

**Molecular Characterization and Genomics Study
of The Clinical Isolates of *Leishmania sp.* From
Indian Kala-azar and Para-KDL Patients**

**Thesis submitted to Jadavpur University for the degree of
Doctor of Philosophy (Science)
In Life Science and Biotechnology**

**By
Nibedeeta Rani Sarraf
M.Sc.**



**DEPARTMENT OF LIFE SCIENCE AND BIOTECHNOLOGY
JADAVPUR UNIVERSITY
KOLKATA-700032, INDIA**

2022



EDUCATION DIRECTORATE WEST BENGAL

BIKASH BHAVAN, (8th Floor)

SALT LAKE, KOLKATA-91

Telephone : (033) 2334-3980, 2334-3004

D.O. No.

Dated 15/09/2022

CERTIFICATE FROM THE SUPERVISOR(S)

This is to certify that the thesis entitled “**Molecular Characterization and Genomics Study of The Clinical Isolates of *Leishmania sp.* From Indian Kala-azar and Para-KDL Patients**” submitted by Smt. Nibedeeta Rani Sarraf who got her name registered on 23/11/2016 for the award of **Ph.D. (Science) degree of Jadavpur University**, is absolutely based upon her own work under the joint supervision of **Dr. Madhumita Manna, WBSES and Prof. Parimal Karmakar** and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

M. Manna

Dr. Madhumita Manna, WBSES
Additional Director of Public Instruction
Higher Education Directorate
Govt. of West Bengal
Bikash Bhavan, Kolkata 700091
Addl. D.P.I. (Admn.) W.B.
Office of the D.P.I. W.B.
Bikash Bhaban, Salt Lake

যাদবপুর বিশ্ববিদ্যালয়
কলকাতা-৭০০০৩২, ভারত



*JADAVPUR UNIVERSITY
KOLKATA-700 032, INDIA

DEPARTMENT OF LIFE SCIENCE AND BIOTECHNOLOGY

CERTIFICATE FROM THE SUPERVISOR(S)

This is to certify that the thesis entitled “**Molecular Characterization and Genomics Study of The Clinical Isolates of *Leishmania sp.* From Indian Kala-azar and Para-KDL Patients**” submitted by Smt. Nibedeeta Rani Sarraf who got her name registered on 23/11/2016 for the award of **Ph.D. (Science) degree of Jadavpur University**, is absolutely based upon her own work under the joint supervision of **Dr. Madhumita Manna, WBSES and Prof. Parimal Karmakar** and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.



Parimal Karmakar, Ph.D.
PROFESSOR
Department of Life Science & Biotechnology
JADAVPUR UNIVERSITY, Kol-32
Email: pkarmakar_28@yahoo.co.in
(M) 9433366323


Prof. Parimal Karmakar

Professor

Department of Life Science & Biotechnology
Jadavpur University, Kolkata-700 032

Established on and from 24th December, 1955 vide Notification No.10986-Edn/IU-42/55 dated 6th December, 1955 under Jadavpur University Act, 1955 (West Bengal Act XXIII of 1955) followed by Jadavpur University Act, 1981 (West Bengal Act XXIV of 1981)

**This thesis is dedicated to my beloved
Parents & Grandparents for being my unparalleled
source of strength and motivation...**

ACKNOWLEDGEMENT

After a long, arduous, and difficult journey through science over the last few years, the time has come to express my gratitude to those without whom I would not have been able to finish my thesis.

First, I would like to acknowledge the Department of Biotechnology (DBT), India, for providing substantial financial support during my PhD tenure.

I express my deep gratitude to my supervisor, Dr Madhumita Manna Ma'am, for her invaluable guidance, encouragement and support throughout this doctoral research. I am obliged to her for her faith in me and for providing the right direction whenever I needed it the most. I have learned a lot from her. Without her help, I could not have finished my thesis successfully. For my doctoral studies, I could not have asked for a better advisor and mentor. Your help developing me as a researcher and a better person is invaluable. I am fortunate to get such an excellent teacher, philosopher and guide in Ma'am. I am thankful to you, Ma'am, for all the suggestions, encouragement and constant inspiration you gave me, and the confidence you had in me that helped a lot during my entire research period. I yet feel that just a thank is not enough, I will ever remain indebted to my guru Dr Madhumita Manna.

I want to express my gratitude to another supervisor, Prof. Parimal Karmakar Sir, for his constructive comments, positive suggestions, and invaluable support during this period. He did all he could to assist me during my PhD tenure. I am very grateful to him.

This thesis will be incomplete without mentioning the name of the respected Sir, Prof. Syamal Roy (Emeritus Scientist, CSIR, IICB Kolkata). I have never seen such a liberal, modest personality with enormous knowledge and perception. I got an opportunity to complete some of the experimental parts of my thesis in his laboratory.

I gained a lot of scientific potentials while working in his laboratory. I will ever remain obliged to Prof. Syamal Roy for his immense support.

I wholeheartedly thank those patients without whom this work has not been possible.

I am very thankful to Dr Anjan Kumar Das, Calcutta National Medical College, for his cooperation during the sample collection.

I extend my deep gratitude to Prof. Partha Saha, Saha Institute of Nuclear Physics, India, for his valuable suggestion regarding my research and for extending his laboratory facilities to me. Now, in Prof. Partha Saha's laboratory, I cordially express my gratitude to Utpal da, Susantho da, Saikat da, Suchi, Indranil and Aditya for their kind cooperation and advice during my research work. I am especially very thankful to Saikat da, who helped me learn the fine details analysis of sequencing data.

I extend my deep gratitude to Dr Saikat Chakrabarti, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata and Dr Anindyajit Banerjee, Tata Translational Cancer Research Centre, Tata Medical Center, Kolkata, for their cooperation and help in the analysis of sequencing data.

The support and help of Dr Subhasis Datta, the Principal, Barasat Govt. College and Dr Saurabh Chakraborti, the Principal, Bidhannagar College, Kolkata is duly acknowledged. I express my gratitude to the Heads of the Dept. and other teachers and non-teaching staff of Post Graduate Depts. of Zoology, Barasat Govt. College & Bidhannagar College, Kolkata.

I take this opportunity to thank all my lab mates, both seniors and juniors. I am grateful to Sanchita di, Supriya di, Sangita di, Kasturi, Sulagna and Payel for their regular help whenever I needed it during this period. My special gratitude to Supriya di, who taught me the meaning of research and some of the laboratory experiments like a teacher.

I express my immense gratitude to my guru Prof. Subir Chandra Das Gupta, who has always been supportive during my college days into my research career.

I am incredibly thankful to all my friends, Kalyani, Arpita, Arup, Areen, Ikbal, Kaushik, Munna, Siraj, Sonali and Nasreen, for their love, support and motivation. A special thanks to Arup, Arpita and Kalyani for being my constant support during the most challenging phases.

Last but not least, I would provide my special gratitude to my family, who have always been my greatest strength my Dada, Dadi, Nani, Maa, Papaji and my sisters Tresa, Jaya, Babu for their unconditional love, support and encouragement. Words are not enough to express the love and sacrifice of the parents. I always knew that you believed in me and wanted the best for me. My younger sisters, Tresa, Jaya and Babu are always there to encourage me. Although my Grandparents are not with us but without their blessing, it would be quite impossible for me to complete my Ph.D. work.

My acknowledgement will never be complete without the special mention of my husband, Ikbal, who stood like a pillar of strength to me. Without his love, inspiration, encouragement and understanding, it would have been impossible for me to finish the work. I am grateful to my in-laws for extending their support.

Above all, I am thankful to the Almighty Creator for showering his blessing over me to complete this mission fruitfully.

Nibedeeta Rani Sarraf

CONTENTS

List of Abbreviations	1
Abstract	5
Introduction	7
Review of Literature	27
Aim and Objective of The Present Study	66
Materials And Methods	72
CHAPTER 1	
Restriction Fragment Length Polymorphism Analysis [RFLP] of the Indian Clinical Isolates of Kala-azar and Para Kala-Azar Dermal Leishmaniasis	96
CHAPTER 2	
Multilocus Sequence Typing (MLST) of <i>Leishmania sp.</i> Isolated from Indian Kala-azar and Para Kala-Azar Dermal Leishmaniasis Patients	111
CHAPTER 3	
Whole Genome Sequence Analysis of the <i>Leishmania donovani</i> strains from Indian VL and para-KDL Clinical Isolates	145
CHAPTER 4	
<i>In Silico</i> Structural Analysis of the Mutated Transporter and Surface Proteins of Indian para-KDL Isolates	182
Discussion	222
References	243
Publications And Proceedings	265

LIST OF ABBREVIATIONS

Ab	Antibody
ABC	ATP Binding Cassette
AIDS	Acquired Immuno Deficiency Syndrome
APC	Antigen Presenting Cell
AmpB	Amphotericin B
BLAST	Basic Local Alignment Search Tool
°C	Degree Celsius
Ci	Curie
CL	Cutaneous Leishmaniasis
CNV	Copy Number Variation
CMI	Cell-Mediated Immunity
ConA	Concanavalin A
Cpm	Count per Minute
DCL	Diffused Cutaneous Leishmaniasis
DNA	Deoxyribonucleic Acid
DP	Discriminatory Power
EDTA	Ethylene Diamine Triacetate
ELISA	Enzyme-Linked Immunosorbent Assay
ERK	Extracellular Regulated Kinase
FCS	Fetal Calf Serum
h	Hour
HAP	Haplotype
H2DCFDA	2',7'-Dichlorohydrofluorescein Diacetate

HIV	Human Immunodeficiency Virus
HRP	Horse Reddish Peroxidase
IFN- γ	Interferon –Gamma
IgG	Immunoglobulin G
IL	Interleukin
Inos	Inducible Nitric Oxide Synthetase
ISPs	Ion PI™ Ion Sphere™ Particles
kDNA	Kinetoplast DNA
Kg	Kilogram
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	<i>Leishmania donovani</i>
LT	<i>Leishmania tropica</i>
LPS	Lipophosphosaccharide
LABCG	ATP-binding cassette G subfamily
M	Molar
mAb	Monoclonal Ab
MAPK	Mitogen-Activated Protein Kinase
μ Ci	Microcurie
MCL	Mucocutaneous Leishmaniasis
μ gm	Microgram
mg	Milligram
MHC	Major Histocompatibility Complex
Min	Minute
μ l	Microlitre

ml	Millilitre
μM	Micromolar
mm	Millimetre
M φ	Macrophage
Mw	Molecular Marker
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NK	Natural Killer
Nm	Nanometre
NO	Nitric Oxide
PBS	Phosphate Buffer Saline
para-KDL	Para Kala-Azar Dermal Leishmaniasis
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pI	Isoelectric Point
Pi	Nucleotide Diversity
PS	Phosphatidylserine
P < 0.05	Significance at 5% level
PKDL	Post Kala-Azar Dermal Leishmaniasis
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAG	Sodium Antimony Gluconate
SNP	Single Nucleotide Polymorphism
SSG	Sodium Stibo Gluconate

ST	Sequence Type
m	Molecular Mass
MIL	Miltefosine
MLST	Multilocus Sequence Typing
MLEE	Multilocus Enzyme Electrophoresis
MLMT	Multilocus Microsatellite Typing
NGS	Next Generation Sequencing
TAE	Tris Acetic acid -EDTA
TGF- β	Tumour Growth Factor beta
Th1	T Helper cell 1
TNF	Tumour necrosis factor
VL	Visceral Leishmaniasis
VAF	Variant Allele Frequency
WGS	Whole Genome Sequencing
WHO	World Health Organization
μ g	Microgram
μ L	Microliter
μ M	Micromole

Abstract

[Index No. 221/16/Life Sc./25]

Title of the Thesis: Molecular Characterization and Genomics Study of the Clinical Isolates of *Leishmania sp.* From Indian Kala-azar and Para-KDL Patients.

Submitted by: Nibedeeta Rani Sarraf


Visceral leishmaniasis (VL) or Kala-azar (KA) is a neglected tropical parasitic disease caused by *Leishmania donovani*, though recent reports confirmed the association of *L. tropica* with the disease. About 10%-20% cases of apparently cured VL patients may develop Post Kala-azar Dermal Leishmaniasis (PKDL). Currently, the sustained presence of active VL in PKDL patients is observed where co-association of PKDL manifestation in active VL is recorded. These cases are referred to as Para Kala-azar Dermal Leishmaniasis (para-KDL). In the Indian subcontinent, 70% of cases reported being unresponsive to the clinically available first-line drug Sodium stibogluconate (SSG). Thus, the emergence of drug-resistant parasites towards the applicable drugs is an alarming signal for future VL elimination. Prior to opt for any drug regimen, the taxonomical identification and detailed characterization of the parasite is a prerequisite for epidemic surveillance and eradication of any disease. The present study has been carried out to examine the structural and functional genomic attributes of the recently collected clinical isolates (n=15) from Indian KA and para-KDL patients. Different PCR based methods and computational tools are used to analyse the parasite genome with restriction analysis of the amplified ITS and hsp70 locus (RFLP), Multilocus Sequence Typing (MLST) of some enzyme coding genes, genome-wide comparison by Whole Genome Sequencing (WGS) and *in silico* structural analysis of some novel mutated genes. The introduction of ten new RFLP markers suggests that only six markers have restriction sites in the ITS amplicon. Among them, Fsp I and Mse I markers unambiguously differentiated *L. donovani* from *L. tropica*. While RFLP patterns of the hsp70 and ITS1 region have confirmed that only one isolate (T5) was *L. tropica*, which is collected from a confirmed KA patient corroborating our earlier observation. To understand the current phylodynamics of the *Leishmania* parasite, a potential molecular approach, namely MLST, has been carried out with fifteen housekeeping gene loci in the genus *Leishmania* at the interspecies and intraspecies level. The results indicate that the genus *Leishmania* displays high nucleotide diversity among the sequenced housekeeping genes coding regions. Notably, only two housekeeping gene loci *enol* and *alat*, were seen amplified in the Indian clinical isolate of *L. tropica*, T5. For the locus *alat*, a unique sequence type ST3 has been found in the T5 isolate. This is reported here for the first time. Neighbour Joining (NJ) phylogeny analysis of the concatenated sequences of *aco*, *alat*, *enol*, *pgm*, *spdsyn* and *hgprt* genes for 11 Indian *L. donovani* isolates and 24 retrieved sequences of other *Leishmania* species from GenBank could differentiate *Leishmania* species complexes and subgenus level with high bootstrap support. In contrast, rooted phylogenetic analysis of individual locus isocitrate dehydrogenase (*icd*) sequence for 38 isolates could demonstrate the sub-continental origin of the *Leishmania donovani* complex.

