Acta Tropica 223 (2021) 106086

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Genome wide comparison of *Leishmania donovani* strains from Indian visceral leishmaniasis and para-kala-azar dermal leishmaniasis patients

Nibedeeta Rani Sarraf [a], Saikat Mukhopadhyay [b], Anindyajit Banerjee [c], Anjan Kumar Das [d,1], Syamal Roy [e], Saikat Chakrabarti [f,***], Madhumita Manna [a,2,**], Partha Saha [b,*]

[a] Department of Zoology, Bidhannagar College, Kolkata, India
[b] Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Homi Bhabha National Institute, Kolkata, India
[c] Tata Translational Cancer Research Centre, Tata Medical Center, Kolkata, India
[d] Department of Pathology, Calcutta National Medical College, Kolkata, India
[e] CSIR-Indian Institute of Chemical Biology, Kolkata, India
[f] Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India

ARTICLE INFO

Keywords:
Visceral leishmaniasis
Post-kala-azar dermal leishmaniasis (PKDL)
para-kala-azar dermal leishmaniasis
Leishmania donovani
Whole genome sequencing
Drug resistance

ABSTRACT

Visceral leishmaniasis (VL) or Kala-azar, primarily caused by *Leishmania donovani*, is a major health concern in many countries including India. Growing unresponsiveness among the parasites toward the available drugs is alarming, and so, it is necessary to decipher the underlying mechanism of such development for designing new therapeutics. Moreover, even after successful treatment, some VL patients develop apparently harmless skin lesions known as post-kala-azar dermal leishmaniasis (PKDL) which may serve as a reservoir of the parasite in the transmission cycle. Furthermore, recent reports of para-kala-azar dermal leishmaniasis (para-KDL) cases having PKDL manifestation with concomitant VL, emphasize the necessity of more attention to address complex nature of the parasite for eradicating the disease effectively. In the present study, whole genome sequencing is performed with sodium stibogluconate (SSG) sensitive and resistant *L. donovani* strains along with SSG sensitive para-KDL strains, derived from the clinical isolates of Indian patients to identify the genomic variations among them. Notably, the analyses of chromosome somy values and genome wide mutation profile in the coding regions reveal distinct clustering of the para-KDL strains with 24 genes being mutated uniquely in this group. Such distinguishing genomic changes among the para-KDL strains could be significant for the parasites to become dermatotropic. Overall, the study reveals a possible correlation of the development of SSG resistance and the transition towards the manifestation of PKDL with chromosome aneuploidy and non-synonymous genetic variations in the coding sequences of the *L. donovani* strains from Indian patients.

1. Introduction

Leishmaniases, caused by at least 20 species of the genus *Leishmania* and spread across the mammalian hosts by the female phlebotomine flies, range from self-healing cutaneous lesion to potentially fatal visceral form and are endemic in more than 90 countries. Visceral leishmaniasis (VL) or Kala-azar (KA), primarily caused by *Leishmania donovani*, is prevailing in more than 60 countries with about 50,000 to 90,000 new cases annually. However, 95% of the VL incidences occur in India, Brazil, Ethiopia, Kenya, Nepal, Somalia, Sudan and South Sudan (Burza et al., 2018; Mann et al., 2021), (WHO). In India, the eastern part, particularly the state of Bihar, is the endemic focus of VL infection, where the periodic infection is common.

For decades, pentavalent antimonial sodium stibogluconate (SSG) has been used as the major drug for treatment of VL. However, growing resistance in the parasite makes its use limited for treatment of VL in the Indian subcontinent, although it remains effective in East Africa (Sundar and Singh, 2018). Miltefosine (MIL), primarily an oral cancer drug,

* Corresponding author: Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Homi Bhabha National Institute, Kolkata, India
** Corresponding author: Department of Zoology, Bidhannagar College, Kolkata, India
*** Corresponding author: Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India
E-mail addresses: saikat@iicb.res.in (S. Chakrabarti), madhumita.manna09@gmail.com (M. Manna), partha.saha@saha.ac.in (P. Saha).
[1] Present address: Department of Pathology, Coochbehar Government Medical College, West Bengal, India
[2] Present address: Higher Education Directorate, Govt. of West Bengal, Bikash Bhavan, Kolkata 700091

https://doi.org/10.1016/j.actatropica.2021.106086
Received 24 February 2021; Received in revised form 29 July 2021; Accepted 30 July 2021
Available online 10 August 2021

has been introduced in 2002 to treat VL patients, especially for the SSG unresponsive cases (Jha, 2006). However, within a decade, MIL resistant cases have started appearing as well (Sundar et al., 2012). Liposomal amphotericin B is another treatment option for VL that is still effective, but the parasites may develop resistance due to its widespread use (Hefnawy et al., 2017). As *L. donovani* causes a major part of the leishmaniasis burden both in terms of morbidity and mortality, particularly in the Indian sub-continent and East Africa (Downing et al., 2011; Sacks et al., 1995), it is essential to understand the underlying mechanism of the emergence and spreading of drug resistance among the parasite populations.

Unfortunately, among the apparently cured VL patients, 10 to 20% in India and 50-60% in Sudan develop post-kala-azar dermal leishmaniasis (PKDL), consisting of painless macular and/or papulonodular skin lesions (Burza et al., 2018). Though the PKDL symptoms may not be very harmful, the lesions on the patients harbour the parasites which can be transmitted by the bites of sand flies (Zijlstra et al., 2017). Therefore, PKDL patients possibly play a critical role in the transmission cycle of the parasites and are considered to be a major concern in the elimination programmes of VL. Hence, it is important to elucidate the mechanism that assists the parasite to become dermatotropic.

The development of PKDL in VL patients is a gradual transition as a proportion of patients having skin lesions are found with concomitant presence of the parasites in bone marrow, lymph node or spleen in several studies (Zijlstra, 2019; Zijlstra et al., 2000; Zijlstra et al., 2003). Such symptoms are referred to as para-kala-azar dermal leishmaniasis (para-KDL). Though VL associated with PKDL was rarely found in the Indian subcontinent previously (Singh et al., 1989; Zijlstra et al., 2003), more recently nine such cases have been reported from Bihar, India (Kumar et al., 2016). Therefore, simultaneous presence of VL in increasing number of PKDL cases requires more attention to understand the pathology, diagnosis and treatment regime of these emerging incidences of para-KDL.

Leishmania parasites can re-shape their genomes rapidly in response to stress (Grunebast and Clos, 2020; Mannaert et al., 2012), implying the importance of genomic variation by which they can promptly acclimatize to changing environmental condition and drug pressure. Though many studies have been carried out using molecular methods to study the population genetics of *Leishmania*, limited information is available on the diversity of the parasite populations in the clinical context and their modification during treatment (Alam et al., 2009; Schonian et al., 2008). The existing genetic markers have poor resolution in this regard, and very little genetic differentiation can be identified using these approaches among *L. donovani* population, particularly in the Indian sub-continent (Downing et al., 2012; Imamura et al., 2016). On the other hand, whole-genome sequencing of different isolates and their comparison have the potential to identify significant variations in the parasitic population. Keeping such objectives in mind, we report here the genome comparison of nine *L. donovani* strains from VL and para-KDL clinical isolates collected from Indian patients between 2010 and 2014. Interestingly, several novel mutations are detected in the parasites from para-KDL patients which could be responsible for gradual transition towards the development of PKDL in the apparently cured VL patients.

