJ.30.1300	LmjF.30.1240	**LbrM.30.1300**	–
AAA family ATPase-like protein	LdBPK_302520	CBZ36273.1	LinJ.30.2520	**LmjF.30.2500**	**LbrM.30.2930**	–
Tryparedoxin-like protein	LdBPK_312010	CBZ36601.1	LinJ.31.2010	**–**	**–**	–
Helicase-like protein, DNA repair	LdBPK_321670	CBZ36898.1	LinJ.32.1670	**LmjF.32.1630**	**LbrM.32.1680**	–
Amastin-like surface protein, putative	LdBPK_341720	CBZ37650.1	LinJ.34.1720	–	**LbrM.34.1870**	–
Endonuclease/exonuclease/phosphatase,	LdBPK_361210	CBZ38559.1	LinJ.36.1210	**LmjF.36.1395**	**LbrM.35.0700**	–
Mitochondrial carrier protein, putative	LdBPK_367030	CBZ39138.1	LinJ.36.7030	**LmjF.36.6785**	–	–
Cytochrome b5-like protein	LdBPK_091580	CBZ32116.1	LinJ.09.1580	LmjF.09.1500	–	LmxM.09.1490
ATP-binding cassette protein subfamily A,	LdBPK_111220	CBZ32385.1	LinJ.11.1230	LmjF.11.1230	–	LmxM.11.1240
Alg9-like mannosyltransferase, putative	LdBPK_120145	CBZ32416.1	LinJ.12.0140	LmjF.12.0160	–	LmxM.12.0160
Nucleotide sugar transporter, putative	LdBPK_150900	CBZ32906.1	LinJ.15.0900	**LmjF.15.0840**	–	LmxM.15.0840
Product = P-type H-ATPase, putative	LdBPK_181490	CBZ33456.1	LinJ.18.1510	LmjF.18.1510	**LbrM.18.1690**	LmxM.18.1520
Product = 4-coumarate:coa ligase-like protein	LdBPK_190950	CBZ33564.1	LinJ.19.0960	LmjF.19.0985	–	LmxM.19.0995
Product = 4-coumarate:coa ligase-like protein	LdBPK_190960	CBZ33565.1	LinJ.19.0960	LmjF.19.0985	–	LmxM.19.0995
Calmodulin, putative	LdBPK_131060	CBZ32605.1	LinJ.13.1060	LmjF.13.1160	–	LmxM.13.1160
EF hand-like protein	LdBPK_131490	CBZ32647.1	LinJ.13.1490	LmjF.13.1450	–	LmxM.13.1450
Fatty acyl CoA synthetase 2, putative	LdBPK_010530	CBZ31140.1	LinJ.01.0530	LmjF.01.0500	–	LmxM.01.0500
Calpain-like cysteine peptidase, putative	LdBPK_201210	CBZ33756.1	LinJ.20.1210	LmjF.20.1180	**LbrM.20.2800**	LmxM.20.1180

(continued on next page)

Table 2 (*continued*)

Functions	*L.* (*L.*) *donovani*[a]	Protein ID of *L.* (*L.*) *donovani*	*L.* (*L.*) *infantum*[a]	*L.* (*L.*) *major*[a]	*L.* (*Viannia*) *braziliensis*[a]	*L.* (*L.*) *mexicana*[a]
Pumilio protein 9, putative	LdBPK_201420	CBZ33777.1	LinJ.20.1420	**LmjF.20.1475**	**LbrM.20.4700**	LmxM.20.1370a
Cyclase associated protein	LdBPK_211110	CBZ33923.1	LinJ.21.1110	LmjF.21.0891	**LbrM.21.0880**	LmxM.21.0891
Hypothetical protein, conserved	LdBPK_211770	CBZ33989.1	LinJ.21.1770	LmjF.21.1530	**–**	LmxM.21.1530
Ring finger protein 138	LdBPK_220070	CBZ34052.1	LinJ.22.0070	LmjF.22.0190	**–**	LmxM.22.0190
Protein–tyrosine phosphatase-like protein	LdBPK_220120	CBZ34057.1	LinJ.22.0120	LmjF.22.0250	**LbrM.22.0190**	LmxM.22.0250
Methylenetetrahydrofolate dehydrogenase	LdBPK_220220	CBZ34067.1	LinJ.22.0220	**LmjF.22.0340**	**LbrM.22.0240**	LmxM.22.0340
A2 protein	LdBPK_220670	CBZ34112.1	LinJ.22.0670	–	**–**	**LmxM.22.0692**
40S ribosomal protein L14, putative	LdBPK_221410	CBZ34182.1	LinJ.22.1410	**LmjF.22.1680**	**LbrM.22.1410**	–
Ubiquitin-activating enzyme, putative	LdBPK_240470	CBZ34435.1	LinJ.24.0470	LmjF.24.0460	**LbrM.24.0710**	LmxM.24.0460
Acylphosphatase, putative	LdBPK_252040	CBZ34834.1	LinJ.25.2040	–	–	LmxM.25.1960
Tagatose-6-phosphate kinase-like protein	LdBPK_252550	CBZ34886.1	LinJ.25.2550	LmjF.25.2440	–	LmxM.25.2440
ATPase beta subunit, putative	LdBPK_052580	CBZ34888.1	LinJ.25.1210	LmjF.25.1170	–	LmxM.25.1170
Ghistone H1 like	LdBPK_271511	CBZ35309.1	LinJ.27.1511	LmjF.27.1605	–	LmxM.27.1605
Tuzin like protein, putative	LdBPK_093020	CBZ35871.1	LinJ.29.3020	LmjF.34.1590	–	LmxM.33.1830
Histone H2A, putative	LdBPK_291850	CBZ35910.1	LinJ.29.1850	–	–	LmxM.21.0915
Histone H2A, putative	LdBPK_291860	CBZ35911.1	LinJ.29.1870	–	–	LmxM.21.0915
Histone H2A, putative	LdBPK_291870	CBZ35912.1	LinJ.29.1870	–	–	LmxM.21.0915
Poly(A) polymerase, putative	LdBPK_292710	CBZ35996.1	LinJ.29.2710	LmjF.29.2600	**LbrM.29.2910**	LmxM.08_29.2600
Phosphoglycan β 1,2 arabinosyltransferase	LdBPK_020190	CBZ31189.1	LinJ.02.0190	LmjF.02.0180	**LbrM.02.0030**	LmxM.33.0510
Ribosomal RNA processing protein, putative	LdBPK_020290	CBZ31199.1	LinJ.02.0290	LmjF.02.0320	LbrM.02.0060	LmxM.02.0320
Amino acid permease 3	LdBPK_310900	CBZ36492.1	LinJ.31.0910	LmjF.31.0870	–	LmxM.30.0880
Sodium stibogluconate resistance protein,	LdBPK_073400	CBZ36733.1	LinJ.31.0950	LmjF.31.0950	–	–
Tuzin-like protein	LdBPK_341160	CBZ37596.1	LinJ.34.1160	LmjF.34.2800	LbrM.34.0500	LmxM.36.1280
Serine acetyltransferase	LdBPK_342710	CBZ37748.1	LinJ.34.2710	LmjF.34.2850	–	LmxM.33.2850
RNA editing associated helicase 2, putative	LdBPK_343010	CBZ37778.1	LinJ.34.3010	–	LbrM.34.3540	LmxM.34.2680
Unc104-like kinesin, putative	LdBPK_344090	CBZ37886.1	LinJ.34.4090	**LmjF.34.3952**	**LbrM.34.4700**	–
Phosphoglycan beta 1,3 galactosyltransferase 6	LdBPK_360010	CBZ38439.1	LinJ.25.2570	LmjF.35.0010	–	LmxM.25.2460
Xylulokinase, putative	LdBPK_360280	CBZ38466.1	LinJ.36.0280	LmjF.36.0260	–	LmxM.36.0260
Aminoalcohol phosphotransferase, putative	LdBPK_030970	CBZ31323.1	LinJ.03.0970	LmjF.36.5900	–	LmxM.03.0821
Peptide deformylase 2 metalloprotease-like	LdBPK_040820	CBZ31420.1	LinJ.04.0820	LmjF.04.0820	–	LmxM.04.0820
Cytochrome b5-like protein	LdBPK_070940	CBZ31808.1	LinJ.07.0940	LmjF.07.0810	**LbrM.07.0965**	LmxM.07.0810
Serine peptidase, clan SF, family S26A	LdBPK_080460	CBZ31890.1	LinJ.08.0460	LmjF.08.0450	–	LmxM.08.0450
Ecotin, putative	LdBPK_150530	CBZ32870.1	LinJ.15.0530	LmjF.15.0510	LbrM.15.0540	–
CFAS, putative	LdBPK_080560	CBZ31900.1	LinJ08.0560	–	LbrM.08.0590	LmxM.08.0545
Kinesin, putative	LdBPK_161550	CBZ33133.1	LinJ.16.1550	–	LbrM.16.1530	–
Uridine kinase-like protein	LdBPK_312560	CBZ36654.1	LinJ.31.2560	**LmjF.31.2785**	**LbrM.31.3370**	–
G-actin binding protein, putative	LdBPK_365830	CBZ39019.1	LinJ.36.5830	LmjF.36.5590	LbrM.35.5860	–
Cytochrome b5-like protein	LdBPK_091580	CBZ32116.1	LinJ.09.1580	LmjF.09.1500	–	–