Interestingly, the intraspecies phylogenetic analysis based on the concatenated sequence of seven housekeeping genes loci (*aco*, *alat*, *asat*, *enol*, *gpi*, *nhl*, *pgm*) form four distinct clusters in the Indian clinical isolates of *L. donovani* studied here. It is the first instance where intraspecies allelic profile analysis has been taken into account to observe any specific sequence type among the Indian *L. donovani* population. One gene, *spermidine synthase* (*spdsyn*), was conserved among all the Indian clinical isolates of *L. donovani* collected for the study.

The other part of the study dealt with the identification and comparison of genome-wide variation by whole genome sequencing of the SSG-sensitive VL and para-KDL, SSG resistant and MIL resistant Indian clinical isolates of *Leishmania sp.* Chromosome copy numbers were estimated using normalized whole chromosome median read depths along with the analyses of non-synonymous SNVs and InDels to identify the possible association of aneuploidy and the contribution of genome-wide mutation profile of the protein coding genes in the development of PKDL and the drug resistance in the Indian VL patients. In the para-KDL isolates, a unique set of 13 genes and in MIL resistant VL isolate (T9), 56 gene mutations homologous to genes with previously known cellular functions have been reported. In the next part, we highlighted the *in silico* structural analyses of the five novel mutated genes (out of 13) in the para-KDL isolates to examine the role of observed mutations in the protein secondary structure, whose functions have previously been observed as a surface protein and transporter. The structural analyses of the observed single nucleotide polymorphisms that lead to a frameshift mutation in the transmembrane domain of encoded surface proteins, including neutral sphingomyelinase activation associated factor-like protein, beta galactofuranosyl transferase and transporters like calcium-translocating P-type ATPase and amino acid transporter aATP11(putative) in the SSG sensitive para-KDL isolates confer the major secondary structure changes in their mutated position. Only one protein, p-glycoprotein e (partial) acting as a surface and a transporter protein showed transversion mutation. The observed mutation present in the ABC transporter transmembrane type-1 domain leads to the replacement of the coil at positions 1238 and 1239 with a helix in the ABC transporter 3rd domain. The study revealed a strong correlation between the development of drug resistance and the clinical manifestation of para-KDL, with special emphasis on aneuploidy and existing mutations.

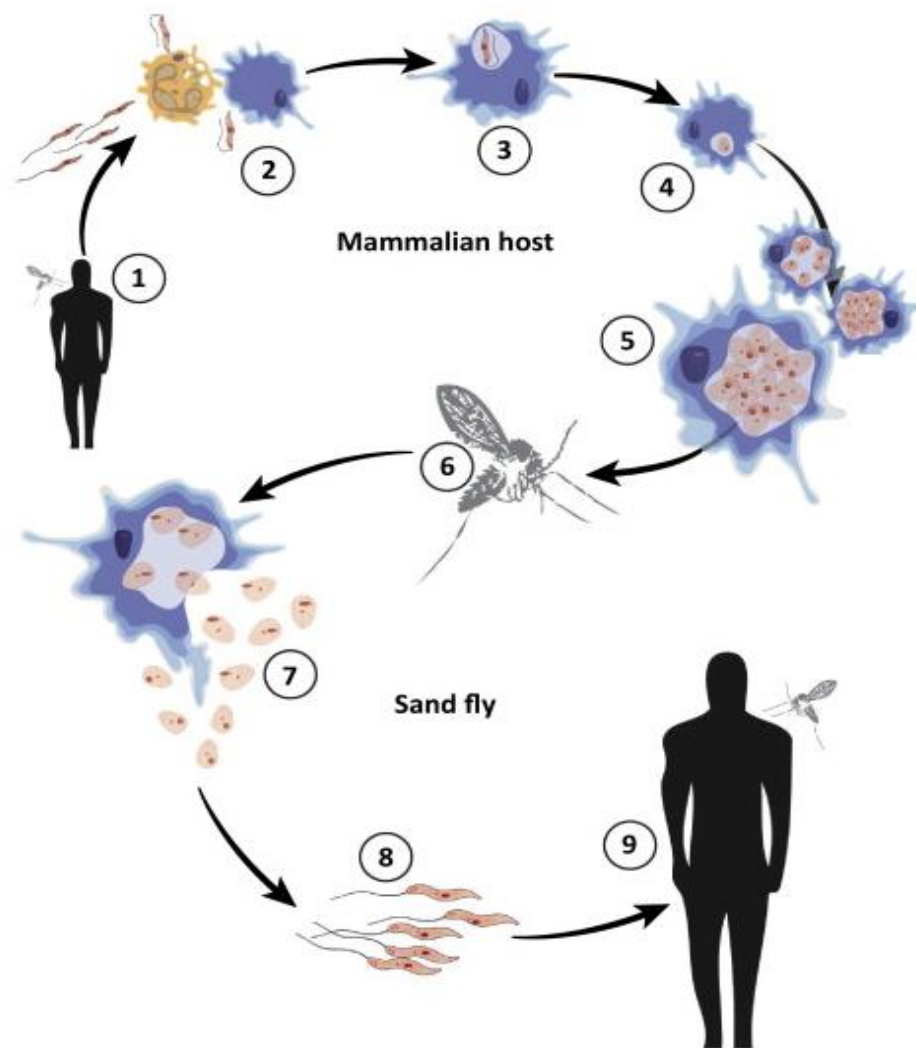
Overall, this clinical isolates-based study is very useful to enhance the epidemiological surveillance of the country and provides detailed information for para-KDL for the first time.

M. Karmakar
15/09/22
Addl. D.P.I. (Admn.) W.B.
Office of the D.P.I. W.B.
Bikash Bhaban, Salt Lake


9/9/22
Parimal Karmakar, Ph.D.
PROFESSOR
Department of Life Science & Biotechnology
JADAVPUR UNIVERSITY, Kolkata-32
Email: pkarmakar_28@yahoo.co.in
(M) 9433366333


Candidate

Supervisors



INTRODUCTION

LEISHMANIASIS: THE VECTOR-BORNE PARASITIC DISEASE

Leishmaniasis is a tropical and subtropical vector-borne disease caused by intracellular protozoan parasite species of the genus *Leishmania* and transmitted by the bite of infective female sandflies of the genera *Phlebotomus* in the Old World and *Lutzomyia* in the New World. The clinical manifestations of the disease are associated with three different forms, i.e., cutaneous, mucosal and visceral, collectively known as Leishmaniasis. In humans, infection is caused by 21 different species of *Leishmania*, all of them morphologically indistinguishable from each other (Figure 1), Several forms of the disease are contagious only to animals (zoonotic), but some can be spread between humans (anthroponotic).

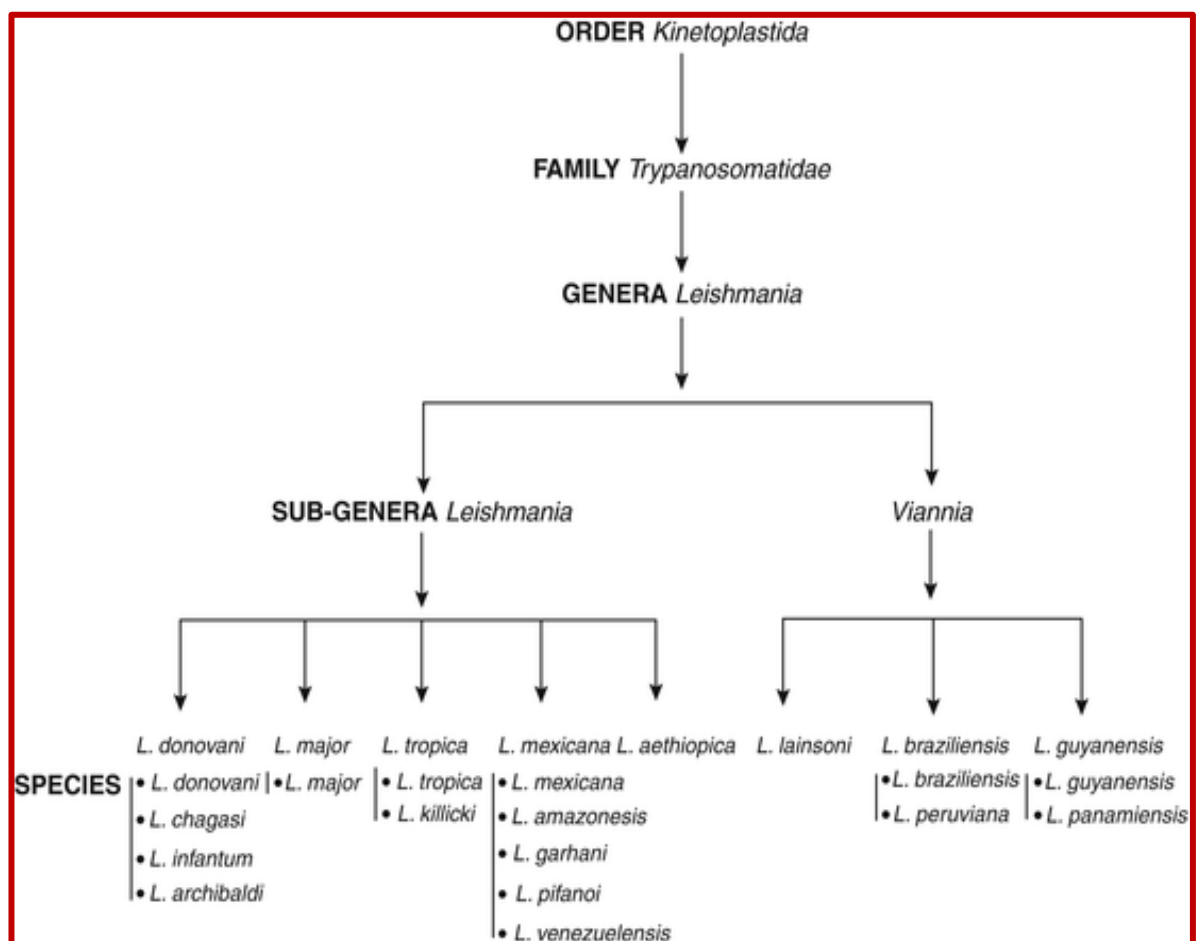


Figure 1. Taxonomy of *Leishmania*

HISTORICAL BACKGROUND:

Leishmaniasis has a long and illustrious history dating back to the first century AD. Pre-Incan pottery from Ecuador and Peru depicted skin lesions and facial abnormalities indicative of Cutaneous and Mucocutaneous Leishmaniasis as early as this period. The occurrence of skin lesions on agricultural labourers returning from the Andes was recorded by Spanish conquistadors between the 15th and 16th centuries. These ulcers were dubbed "white leprosy," "Andean sickness," or "valley sickness" because they resembled leprosy lesions. In the mid-eighteenth century, records from Africa and India referred to the disease today known as Visceral Leishmaniasis as 'Kala-azar' or 'black fever'. After examining a Turkish patient in 1756, Alexander Russell significantly contributed to the discovery of Leishmaniasis. It, according to Russell, leaves an impression that remains throughout life and has a livid colour for many months. Russell called this disease 'Aleppo boil'.



Figure 2. Dr William Leishman and Dr Charles Donovan

The disease became known as Leishmaniasis after William Leishman, a doctor serving with the British Army in India who developed one of the earliest strains of *Leishmania* in 1901. In Dum Dum, a town near erstwhile Calcutta, now Kolkata, Leishman discovered oval bodies in the spleen of a British soldier who was experiencing spells of fever, anaemia, muscular atrophy and swelling of the spleen. Leishman described this illness as ‘Dum-Dum fever’ and published his findings in 1903. Charles Donovan also recognised these symptoms in other kala-azar (Hindi: kala=black, azar=sickness) patients and published his discovery a few weeks after Leishman. Later officially, the stalwart parasitologist, Sir Ronald Ross, gave the name *Leishmania donovani* by honouring both Leishman and Donovan for the discovery of the parasite. After examining the parasite using Leishman's stain, these amastigotes were known as Leishman-Donovan (LD) bodies.

CLASSIFICATION:

The genus *Leishmania* is divided into two subgenera - *Leishmania* and *Viannia*, depending on the distribution of parasites in the gut of the sandfly vector. In the subgenera *Leishmania*, the promastigotes develop in the midgut and foregut of the insect (Suprasyplaria section). In contrast, in the subgenera *Viannia*, parasites are restricted to the sandfly's hindgut (Peripsyllaria Section). The parasite exists in two forms- one is a flagellate extracellular promastigote in the sandfly vector, and another is the aflagellate intracellular amastigote within mononuclear phagocytes of the vertebrate host. Species of the subgenus *Leishmania* belong to the *Leishmania donovani* complex, *L. tropica* complex, and *L. mexicana* complex. On the other hand, the *L. braziliensis* complex belongs to subgenera *Viannia*.

SYSTEMATIC POSITION (Levine et al. 1980):

Kingdom: Protista (Hakel 1886)

Subkingdom: Protozoa (Goldfuss 1817)

Phylum: Sarcomastigophora (Honigberg and Balamua 1963)

Subphylum: Mastigophora (Deising 1866)

Class: Zoomastigophora (Calkins 1909)

Order: Kinetoplastida (Honigberg 1963)

Suborder: Trypanosomatina (Kent 1880)

Family: Trypanosomatidae (Dolfein 1901)

Genus: *Leishmania* (Ross 1903)

MORPHOLOGY:

The parasite exists in two forms: Amastigotes and Promastigotes. The morphology of *Leishmania* protozoa varies by species and throughout the life cycle (Peters et al., 1987; Despommier et al., 1994). *Leishmania* is a unicellular, dimorphic, eukaryotic parasite infecting a wide range of vertebrates throughout the tropical and subtropical world with a broad disease spectrum.

Amastigote stage: when it is in the human host (intracellular).

1. Amastigote is a non-flagellar form of the parasite.
2. The parasite at the amastigote stage is found in man and other mammalian hosts.

3. They are found inside monocytes, polymorphonuclear leukocytes or endothelial cells.
4. Amastigotes are small, round to oval bodies, measuring 2-3 μm in length.
5. They are also known as LD (Leishman Donovan) bodies.
6. The nucleus is less than 1 μm in diameter, oval or round and is usually situated in the middle of the cell.
7. A rod-shaped kinetoplast lies at the right angle to the nucleus. It comprises DNA containing the body and a mitochondrial structure.
8. Axoneme arises from the kinetoplast and extends to the margin of the body. It represents the foot of the flagellum.
9. Vacuole, a clear unstained space, lies alongside the axoneme.
10. In a Giemsa-stained preparation, the cytoplasm surrounded by a limiting membrane appears pale blue. The nucleus is relatively larger and stained red. The kinetoplast stained deep red. Amastigote divides by binary fission at 37°C.