2. Materials and methods

2.1. Ethics statements

Collection of Bone marrow or splenic aspirates from VL and para-KDL patients were approved by the Ethical Committee of the Calcutta National Medical College, Kolkata, India. The written consent was obtained from each patient or guardian (for a minor patient) prior to the study.

2.2. Collection of isolates and culture of parasite

Nine clinical isolates from VL and para-KDL patients of India (Table S1) including five from our previous studies (Khanra et al., 2012; Khanra et al., 2017; Khanra et al., 2016) were analysed in the present work. The bone marrow aspirates with demonstrable presence of the parasite were collected from the three patients having PKDL-like skin patches (Table S1) and used in the present study. Such patients having simultaneous symptom of VL and PKDL were termed as para-KDL cases (Zijlstra et al., 2000; Zijlstra et al., 2003). The symptoms of these patients were fever and hepatosplenomegaly associated with macular patches on skin. The occurrence of amastigotes in the bone marrow or splenic aspirates of the patients was established by Giemsa staining as well as transformation into promastigotes in M199 (Sigma-Aldrich) culture medium supplemented with 10% Fetal Bovine Serum (Invitrogen-Thermo Fisher) at 22°C. In vitro drug sensitivity of the intracellular amastigotes was evaluated (Khanra et al., 2017; Khanra et al., 2016) in murine macrophage cell line RAW 264.7, which was maintained in RPMI-1640 medium supplemented with 10% FBS at 37°C and in a humid atmosphere containing 5% CO_2. All promastigotes used in the present study were cloned as described elsewhere (Bhattacharyya et al., 2002; Pal et al., 2001).

2.3. Whole genome sequencing and analysis

Isolation of genomic DNA from promastigotes of the cloned strains was carried out using QIAamp DNA isolation kit (Qiagen) and then characterized by ITS1 sequencing and PCR-RFLP methods (Alam et al., 2009; Khanra et al., 2012; Khanra et al., 2016). High quality RNA free DNA libraries were prepared using Ion plus Fragment Library Kit (Thermo-Fisher). The Libraries were quantified by Qubit DNA HS kit and their quality-check was carried out in Agilent 2100 Bioanalyzer. Enriched and template-positive Ion PI™ Ion Sphere™ Particles (ISPs) containing clonally amplified DNA up to 200 bp was prepared from the qualified DNA library by OneTouch™ 2 System (Thermos-Fisher). The enriched library was sequenced in Ion Proton Semiconductor Sequencer using Ion PI™ Chip V3 and Ion PI™ Hi-Q™ Sequencing 200 reagents. The adapter sequences were removed from the reads and the data were trimmed based on Phred quality score followed by alignment to the *L. donovani* BPK282 reference strain using Ion Torrent Suite Version 4.2. The SNPs analysis was carried out using low frequency variant detection option in CLC genomics workbench 12 (Qiagen). In addition, the in-house perl scripts were also used to analyse the SNP data derived from the sequencing platform. To eliminate the pseudo variants from the analyses, at least 100x or more coverage at a specific position was ascertained and variant allele frequencies (VAF, fraction of mutant reads) of 25 % or above were considered. For the present study, the non-synonymous mutations in the protein coding genes were only considered.

2.4. Somy estimation

The chromosomal somy values were estimated based on variability of sequence depth and median read depth of the sample (Downing et al., 2011; Dumetz et al., 2017; Imamura et al., 2016). Read depth for each position of chromosome in a sample was estimated using "Samtools Depth" in galaxy web server version 1.9 (https://usegalaxy.org/). To calculate the median read depth, the first 7,000 bases and last 2,000 bases of each chromosome were excluded because these regions tend to be repetitive telomeric regions and their depths were not reliable (Domagalska et al., 2019). Similarly, certain over-amplified chromosome regions were excluded from median depth calculation (Table S3). The resulting raw median read depth of an individual chromosome was normalized against the median read depths of the 15 chromosomes - 7 neighbouring chromosome on each side and the chromosome itself (Dumetz et al., 2017). The chromosome somy value S was finally

obtained by dividing 2 × raw median depth value of the chromosome with its normalized median depth value. The range of monosomy, disomy, trisomy, tetrasomy, and pentasomy was defined as described previously (Dumetz et al., 2017). The somy values of each chromosome in a sample were plotted to create the clustered heatmap using the Rplot of heatmap2.

2.5. Gene clustering and pathway analyses

Principal component analysis (PCA) was performed based on the mutation profile in the coding regions as well as on the genes that contained those mutations in the individual strains using ClustVis (Metsalu and Vilo, 2015). Further, the Venn-diagrams were generated to analyse the overlap of the mutated genes of individual strains, using the webtool to calculate and draw custom Venn Diagram at http://bioinformatics.psb.ugent.be/webtools/Venn/. In order to understand the functional involvement of mutated genes of individual strains, the pathway analysis was carried out by mapping the respective sets of the mutated gene of each strains onto the *L. donovani* specific KEGG pathway database (Du et al., 2014).

3. Results

3.1. Whole genome sequencing

To identify the genetic variations that could lead to manifestation of PKDL and development of SSG resistance in Indian VL patients, whole genome sequencing (WGS) of nine *L. donovani* strains from the clinical isolates (Table S1) were carried out on Ion Proton plus platform followed by mapping to the reference genomic sequence of LdBPK282/Oc14 (Downing et al., 2011). Three strains each of SSG sensitive VL (VL-SSG-S), SSG resistant VL (VL-SSG-R) and SSG sensitive para-KDL (para-KDL-S) were isolated from the bone marrow or splenic aspirates of the patients from Bihar-West Bengal area of India during 2010-2014 (Table S1) and cloned. The strains were genotyped by PCR-RFLP or PCR-ITS1 sequencing methods and their drug sensitivities were