[a] The IDs represent the GeneDb IDs of the corresponding leishmanial species. Bold highlighted — pseudogenes.

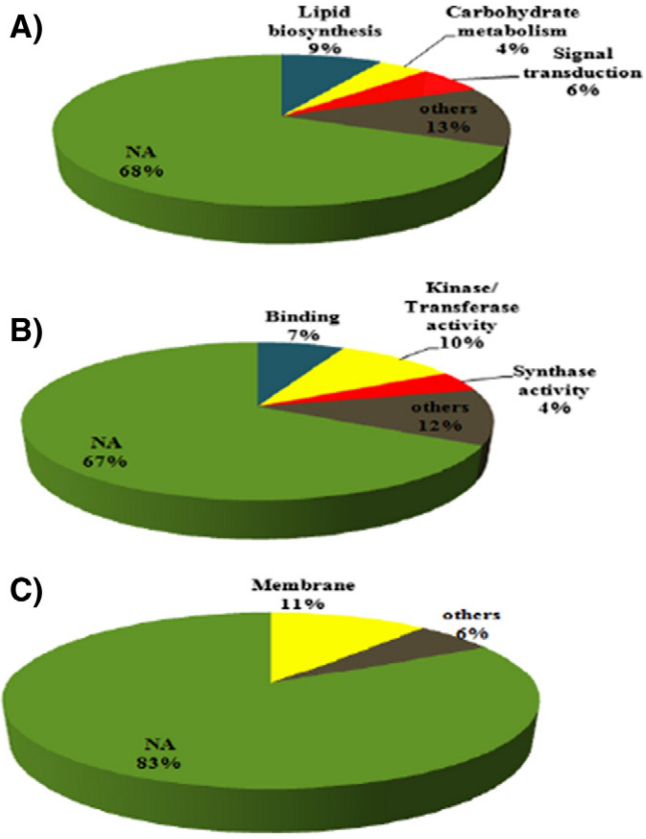

Fig. 2. Gene Ontology predicted for the genes differentially distributed between *L.* (*L.*) *donovani* and four other leishmanial species. The pie chart shows the most represented functions under three categories: A) biological process, B) molecular function and C) cellular component. The biological process predicted for these proteins is mainly involved in lipid biosynthesis, signal transduction and carbohydrate metabolism. The molecular function shows that these proteins are involved greatly in binding and some proteins contain enzymatic activity and hence involved in signaling pathways. Cellular component shows that these proteins are mainly membrane proteins. NA indicates that the GO predictions were not available.

protease in *L.* (*L.*) *donovani*, which was discussed below. The important surface protein of leishmanial species leishmolysin GP63, a metalloprotease ubiquitously distributed in trypanosomatids, plays a myriad of functions and was found conserved in all leishmanial species including *Herpetomonas samuelpessoai*, an insect trypanosomatid (Pereira et al., 2010).We found that a molecule of GP63(gene ID: LdBPK_100520), leishmanolysin metallo-peptidase, clan MA (M), and family M8 (gene ID: LdBPK_100520) specific in *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum*, released from the surface by proteolysis might participate in different stages of the parasite life cycle (Elias et al., 2006). The enzyme has a physiological function in the promastigote stage of these parasites (Schneider and Glaser, 1993) which indicates its role in the initial stage of parasite infection and its apparent role of interaction with macrophages in visceral organs needs further experimentations.

Two genes which encode amastin like surface proteins (*L.* (*L.*) *donovani* gene IDs: LdBPK_341720, LdBPK_342660) located at chromosome 34 found to be specific to *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum* probably indicates the involvement of this protein in visceralization. The comparative studies also identified two other significant gene families of pathogenesis 1) amastin gene family and 2) A2 gene family. The *L.* (*L.*) *donovani* specific amastin genes being the important pathogenesis factor was compared with the other leishmanial species and discussed in the following sections. Also, A2 protein encoded by A2 gene family, the only documented protein to be directly involved in visceralization of the *L.* (*L.*) *donovani* was detailed in later sections.