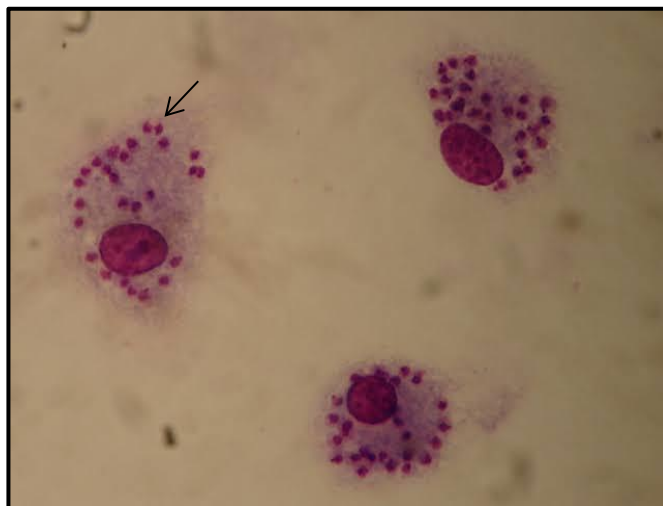


Figure 3. Amastigotes in infected macrophage

Promastigote stage: when it exists in the vector (extracellular).

1. Promastigotes are discovered in the sand fly (vector) digestive tract and culture media.
2. Promastigotes that have fully matured are long, slender, and spindle-shaped. They are 15 to 25 μ m long and 1.5 to 3.5 μ m wide.
3. There is a solitary nucleus in the centre.
4. The kinetoplast is transversely located near the anterior end of the kinetoplast.
5. A single, fragile flagellum measuring 5-10 μ m in length and protruding from the front. It can be the same length as the body or even longer. There is no undulating membrane because the flagellum does not bend around the parasite's body. With Leishman stain, the cytoplasm appears blue, the nucleus pink or violet and the kinetoplast bright red.
6. At 27°C, promastigote multiplies by binary fission.



Figure 4. *Leishmania* promastigotes

ULTRASTRUCTURE OF THE PARASITE:

➤ Plasma Membrane

The plasma membrane of all stages of *Leishmania* is a typical single trilaminar membrane 2-4 nm wide, underneath which are microtubules and desmosomes that help to maintain the flagellate integrity. The membranes of amastigotes and promastigotes have a polysaccharide composition. The polysaccharides remaining on the membranes of *L. donovani* promastigotes contain the sugar N-acetyl glucosamine and 14 and 16 glucan-like D-glucose-like units.

➤ Flagellum and the basal body apparatus

In *Leishmania* promastigotes (Figure 5), the flagellum's typical axonemal structure can be seen, which is known as the flagellum. The flagellum structure comprises the nine pairs of peripheral axonemal doublets ringed and encircled by a centre pair. The centre doublet comes from the first basal plate, which is located distant from the basal body. The paraxial rod of the flagellum is a para crystalline or lattice-like structure that runs parallel to the axoneme. While in the case of *Leishmania* amastigote, no such structure has been seen. Inside the sandfly, the flagella of some promastigotes are modified during the attachment and circulate from hindgut to foregut. The structure of the flagellate basal body is typical of other eukaryotic flagellates.

➤ Kinetoplast DNA (kDNA):

The kinetoplast structurally looks like a tiny rod-shaped body posterior to the basal body. The kinetoplast is made up of kDNA, which is densely packed DNA. The kDNA molecule comprises hundreds of 1-2.5 kbp circular DNA molecules called minicircles and 20-40 kbp circular DNA molecules called maxicircles. The maxicircle is similar to mitochondrial DNA in terms of function. The fibrous kDNA disc is found within the mitochondrial membranes that run parallel to the cell's mitochondrion. Plate-like cristae protrude from the

membrane's inner wall, encircling the kinetoplast disc. *Leishmania*'s DNA disc is slightly concave, measuring 0.083 mm deep and 0.5-1 mm in diameter. Fibrillar arrangement varies by species, but the fibrils are 2.5 nm broad and orientated isotopically.

➤ **Structure of nucleus:**

Depending on the species, the nucleus of *Leishmania* includes 34 to 36 chromosomes. The cell's mitochondrial DNA (mtDNA) is represented by kDNA, which accounts for 10-20% of the total DNA (Simpson, 1987). It's a circular network of DNA that's separated into two layers: homogeneous maxicircles (25-50 molecules, 20 kb) and heterogeneous minicircles (0.8 kb), with 104 copies. Maxicircle is mtDNA's functional partner and plays a role in editing Uracil moieties to mRNA nucleotides (Leon et al., 1996). The minicircles encode guide RNAs (gRNAs) for cytoplasmic oxidative subunit III mRNA editing (Sturm and Simpson, 1990).

➤ **Cytoplasm:**

Ultrastructural studies of *Leishmania* show the presence of cytoplasmic membrane-associated organelles such as ER and Golgi apparatus, like the characteristic of other eukaryotes. The ER of both rough and smooth types is found throughout the cytoplasm. Golgi products play an important role during intracellular digestion and are packaged as lysosomes. A variety of membrane-bound vacuoles is observed throughout the cytoplasm of *Leishmania*. Inclusion vacuoles are membrane-bound and contain particulate material scattered peripherally. Such vacuoles are 0.2-0.4mm in diameter. Inclusions of lipid bodies observed which are not bound by a membrane are up to 0.5mm in diameter. They are present in both amastigotes and promastigotes (Molyneux et al. 1975). Other structures that occur within the cytoplasm are microbodies or peroxisomes. These structures contain oxidative enzymes that play a role in the oxidative metabolism of the cell.

They are membrane-bound structures with a single unit membrane and have a crystalloid pore of the dense plate (Vickennan 1976)

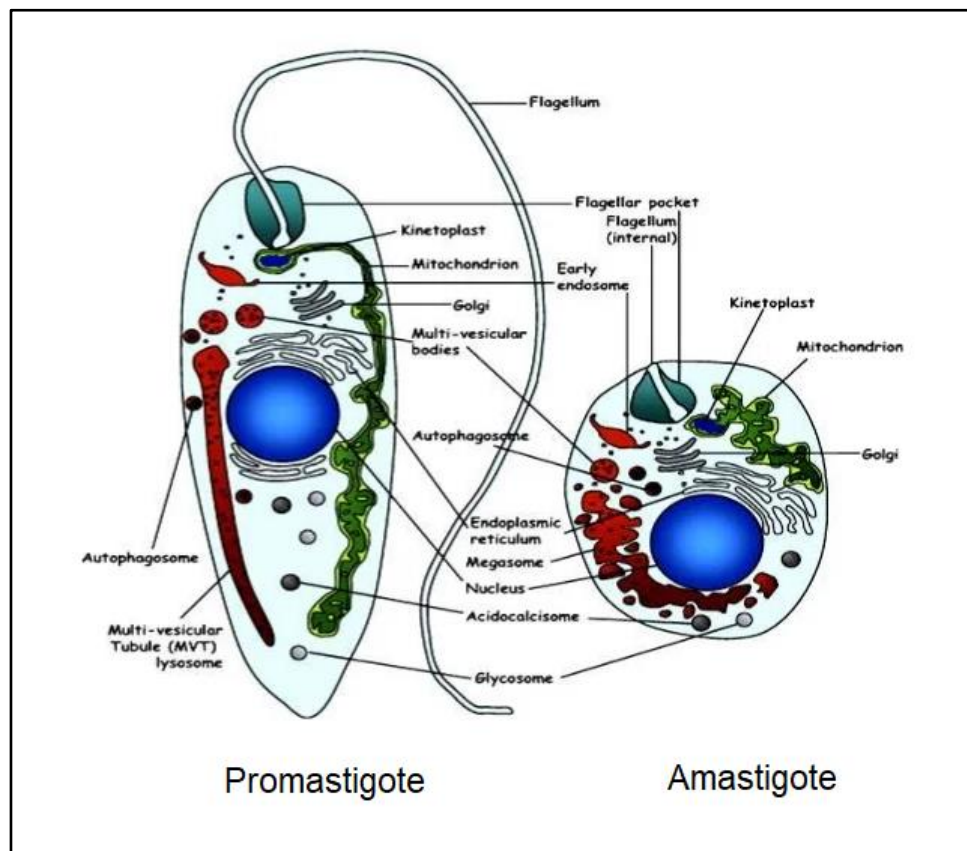


Figure 5. Ultrastructure of *Leishmania*

VECTOR OF LEISHMANIASIS: Leishmaniasis is a vector-borne disease transmitted



Phlebotomus

by sandflies. Sandflies are divided into two genera: the Old World 'Phlebotomus' and the New World 'Lutzomyia' (Killick-kendrick, 1990). In India, the primary vector is *Phlebotomus argentipes*; some authors also reported *P. papatasi* as a secondary vector (Rassi et al., 2011).

LIFE CYCLE:

Leishmania species have a sexually dimorphic life cycle that alternates between two hosts. Primary reservoirs host humans and secondary host sandflies with two different life forms, promastigotes and amastigotes. The metacyclic pre-flagellar promastigote form is transmitted from the sandfly to the vertebrate host when the host eats a blood meal. Invertebrate hosts, the promastigote is phagocytosed by macrophages, which fuse with lysosomes.

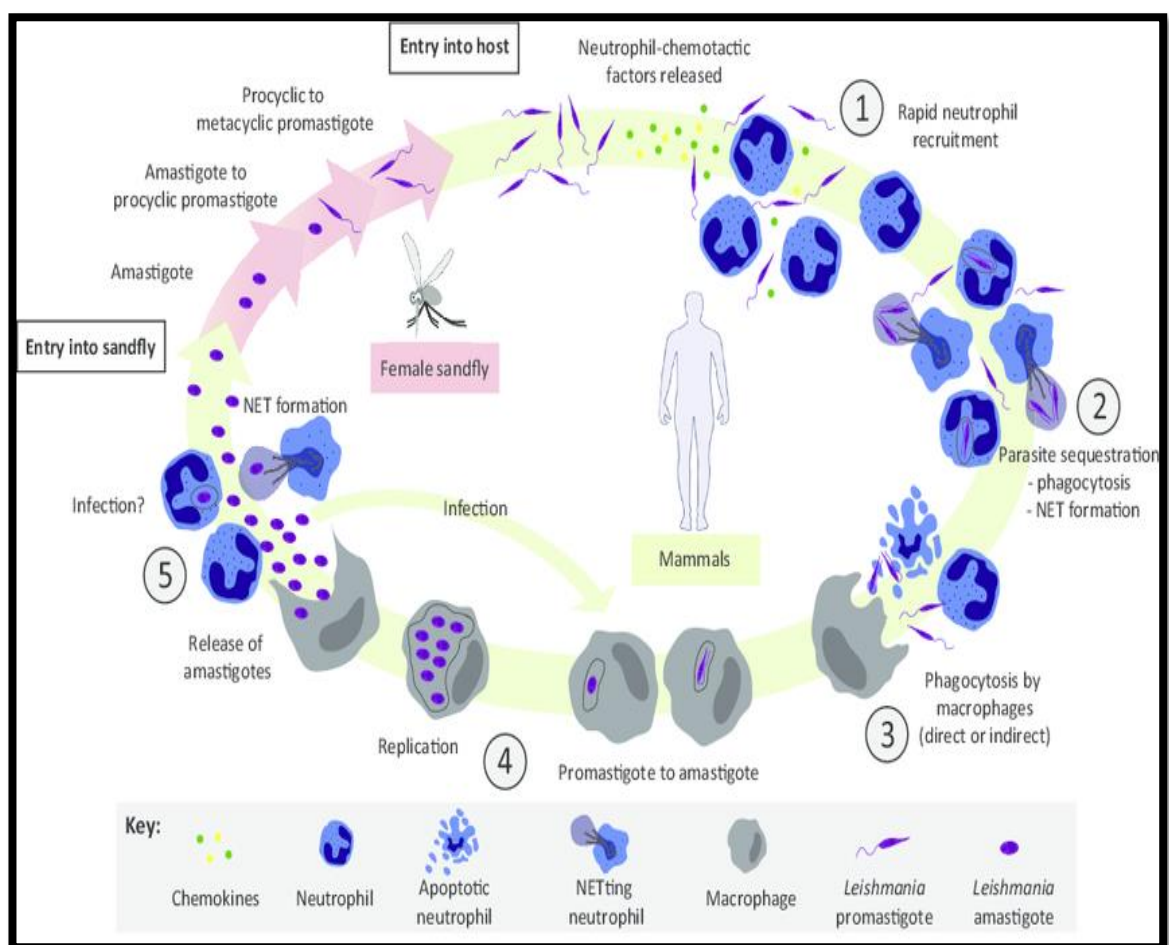


Figure 6. Schematic representation of the life cycle of the *Leishmania* parasite.

In addition, parasites prevent macrophages from producing reactive oxygen intermediates and effectively transform into amastigotes. These amastigotes proliferate by binary fission and replenish macrophages. Eventually, the amastigote is released from infected

macrophages and resumes this replication cycle with new macrophages. Amastigotes to promastigotes conversion occur in the gut of the sandflies after ingestion of infected macrophages during host bite, after replication by binary fission parasite adhere to the sandfly gut epithelium to avoid being excreted. The promastigotes then stop dividing and undergo terminal differentiation into metacyclic promastigotes, infective for the vertebrate hosts (Chang et al., 1985). Figure 6, gives a schematic life cycle representation of the *Leishmania* parasite.

RESERVOIR HOSTS:

In most cases, a primary reservoir for a particular *Leishmania* organism exists in a given region, but secondary and accidental hosts may also be implicated. Domestic dogs were the principal reservoir for VL infection caused by *L. infantum* in the Old World. Secondary reservoir for *L. infantum*, carried by wildlife hosts such as foxes, jackals, wolves, and raccoons (Magill, 1995). Many rodents serve as reservoirs for *L. major* across Africa, Asia, Arabia and hyraxes serve as reservoirs for *L. aethiopica* in East Africa. Domestic dogs, out of all the possible animal hosts, play the most crucial role in disease transmission to humans due to the close relationship between humans and dogs as pets (WHO 1991).

Disease types with their causative agents, reservoirs and vectors (Dey et al., 2007).