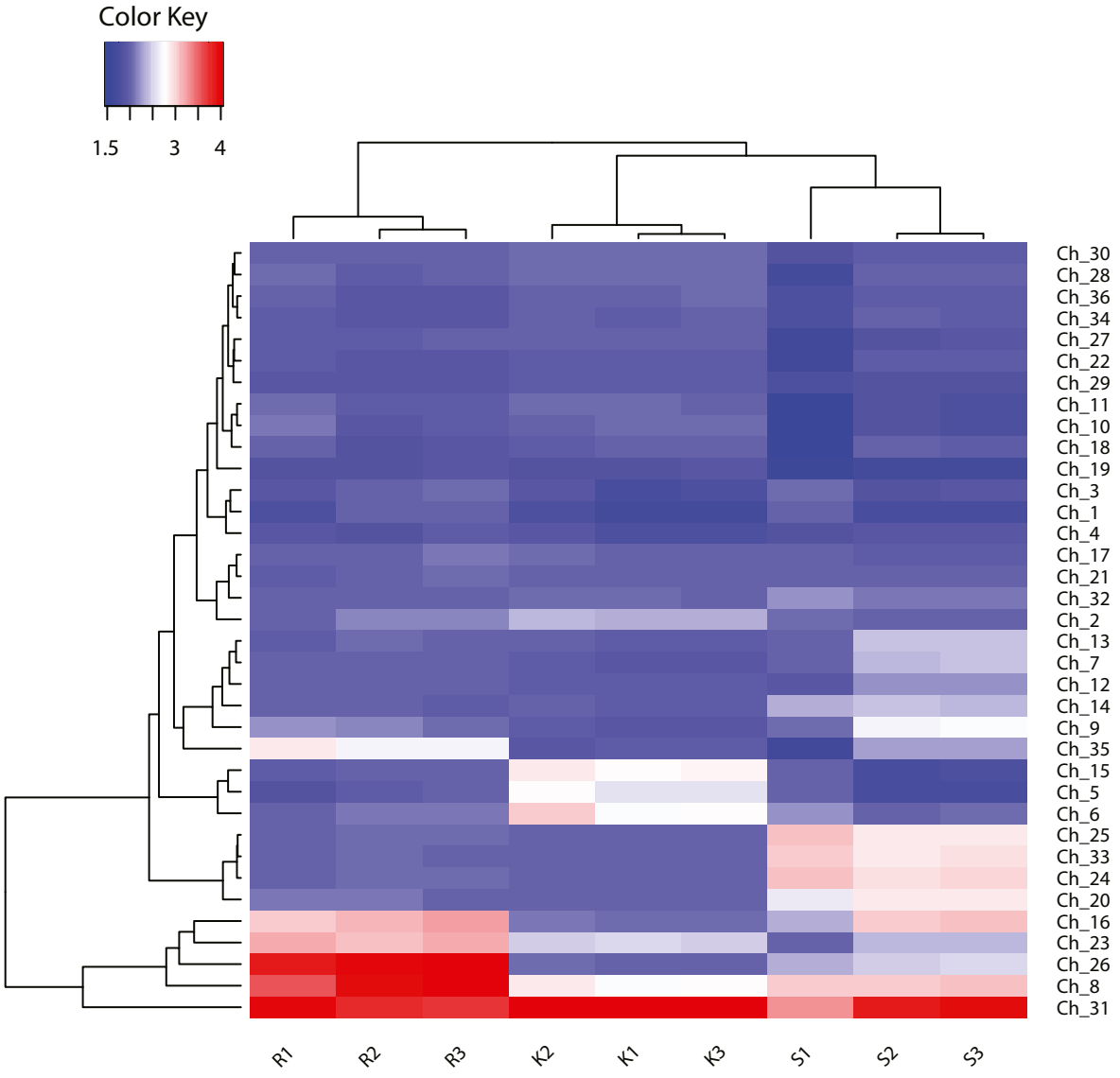


Fig. 1. Comparison of aneuploidy profiles of the *L. donovani* strains obtained from clinical isolates. Chromosome numbers are indicated on the right side of the heatmap and the study codes of the *L. donovani* isolates are mentioned across the bottom of the map. The dendrogram on the top indicates the clustering of the strains based on similarity of aneuploidy profile. Similarly, the dendrogram on the left represents the clustering the chromosomes. The insert shows the colour code that is used to indicate the ploidy of chromosomes.

confirmed. A mean depth coverage of 117-fold was obtained for each strain in WGS, and on an average, 97% of the sequence reads could be mapped to the reference genome (Table S2).

3.2. Aneuploidy

To evaluate the possible association of aneuploidy with the development of PKDL and the drug resistance in the Indian VL patients, chromosome copy numbers were estimated using normalized whole chromosome median read depths (Domagalska et al., 2019; Downing et al., 2011; Dumetz et al., 2017). The estimated somy values (Table S3) showed that each strain contained on an average 30 disomic chromosomes. In para-KDL-S strains, number of trisomic and tetrasomic chromosomes were estimated to be 4 and 1, respectively, whereas the numbers were 6 and 1, respectively, in VL-SSG-S strains. Notably, the proportion of tetrasomy was estimated to be more in VL-SSG-R strains.

Analysis of the somy values revealed the presence of significant variability in aneuploidy across the strains, though the SSG sensitive (VL and para-KDL) and resistant strains were broadly clustered distinctly (Fig. 1 and Table S3). Considering somy value 2±0.5 as disomy, 18 chromosomes were found to be disomic in all strains (Chromosomes 1-4, 7, 12-14, 17, 19, 21, 22, 27-30, 32, 34, 36). On the other hand, chromosomes 8 and 31 showed variable somy (trisomy to pentasomy) across the strains as reported earlier (Downing et al., 2011). Other chromosomes showed characteristics changes in somy among the analysed strains. As shown in Fig. 1, among the SSG sensitive strains, VL-SSG-S (S1, S2 and S3) and para-KDL-S (K1, K2 and K3) strains were further grouped separately. However, S1 strain showed unique aneuploidy profile compared to other two VL-SSG-S strains, though all are collected from the same geographical region. Distinct changes were observed in S1 strain compared to S2 and S3 strains for Chromosomes 9 and 16 (trisomy in S2 and S3 to disomy in S1), chromosomes 10, 11 and 18 (disomy in S2 and S3 to monosomy in S1), and chromosome 31 (tetrasomy in S2 and S3 to trisomy in S1) (Table S3). Also, VL-SSG-S and para-KDL-S strains could be distinguished based on significant change of average chromosome somy values ($p<0.05$) with chromosomes 5, 6 and 15 showing disomy to trisomy and chromosomes 16, 20, 24, 25 and 33 showing trisomy to disomy, respectively (Fig. 1 and Table 1). Similarly, distinct changes could also be observed between VL-SSG-S and VL-SSG-R strains, with a greater number of chromosomes showing varying average somy values (Fig. 1 and Table 2). Strikingly, for certain chromosomes, unique variations were noted when somy values of para-KDL-S and VL-SSG-R strains were compared with that of VL-SSG-S strains (Table 3). As shown in Table 3, for chromosomes 5, 6, 15 and 16, overall somy changes were observed only between VL-SSG-S and para-KDL-S strains. On the other hand, somy variation was found in chromosomes 8, 23, 26 and 35 only between VL-SSG-S and VL-SSG-R strains (Table 3). Thus, differential somy changes in SSG sensitive strains could be important for transition to PKDL or development of the drug resistance.

Table 1
Distinction between SSG sensitive para-KDL and VL strains based on average chromosome ploidy.