L. (L.) donovani specific amastins and its gene locations with respect to other leishmanial species

Amastin, encoded by a large gene family was a transmembrane glycoprotein initially documented in amastigote stage of trypanosomatid parasites and subsequently documented as surface proteins expressed more in leishmanial species than trypanosoma species (Jackson, 2010). The exact function of amastin and tuzin, the gene family found to be associated with amastin of unknown functions is yet to be classified. The significance of amastin gene family in pathogenesis of leishmanial species has been partially reported in earlier studies (Rochette et al., 2005). Amastin proteins of leishmanial species were mainly found in chromosome 8, 24, 28, 29, 30 and 34 of which chromosome 34 was most represented and chromosomes 28 and 29 were the least represented. Chromosome 34 contains mainly the amastin genes (6 copies) which were specific for visceralization causing protozoans. One amastin gene was found to be specific for *L.* (*L.*) *donovani*, 2 copies of amastin genes were found conserved in *L.* (*L.*) *major*, *L.* (*L.*) *infantum*, *L.* (*L.*) *mexicana* and *L.* (*L.*) *donovani*, one amastin gene was found present in all leishmanial species except *L.* (*L.*) *major* and one amastin gene was found conserved in all leishmanial species. Interestingly chromosome 8 contains 5 amastin gene locations out of which 3 were *L.* (*L.*) *donovani* specific amastin genes, one is found in *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum* and one found in all leishmanial species except *L.* (*Viannia*) *braziliensis* where it was absent. This clearly indicates that the amastin gene concentrated on chromosome 8 might be the visceralization factor of *L.* (*L.*) *donovani*. No *L.* (*L.*) *donovani* specific amastin gene family was detected in chromosomes 24, 28 and 30 during our analysis and these amastin genes were well conserved in all leishmanial species except *L. braziliensis* where the amastins were absent. Also a gene from chromosome 29 encodes a specific amastin protein which was found absent in *L.* (*L.*) *major* and *L.* (*Viannia*) *braziliensis*. In total *L.* (*L.*) *donovani* contains 4 specific amastin genes (gene IDs: LdBPK_341700, LdBPK_080710, LdBPK_080780, LdBPK_080790) the importance of this specific gene expression in virulence and pathogenesis needs future investigations. Altogether the comparative study shows that amastin gene family was represented more in *L.* (*L.*) *donovani* and less in *L.* (*Viannia*) *braziliensis*. The probable explanation could be the possible involvement of this protein in visceralization and extreme evolutionary diversification of *L.* (*Viannia*) *braziliensis* and *L.* (*L.*) *donovani*. The complete comparison and chromosomal location of amastin gene family from leishmanial species were picturized in Fig. 3.

Analysis of A2 gene family with respect to copy number variations

L. (*L.*) *donovani* disease tropism differs from other leishmanial species depending mainly on visceralization. The gene mainly involved in the phenomena was the A2 gene family with its role in survival of the parasite in visceral organs. Studying the evolutionary pattern of this important gene may give insight into the difference in its mode of expression, leading to difference in mechanism of pathogenesis of *L.* (*L.*) *donovani* from other leishmanial species.

The research on *L.* (*L.*) *donovani* till now presents only one specific gene family A2 which was directly involved in disease tropism (Zhang and Matlashewski, 2001). In contrast, *L.* (*L.*) *major* contained only A2 pseudogenes (Zhang et al., 2003) and were completely absent in *L.* (*Viannia*) *braziliensis*. The function of A2 protein in *L.* (*L.*) *donovani* might be to relieve the stress in visceral organs following infection (McCall and Matlashewski, 2010). To study the evolution, A2 protein sequences of *L.* (*L.*) *donovani* were collected from literature (Oliveira et al., 2011). HMMER and Blastp result revealed that the A2 protein had clear homology with Stage Specific S Antigen (SSSA) of other leishmanial species. A2 gene family was specifically expressed in the amastigote stage inside host macrophage, though the exact function of the protein coded by this gene family was unclear (Charest and Matlashewski, 1994). The multiple sequence alignment showed that the A2 protein sequence was conserved in all leishmanial species though the length of the protein was 5 fold less represented in *L.* (*L.*) *donovani*, may be due to deletion events. The multiple sequence alignment was given in (Supplementary Fig. S1). The clear homology and exact length with S-antigen homolog indicates that in course of evolution, part of A2 protein might have got lost in other leishmanial species, leaving the SSSA part to play a role in pathogenesis. So the important role of SSSA in *L.* (*L.*) *donovani* remains to be uncovered in future. Also significant homology found between the A2 protein sequence of *L.* (*L.*) *donovani* species and uncharacterized protein sequences of *Streptomyces ambofaciens* and *Thermobispora bispora* indicates that these proteins might have originated from common ancestor and as evolution progresses these proteins diverged depending on the host as well as the environment and formed a significant protein to play a key role in pathogenesis. The evolutionary closeness was further confirmed by phylogenetic analysis (data not shown). Moreover A2 protein from *L.* (*L.*) *mexicana*, yet not declared as non-functional protein coded by a pseudogene, was very much

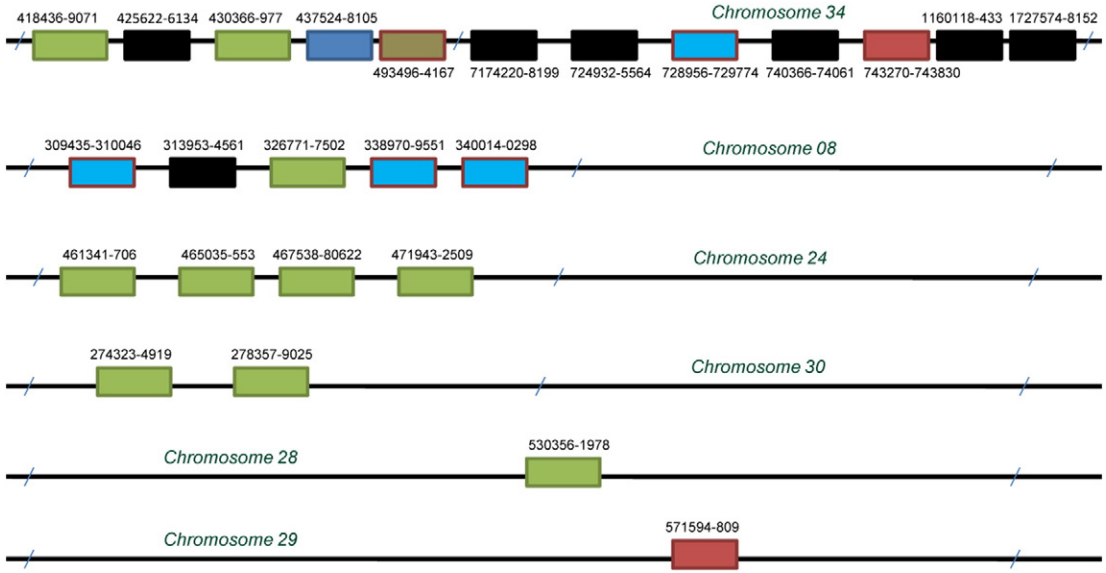

Fig. 3. Comparison of amastin gene family in five different Leishmanial species. The straight lines indicate the chromosomes and the boxes denote the particular amastin gene. The numbers are given above and below to differentiate the chromosomal locations of the gene in the forward and reverse strands. Different colors are used to differentiate the amastin gene conserved or differentiated among the leishmanial species. The box indicates the amastin genes conserved in all leishmanial species. The box indicates the amastin gene conserved between *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum* but absent in other three leishmanial species. The box indicates the amastin gene conserved between *L.* (*L.*) *donovani*, *L.* (*L.*) *infantum* and *L.* (*L.*) *mexicana* but absent in other two leishmanial species. The box indicates the genes conserved in all leishmanial species except *L. braziliensis*. The box indicates the amastin gene expressed specifically in *L.* (*L.*) *donovani* and absent in remaining four leishmanial species. The box indicates the amastin gene conserved in all leishmanial species except *L.* (*L.*) *major*. The numbers over the boxes indicate the exact chromosomal location of amastin genes in *L.* (*L.*) *donovani* genome.