Clinical Symptom	Parasite	Reservoir	Vector
Visceral leishmaniasis (Kala-azar)	<i>Leishmania donovani</i> <i>Leishmania infantum</i> <i>Leishmania chagasi</i>	Human Dog, Fox Jackal	<i>Phlebotomus argentipes</i> <i>Phlebotomus ariasi</i> <i>Lutzomyia sp.</i>
Post kala-azar dermal leishmaniasis	<i>Leishmania donovani</i>	-	-
Cutaneous leishmaniasis	<i>Leishmania major</i> <i>Leishmania tropica</i> <i>Leishmania aethiopica</i>	Rodent Human Hyrax	<i>Phlebotomus duboscqi</i> <i>Phlebotomus sergenti</i> <i>Phlebotomus longipes</i>

CLINICAL MANIFESTATIONS:

Leishmaniasis consists of six clinical syndromes. Among them, one is an intermediate form between VL and PKDL: Visceral Leishmaniasis (VL), Cutaneous Leishmaniasis (CL), Mucocutaneous Leishmaniasis (MCL), Leishmaniasis Recidiva, Post Kala-azar Dermal Leishmaniasis (PKDL), and Para-Kala-Azar Dermal Leishmaniasis i.e., para-KDL (Intermediate form).

Kala-azar (KA) or Visceral Leishmaniasis (VL):

VL is caused by the *Leishmania donovani* complex composed of *L. donovani donovani* in India, Africa (Naik et al., 1979; Bettini et al., 1981); in the Mediterranean region, *L. donovani infantum* (Rocha, 1982) and South American *L. donovani chagasi* (Napier and Dasgupta, 1930). It was later reported that *Leishmania tropica* could also cause KA (Sacks, 1995; Khanra et al., 2012). In India, 90% of KA cases are reported from the Bihar state (Khanra et al., 2012). VL is the most severe form of the disease and becomes fatal if left untreated. In its classical description, VL is characterized by fever, anaemia, hepatosplenomegaly and leukopenia, thrombocytopenia, and hypergammaglobulinemia (Ghosh et al., 1980). The parasite invades the host cell, which is present and proliferates to attack the internal organs (spleen, liver, lymph nodes, bone marrow).



Typical Characteristics of Visceral Leishmaniasis

Cutaneous Leishmaniasis (CL):

Generally, two forms are visible: Localized Cutaneous Leishmaniasis (LCL) and the second is Diffused Cutaneous Leishmaniasis (DCL). The disease manifests as a progression of relatively benign self-healing lesions confined to the skin in the LCL case. At the same time, DCL disease is characterized by 'anergy' lesions. The DCL lesions are painless and seldom distribute over the entire body. Lesions can affect any part of the body surface, but most often, there are exposed areas with access to the sandfly. Old World Cutaneous Leishmaniasis causative agents include *L. major* and *L. tropica*, and *L. aethiopica*, and the new world cutaneous leishmaniasis, Caused by the *L. braziliensis* complex. Approximately 11 million new cases of CL occur each year, 90% of which occur in Afghanistan, Brazil, Iran, Peru, Saudi Arabia, and Syria (Kumar et al., 2007). The chronic condition is an important feature of this disease. During *L. mexicana* infection, several lesions develop in the pinna and are associated with prolonged tissue destruction, a phenotype commonly known as 'Chiclero's ulcer'. It should also be noted that Kala-azar and CL are not found in the same region. For example, in India, Kara-azar is found primarily in the humid eastern part of the country. At the same time, oriental wounds, i.e., cutaneous leishmaniasis by *L. tropica*, are confined to the western part of the country (Kumar et al., 2007).



Typical characteristics of a) LCL & b) DCL

Mucocutaneous Leishmaniasis (MCL): This type of leishmaniasis is commonly referred to as 'Espundia,' a South American illness. *Leishmania braziliensis* is the most common cause of the disease. MCL is found in Bolivia, the Brazilian state of Bahia, and Peru,



Typical characteristics of MCL

accounting for 90% of all cases. Hematogenous or lymphatic spread of infected macrophages to pharyngeal, buccal, or nasal tissues is the hallmark of the disease. MCL can potentially address serious issues such as stable face mutilation for patients due to a lack of effective treatment in the early stages of illness.

Leishmaniasis recivida:

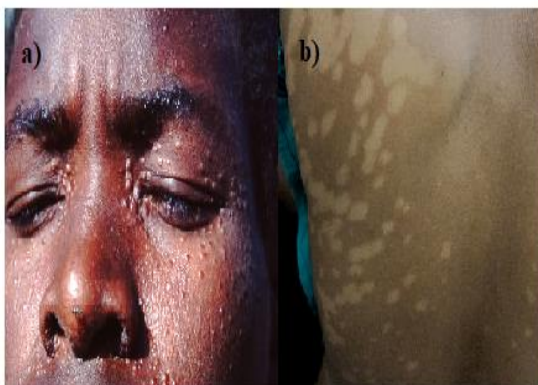


Typical characteristics of Leishmaniasis recivida

One type of hypersensitive dermal response is mainly connected with a chronic recurrence of *L. tropica* in the Middle East. Like DCL, the disease is not easily cured and is visible by the presence of nodular lesions.

Post Kala-azar Dermal Leishmaniasis (PKDL):

Post-Kala-azar Dermal Leishmaniasis (PKDL) is a common sequel in the apparently cured visceral leishmaniasis (VL) patients caused by *Leishmania donovani*. Due to its potential



Typical Characteristics of PKDL a) Papular lesion
b) Macular rash

role in infection is considered a public health problem in VL endemic areas. Clinical features include a painless hypopigmentation patch on skin and extremities, papules, or nodules in healthy individuals. The immune response to skin parasites determines this symptom.

This may have persisted since the previous VL episode. Because among the apparently cured KA patients, 10% to 20% in India and 50% to 60% in Sudan develop PKDL (Burza et al., 2018). Though the symptom might not seem very harmful, the lesions on the patient harbour parasites, which could be transmitted by the bites of sand flies (Zijlstra et al., 2017).

Para-Kala-Azar Dermal Leishmaniasis (para-KDL):

In the recent past, Zijlstra et al., 2000; Zijlstra et al., 2003 reported the occurrence of PKDL manifestation in active VL patients, i.e., patients having PKDL-like skin patches along with amastigotes in their visceral organs, these are therefore called 'para-Kala-azar Dermal Leishmaniasis' (para-KDL) cases. It comprises an intermediate position between active VL and PKDL. Such rare cases were reported in East Africa and the Indian subcontinent (Kumar et al., 2016).

EPIDEMIOLOGICAL DISTRIBUTION:

Global Distribution

Leishmaniasis are widespread all over the tropical and sub-tropical regions of Africa, Asia, the Mediterranean, Southern Europe (Old World) and South and Central America (New World) (Fig.13). Despite significant efforts, it is difficult to presume the exact magnitude of leishmaniasis on public health. At the same time, many cases have not been reported or misdiagnosed. It is estimated that approximately 12 million people are infected in 98 countries (Alvar et al., 2012). Leishmaniasis is associated with the scarcity of financial development and several environmental changes, such as urbanisation, deforestation, and migration of people into endemic areas. Annually the occurrence of CL and VL cases are estimated to be 11.5 million and 500,000, respectively.

95% of the VL incidences occur in India, Brazil, Ethiopia, Kenya, Nepal, Sudan and South Sudan (Burza et al., 2018; Mann et al., 2021). Approximately 1.5% to 9% of AIDS patients in the Mediterranean basin develop VL, and 30% to 70% of adult visceral leishmaniasis cases are related to HIV infection (WHO 2019).

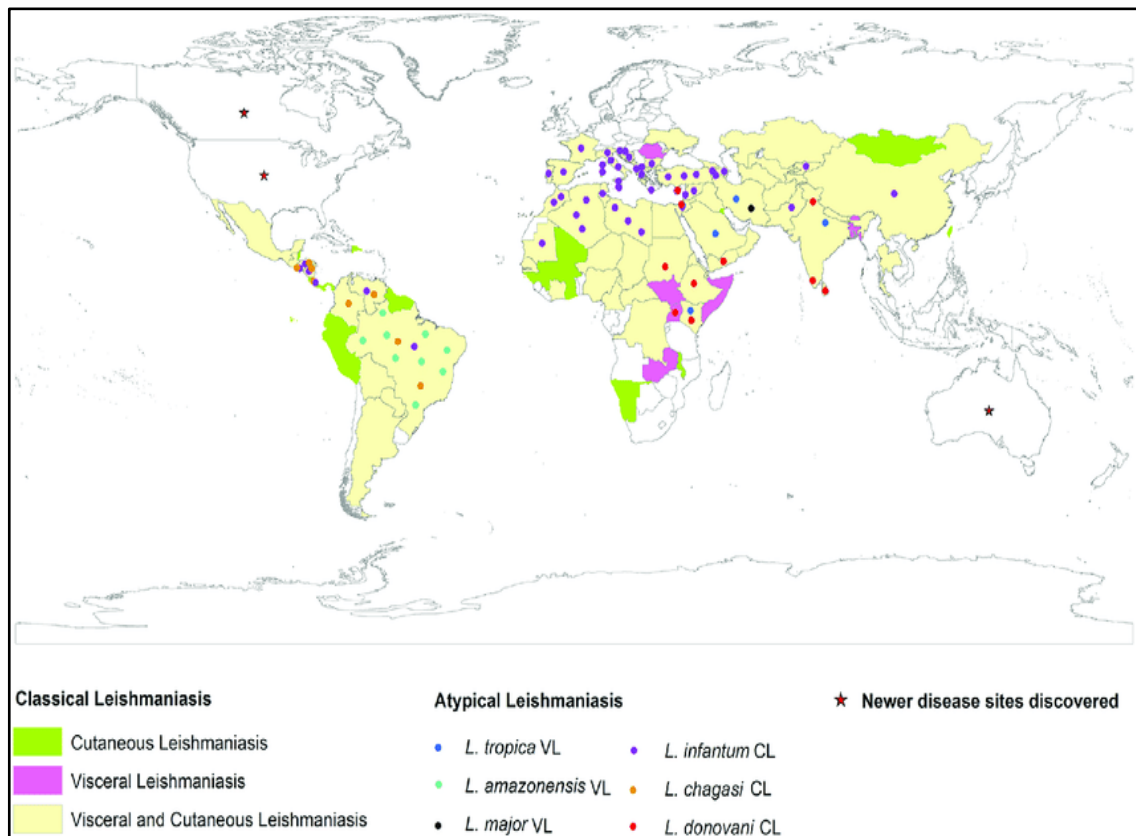


Figure 7. Global Distribution of Leishmaniasis

Indian Scenario

According to the medical records, the first Kala-azar (KA) outbreak occurred in 1824 in the Jessore region of Bengal, now Bangladesh, killing around 45,000 people over three years. Sanyal et al. have detailed epidemiologically significant characteristics of this disease in the past (1994). In the past, most of the pioneering work was done by the British Army. Sandflies were also destroyed after independence due to the massive use of DDT as part of the National Malaria Elimination Program, and Kala-azar has become a sickness of the past.

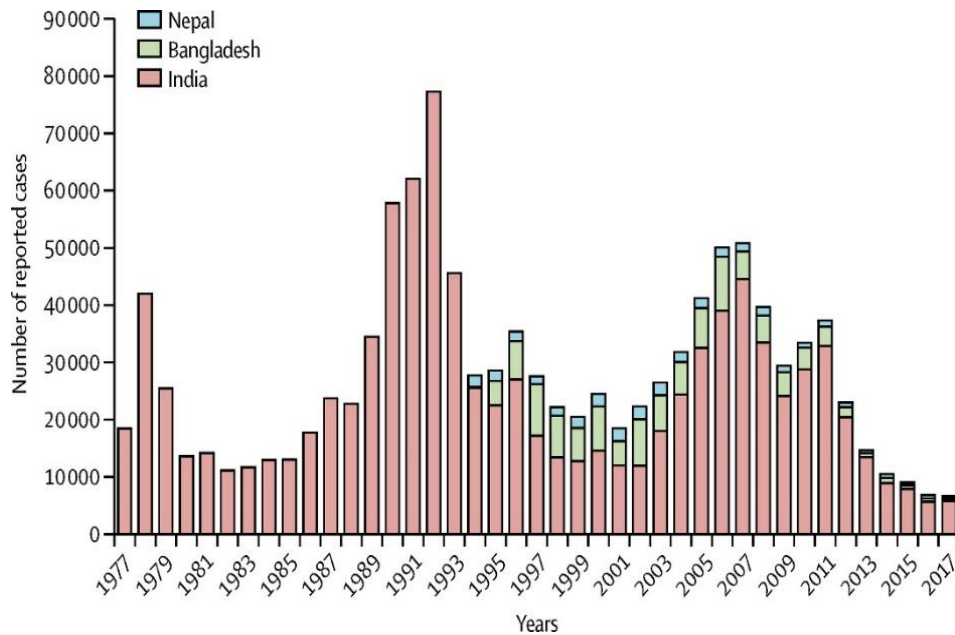


Figure 8. VL Cyclical Epidemiological Patterns in South Asia

A significant epidemic then broke out in 1993 in Bihar and the neighbouring states of West Bengal and Uttar Pradesh. KA has become endemic to large areas of eastern India, mainly in Bihar, Jharkhand, West Bengal and Uttar Pradesh. In the All India Institute of Medical Sciences, between July 1997 and December 1998, the sub-Himalayan (Kumaun) region of northern India accounted for new cases of KA, this is an entirely new outbreak in India (Singh et al., 1999). The disease has spread from Bihar, West Bengal, Uttar Pradesh, Assam and Tripura to as far away as Chennai in Tamil Nadu and urban areas of Mumbai (Bhatia et al., 1995). In Bihar alone, 44 million people from 28 districts and about 3.5 million in West Bengal are at risk of KA. The most alarming part of the new cases is that most reported KA cases are unresponsive to antimonial treatment (Thakur, 1993). While in the year 2008, Miltefosine resistance cases have also been reported (Arif et al., 2008; Khanra et al., 2012).