Chromosome	Mean_K[a]	Mean_S[a]	P value	Mean_K/Mean_S
15	2.81	1.84	0.000	1.53
5	2.64	1.81	0.001	1.46
6	2.85	2.07	0.002	1.38
11	2.03	1.69	0.050	1.20
2	2.33	2.01	0.000	1.16
19	1.87	1.63	0.004	1.14
29	1.95	1.82	0.013	1.07
30	2.05	1.92	0.019	1.06
8	2.79	3.06	0.007	0.91
26	2.02	2.41	0.003	0.84
14	1.95	2.35	0.000	0.83
16	2.08	2.81	0.042	0.74
20	2.00	2.79	0.001	0.72
25	2.01	2.94	0.000	0.68
24	2.02	2.99	0.000	0.68
33	1.98	2.96	0.000	0.67

[a] K, SSG sensitive para-KDL strains. S, SSG sensitive VL strains

Table 2
Distinction between SSG resistant and sensitive VL strains based on average chromosome ploidy.

Chromosome	Mean_R[a]	Mean_S[a]	P value	Mean_R/Mean_S
26	3.96	2.41	0.000	1.64
23	3.16	2.24	0.002	1.41
35	2.74	2.05	0.037	1.33
8	3.83	3.06	0.005	1.25
19	1.86	1.63	0.003	1.14
14	1.97	2.35	0.000	0.84
20	2.05	2.79	0.001	0.74
25	2.02	2.94	0.000	0.69
33	2.01	2.96	0.000	0.68
24	2.03	2.99	0.000	0.68

[a] R, SSG resistant VL strains. S, SSG sensitive VL strains

Table 3
Distinct change of chromosome ploidy among SSG sensitive VL and para-KDL strains and SSG resistant VL strains.

Chromosome	Mean_S[a]	Mean_K[a]	Mean_R[a]
5	1.81	2.64	1.93
6	2.07	2.85	2.06
15	1.84	2.81	2.00
16	2.81	2.08	3.15
8	3.06	2.79	3.83
23	2.24	2.46	3.16
26	2.41	2.02	3.96
35	2.05	1.93	2.74

[a] S, SSG sensitive VL strains; K, SSG sensitive para-KDL strains, R, SSG resistant VL strains

3.3. Mutation analysis in coding sequences

To estimate the contribution of genome-wide mutation profile of the protein coding genes in the development of PKDL and the drug resistance, the analysis of 3522 unique non-synonymous mutations (InDels; multi-nucleotide variations, MNVs and single nucleotide variations, SNVs) across the strains (Table S4) was performed. The mutation burden was found to be the least in R2 and the most in R3. Overall, on an average only 2.8% mutations were homozygous (Table S4). However, significantly low percentage of homozygous mutation was observed in para-KDL-S strains compared to that in VL-SSG-S strains ($P<0.02$ with z-statistic 2.495). But, homozygous mutations in four genes were detected in the three para-KDL-S strains only (Table S4), suggesting their possible association with the development of the symptom.

The extent of similarities among the different strains based on the pair wise analysis of non-synonymous mutation profiles in the coding sequences and profiles of the affected genes were shown in Fig. 2. In the heat-maps, each coloured square represented fraction of similarity between the strain mentioned on the right and corresponding strain indicated at the bottom. As shown, the para-KDL-S strains were found to form a distinct cluster in the heat-maps based on total mutation profiles in coding sequences as well as profiles of mutated genes (3 × 3 cluster in the top left corners of Fig. 2A and B). The para-KDL-S strains also showed fewer similarities with the other strains (mainly lower similarity squares in 6 × 3 rectangles of R1-S3 × K1-K3 strains in Fig. 2A and B). Similarly, VL-SSG-S strain S1 and VL-SSG-R strain R3 showed less resemblance with most of the other strains.

To evaluate the significance of genetic differences across the strains, the principal component analysis (PCA) was performed based on the frequency of occurrence of the non-synonymous mutation among the genes in the individual strains (Fig. 3 and Table S5). As shown in Fig. 3A, two most significant principal components resulted in distinct clustering

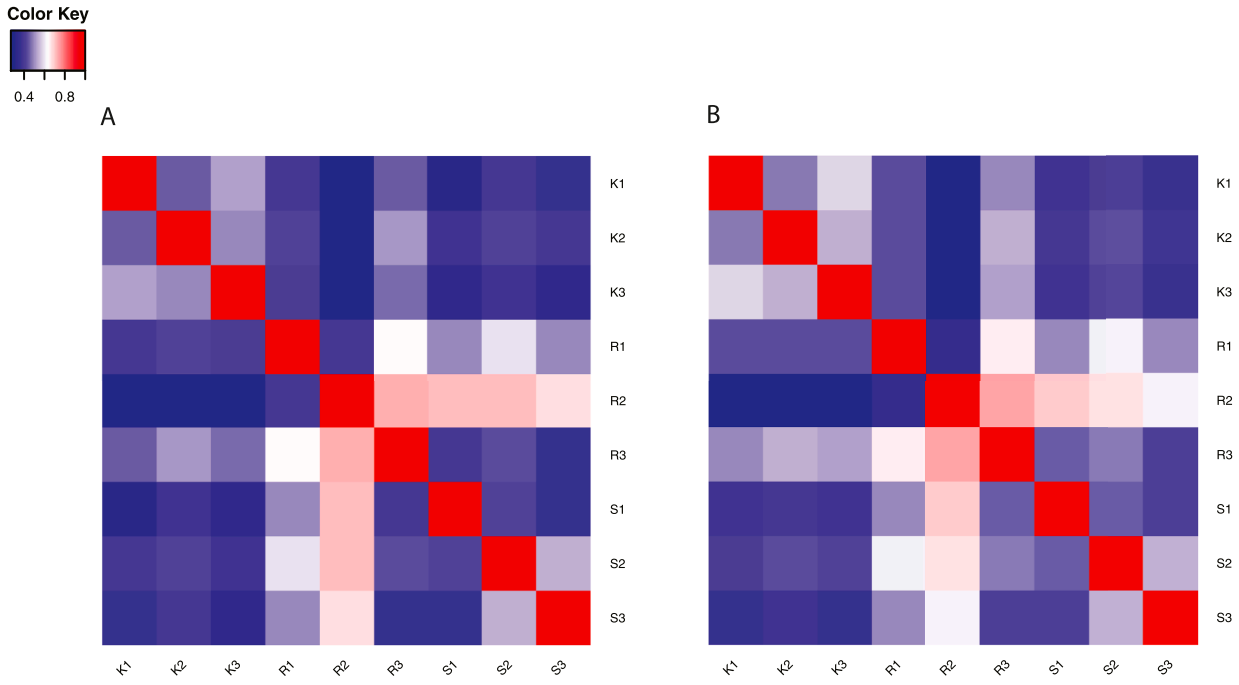


Fig. 2. Comparison between mutation profiles of the *L. donovani* strains obtained from clinical isolates. The heat-map visualizes the similarities between all mutations in the coding sequences of the sequenced strains. Each coloured square in the map represents the fraction of similarity between strain mentioned on the right and that listed at the bottom. The colour code representing the fraction of similarity is shown. A. Comparison based on the complete mutation profiles. B. Comparison based on the profiles of genes that were mutated.