conserved with the uncharacterized protein sequences of S*treptomyces ambofaciens* and *Thermobispora bispora* indicating that it was preserved throughout the evolution but the exact role of these protein in *L.* (*L.*) *mexicana* pathogenesis remains unclear. The copy number of A2 gene also plays a significant role in pathogenesis of particular leishmanial species. A2 gene family has become a pseudogene in *L.* (*L.*) *major* genome (Zhang et al., 2003) and was completely absent in *L.* (*Viannia*) *braziliensis*. A2 family genes that are abundantly expressed in amastigote stage of leishmanial species known to be the primary factor for visceralization and virulence. Previous studies demonstrate the crucial role of A2 genes (Zhang and Matlashewski, 2001). A2 gene family present in chromosome 22, consists of a 5′A2 rel, 3′A2 rel, internal A2 rel and A2 genes and organized as 5′A2rel, A2 gene, internal A2 rel and 3′A2 rel (Zhang et al., 2003; Zhang and Matlashewski, 2001). All the available genes in A2 family consisting of 5′A2rel, 3′A2rel, internal A2rel and A2 gene sequences from leishmanial genomes were retrieved from NCBI GenBank and GeneDb. A2 gene sequence was not available for *L.* (*L.*) *major* and *L.* (*Viannia*) *braziliensis* in the databases while 5′A2rel, 3′ A2rel and internal A2rel gene sequences were not available for the species other than *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum*. To understand the organization of A2 gene family within the leishmanial species, A2 genes were searched in the genome of *L.* (*L.*) *donovani* using ACT (Carver et al., 2005). The results of respective A2 gene against respective genome shows that, in *L.* (*L.*) *infantum* single copy of *A2* gene is present but in *L.* (*L.*) *mexicana* two copies of A2 gene are present which are adjacent to each other with very little sequence differences. In *L.* (*L.*) *donovani* there are multiple copies of partial A2 gene scattered throughout the chromosome 22. The location of A2 genes on the chromosome 22 of four leishmanial species are depicted in (Fig. 4). But the presence of two A2 genes and its role in *L.* (*L.*) *mexicana* remains unclear. *L.* (*L.*) *donovani* genome evolved with 5 copies of A2 gene as a result of duplication remains the most severe form of visceral leishmaniasis. The copy number difference of A2 gene in different leishmanial species was depicted in (Supplementary Fig. S2).

L. (L.) donovani genes with high-variation

Cysteine peptidases were considered to be the important molecule in leishmanial pathogenesis (Vinet et al., 2009). A gene (gene ID: LdBPK_201210) encoding calpain-like cysteine peptidase, putative in *L.* (*L.*) *donovani* was identified and interestingly the comparison of this protein with the other leishmanial proteins showed that it contains repeat motifs at different locations strengthening the anticipation of this protein to have a vital role in parasite survival. Also 11 other proteins recently documented as proteins under positive selection in *L.* (*L.*) *donovani* (Downing et al., 2011) were compared and listed in (Supplementary Table S4).

Repeat analysis identifies huge differences in gene coding for surface proteins between L. (L.) donovani and other leishmanial species

Amino acid repeat-containing proteins have a broad range of functions and their identification was of relevance to many experimental biologists. The involvement of these proteins in immune evasion had been shown in protozoan parasites such as the kinetoplastid and *Plasmodium* species, probably by influencing virulence and pathogenicity (Goto et al., 2010). Leishmanial proteomes were enriched with amino acid repeats, approximately 3–4% proteins were repeat containing proteins and the probable role of these proteins might be to mediate host parasite interaction (Kedzierski et al., 2004; Depledge et al., 2007). Identification of repeat-containing proteins provides researchers with a defined subset of proteins which can be analyzed by expression profiling and functional characterization, thereby facilitating study of pathogenicity and virulence factors in the parasitic protozoa.

Total repeat containing protein sequences were collected from the nearby species of *L.* (*L.*) *donovani* from Repseq database (Depledge et al., 2007). The comparative analysis of number and type of repeat containing protein sequences in 4 leishmanial species are illustrated here (Fig. 5). Many virulence factors were already reported in leishmanial species in which surface proteins were of special interest. Comparative analysis of genes encoding the repeat containing surface proteins and other repeat proteins known to be involved in virulence was given in (Table 3). Among the surface proteins, surface antigen proteins and Proteophosphoglycan were well represented repeat proteins of leishmanial species. In *L.* (*L.*) *donovani*, two surface proteins (protein IDs: CBZ33953.1 and CBZ31356.1) from chromosome 12 and 4 were duplicated many times in other leishmanial species. Also calpain like cysteine peptidases from *L.* (*L.*) *donovani* (protein ID: CBZ35211.1), were represented more in other leishmanial species. The genes which encode other repeat containing proteins which

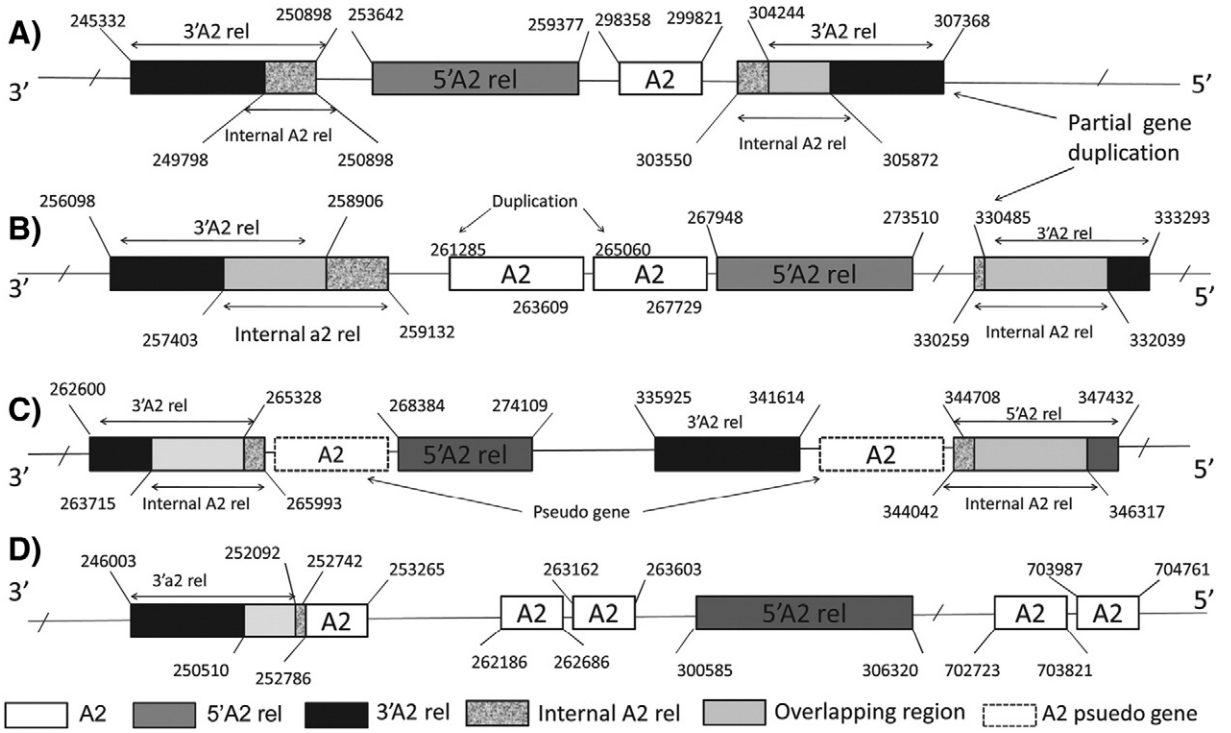

Fig. 4. Location of A2 gene family in chromosome 22 of four leishmanial species. Chromosomal location of A2 gene in four leishmanial species is shown, except for *L.* (*Viannia*) *braziliensis* where A2 gene is completely absent. A, B, C and D represent the chromosomal locations of *L.* (*L.*) *infantum*, *L.* (*L.*) *mexicana*, *L.* (*L.*) *major* and *L.* (*L.*) *donovani* respectively. Different grayscale representations as given in the figure are shown to differentiate and locate the 5′A2rel, A2 gene, internal A2 rel and 3′A2 rel on chromosome 22 of four leishmanial species.