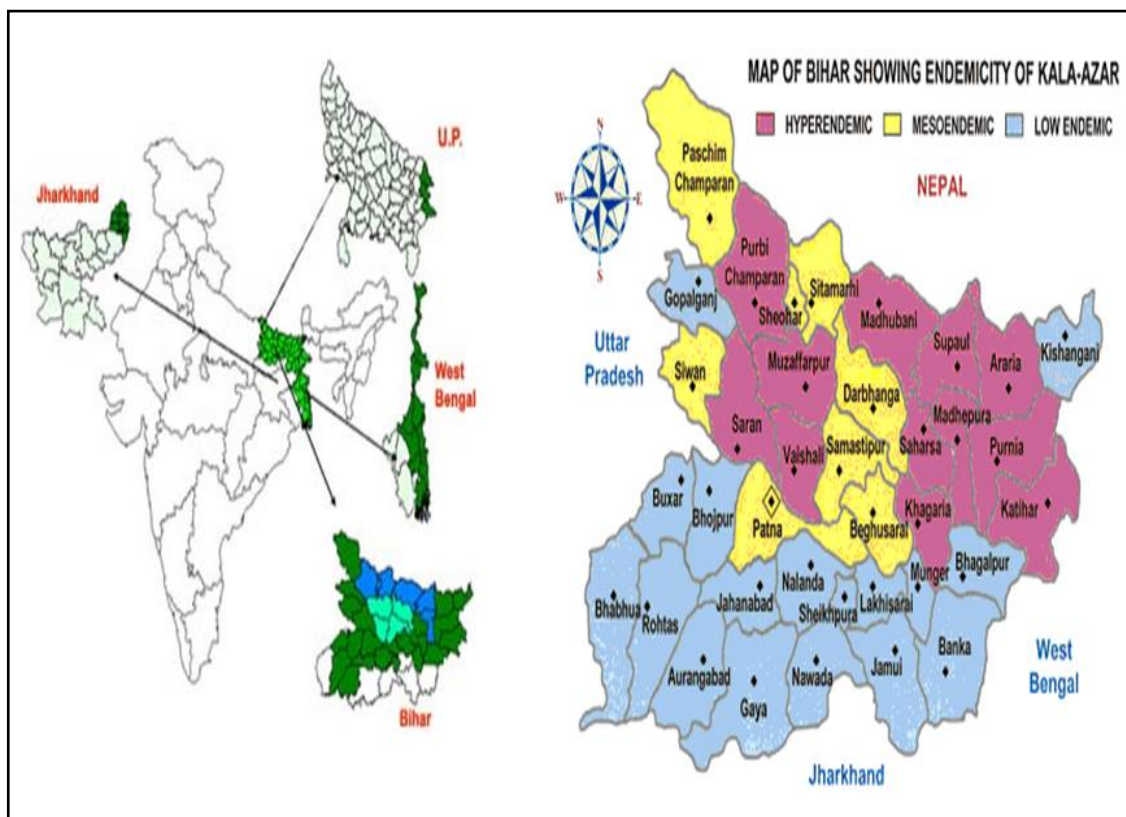


Figure 9. The distribution of KA in India and Bihar is the most endemic region

Leishmaniasis / HIV CO-INFECTION:

The Leishmania/HIV co-infection has emerged because of the growing overlap among leishmaniasis (in particular visceral, hardly cutaneous) and AIDS, that's because of the spread of the AIDS pandemic to rural regions and that of visceral leishmaniasis to the suburban areas (Desjeux et al., 2001). *Leishmania* and HIV co-infection have developed as severe illness conditions. Leishmania/HIV co-infection cases have been documented worldwide, with the worst effects occurring in Southern Europe and Brazil. VL is hastening the onset of AIDS in HIV-positive individuals. In endemic areas, HIV, on the other hand, raises the risk of VL infection by more than 100 times.

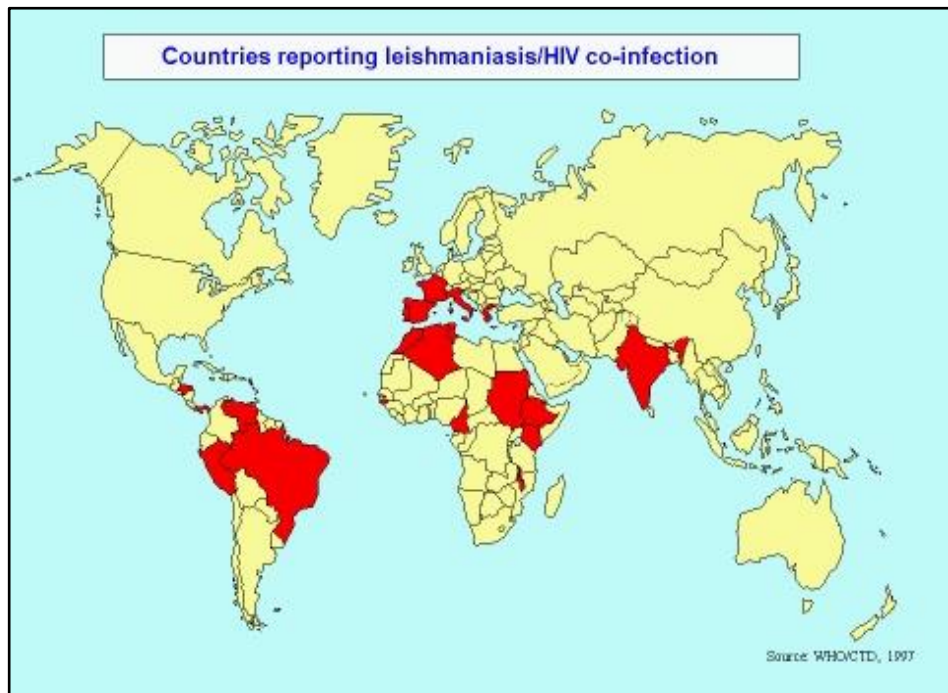


Figure 10. Leishmaniasis / HIV co-infection Distribution

Multiple environmental, host or parasite-related variables have contributed to the global spread of the leishmania-HIV coinfection (Lindoso et al., 2016). Most VL cases recorded on the Mediterranean coast involve *L. donovani* and *L. infantum* (mainly in Spain, Italy and France). In some HIV-endemic locations, *L. braziliensis* (Hernandez et al., 1995), *L. aethiopica* (Berhe et al., 1995), *L. tropica* (Magill et al., 1993), and *L. Major* (Gillis et al., 1995) have also been found to cause VL.



REVIEW OF THE LITERATURE

THE DISEASE LEISHMANIASES:

Leishmaniasis are a vector-borne disease caused by the protozoan parasite of the genus *Leishmania*. This protozoan parasite is harmful to humans due to its wide range of clinical manifestations, ranging from minor, self-healing cutaneous lesions to severe mucocutaneous ulcers or visceral infection, which can be fatal, if left untreated. More than 20 species exist in 88 countries belonging to the Trypanosomatidae family. The *Leishmania* parasites that cause leishmaniasis are spread through the bite of an infected female sandfly. The life cycle of a parasite comprises two different hosts. In the vertebrate host macrophages, the parasite lives as an amastigote and the digestive system of sand flies thrive as a promastigote. However, *Leishmania*'s sexual cycle was not seen during the promastigote and amastigote stages. According to previous research, global mortality and morbidity from leishmaniasis are increasing at an alarming rate (WHO, 2010). Because leishmaniasis is considered a Third World endemic disease, pharmaceutical corporations have shown little interest in developing low-toxicity safe medications and effective vaccinations for combating the disease. As a result, deciphering the parasite's genomic sequence was considered a way to improve techniques and speed up the study of the biology of this formidable human parasite and its interaction with its host. Even though *Leishmania donovani* has been extensively investigated, its genome is still poorly annotated due to its unevenness. A complete assembly is required to fully comprehend the parasite's biology, metabolic pathways, tissue tropism and disease pathology (Lypaczewski et al., 2018).

MODES OF TRANSMISSION:

Globally, the most frequent mode of transmission of leishmaniasis is vector-borne transmission. Other modes of transmission may be becoming more significant in HIV-positive patients.

◆ Vector-borne Transmission

Infection with the *Leishmania* parasite is spread through the bites of infected female phlebotomine sandflies, which feed on blood to generate eggs. The protozoa grow inside the sandfly and are passed on when the sandfly eats a healthy person's blood. The epidemiology of leishmaniasis is determined by parasite and sandfly species features, local ecological factors of transmission locations, current and previous parasite exposure in the human population. *Leishmania* parasites have been detected in natural reservoir hosts in more than 70 animal species, including humans (WHO, 2021). *Leishmania* protozoa multiply inside a healthy person's white blood cells and cause sickness.

◆ Sharing of Needle

In the year 1993, 18,347 cases of Acquired Immune Deficiency Syndrome (AIDS) and 200 cases of Human Immune Virus (HIV)-associated leishmaniasis were reported in Spain, with more than 85% of the cases occurring among IVUDs (Alvar, 1994). In 1998, seven cases of transfusion-associated leishmaniasis were documented, including four children who had had several transfusions and three who got blood from a single donor (WHO, 1998).

◆ Congenital Transmission

Congenital transmission of visceral leishmaniasis or KA from an asymptomatic mother to her 16-month-old child has been reported in Germany who was admitted to Children's Hospital of Stuttgart in Germany (Meinecke et al., 1999; Osorio et al., 2012).

GENOME ORGANIZATION OF *Leishmania*:

Leishmania spp. contains a haploid genome with 32,816,678 bp arranged into 36 chromosomes, 911 RNA genes, and 39 pseudo-genes. A total of 8272 genes encodes proteins. The New World *Leishmania spp.* has 34 or 35 chromosomes, whereas the Old World *Leishmania spp.* has 36. In *L. major*, *Trypanosoma brucei*, and *Trypanosoma cruzi*, producer protein genes encode long polycistronic genes lacking transcription factors. Chromosomes 8 and 29, as well as chromosomes 30 and 36, have linkage groups in *L. mexicana* and chromosomes 20 and 36 have a linkage group in *L. braziliensis* (Ivens et al., 2005). Proteins from *Leishmania spp.* are produced during translation or after the end of replication. The mechanism involves eukaryotic RNA polymerase II for transcription regulation which differs from the other mechanisms, although they exhibit a chromatin remodelling process. The *Leishmania* genome lacks a sub-telomeric region (species-specific genes) and a transposable element in contrast to other Trypanosomatidae members. The genome of *Leishmania spp.* is divided into two parts: the nucleus, which contains chromosomal and episomal DNA and the kinetoplasts, which contain self-replicating DNA molecules. In addition, virus-like particles can be seen in the cytoplasm (Ivens et al., 2005). Ultracentrifugation was used to separate and study the kinetoplasts, while earlier pulsed-field gel electrophoresis was used to study the chromosomes (PFGE).

Nowadays, third-generation sequencing technologies, including ion semiconductor sequencing, have made large genome sequencing feasible by mapping short sequence reads and producing a total data output of 10–1,000 MB, depending on the type of ion semiconductor chip employed (Gupta et al., 2020).

DIAGNOSIS OF VISCERAL LEISHMANIASIS (VL):

This vector-borne disease, visceral leishmaniasis, can infect up to 350 million individuals worldwide. Due to various clinical symptoms, VL is the most severe form of the disease. Because the clinical symptoms of VL are similar to those of various other febrile diseases like malaria and enteric fever, a correct diagnosis is crucial for the treatment. The gold standard for VL diagnosis is still invasive and dangerous, including parasites from splenic and bone marrow aspirate stained preparations. Serological tests, for example, the Direct Agglutination Test (DAT), rK39 in ELISA or rapid immunochromatographic format and immunoblotting have issues with a significant proportion of asymptomatic individuals being positive and their inability to diagnose relapses because these tests remain positive for months to years after cure. The most popular molecular approach for successfully diagnosing and differentiating species is PCR. A list of various diagnostic assays related to VL, in brief, is given below and has also been presented in Table 1 (Srivastava et al., 2011).

➤ Parasitological diagnosis of VL

The symptoms and clinical manifestation of visceral leishmaniasis, such as splenomegaly, hepatomegaly and a high undulating fever, are used to make a preliminary diagnosis. However, these factors are insufficient to distinguish VL from other comparable illnesses, such as malaria, relapsing fever, liver abscess and trypanosomiasis. Parasitological methods remain the gold standard in diagnosing visceral leishmaniasis because of their excellent specificity. In tissue smears from lymph nodes, bone marrow or spleen, the amastigotes, called LD bodies, can be detected. The cytoplasm appears pale blue in preparations stained with Giemsa or Leishman stain, with a comparatively large nucleus staining red. A kinetoplast is a deep red or violet rod-like body in the same plane as the nucleus but at a right angle. Splenic smears have a sensitivity range of 93.1–98.7%.

The sensitivity of bone marrow and lymph node smears ranges from 52–85% 11 and 52–58 %, respectively (Zijlstra et al., 1992; Siddig et al., 1998).

➤ **Serological diagnosis of VL**

The Indirect Fluorescent Antibody Test (IFAT) is based on detecting antibodies present in the early stages of infection and disappearing six to nine months after the cure. If low titres of antibodies persist, it is a sign of a possible recurrence. The sensitivity and specificity are 96% and 98% respectively (Sassi et al., 1999).

Enzyme-Linked Immunosorbent Assay (ELISA)-The antigen employed determines the sensitivity and specificity of ELISA. Antigen rk39 had the most promising results, with a sensitivity and specificity of 100% and 96% respectively (Houghton et al., 1998).

Immunoblotting-This test is more sensitive than the IFAT and ELISA tests but is also more expensive and time-consuming. It also necessitates a high skill level; thus, it is only employed in diagnosing VL in a few cases (Srivastava et al., 2011).

Direct Agglutination Test (DAT)- In this test, promastigotes stained with Coomassie brilliant blue are incubated with patient sera and agglutination is seen after overnight incubation. It is a particular, sensitive, low-cost and straightforward test. Initially, aqueous antigen was utilised, but it had the disadvantage of requiring a cold chain and having a short shelf life. Its sensitivity and specificity estimate of 94.8% (Chappuis et al., 2006).

Antigen Detection -Antigen detection is a fantastic way to diagnose an infection. Antibody-based immunodiagnostic assays are less specific. Antigen levels are thought to have a strong relationship with parasite load. In the case of HIV-VL coinfection, this method of diagnosis should be a better alternative to antibody detection (Singh et al., 2002; Srivastava et al., 2011).

Table 1: Various diagnostic assays for VL (Srivastava et al., 2011).

Assay	Sensitivity	Specificity	Comments
Parasitological diagnosis: microscopic examination	Splenic aspirate 93.1-98.7%; Bone marrow aspirate 52-85%; Lymph node aspirate 52-58 %	100%	Restricted to the endemic area where clinicians are familiar with signs and symptoms and for culture, sophisticated laboratories are required.
Agglutination test (using urine)	64-100%	100%	The ideal method of diagnosing an infectious agent; antigen level correlates with parasite load.
Indirect Fluorescent Antibody Test	96%	98%	Requires equipped laboratory setup
Direct Agglutination Test	94.8% (95% CI, 92.7-96.4)	97.1% 95% CI, 93.9-98.7)	Useful for epidemiology studies
Immuno-chromatic Strip Test	93.9% (95% CI, 87.7-97.1)	95.3% (95% CI, 88.8-98.1)	Used for screening by using rk39 strip, not valuable for the treated patient
ELISA	100%	96%	It cannot be used in a field setting.
PCR	70-100%	85-99%	Standardisation is a hindrance, lab to lab variation and contamination.