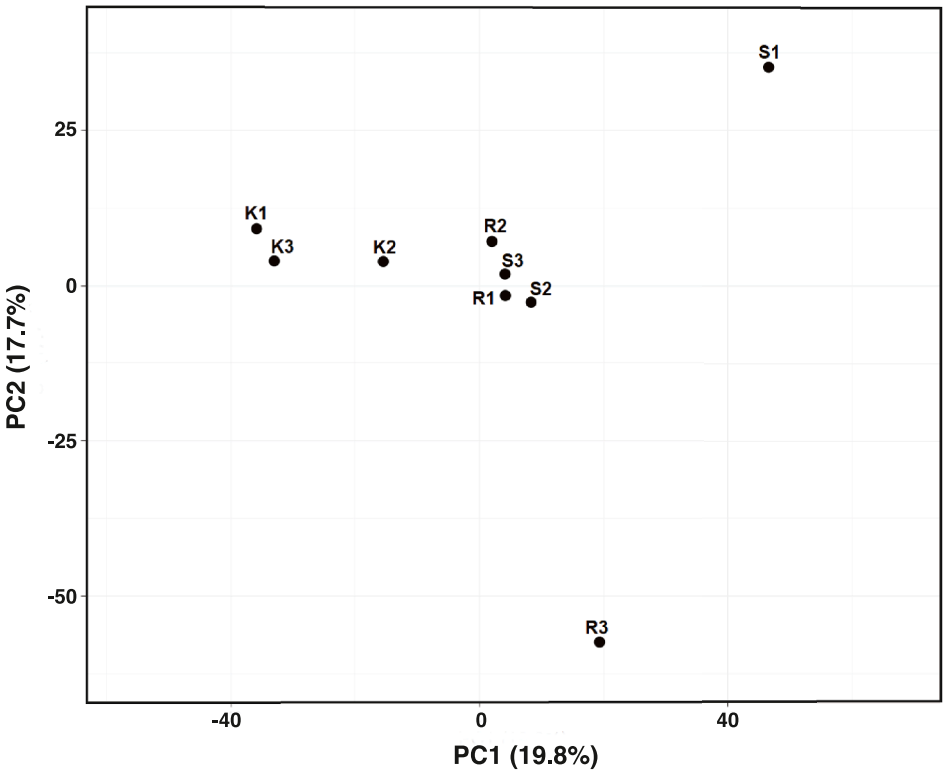
of para-KDL strains with slight deviation of K2 from K1 and K3, indicating an overall covariance of the mutation profiles. The identification of comparable number of mutations among the para-KDL strains and distinctly higher number of para-KDL specific mutations (Table S4 and Fig. 4A and D; Table S6) justified such a separate clustering. Fig. 3 also depicted that VL-SSG-S strains S2 and S3 formed a cluster with VL-SSG-R strains R1 and R2. Strikingly, VL-SSG-S strain S1 and VL-SSG-R strain R3 distinctly deviated from the clustering groups. The comparative mutation analysis across the R strains clearly revealed that the number of mutation present in R3 were quite different from that in R1 and R2 (Fig. 4B, Table S6). These data supported R3 being a far distant strain in comparison to R1 and R2 strains (Fig. 3). Similarly, the mutational information among the S strains showed comparatively higher mutation burden in S1 strain (Fig. 4C, Table S6), which was clearly reflected in the PCA analysis plot (Fig. 3). In order to validate the clustering analysis performed with the mutation profiles, the affected mutated genes were further considered for the principal component analysis. As represented in Fig. 3B, the clustering analysis using the mutated gene sets highly correlated with that using the total mutation datasets.

The mutational analysis at the genomic level revealed the presence of 61 genes, which were found to get mutated across the all strains studied. Further, the pathway analyses revealed that most of these 61 genes are involved in different metabolic processes (Fig. S1 and Table S7). Interestingly, the analysis also revealed the occurrence of 24 genes which were mutated only in the para-KDL-S strains, with 13 of them being homologous to genes with previously known function (Table 4 and Table S8). As shown in Table 4, various cellular processes including metabolism, ion transport, repair and translation could be affected due to the mutations, possibly contributing significantly towards the development of PKDL.

4. Discussion

In the present study, nine *L. donovani* strains including three SSG sensitive VL strains, three SSG resistant VL strains, and three para-KDL strains derived from the patients' isolates of Bihar and West Bengal region in India are subjected to whole genome sequencing. All the strains used in the study are promptly derived from the patients' isolates and cultured under identical conditions for similar number of passages prior to isolation of DNA for WGS to minimize variable effects due to environmental conditions. Also, distinct clustering of the strains derived from similar disease condition based on aneuploidy is observed in our study. Therefore, the sequenced parasites can safely be said to represent that of the corresponding original patient isolates, as observed previously (Downing et al., 2011; Zackay et al., 2018).

Like other clinical isolates, extensive aneuploidy is observed across the strains with distinct changes in VL-SSG-R and para-KDL-S strains compared to VL-SSG-S strains. In the absence of transcriptional regulation of gene expression in kinetoplastida parasites, variation of gene dosage via aneuploidy and mRNA turnover is critical to maintain a differential pool of messages for adopting prompt survival strategy in response to stressful and changing environments in different hosts (Dumetz et al., 2017). Higher somy of a chromosome increases the mean gene expression from it, and expectedly, the gene expression from the trisomic chromosomes has been reported to be significantly higher than that from disomic chromosomes (Iantorno et al., 2017; Ubeda et al., 2008). Interestingly, trisomy is noticed for chromosomes 5 and 6 exclusively in the para-KDL strains. Several essential genes are located on chromosomes 5 and 6, whose protein products are involved in various metabolic activity (Manzano et al., 2017; Tovar et al., 1998; Vickers and Beverley, 2011), including three ABCG (ATP-binding cassette G) subfamily of transporters – LABCG1 (LdBPK_060080), LABCG2 (LdBPK_060090) and LABCG3 (LdBPK_060100) on chromosome 6. Recent studies show the association of LABCG2 transporter with polarization of parasitized macrophage to M2 phenotype leading to *Leishmania* survival and PKDL manifestation (Mukhopadhyay et al., 2015). But the three ABCG subfamily of transporters on chromosome 6 are deleted in the para-KDL strains studied here. However, three more copies of the similar genes - LABCG4 (LdBPK_150950), LABCG5 (LdBPK_230430) and LABCG6 (LdBPK_363040), are present on chromosomes 15, 23 and 36, respectively, without any change. Intriguingly, chromosome 15 is trisomic in the para-KDL strains, implicating higher