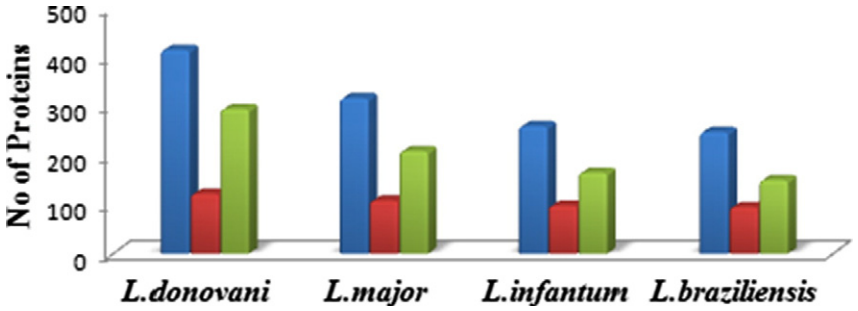

Fig. 5. Analysis of *L.* (*L.*) *donovani* proteins containing repeats. *L.* (*L.*) *donovani* proteome contains 415 repeat containing proteins out of which are 293 functional proteins and 122 are hypothetical. Similarly *L.* (*L.*) *major* contains 317 repeat containing proteins out of which are functional proteins and are hypothetical. *L.* (*L.*) *infantum* contains 261 repeat containing proteins out of which are functional proteins and are hypothetical. *L.* (*Viannia*) *braziliensis* contains 248 repeat containing proteins out of which are functional proteins and are hypothetical. The blue bars represent the total number of repeat proteins, red bars represent the functional proteins and green bars represent the hypothetical proteins. *L.* (*L.*) *donovani*, *L.* (*L.*) *infantum*, *L.* (*L.*) *major*, and *L.* (*Viannia*) *braziliensis* are represented as *L. donovani*, *L. infantum*, *L. major* and *L. braziliensis* respectively.

were functional homologs or hypothetical proteins of other leishmanial species were also compared and tabulated. The most common type of motif found in repeat proteins was Single Repeat Regions (SRR). The complete list of comparative analysis of genes encoding repeat proteins was given in (Supplementary Tables S5, S6). The frequency of occurrence of particular amino acid was important for a repeat containing protein to be involved in pathogenesis directly or indirectly. Alanine the common amino acid was occurring more in repeat containing proteins also, though there was no relation reported between alanine and pathogenesis. Leucine was represented more in these proteins and the significance of this amino acid in virulence or pathogenesis of many organisms was well documented in literature (Kedzierski et al., 2004). Also frequency of serine was found more, though the exact involvement of this amino acid in pathogenesis was unknown. The amino acid composition of the repeat containing proteins was given in (Supplementary Fig. S3).

Discussion

The aim of this study is to compare and reveal the species specific differences of *L.* (*L.*) *donovani* with some other important leishmanial species. The difference between *L.* (*L.*) *donovani* and other leishmanial species at the genome level is completely studied. Though genome of leishmanial species are highly conserved as reviewed in the introduction, few important differences were identified in the genome of *L.* (*L.*) *donovani*. The comparative genome analysis of *L.* (*L.*) *donovani* with the other leishmanial species identified 55 species specific genes. Since the proteins encoded by these genes are species specific, the importance of these proteins in disease representation, disease progression, pathogenesis and virulence can be evaluated. To read the functions of these genes, Gene Ontology was done using Amigo tool available at GO database. Major portions of the functionally specific proteins fall into two classes 1) Ribosomal protein and 2) Surface/surface-like proteins. Though the specificity of ribosomal proteins was not urged, the probable involvement of surface proteins in *L.* (*L.*) *donovani* disease spectrum is revealed in our study. In addition some genes involved in sugar metabolism was also identified and the probable role of these proteins in pathogenesis was discussed. Importantly a gene which encodes cathepsin like cysteine protease with its established role in pathogenesis was identified as specific gene of *L.* (*L.*) *donovani* which can be verified as important drug target in future. Also one other gene which encodes Apical Membrane Antigen 1 (AMA1), though less research was done in leishmanial species regarding this protein and the involvement of this protein in pathogenesis is completely undefined, the involvement of this protein in virulence and pathogenesis in organisms such as plasmodium and *Toxoplasma gondii* attracts the future interest to investigate the importance of this protein in leishmanial pathogenesis. We hypothesize the specific AMA1 of *L.* (*L.*) *donovani* might follow a similar mechanism documented in apicomplexon parasites to invade the host cells though the mediators involved in host–parasite interaction are RON2 in the case of apicomplexon parasites and may be cholesterol in the case of *L.* (*L.*) *donovani*.

Matching our expectation, many genes which are found conserved with *L.* (*L.*) *infantum* showed less homology with remaining three leishmanial species. Apart from this, many genes were identified that are

Table 3
Repeat analysis of *L.* (*L.*) *donovani*.[a]