➤ **Molecular biological diagnosis of VL**

Amplification techniques, including PCR based assays, are the mainstay of molecular diagnostics, notably for HIV-VL co-infections, with different primers targeting numerous multicopy genes such as rRNA genes and kinetoplast DNA (kDNA) minicircles. A comparative overview of sensitivities and specificities of different types of PCR based diagnostic assays targeting multiple regions of the *Leishmania* genome has been presented in Table 2. The time-consuming and risky technique of bone marrow and splenic aspiration is replaced by PCR based technique from blood samples. Recently, VL and PKDL detection using epitope-specific PCR and oligo chromatographic dipstick assays has been described (Alvar et al., 2004).

For clinical samples, real-time PCR has allowed for determining parasite burden with a high degree of diagnostic accuracy. For blood samples, however, the ITS1 PCR positivity was much higher than that obtained by kDNA PCR-hybridization, demonstrating that PCR sensitivity varies depending on the biological sample examined. A nucleic acid-based reverse transcriptase-loop-mediated isothermal amplification (RT-LAMP) assay has recently been developed as a point-of-care diagnostic tool (Alvar et al., 2004).

The following are the main ways for nucleic-acid-based detection:

1. Hybridization with DNA probes.
2. For the detection of DNA, amplification procedures such as PCR are used.
3. RT-PCR (reverse-transcriptase polymerase chain reaction) for RNA detection

Table 2: List of the different targets used for PCR protocols to diagnose VL.

Target (bp) for PCR	Sensitivity (%)	Specificity (%)	Reported by
kDNA (nr): <i>L. donovani</i>	90	100	Nuzum et al., 1995
kDNA (204 bp): <i>L. donovani</i>	82.3	100	Singh et al., 1999
kDNA (600 bp): <i>L. donovani</i>	96	96	Salotra et al., 2001
kDNA (790 bp): <i>Leishmania</i> species	100	100	Pal et al., 2004
kDNA (600 bp): <i>L. donovani</i>	99	100	Maurya et al., 2005
SSU-rRNA (nr): <i>Leishmania</i> species	70	100	Osman et al., 1997
SSU-rRNA n-PCR (358 bp): <i>L. infantum</i>	95.4	100	Cruz et al., 2002
SSU-rRNA	73.2 by conventional PCR; 83.9 by ELISA PCR	87.2	De Doncker et al., 2005
SSU-rRNA n-PCR (358 bp): <i>L. infantum</i>	79	100	Cruz et al., 2006
SSU-rRNA real-time: <i>L. infantum</i>		100	Bossolasco et al., 2003
SSU-rRNA oligoC Test	93.2	98.3	Deborggraeve et al., 2008
MedRNA (180 bp): <i>L. donovani</i>	96.8	100	Adhya et al., 1995
PCR mini-exon gene (450 bp): <i>L. donovani</i>	NR	100	Katakura et al., 1998
n-PCR (100 bp): <i>L. infantum</i>	100	100	Fisa et al., 2002

BMA: bone marrow aspirate; bp: base pair; kDNA: kinetoplast DNA; medRNA: multicopy mini-exon RNA; n-PCR: nested PCR; NR: not reported; SSU-rRNA: small subunit ribosomal RNA.

➤ **Diagnosis of HIV-VL coinfection:**

VL is a common opportunistic infection in AIDS patients, and atypical clinical manifestations of VL in HIV-positive individuals can be challenging to diagnose. Leishmaniasis can cause gastrointestinal symptoms in some patients (stomach, duodenum or colon), lung, tonsil and skin are also involved. The sensitivity of antibody-based immunologic diagnostics like the IFA test and ELISA for HIV patients is low. The presence of *Leishmania* amastigotes in the bone marrow can often be demonstrated in these patients due to the high parasite load. Still, there have been well-documented cases in the literature where amastigotes were not verifiable in bone marrow but were found in unexpected places such as the stomach, colon or lungs. For these individuals, PCR analysis of whole blood or its buffy coat preparation may prove to be a valuable screening test avoiding the necessity for painful treatments.

DIAGNOSIS OF POST-KALA AZAR DERMAL LEISHMANIASIS (PKDL):

The disease Post Kala-Azar Dermal Leishmaniasis (PKDL) is endemic to various parts of India, Nepal, Bangladesh, and eastern Africa. It is generally observed in apparently cured, inadequately treated or untreated cases of visceral leishmaniasis (Sudan, Ethiopia, Kenya). The disease usually appears on the face and upper body as a varied combination of hypopigmented patches, erythematous succulent papulo-plaques and nodular lesions, occasionally extending to the extremities, genitalia, and tongue. Photosensitivity, verrucous, hypertrophic, xanthomatous and ulcerative lesions are common morphologies and presentations, especially in endemic locations. Recognising a wide range of mucocutaneous alterations helps clinicians during the early initiation of treatment and limits disease from spreading in the community (Zijlstra et al., 2016).

The pattern of symptoms determines the differential diagnosis. However, lepromatous leprosy is the most closely related disease (Kumar et al.,2021). Although PKDL does not cause significant morbidity at first, affected patients can act as a disease reservoir. To restrict disease transmission in the community, early detection of active cases is necessary to identify. A differential diagnosis of PKDL has been given in Table 3.

Table 3: Differential diagnosis of PKDL (Kumar et al.,2021).

Lesions of PKDL	Differential diagnosis
Hypopigmented macules and patches	Lepromatous leprosy (glove and stocking anaesthesia, peripheral nerve thickening, and madarosis), pityriasis Versicolor (scaling, perifollicular lesions), vitiligo Vulgaris (depigmented lesions, leukotrichia and Koebnerization), progressive macular hypopigmentation (central trunk), hypopigmented mycosis fungoides (sun covered areas, older adults)
Papules	Lepromatous leprosy, xanthoma (yellow-orange lesions), maculopapular drug reaction (acute onset)
Plaque and nodular lesions	Lepromatous leprosy, sarcoidosis (lung involvement, elevated angiotensin converting enzyme level), cutaneous T cell lymphoma (elderly population, ulcerated plaques),
Photosensitive form	Systemic lupus erythematosus (joint and systemic involvement), Rosacea

In PKDL, depending on the progression of the lesion, the clinical manifestation of the disease is divided into three clinical grades as follows:

- Grade 1-Scattered maculopapular or nodular lesions on the face.
- Grade 2-Dense nodular or maculopapular lesions on the face with extension to the chest, back, upper arms, and legs.
- Grade 3- All over the body are affected, including hands and feet, with dense nodular or maculopapular lesions.

DIAGNOSIS OF PARA-KALA AZAR DERMAL LEISHMANIASIS (para-KDL):

Para-KDL cases are intermediated positions between the VL and PKDL, where PKDL lesions are present when the patients still suffer from VL or Kala-azar. The parasites can be found on the skin, lymph nodes and bone marrow. In such patients, the spleen is frequently larger than in Post Kala-azar Dermal Leishmaniasis patients but smaller than in Kala-azar patients. The active association of VL with PKDL are rare and restricted to a few geographical locations. According to some sources, active VL in PKDL was rarely documented in the Indian subcontinent, although it was reported in East Africa. It is still unclear how the PKDL manifested in VL patients after or without treatment (Zijlstra et al., 2003).

TREATMENT OF THE DISEASE LEISHMANIASIS:

An accurate parasitological diagnosis is required to determine the proper treatment for Leishmaniasis. Few infections, mainly minor cutaneous lesions caused by *L. major*, heal on their own and induce immunity to re-infection. Thus, treatment is usually not preferred unless the lesions do not heal (approx. six to nine months).

While some drugs are clinically used until the disease is cured, for example, to treat the chronic lesions caused by *L. tropica*, pentavalent antimonials are administered intramuscularly or intravenously at 10- 20 mg/day until the patient is not getting cured. Currently, the sodium stibogluconate, meglumine antimoniate, Miltefosine (an oral cancer drug), pentamidine, and amphotericin B (Bristol-Myers Squibb) and its lipid preparation, AmBisome® are clinically in use for the treatment of leishmaniasis. Available treatment approaches include several drawbacks, including adverse events, the need for frequent injections, high costs and little efficacy. Another risk is the emergence of drug resistance. Hence, despite its long history, leishmaniasis is still a complex condition to cure. The series of conventional medications and combination chemotherapy are options to strengthen the immune system and continue to overcome these limits. Newer modes of therapy, such as immunotherapy or immunochemotherapy, have recently been tested due to a better understanding of disease pathophysiology. Recently, novel treatment targets in leishmania's metabolic pathways are being investigated (Pradhan et al., 2021). The second line medicine, Amphotericin B, is used to be administered in patients who are not responding to antimonial or have developed antimony resistance with caution to avoid major adverse effects. Amphotericin B in liposomes has been proven to be quite effective due to its oral efficacy and short course. Miltefosine (Hexa-decyl phosphocholine) is utilised in VL and CL but its main drawbacks were the arrival of drug resistance and teratogenicity (Arif et al., 2008). The use of azoles (antifungal agents) such as itraconazole, ketoconazole and fluconazole are also investigated as an antileishmanial property. During the multicentre trial, Paromomycin was more effective in Indian VL patients than in the Sudanese VL population. All these treatments are enlisted in Table 4 (Pradhan et al., 2021).

Table 4: List of antileishmanial drugs (Pradhan et al., 2021).

DRUGS	ROUTE OF ADMINISTRATION	DOSAGE	SIDE EFFECTS	REMARKS
Pentavalent Antimonials	IM or IV IM- Intramuscular; IV- Intravenous	20mg/kg/day (28–30 days)	Cardiotoxicity, pancreatitis,	Easy Availability, cost-effective
Amphotericin B	IV	0.75–1mg/kg/day for 15–20 days, every day or alternately	Renal toxicity, injection related reactions, hypokalemia	Primary resistance is not common
Liposomal Amphotericin	IV	single dose 3-5mg/kg/dose	Chills and rigors during injection,	High efficacy; low toxicity
Miltefosine	Oral	100-150mg/day for 28 days	Teratogenicity, Gastrointestinal problems	Cost Effective
Paromomycin	IM (VL) or topical CL	15mg/day (21 days) or 20mg/kg (17 days)	Renal and liver toxicity	Cost Effective

Combination chemotherapy was developed to minimise drug resistance, enhance compliance and shorten treatment duration with lower therapeutic costs. Sodium stibogluconate/meglumine antimoniate plus paromomycin, liposomal amphotericin B plus miltefosine, miltefosine plus paromomycin and liposomal amphotericin B plus paromomycin are some of the available combination therapies. Figure 11 represents the chemical structures of some antileishmanial drugs.

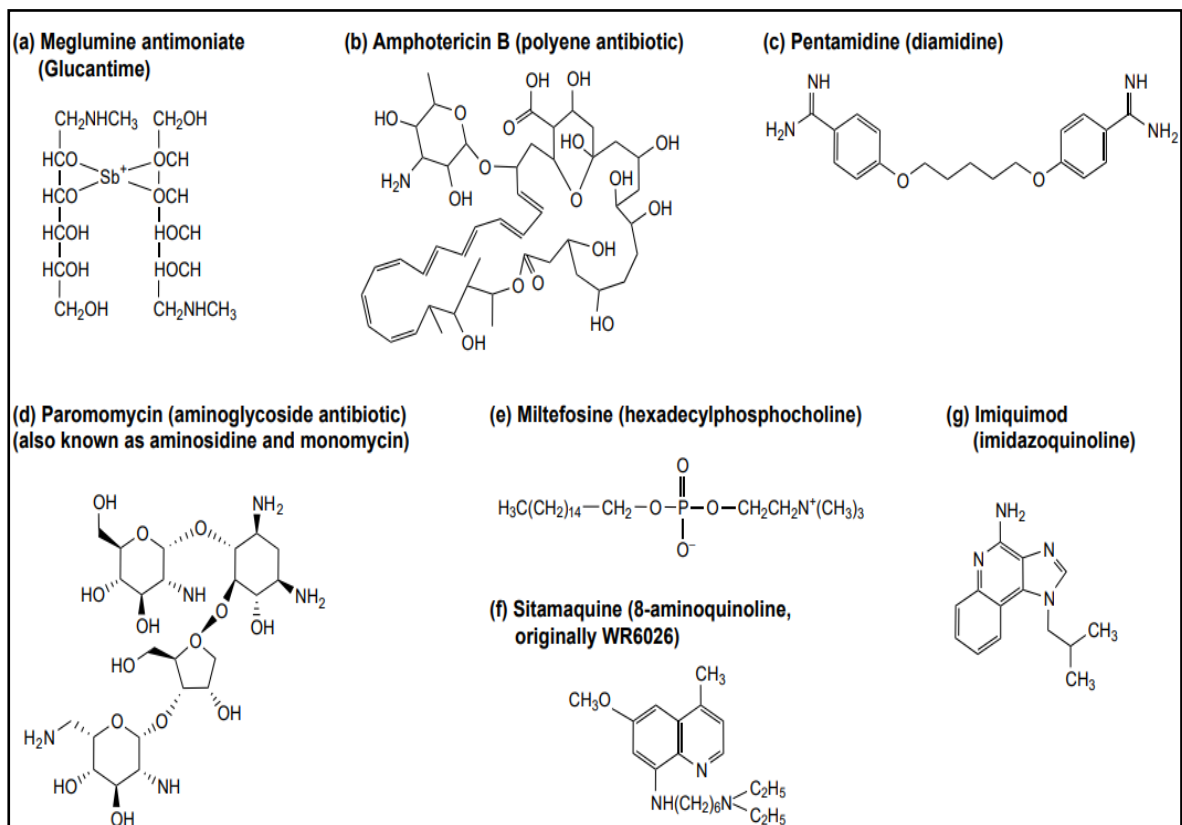


Figure 11. Chemical structures of some antileishmanial drugs.

THE FACTORS INVOLVED IN DRUG RESISTANCE:

The current knowledge about the epidemiology and transmission of Leishmaniasis reflects that the acquired drug resistance appears to be a crucial factor in anthroponotic disease foci such as *Leishmania donovani* in Bihar State, India (Sundar et al., 2001).