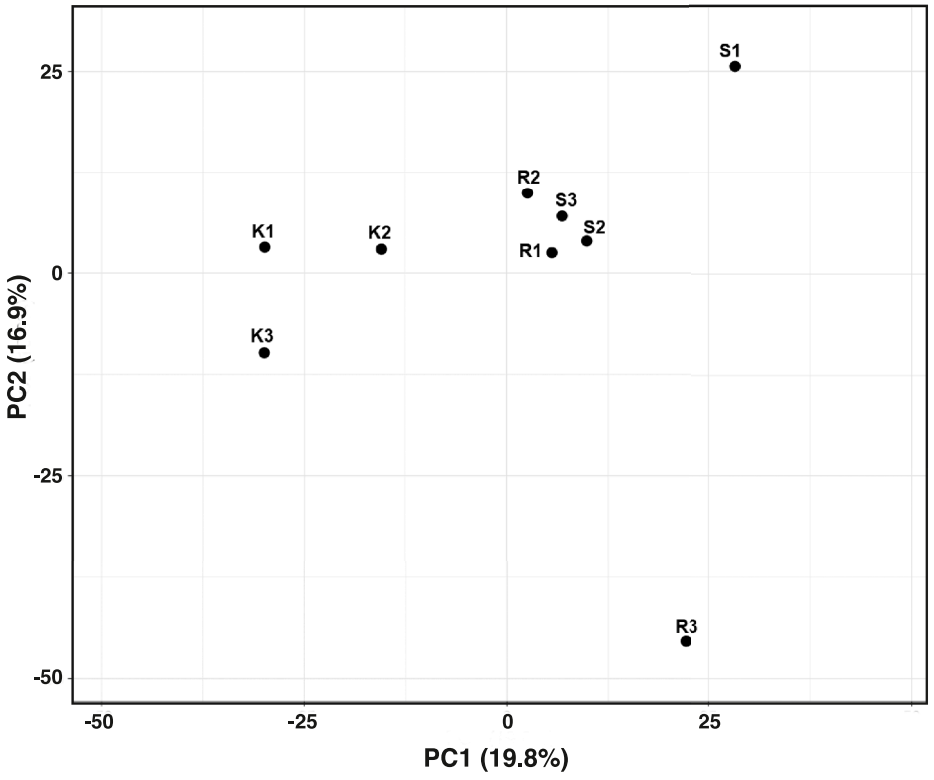


Fig. 3. Principal component analysis (PCA) based on (A) the mutational data types in protein coding genes and (B) the mutated genes in the *L. donovani* strains from the clinical isolates. In both the clustering analyses, two most significant PCs resulted in formation two clusters – one consisting of para-KDL-SSG-S strains (K1, K2 and K3) and the other including VL-SSG-S strains S2 and S3 and VL-SSG-R strains R1 and R2. Strikingly, VL-SSG-S strain S1 and VL-SSG-R strain R3 were independently and distantly located from the two clusters in the PCA plot.

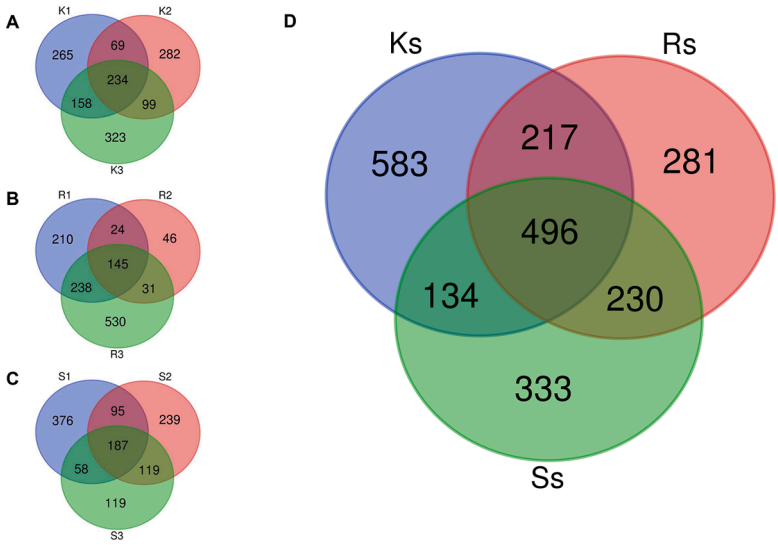


Fig. 4. Venn diagram of intra- and inter-strain specific clustering analysis. A to C. Intra-strain clustering analysis of para-KDL-SSG-S (K1, K2 and K3) strains, VL-SSG-R (R1, R2 and R3) strains and VL-SSG-S (S1, S2 and S3) strains. D. Inter strain clustering analysis using K's R's and S's strains simultaneously.

Table 4
List of genes mutated only in para-KDL strains and encoding previously known proteins.

	Changes in coding sequence		Encoded protein	
GeneDB ID	Nucleotide	Amino acid		
LDBPK_040010	2944delA	Met982fs	calcium-translocating P-type ATPase	Ion Transport
LDBPK_080830	216delA	Ala73fs	mitochondrial DNA polymerase beta	Repair
LDBPK_120610	2548C>T	Gln850stop	serine/threonine protein phosphatase with EF-Hand domain	Phosphatase
LDBPK_130390	4046T>C	Ile1349Thr	RNA helicase, putative	Helicase
LDBPK_210010	154C>A	Pro52Thr	phosphoglycan beta 1,3 galactosyltransferase 2	Metabolism
LDBPK_210700	1685C>T	Ala562Val	phosphoglucomutase, putative	Metabolism
LDBPK_231970	760delC	Leu254fs	neutral sphingomyelinase activation associated factor-like protein	Metabolism
LDBPK_250010	112delT	Tyr38fs	beta galactofuranosyl transferase	Metabolism
LDBPK_310600	810delT	Phe270fs	amino acid transporter aATP11, putative	Transport
LDBPK_311290	3539G>C	Ser1180Thr	p-glycoprotein e, partial	Metabolism
LDBPK_321520	5626delG	Val1876fs	phosphatidylinositol 3-related kinase, putative	Metabolism
LDBPK_332650	1255G>A	Ala419Thr	UDP-N-acetylglucosamine pyrophosphorylase, putative	Metabolism
LDBPK_360270	181C>A	Pro61Thr	EIF3-interacting protein-like protein	Translation

expression of LABCG4 possibly leading to the survival of the parasite.

In our study, distinct clustering of para-KDL strains in pairwise analysis as well as principal component analysis strongly suggest the covariance of the unique sets of mutations. Further identification of 24 genes that are mutated only in the three para-KDL strains, also supports the commonalities in mutation burden. Non-synonymous mutations present in the coding sequences are crucial, because they can change the structure-function properties of the encoded proteins. Moreover, mutations in the four out of these 24 genes are homozygous in nature, raising the probability of more pronounced effect on the functionality of the gene-products. Also, thirteen of them are homologous to genes with previously known cellular function including metabolism, transport, repair and RNA activity. Thus, the appearance of the unique set of mutations in the para-KDL strains, as identified in the present study, can be important for viscerotropic to dermatotropic transition of the parasites.

The isolated positions of VL-SSG-S strain S1 and VL-SSG-R strain R3 in the PCA plots and their distinct aneuploidy profiles imply unique mutation burden for each of these strains. However, the strains are obtained from the isolates of the patients having close geographical locations, implying a possible role of divergent host immune responses for the development of the unique mutation profiles and aneuploidy patterns in the strains.