L. (*L.*) *donovani*[a]	*L.* (*L.*) *major*[a]	*L.* (*L.*) *infantum*[a]	*L.* (*Viannia*) *braziliensis*[a]	Length	Type	Function
LdBPK_211410	LmjF12.0730	**LinJ12.0690**	–	477	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.0740	**LinJ12_v4.0668**	–	477	SRR	Surface antigen protein
LdBPK_211410	LmjF12.0755	**LinJ12_v4.0663**	–	477	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.0760	**LinJ12_v4.0665**	–	477	SRR	Surface antigen protein 2
LdBPK_311480	LmjF12.0765	**LinJ12.0690**	LbrM12.0750	412	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.0830	**LinJ12_v4.0668**	–	477	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.0850	**LinJ12_v4.0668**	–	477	SRR	Surface antigen protein 2
LdBPK_311480	LmjF12.0870	**LinJ12_v4.0668**	–	412	SRR	Surface antigen protein 2
LdBPK_211410	LmjF21.1170	LinJ21.1410	LbrM21.1370	477	SRR	Surface antigen-like protein
LdBPK_311490	LmjF31.1450	LinJ31.1490	–	302	SRR	Surface membrane protein gp46-protein
LdBPK_040200	LmjF04.0210	LinJ04.0200	–	285	SRR	Surface antigen-like protein
LdBPK_211410	LmjF12.0860	–	–	477	SRR	Surface antigen protein
LdBPK_211410	LmjF12.0910	–	–	477	SRR	Promastigote surface antigen protein
LdBPK_211410	LmjF12.0920	–	–	477	SRR	Promastigote surface antigen protein
LdBPK_211410	LmjF12.0960	–	–	477	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.0990	–	–	477	SRR	Surface antigen protein
LdBPK_312750	LmjF12.1005	–	–	670	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.1020	–	–	477	SRR	Surface antigen protein
LdBPK_211410	LmjF12.1040	–	–	477	SRR	Surface antigen protein
LdBPK_211410	LmjF12.1060	–	–	477	SRR	Surface antigen protein
LdBPK_211410	LmjF12.1070	–	–	477	SRR	Surface antigen protein 2
LdBPK_211410	LmjF12.1090	–	–	477	SRR	Surface antigen protein
LdBPK_311490	LmjF31.1460	–	–	302	SRR	Surface membrane protein gp46 protein
LdBPK_363500	–	LinJ36.3500	–	2533	SAAR	Hypothetical transmembrane protein
LdBPK_322420	LmjF32.2270	LinJ32.2420	LbrM32.2500	341	SRR	Membrane associated protein-like protein
LdBPK_040170	–	–	LbrM04.0670	349	SRR	Surface antigen-like protein
LdBPK_040170	–	–	LbrM04.1260	349	SRR	Surface antigen-like protein
LdBPK_040170	–	–	LbrM04.1270	349	SRR	Surface antigen-like protein
LdBPK_040170	–	–	LbrM04.1340	349	SRR	Surface antigen-like protein
LdBPK_050240	LmjF05.0240	LinJ05.0240	–	433	SRR	Viscerotropic leishmaniasis antigen
LdBPK_040430	LmjF20.1180	**LinJ20.1210**	–	855	SRR	Calpain-like cysteine peptidase
LdBPK_270510	LmjF27.0490	LinJ.27.0500	LbrM.27.0600	5550	SRR	Calpain-like cysteine peptidase
LdBPK_270510	LmjF27.0500	**LinJ27.0500**	LbrM27.0600	5550	SRR	Calpain-like cysteine peptidase
LdBPK_270510	–	–	LbrM27.2140	5550	SRR	Calpain-like cysteine peptidase
LdBPK_270510	–	–	LbrM28.2100	5550	SRR	Calpain-like cysteine peptidase
LdBPK_350490	LmjF35.0500	LinJ35.0490	LbrM34.0520	453	SRR	Proteophosphoglycan ppg3
LdBPK_350500	–	LinJ35.0500	–	392	SRR	Proteophosphoglycan ppg3
LdBPK_350490	–	**LinJ35.0530**	–	453	SRR	Proteophosphoglycan 5
LdBPK_311480	–	**LinJ35.0540**	–	413	SRR	Proteophosphoglycan 5
LdBPK_020200	–	**–**	LbrM02.0240	993	TR	Phosphoglycan beta 1
LdBPK_060810	–	**–**	LbrM34.0530	3343	SRR	Proteophosphoglycan ppg4
LdBPK_080630	–	**–**	LbrM34.0540	2883	SRR	Proteophosphoglycan ppg4
LdBPK_350500	–	**–**	LbrM34.0550	392	SRR	Proteophosphoglycan ppg3
LdBPK_280600	–	**LinJ10.0520**	–	566	TR	GP63-3

[a] The IDs represent the GeneDb IDs of the corresponding leishmanial species. Bold highlighted — least homologous/pseudogenes.

found in *L.* (*L.*) *donovani* and few leishmanial species but absent or encode non functional proteins in other leishmanial species.

The comparative analysis list contains genes which encode many surface proteins noticeably amastins, amastin like surface proteins, peptidases etc., and these proteins are highly linked with pathogenesis, which may be experimentally verified further for complete understanding of distinct mechanism of *L.* (*L.*) *donovani* pathogenesis. As amastins have significant role in pathogenesis of any leishmanial species, the amastin comparison was done among all the five leishmanial species which identified few *L.* (*L.*) *donovani* specific amastin gene locations. Amastin gene family is represented more in *L.* (*L.*) *donovani* though the exact cause is not known. Also few genes which encode proteins that are not surface proteins but linked to pathogenesis directly or indirectly are discussed to their relevance. In addition, we identified a gene encoding a repeat containing protein, calpain like cysteine peptidases was identified and this particular protein is conserved in all leishmanial

species except in *L.* (*Viannia*) *braziliensis* where possibility of pseudogene presence was reported. Totally 12 genes which were under evolutionary selection were compared. Unexpectedly none of these genes were found in *L.* (*Viannia*) *braziliensis*. It is inferred that the possibility of deletion events as the evolution of *L.* (*Viannia*) *braziliensis* is fast and diverged unlike other leishmanial species because of the presence of transposable elements. Previously the comparative analysis of three leishmanial species *L.* (*L.*) *major*, *L.* (*L.*) *infantum and L.* (*Viannia*) *braziliensis* identified few genes which are specific to each leishmanial species (Peacock et al., 2007). The availability of genome sequence of two new leishmanial species *L.* (*L.*) *donovani and L.* (*L.*) *mexicana* and the complete comparative study of five species of leishmanial in the present study eliminated few genes reported as species specific genes for three leishmanial species *L.* (*L.*) *major*, *L.* (*L.*) *infantum* and *L.* (*Viannia*) *braziliensis* by Peacock et al. [data not shown], as these genes are identified in *L.* (*L.*) *donovani*.

The only specific gene family till now documented to be responsible for visceralization, the A2 gene family was analyzed for its evolutionary divergence. A2 gene family encodes A2 protein which was highly homologous to Stage Specific S Antigen of other leishmanial species, though the exact relevance of this stage specific expression is completely not understood. Also A2 gene copy number difference between leishmanial species was reported which clearly identified high copy number in *L.* (*L.*) *donovani* and *L.* (*L.*) *infantum*. The presence of A2 protein in *L.* (*L.*) *mexicana* was ambiguous which needs future verifications.

Specific amino acid repeats play a direct role in virulence from prokaryotes to eukaryotes. Complete *L.* (*L.*) *donovani* repeat containing proteins were identified and compared with the other leishmanial species in the present study.

Altogether the current study shows the complete analysis of the recent draft genome of *L.* (*L.*) *donovani*. The genes identified as *L.* (*L.*) *donovani* species specific can be experimented in the future to explore the complexity of *L.* (*L.*) *donovani* genome information and probably if some of these genes are established to be involved in pathogenesis, it can be a clear target for anti-leishmanial therapy.

Materials and methods

Data collection

The genome sequence of *L.* (*L.*) *donovani* and all other sequenced leishmanial species including *L.* (*L.*) *infantum*, *L.* (*L.*) *major*, *L.* (*Viannia*) *braziliensis* and *L.* (*L.*) *mexicana* was retrieved from NCBI GenBank and GeneDb (Benson et al., 2011; Logan-Klumpler et al., 2012). The total protein sequences of all other four leishmanial species were collected from Trityp database (Aslett et al., 2010).