Specific assessable parameters primarily ascertain the arrival of drug-resistant genotypes in the parasitic population:

- (i) The amount of drug dose and its interval of used,
- (ii) The probability between the arrival of drug-sensitive to resistant infection,
- (iii) The time interval of infection in the respective individuals and
- (iv) The multiplication and transmission rate of the pathogen which is experienced by parasites to being resistant in the absence of drugs
- (v) Genomic variation in the parasites due to drug pressure

Host-related factors:

The immune status of leishmaniasis patients has long been known to affect the applied drug efficacy. Experimentally the lack of activity of pentavalent antimonial has been checked in immunodeficient mouse models, for which the effects are probably due to the deficiencies of both Th1-cell-mediated and macrophage responses (Murray et al., 1989). Several experimental models have shown that the antileishmanial activity of pentamidine is T-cell dependent, whereas those of amphotericin B and miltefosine are T-cell independent (Murray et al., 1993). A few reports are available on the expression of ABC transporters in laboratory isolates of in vitro-developed SSG-R strains of *Leishmania* amastigotes (El Fadili et al., 2005) or on amastigotes from field isolates of antimony-resistant *L. donovani* (Vergnes et al., 2007).

Leishmania related factors:

Biochemical and molecular distinctions have been used to describe the species and these differences give a framework for phylogenetic analysis and improved species identification and diagnosis methods. Given the known biochemical and molecular distinctions between species, it is somewhat unsurprising that *Leishmania* species differ in their innate sensitivity to various clinically used drugs. Multiple processes can reduce drug levels at the target site of action, including decreased uptake, increased export and inactivation of the drugs by metabolism or sequestration. Similarly, changes in primary target levels can occur due to reduced target affinity for the drug or complete loss of target, which is frequently associated with a bypass mechanism. Inhibition of a primary target often triggers complex downstream cascades that lead to cell damage and death. Many antiparasitic medications, for example, undergo futile-redox cycling, creating reactive oxygen species that can cause peroxidative damage to membrane lipids, proteins or DNA. As a result, drug resistance may be linked to the overexpression of specific repair systems. Typically, many mechanisms are involved (Croft et al., 2003). Intracellular *Leishmania* stages actively monitor arginine levels within the host macrophage's phagolysosome and initiate a MAP kinase-dependent response when arginine levels are lowered. This response may have an impact on macrophage responses and is necessary for intracellular parasite replication (Goldman-Pinkovich et al., 2016).

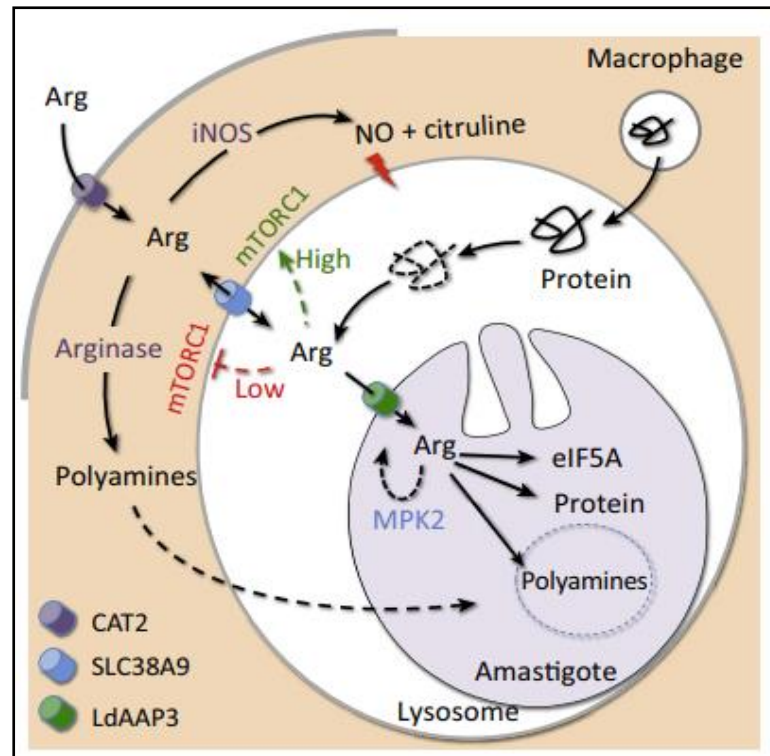


Figure 12. *Leishmania Amastigotes Arginine (Arg) Salvage Pathway.* This pathway is essential for intracellular growth and may modulate metabolic and signalling pathways in host macrophages (Goldman-Pinkovich et al., 2016). iNOS stands for inducible nitric oxide synthase, and LdAAP3 stands for high-affinity arginine transporter

Protein synthesis, polyamine production which begins in the glycosome (broken circle), and hypusine modification of the translation factor eIF5A depend on Arg absorption in *Leishmania*. The rate of protein degradation, endosome fluid phase uptake and import via cationic amino acid2 (CAT2) and SLC38A9 Arg transporters can all influence Arg availability in the phagolysosome. Inactivation of mTORC1 signalling complexes on the phagolysosome membrane, polarisation of the macrophage host towards an M2 phenotype, and increased expression of arginase and polyamine synthesis, rather than synthesis of microbicidal nitric oxide (NO), may result from intracellular amastigotes depleting phagolysosome pools of Arg (Figure 12).

MECHANISMS OF ACTION AND RESISTANCE OF DIFFERENT DRUGS:**Pentavalent antimonial**

The first line of treatment for Leishmaniasis is pentavalent antimonial [Sb (V)]. Antimonial has disrupted key metabolic processes such as fatty acid oxidation, glycolysis and energy metabolism. However, the drug's exact mechanism of action is still uncertain. To make matters worse, there is the issue of ‘drug resistance’. Many attempts have been made to understand the mechanism of drug resistance in *Leishmania* but the exact nature of the phenomena remains unknown.

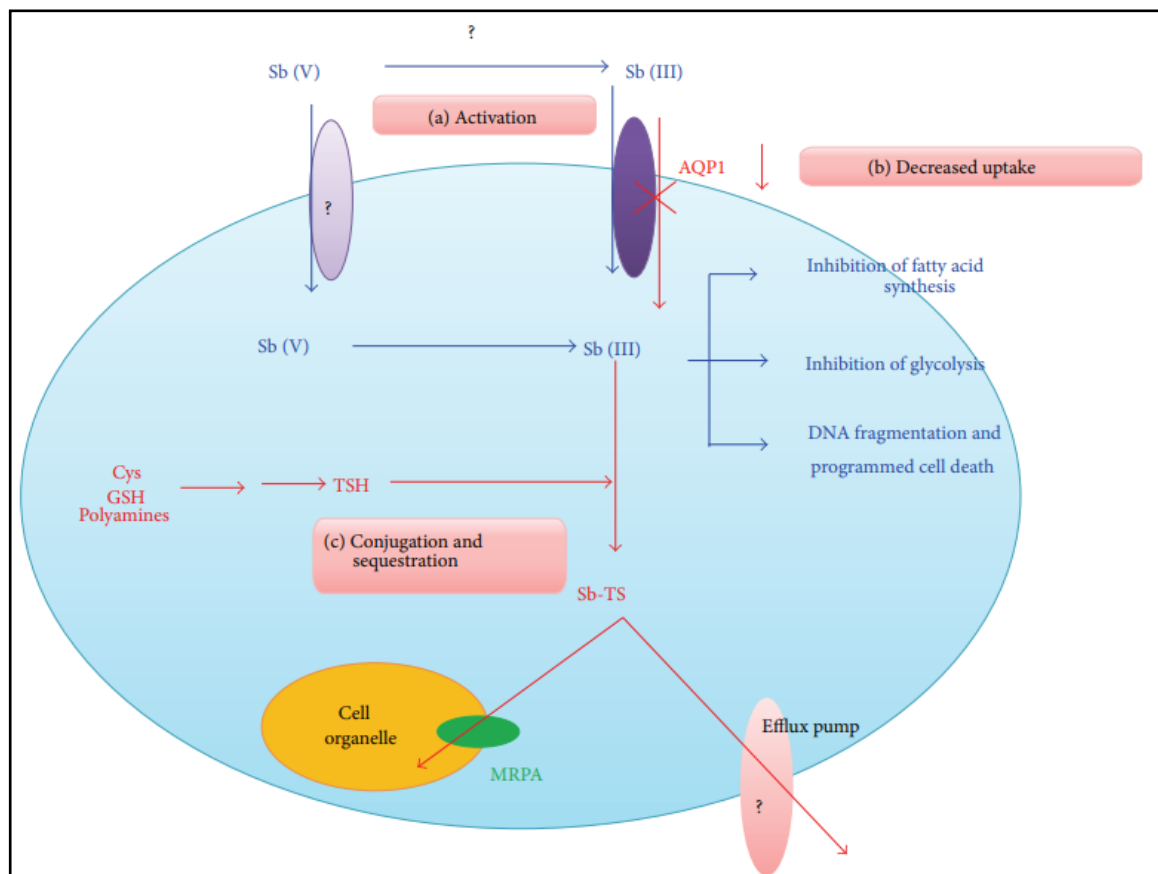


Figure 13. The mechanisms of antimony resistance in *Leishmania* (Kaur et al., 2016).

Here are the following steps which explain the mechanisms of antimony resistance in *Leishmania* (Figure 13):

(a) **Activation:** Sb (V) to Sb (III) conversion is inhibited in Resistant parasite cells (R-strain), inhibition either occurring extracellularly (by unknown enzyme inactivation) or intracellularly (by inhibiting enzymes like ACR2 or TDR1).

(b) **Decreased uptake:** Sb uptake into the cell reduces by decreasing expression of AQP1 hence, conferring resistance.

(c) **Conjugation and sequestration:** As per the increase in thiol levels [such as cysteine (Cys), GSH, TSH, and polyamines] in the cell results, its conjugation with Sb to form an Sb-thiol complex (Sb-TS). Sequestration of the Sb-TS complex inside the cell organelle or extrusion from the cell leads to a lowering of the intracellular amount of antimony. Blue lines show the possible drug action in sensitive *Leishmania* strains, while red lines depict the probable routes to achieve resistance as observed in resistant cells.

The study has focused on the mechanism of antimony resistance in *Leishmania* genomics. According to genomic research, gene amplification has been identified as one of the critical mechanisms for the treatment of drug resistance in *Leishmania*. Various studies have compared the proteomes of resistant and sensitive strains of *Leishmania*. Overexpression of metabolically essential proteins like carboxypeptidase, enolase, fructose-1, 6-bisphosphatase, Hsp70, and Hsp83, as well as downregulation of a kinetoplast membrane protein (KMP11) and calcineurin, were discovered in the resistant strains (Kaur et al., 2016).

A brief account of the mechanism of action of these antileishmanial drugs has been given in Table 1.

Table 5: List of antileishmanial drugs and mechanism of action.

Generic name of the drug (chemical type)	Mechanism of action
Pentavalent antimonials: Meglumine antimoniate (Glucantime) Sodium stibogluconate (Pentostame)	The structure of sodium stibogluconate is still unknown despite its use for over 50 years. Activated within the amastigote, but not in the promastigote, by conversion to a lethal trivalent form. The activation mechanism is not known. The antileishmanial activity might be due to action on the host macrophage.
Amphotericin B (polyene antibiotic)	Complexes with 24-substituted sterols, such as ergosterol, in the cell membrane, thus causing pores which alter ion balance and result in cell death.
Pentamidine (diamidine)	Accumulated by the parasite, effects include binding to kinetoplast DNA. The primary mode of action is uncertain.
Paromomycin (an aminoglycoside antibiotic) (also known as aminosidine or monomycin)	In bacteria, paromomycin inhibits protein synthesis by binding to 30S subunit ribosomes, causing misreading and premature termination of mRNA translation. In <i>Leishmania</i> , paromomycin also affects mitochondrion.
Miltefosine (hexadecylphosphocholine)	Primary effect uncertain, possible inhibition of ether remodelling, phosphatidylcholine biosynthesis, signal transduction and calcium homeostasis.
Sitamaquine (8-aminoquinoline, originally WR6026)	Unknown, might affect mitochondrial electron transport chain.
Imiquimod (imidazoquinoline)	Stimulates nitric oxide production from macrophages.

Miltefosine

The proposed mechanisms include the killing of *Leishmania* parasites and several immunomodulatory effects, which are as follows:

- (i) Platelet aggregation factor (PAF) receptor, increasing production of interleukin (IL)-12, enhancement of interferon-gamma (IFN- γ) receptor, which in turn lowers the production of T-helper (Th) cell type 2 cytokines (such as IL-4, -5, -10, and -13),
- (ii) Activation of IFN- γ reversing sphingosine-1-phosphate (SPH-1) inhibition of signal transducer and activator of transcription 1 (STAT-1), which is translocated to the nucleus and involved in the stimulation of the host cellular immunity,
- (iii) Activation of p38 Mitogen-Activated Protein Kinase (p38MAPK), which is initially inhibited by *Leishmania* and
- (iv) Inhibition of PI3 kinase phosphorylation of protein kinase B (Akt), which is initially stimulated by the parasite. Red lines indicate an inhibitory effect, while green arrows indicate a stimulatory effect.

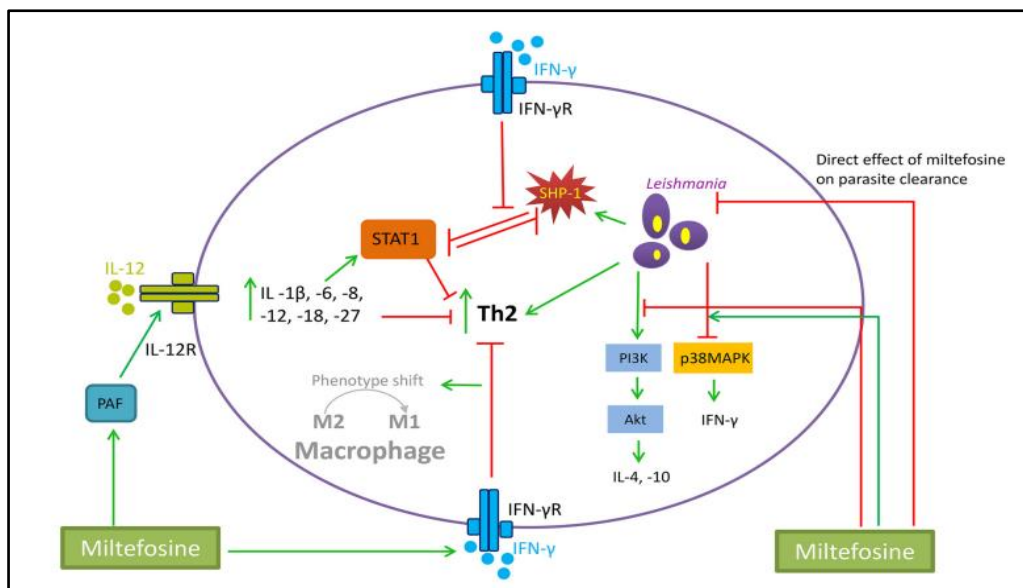


Figure 14. Proposed mechanisms of action for Miltefosine (Palic et al., 2019).