The association of *L. donovani* with viscerotropism as well as dermatotropism in Sri Lanka has been reported (Ranasinghe et al., 2012) with differential copy number variation (CNV) of some genes as a possible contributing factor causing visceral and cutaneous (CL) manifestation (Zhang et al., 2014). Significant CNVs are also observed for some of these gene loci in the Indian strains studied here (Tables S9 and S10). However, no differential CNV is observed among the Indian strains except for one major facilitator superfamily protein in chromosome 29 (LdBPK_291620), which is absent in SSG-R-VL strains. The absence of ABC transporter like protein (LdBPK_111220) and phosphoglycerate kinase B (LdBPK_200120) is the other notable feature found among the Indian strains. Moreover, similar average copy number of A2 and A2rel repeat cluster (LdBPK_220670) is found in all the Indian strains (Table S9), though its lower copy number in CL-causing *L. donovani* strains of Sri Lanka is associated with the impaired survival of the parasite in visceral organs (Zhang et al., 2014). Among the genes that are mutated in CL-causing *L. donovani* strains of Sri Lanka (Lypaczewski et al., 2018), two are mutated in all the Indian strains, though the mutation profiles are different (Table S9). On the other hand, 24 genes are mutated specifically in the Indian para-KDL strains. Therefore, genomic variations exist in *L. donovani* strains between two neighbouring countries in the same continent and the findings strongly imply the presence of different underlying mechanisms for the development of PKDL like manifestation in India and CL in Sri Lanka, even though both are skin manifestations caused by the same species of the parasite.

Our whole genome sequencing data reveal distinguishing genomic changes among the para-KDL strains, which possibly contribute for survival of the persistent parasites in the skin with gradual removal from the visceral area. The elucidation of structural and functional alterations

of the encoded proteins due to the specific genomic modifications will reveal the possible molecular details of the parasite persistence leading to coexistence of PKDL and VL in the para-KDL patients of Indian subcontinent. Nevertheless, the analysis of genomic sequences of *L. donovani* strains in the present study reveals a strong correlation of development of para-KDL with chromosome aneuploidy and specific non-synonymous genetic variation in the coding sequences. Further studies are needed to understand the 'complex rules' revealing how such variation of para-KDL isolates at the genome level dictates tropism and PKDL development.

Data Accession

The sequence data related to the study are available under the Bio-Project Accession number PRJNA598250.

CRediT authorship contribution statement

Nibedeeta Rani Sarraf: Conceptualization, Resources, Methodology, Investigation, Writing – review & editing. **Saikat Mukhopadhyay:** Resources, Methodology, Investigation, Software, Data curation. **Anindyajit Banerjee:** Software, Resources, Validation, Data curation. **Anjan Kumar Das:** Resources. **Syamal Roy:** Supervision. **Saikat Chakrabarti:** Supervision, Software, Writing – original draft. **Madhumita Manna:** Conceptualization, Supervision, Writing – original draft. **Partha Saha:** Conceptualization, Supervision, Supervision.

Declaration of Competing Interest

The authors hereby declare that there is no conflict of interest.

Acknowledgement

This work was supported by the intramural grant to Saha Institute of Nuclear Physics from the Department of Atomic Energy, Government of India; and the extramural grant to MM and AKD from the Department of Biotechnology, Government of India (grant number BCIL/NER-BPMC/2013). NRS was a research fellow supported by the grant from the Department of Biotechnology, Government of India. SR acknowledges the Emeritus Fellowship of the Indian Council of Medical Research, Government of India.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.actatropica.2021.106086.

References

Alam, M.Z., Kuhls, K., Schweynoch, C., Sundar, S., Rijal, S., Shamsuzzaman, A.K., Raju, B.V., Salotra, P., Dujardin, J.C., Schonian, G., 2009. Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of *Leishmania donovani* strains in the Indian subcontinent. Infect Genet Evol 9, 24–31.

Bhattacharyya, A., Mukherjee, M., Duttagupta, S., 2002. Studies on stibanate unresponsive isolates of *Leishmania donovani*. J Biosci 27, 503–508.

Burza, S., Croft, S.L., Boelaert, M., 2018. Leishmaniasis. Lancet 392, 951–970.

Domagalska, M.A., Imamura, H., Sanders, M., Van den Broeck, F., Bhattarai, N.R., Vanaerschot, M., Maes, I., D'Haenens, E., Rai, K., Rijal, S., Berriman, M., Cotton, J. A., Dujardin, J.C., 2019. Genomes of *Leishmania* parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. PLoS Negl Trop Dis 13, e0007900.

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.

Downing, T., Stark, O., Vanaerschot, M., Imamura, H., Sanders, M., Decuypere, S., de Doncker, S., Maes, I., Rijal, S., Sundar, S., Dujardin, J.C., Berriman, M., Schonian, G., 2012. Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the *Leishmania donovani* species complex. Infect Genet Evol 12, 149–159.

Du, J., Yuan, Z., Ma, Z., Song, J., Xie, X., Chen, Y., 2014. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. Mol Biosyst 10, 2441–2447.

Dumetz, F., Imamura, H., Sanders, M., Seblova, V., Myskova, J., Pescher, P., Vanaerschot, M., Meehan, C.J., Cuypers, B., De Muylder, G., Spath, G.F., Bussotti, G., Vermeesch, J.R., Berriman, M., Cotton, J.A., Volf, P., Dujardin, J.C., Domagalska, M. A., 2017. Modulation of Aneuploidy in *Leishmania donovani* during Adaptation to Different In Vitro and. In Vivo Environments and Its Impact on Gene Expression. mBio 8.

Grunebast, J., Clos, J., 2020. *Leishmania*: Responding to environmental signals and challenges without regulated transcription. Comput Struct Biotechnol J 18, 4016–4023.

Hefnawy, A., Berg, M., Dujardin, J.C., De Muylder, G., 2017. Exploiting Knowledge on *Leishmania* Drug Resistance to Support the Quest for New Drugs. Trends Parasitol 33, 162–174.

Iantorno, S.A., Durrant, C., Khan, A., Sanders, M.J., Beverley, S.M., Warren, W.C., Berriman, M., Sacks, D.L., Cotton, J.A., Grigg, M.E., 2017. Gene Expression in *Leishmania* Is Regulated Predominantly by Gene Dosage. mBio 8.

Imamura, H., Downing, T., Van den Broeck, F., Sanders, M.J., Rijal, S., Sundar, S., Mannaert, A., Vanaerschot, M., Berg, M., De Muylder, G., Dumetz, F., Cuypers, B., Maes, I., Domagalska, M., Decuypere, S., Rai, K., Uranw, S., Bhattarai, N.R., Khanal, B., Prajapati, V.K., Sharma, S., Stark, O., Schonian, G., De Koning, H.P., Settimo, L., Vanhollebeke, B., Roy, S., Ostyn, B., Boelaert, M., Maes, L., Berriman, M., Dujardin, J.C., Cotton, J.A., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5.