Comparative analysis

The total protein sequence of *L.* (*L.*) *donovani* was simultaneously searched against the proteomes of other leishmanial species using HMMER package (Eddy, 2009) and Psi-blast (Altschul et al., 1997) and it was further confirmed by Bioedit local blastp package (Hall, 1999). The matches of *L.* (*L.*) *donovani* which was showing greater than 30% homology or e-value lesser than e^{-05} were eliminated using in-house Perl script and the remaining proteins coded by *L.* (*L.*) *donovani* genes which are completely non-homologous in all leishmanial species were considered species specific genes of *L.* (*L.*) *donovani*. The stringent e value is set to completely eliminate any homologous genes. The identified species specific genes were supposed to length criteria of greater than 90 codons and the genes which code proteins less than 30 amino acids were manually discarded, as it is very difficult to justify the proteins which are coded by less than 90 codons. Further *L.* (*L.*) *donovani* genes were individually compared with the other four leishmanial species and the genes which were present between *L.* (*L.*) *donovani* and any leishmanial species but absent in other leishmanial species were listed. All the possible combinations were done between *L.* (*L.*) *donovani* and other four leishmanial species and compared. In addition, the proteins encoded by these genes were further annotated by transferring functions of already assembled and annotated *L.* (*L.*) *major*, *L.* (*L.*) *infantum* and *L.* (*Viannia*) *braziliensis* proteins using blastp searches. Also annotations of *L.* (*L.*) *donovani* specific genes were done by searching the GO database using Amigo package (Carbon et al., 2009). Functionally homologous genes which show specificity between *L.* (*L.*) *donovani* and other species were also functionally annotated by Gene Ontology using Amigo tool. In addition, genes which were continuously under change (i.e.) genes under positive selection were detected using Psi-blast and blastp matches. The genes meeting these criteria were retrieved using in house Perl scripts. The

amastin gene sequences from different leishmanial species were collected and HMMER searches identified the homologous amastin genes in *L.* (*L.*) *donovani*. The identified amastin genes were compared and localized by searching all the leishmanial genome in parallel using ACT software (Carver et al., 2005).

Multiple sequence alignment (MSA)

A2 protein sequences and gene sequences were collected from NCBI. Blastp using NCBI blast (Altschul et al., 1990) and Fasta (Lipman and Pearson, 1985) format of those gene sequences identified the nearby homologous sequences from different species. Multiple sequence alignment was done using MUSCLE (Edgar, 2004). Also the location of the A2 gene family was identified by searching all the leishmanial species using Artemis software (Carver et al., 2005).

Repeat analysis

The entire *L.* (*L.*) *donovani* proteins were analyzed for the presence of amino acid repeats using Repseq database (Depledge et al., 2007). The entire repeat sequences from *L.* (*L.*) *infantum*, *L.* (*L.*) *major*, and *L.* (*Viannia*) *braziliensis* and Trypanosoma species were retrieved from Repseq database and checking the presence of the homologous sequence in *L.* (*L.*) *donovani* using HMMR, Psi-Blast and Blastp revealed the most probable repeat sequence in *L.* (*L.*) *donovani*. Further the proteins identified as species specific were also checked for the presence of repeats. Altogether this will reveal the complete repeat containing proteins of *L.* (*L.*) *donovani*.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mgene.2014.10.003.

Acknowledgments

This work was funded by the Ministry of Human Resource Development, Government of India. Also the authors acknowledge the lab members for help at various stages of the work.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., Gardner, M.J., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Logan, F.J., Miller, J.A., Mitra, S., Myler, P.J., Nayak, V., Pennington, C., Phan, I., Pinney, D.F., Ramasamy, G., Rogers, M.B., Roos, D.S., Ross, C., Sivam, D., Smith, D.F., Srinivasamoorthy, G., Stoeckert Jr., C.J., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., Wang, H., 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 38, D457–D462.
Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2011. GenBank. Nucleic Acids Res. 39, D32–D37.
Bringaud, F., Ghedin, E., Blandin, G., Bartholomeu, D.C., Caler, E., Levin, M.J., Baltz, T., El-Sayed, N.M., 2006. Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. Mol. Biochem. Parasitol. 145, 158–170.
Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288–289.
Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21, 3422–3423.
Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R.W., Alvar, J., Boelaert, M., 2007. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? Nat. Rev. Microbiol. 5, 873–882.
Charest, H., Matlashewski, G., 1994. Developmental gene expression in *Leishmania donovani*: differential cloning and analysis of an amastigote-stage-specific gene. Mol. Cell. Biol. 14, 2975–2984.
Dalton, J.P., Neill, S.O., Stack, C., Collins, P., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D., Khaznadji, E., Moire, N., Brennan, G., Mousley, A., Kreshchenko, N., Maule, A.G., Donnelly, S.M., 2003. *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. Int. J. Parasitol. 33, 1173–1181.
Depledge, D.P., Lower, R.P., Smith, D.F., 2007. RepSeq—a database of amino acid repeats present in lower eukaryotic pathogens. BMC Bioinforma. 8, 122.
Desjeux, P., 2001. The increase in risk factors for leishmaniasis worldwide. Trans. R. Soc. Trop. Med. Hyg. 95, 239–243.
Dolezal, P., Aili, M., Tong, J., Jiang, J.H., Marobbio, C.M., Lee, S.F., Schuelein, R., Belluzzo, S., Binova, E., Mousnier, A., Frankel, G., Giannuzzi, G., Palmieri, F., Gabriel, K., Naderer, T., Hartland, E.L., Lithgow, T., 2012. Correction: *Legionella pneumophila* secretes a mitochondrial carrier protein during infection. PLoS Pathog. 8.
Downing, T., Imamura, H., Decuypere, S., Clark, T.G., Coombs, G.H., Cotton, J.A., Hilley, J.D., de Doncker, S., Maes, I., Mottram, J.C., Quail, M.A., Rijal, S., Sanders, M., Schonian, G., Stark, O., Sundar, S., Vanaerschot, M., Hertz-Fowler, C., Dujardin, J.C., Berriman, M., 2011. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res. 21, 2143–2156.

Eddy, S.R., 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. 23, 205–211.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Elias, C.G., Pereira, F.M., Silva, B.A., Alviano, C.S., Soares, R.M., Santos, A.L., 2006. Leishmanolysin (gp63 metallopeptidase)-like activity extracellularly released by *Herpetomonas samuelpessoai*. Parasitology 132, 37–47.
Goto, Y., Carter, D., Guderian, J., Inoue, N., Kawazu, S., Reed, S.G., 2010. Upregulated expression of B-cell antigen family tandem repeat proteins by *Leishmania* amastigotes. Infect. Immun. 78, 2138–2145.
Guerin, P.J., Olliaro, P., Sundar, S., Boelaert, M., Croft, S.L., Desjeux, P., Wasunna, M.K., Bryceson, A.D., 2002. Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda. Lancet Infect. Dis. 2, 494–501.
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.
Jackson, A.P., 2010. The evolution of amastin surface glycoproteins in trypanosomatid parasites. Mol. Biol. Evol. 27, 33–45.
Joshi, P.B., Kelly, B.L., Kamhawi, S., Sacks, D.L., McMaster, W.R., 2002. Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. Mol. Biochem. Parasitol. 120, 33–40.
Kedzierski, L., Montgomery, J., Bullen, D., Curtis, J., Gardiner, E., Jimenez-Ruiz, A., Handman, E., 2004. A leucine-rich repeat motif of *Leishmania* parasite surface antigen 2 binds to macrophages through the complement receptor 3. J. Immunol. 172, 4902–4906.
Khan, M.G., Alam, M.S., Podder, M.P., Itoh, M., Jamil, K.M., Haque, R., Wagatsuma, Y., 2010. Evaluation of rK-39 strip test using urine for diagnosis of visceral leishmaniasis in an endemic area in Bangladesh. Parasites Vectors 3, 114.
Lipman, D.J., Pearson, W.R., 1985. Rapid and sensitive protein similarity searches. Science 227, 1435–1441.
Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J.A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., Phan, I., Farris, C., Mitra, S., Ramasamy, G., Wang, H., Tivey, A., Jackson, A., Houston, R., Parkhill, J., Holden, M., Harb, O.S., Brunk, B.P., Myler, P.J., Roos, D., Carrington, M., Smith, D.F., Hertz-Fowler, C., Berriman, M., 2012. GeneDB—an annotation database for pathogens. Nucleic Acids Res. 40, D98–D108.
Loughman, J.A., Caparon, M.G., 2006. A novel adaptation of aldolase regulates virulence in *Streptococcus pyogenes*. EMBO J. 25, 5414–5422.
Lustigman, S., Zhang, J., Liu, J., Oksov, Y., Hashmi, S., 2004. RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. Mol. Biochem. Parasitol. 138, 165–170.
McCall, L.I., Matlashewski, G., 2010. Localization and induction of the A2 virulence factor in *Leishmania*: evidence that A2 is a stress response protein. Mol. Microbiol. 77, 518–530.
Mottram, J.C., Coombs, G.H., Alexander, J., 2004. Cysteine peptidases as virulence factors of *Leishmania*. Curr. Opin. Microbiol. 7, 375–381.
Moyrand, F., Fontaine, T., Janbon, G., 2007. Systematic capsule gene disruption reveals the central role of galactose metabolism on *Cryptococcus neoformans* virulence. Mol. Microbiol. 64, 771–781.
News, B., 2006. Hope for Tropical Disease Vaccine.
Oliveira, T.M., de Vasconcelos, E.J., Nakaghi, A.C., Defina, T.P., Jusi, M.M., Baldani, C.D., Cruz, A.K., Machado, R.Z., 2011. A novel A2 allele found in *Leishmania* (*Leishmania*) *infantum chagasi*. Rev. Bras. Parasitol. Vet. 20, 42–48.
Peacock, C.S., Seeger, K., Harris, D., Murphy, L., Ruiz, J.C., Quail, M.A., Peters, N., Adlem, E., Tivey, A., Aslett, M., Kerhornou, A., Ivens, A., Fraser, A., Rajandream, M.A., Carver, T., Norbertczak, H., Chillingworth, T., Hance, Z., Jagels, K., Moule, S., Ormond, D., Rutter, S., Squares, R., Whitehead, S., Rabbinowitsch, E., Arrowsmith, C., White, B., Thurston, S., Bringaud, F., Baldauf, S.L., Faulconbridge, A., Jeffares, D., Depledge, D.P., Oyola, S.O., Hilley, J.D., Brito, L.O., Tosi, L.R., Barrell, B., Cruz, A.K., Mottram, J.C., Smith, D.F., Berriman, M., 2007. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat. Genet. 39, 839–847.
Pereira, F.M., Santos-Mallet, J.R., Branquinha, M.H., d'Avila-Levy, C.M., Santos, A.L., 2010. Influence of leishmanolysin-like molecules of *Herpetomonas samuelpessoai* on the interaction with macrophages. Microbes Infect. 12, 1061–1070.
Pourshafie, M., Morand, S., Virion, A., Rakotomanga, M., Dupuy, C., Loiseau, P.M., 2004. Cloning of S-adenosyl-L-methionine:C-24-Delta-sterol-methyltransferase (ERG6) from *Leishmania donovani* and characterization of mRNAs in wild-type and amphotericin B-resistant promastigotes. Antimicrob. Agents Chemother. 48, 2409–2414.
Pucadyil, T.J., Chattopadhyay, A., 2006. Cholesterol: a potential therapeutic target in *Leishmania* infection? Trends Parasitol. 23.
Pucadyil, T.J., Tewary, P., Madhubala, R., Chattopadhyay, A., 2004. Cholesterol is required for *Leishmania donovani* infection: implications in leishmaniasis. Mol. Biochem. Parasitol. 133, 145–152.
Rochette, A., McNicoll, F., Girard, J., Breton, M., Leblanc, E., Bergeron, M.G., Papadopoulou, B., 2005. Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania* spp. Mol. Biochem. Parasitol. 140, 205–220.
Rodriguez, N.E., Gaur, U., Wilson, M.E., 2006. Role of caveolae in *Leishmania chagasi* phagocytosis and intracellular survival in macrophages. Cell. Microbiol. 8, 1106–1120.
Sakanari, J.A., Nadler, S.A., Chan, V.J., Engel, J.C., Leptak, C., Bouvier, J., 1997. *Leishmania major*: comparison of the cathepsin L- and B-like cysteine protease genes with those of other trypanosomatids. Exp. Parasitol. 85, 63–76.
Santos, J.M., Ferguson, D.J., Blackman, M.J., Soldati-Favre, D., 2011. Intramembrane cleavage of AMA1 triggers *Toxoplasma* to switch from an invasive to a replicative mode. Science 331, 473–477.
Schneider, P., Glaser, T.A., 1993. Characterization of a surface metalloprotease from *Herpetomonas samuelpessoai* and comparison with *Leishmania major* promastigote surface protease. Mol. Biochem. Parasitol. 58, 277–282.
Tewary, P., Veena, K., Pucadyil, T.J., Chattopadhyay, A., Madhubala, R., 2006. The sterol-binding antibiotic nystatin inhibits entry of non-opsonized *Leishmania donovani* into macrophages. Biochem. Biophys. Res. Commun. 339, 661–666.
Toledo, J.S.d., Vasconcelos, E.J.R., Ferreira, T.R., Cruz, A.K., 2010. Using genomic information to understand *Leishmania* biology. Open Parasitol. J. 4, 156–166.
Tonkin, M.L., Roques, M., Lamarque, M.H., Pugniere, M., Douguet, D., Crawford, J., Lebrun, M., Boulanger, M.J., 2011. Host cell invasion by apicomplexan parasites: insights from the co-structure of AMA1 with a RON2 peptide. Science 333, 463–467.
Villanueva, M.S., Williams, S.P., Beard, C.B., Richards, F.F., Aksoy, S., 1991. A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. Mol. Cell. Biol. 11, 6139–6148.
Vinet, A.F., Fukuda, M., Turco, S.J., Descoteaux, A., 2009. The *Leishmania donovani* lipophosphoglycan excludes the vesicular proton-ATPase from phagosomes by impairing the recruitment of synaptotagmin V. PLoS Pathog. 5, e1000628.
Zhang, W.W., Matlashewski, G., 2001. Characterization of the A2–A2rel gene cluster in *Leishmania donovani*: involvement of A2 in visceralization during infection. Mol. Microbiol. 39, 935–948.
Zhang, W.W., Mendez, S., Ghosh, A., Myler, P., Ivens, A., Clos, J., Sacks, D.L., Matlashewski, G., 2003. Comparison of the A2 gene locus in *Leishmania donovani* and *Leishmania major* and its control over cutaneous infection. J. Biol. Chem. 278, 35508–35515.