Jha, T.K., 2006. Drug unresponsiveness & combination therapy for kala-azar. Indian J Med Res 123, 389–398.

Khanra, S., Datta, S., Mondal, D., Saha, P., Bandopadhyay, S.K., Roy, S., Manna, M., 2012. RFLPs of ITS, ITS1 and hsp70 amplicons and sequencing of ITS1 of recent clinical isolates of Kala-azar from India and Bangladesh confirms the association of *L. tropica* with the disease. Acta Trop 124, 229–234.

Khanra, S., Sarraf, N.R., Das, A.K., Roy, S., Manna, M., 2017. Miltefosine Resistant Field Isolate From Indian Kala-Azar Patient Shows Similar Phenotype in Experimental Infection. Sci Rep 7, 10330.

Khanra, S., Sarraf, N.R., Das, S., Das, A.K., Roy, S., Manna, M., 2016. Genetic markers for antimony resistant clinical isolates differentiation from Indian Kala-azar. Acta Trop 164, 177–184.

Kumar, R., Das, V.N., Topno, R.K., Pal, B., Imam, A., Agrawal, K., Singh, A., Murti, K., Lal, C.S., Verma, N., Das, P., Pandey, K., 2016. Para-kala-azar dermal Leishmaniasis cases in Indian subcontinent - A case series. Pathog Glob Health 110, 326–329.

Lypaczewski, P., Hoshizaki, J., Zhang, W.W., McCall, L.I., Torcivia-Rodriguez, J., Simonyan, V., Kaur, A., Dewar, K., Matlashewski, G., 2018. A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. Sci Rep 8, 16549.

Mann, S., Frasca, K., Scherrer, S., Henao-Martínez, A.F., Newman, S., Ramanan, P., Suarez, J.A., 2021. A Review of Leishmaniasis: Current Knowledge and Future Directions. Curr Trop Med Rep 1–12.

Mannaert, A., Downing, T., Imamura, H., Dujardin, J.C., 2012. Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. Trends Parasitol 28, 370–376.

Manzano, J.I., Perea, A., Leon-Guerrero, D., Campos-Salinas, J., Piacenza, L., Castanys, S., Gamarro, F., 2017. *Leishmania* LABCG1 and LABCG2 transporters are involved in virulence and oxidative stress: functional linkage with autophagy. Parasit Vectors 10, 267.

Metsalu, T., Vilo, J., 2015. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res 43, W566–W570.

Mukhopadhyay, D., Mukherjee, S., Roy, S., Dalton, J.E., Kundu, S., Sarkar, A., Das, N.K., Kaye, P.M., Chatterjee, M., 2015. M2 Polarization of Monocytes-Macrophages Is a Hallmark of Indian Post Kala-Azar Dermal Leishmaniasis. PLoS Negl Trop Dis 9, e0004145.

Pal, S., Mandal, A., Duttagupta, S., 2001. Studies on stibanate resistant *Leishmania donovani* isolates of Indian origin. Indian J Exp Biol 39, 249–254.

Ranasinghe, S., Zhang, W.W., Wickremasinghe, R., Abeygunasekera, P., Chandrasekharan, V., Athauda, S., Mendis, S., Hulangamuwa, S., Matlashewski, G., Pratlong, F., 2012. *Leishmania donovani* zymodeme MON-37 isolated from an autochthonous visceral leishmaniasis patient in Sri Lanka. Pathog Glob Health 106, 421–424.

Sacks, D.L., Kenney, R.T., Kreutzer, R.D., Jaffe, C.L., Gupta, A.K., Sharma, M.C., Sinha, S. P., Neva, F.A., Saran, R., 1995. Indian kala-azar caused by *Leishmania tropica*. Lancet 345, 959–961.

Schonian, G., Mauricio, I., Gramiccia, M., Canavate, C., Boelaert, M., Dujardin, J.C., 2008. Leishmaniases in the Mediterranean in the era of molecular epidemiology. Trends Parasitol 24, 135–142.

Singh, N.K., Agrawal, S.K., Jha, T.K., 1989. Visceral kala-azar associated with post-kala-azar dermal leishmaniasis. J Assoc Physicians India 37, 191.

Sundar, S., Singh, A., 2018. Chemotherapeutics of visceral leishmaniasis: present and future developments. Parasitology 145, 481–489.

Sundar, S., Singh, A., Rai, M., Prajapati, V.K., Singh, A.K., Ostyn, B., Boelaert, M., Dujardin, J.C., Chakravarty, J., 2012. Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use. Clin Infect Dis 55, 543–550.

Tovar, J., Wilkinson, S., Mottram, J.C., Fairlamb, A.H., 1998. Evidence that trypanothione reductase is an essential enzyme in *Leishmania* by targeted replacement of the tryA gene locus. Mol Microbiol 29, 653–660.

Ubeda, J.M., Legare, D., Raymond, F., Ouameur, A.A., Boisvert, S., Rigault, P., Corbeil, J., Tremblay, M.J., Olivier, M., Papadopoulou, B., Ouellette, M., 2008. Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. Genome Biol 9, R115.
Vickers, T.J., Beverley, S.M., 2011. Folate metabolic pathways in *Leishmania*. Essays Biochem 51, 63–80.
WHO, Leishmaniasis [https://www.who.int/news-room/fact-sheets/detail/leishmaniasis].
Zackay, A., Cotton, J.A., Sanders, M., Hailu, A., Nasereddin, A., Warburg, A., Jaffe, C.L., 2018. Genome wide comparison of Ethiopian *Leishmania donovani* strains reveals differences potentially related to parasite survival. PLoS Genet 14, e1007133.
Zhang, W.W., Ramasamy, G., McCall, L.I., Haydock, A., Ranasinghe, S., Abeygunasekara, P., Sirimanna, G., Wickremasinghe, R., Myler, P., Matlashewski, G., 2014. Genetic analysis of *Leishmania donovani* tropism using a naturally attenuated cutaneous strain. PLoS Pathog 10, e1004244.
Zijlstra, E.E., 2019. Biomarkers in Post-kala-azar Dermal Leishmaniasis. Front Cell Infect Microbiol 9, 228.
Zijlstra, E.E., Alves, F., Rijal, S., Arana, B., Alvar, J., 2017. Post-kala-azar dermal leishmaniasis in the Indian subcontinent: A threat to the South-East Asia Region Kala-azar Elimination Programme. PLoS Negl Trop Dis 11, e0005877.
Zijlstra, E.E., Khalil, E.A., Kager, P.A., El-Hassan, A.M., 2000. Post-kala-azar dermal leishmaniasis in the Sudan: clinical presentation and differential diagnosis. Br J Dermatol 143, 136–143.
Zijlstra, E.E., Musa, A.M., Khalil, E.A., el-Hassan, I.M., el-Hassan, A.M., 2003. Post-kala-azar dermal leishmaniasis. Lancet Infect Dis 3, 87–98.