Amphotericin B

The second line of defence against leishmaniasis is Amphotericin B (AmB). This polyene medication has a high affinity for membrane-bound ergosterol, causing small membranous pores to develop, altering the membrane permeability towards cations, water and glucose molecules. *Leishmania* does not have a lot of AmB-resistance in the field or clinical isolates. Instead, other studies show that AmB susceptibility is unaltered even after repeated medication administration. A patient infected with an AmB-resistant strain of *L. donovani* was recently reported in India, raising the prospect of more instances in the future (Kaur et al., 2016).

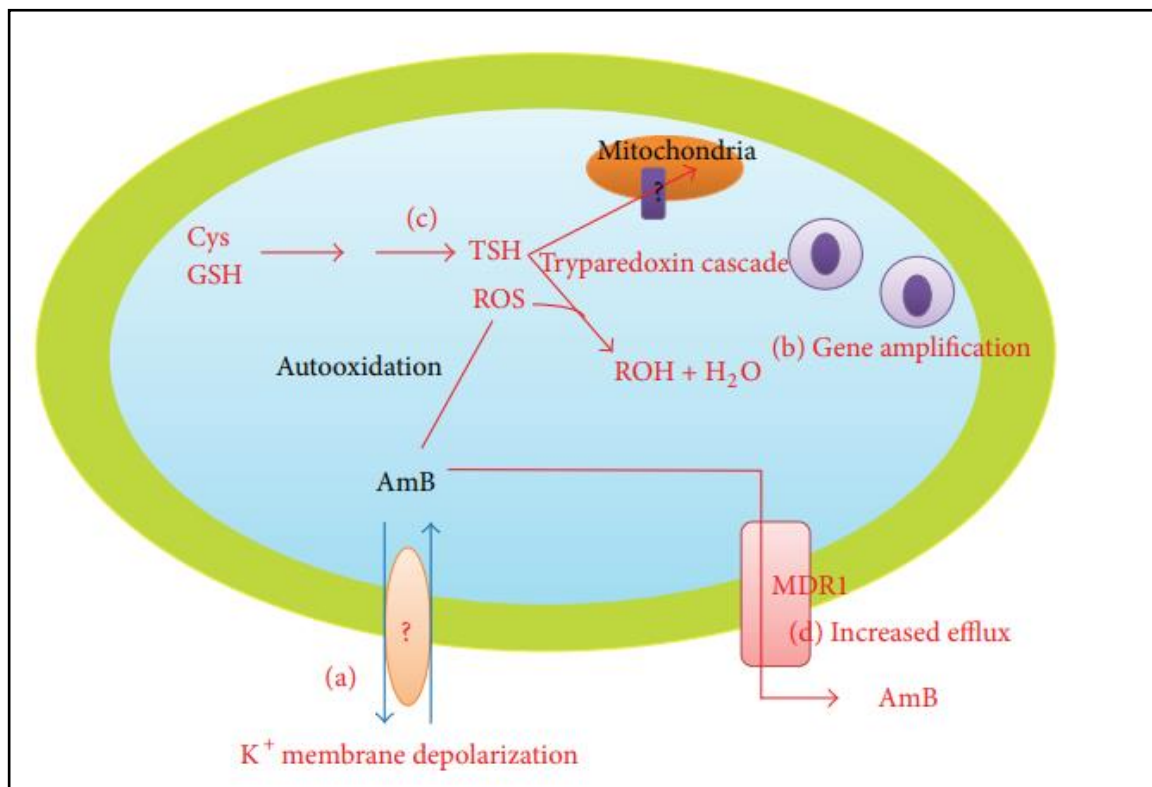


Figure 15. The mechanisms of amphotericin B resistance in *Leishmania* (Kaur et al., 2016). (a) Change in membrane fluidity results in blocking the drug entry inside the cell. The membrane transporters or factors responsible for such changes in membrane depolarization are still unknown. (b) Gene amplification of genes to confer resistance.

(c) Activation of trypanothione cascade to prevent the oxidative damage caused by the drug.

(d) Drug efflux through various membrane-bound pumps like MDR1.

Pentamidine

In cases with Sb resistance, pentamidine (PMD) has been utilised as an alternate treatment for VL. Its precise mechanism of action is unknown. However, it has been shown to block the enzyme S-adenosyl-L-methionine decarboxylase, interfere with polyamine production, and lower mitochondrial membrane potential. As a result, parasite mitochondria appear to be the drug's primary target.

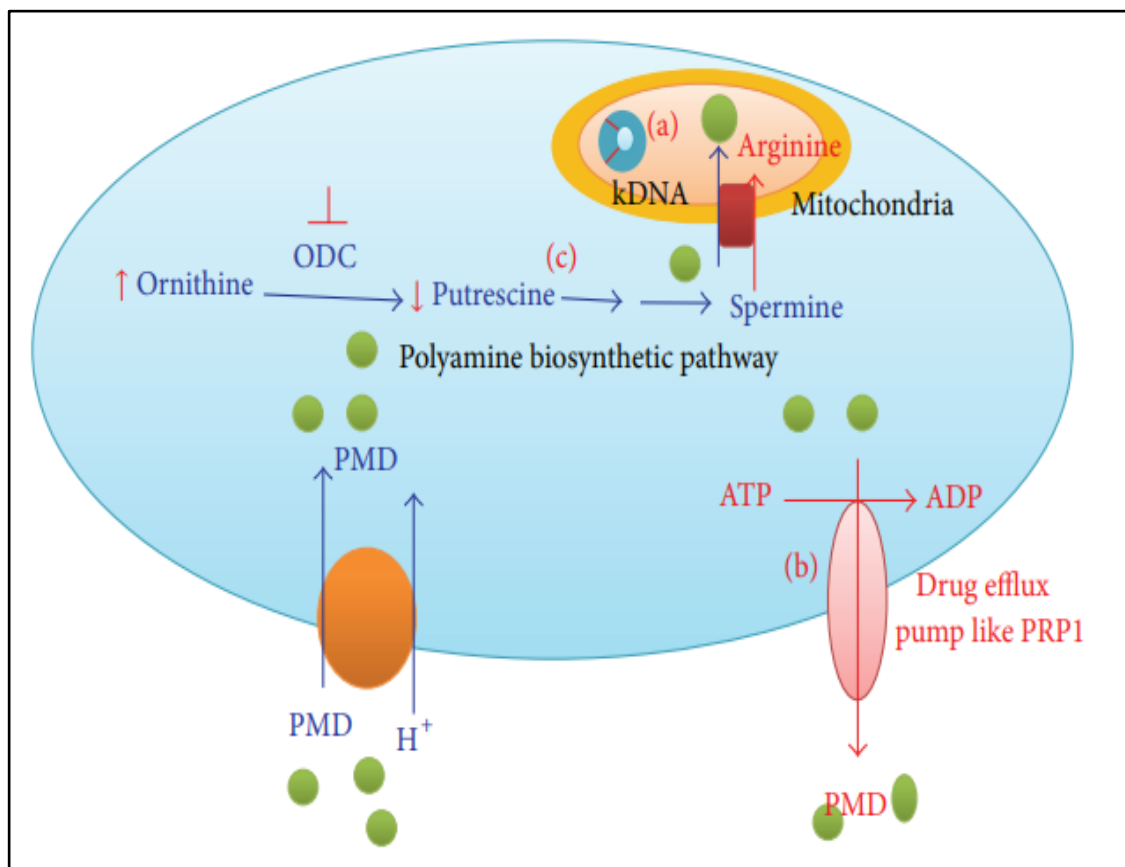


Figure 16. The mechanisms of Pentamidine resistance in *Leishmania* (Kaur et al., 2016). (a) Change in kDNA sequence confers resistance. The exact mechanism of conferring resistance remains unknown. (b) Presence of drug efflux pumps like PRP1 to remove the drug molecules from the cell and thus protect it from damage. (c) Reduced uptake of pentamidine in mitochondria due to altered polyamine biosynthetic pathways and

lowered membrane potential. Blue lines indicate the possible drug action in sensitive *Leishmania* strains, while red lines depict the possible routes to achieve resistance as observed in resistant cells.

DIFFERENT METABOLIC PATHS FOR INTRACELLULAR SURVIVAL OF *Leishmania sp.*

The morphological changes including rounding up and retraction of the flagellum, are associated with promastigote to amastigote differentiation. It has been more challenging to define the extent to which intracellular metabolism and related functions are altered. These parasites proliferate in acidified vacuoles in phagocytic host cells such as macrophages, dendritic cells and neutrophils in the mammalian host. Recently, the nutritional composition of the *Leishmania* parasitophorous vacuole and the metabolic pathways required for virulence by these parasites have been reported (McConville et al., 2011). The virulence phenotype of *Leishmania* mutants has proven to be very valuable in identifying carbon sources and nutritional salvage pathways required for parasite persistence. *Leishmania* is exposed to varying amounts of oxidative stress during the innate and adaptive immunological phases of infection. The oxidative defences of these parasites are carefully tuned to cope with these challenges. According to gene deletion studies, impairment of any of these systems leads to virulence attenuation.

Recent genome-wide transcriptomic and proteomic studies show that transcription (mediated by polymerase II) and protein synthesis are both downregulated globally in axenic, *in vitro* amastigotes culture, implying that intracellular stages enter a slow-growth state with lower metabolic requirements than extracellular promastigotes (Alcolea et al., 2010; Lahav et al., 2011).

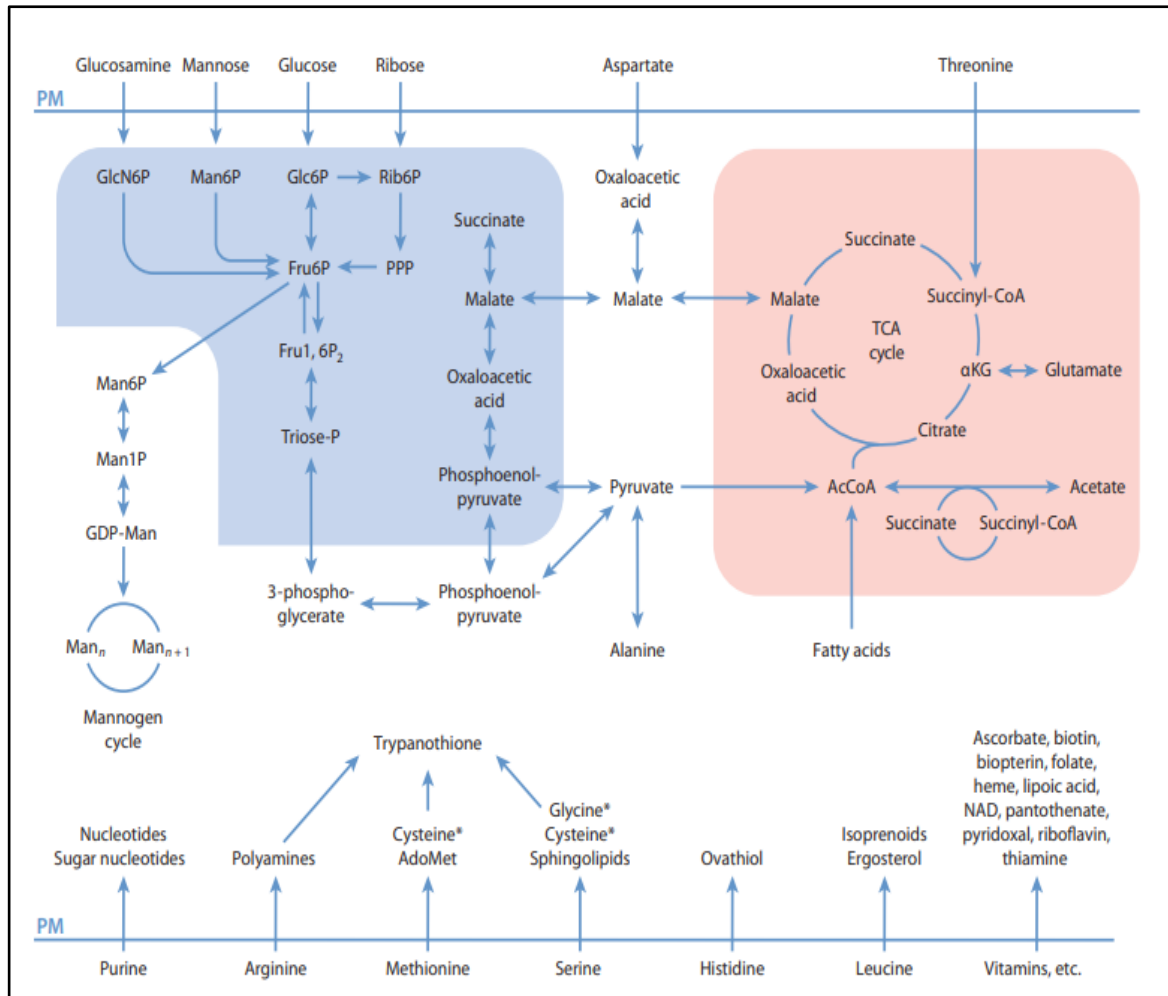


Figure 17. Metabolic pathways associated with core metabolism. *Metabolism is required for amastigote survival in macrophages. Asterisks indicate the presence of de novo pathways of synthesis that are insufficient to sustain normal growth without salvage from the PV/medium. Other essential amino acids that must be salvaged from the PV include Lys, Ile, Val, Phe, Tyr, and Trp. Several pathways required for amastigote survival are compartmentalized in glycosomes (green) or a single mitochondrion (red) (McConville et al., 2011).*

Abbreviations: AcCoA, acetyl-coenzyme A; Mann, mannogen oligosaccharides; PM, plasma membrane; PPP, pentose phosphate pathway; PV, parasitophorous vacuole; Triose-P, triose phosphate.

STRATEGIES AVAILABLE TO COMBAT DRUG RESISTANCE:

Improved methods to monitor drug resistance that determines either the

- (i) Parasite isolates' phenotypic sensitivity
- (ii) Molecular changes suggest changes in the drug target or mechanisms that affect the active drugs at the intraspecies level.

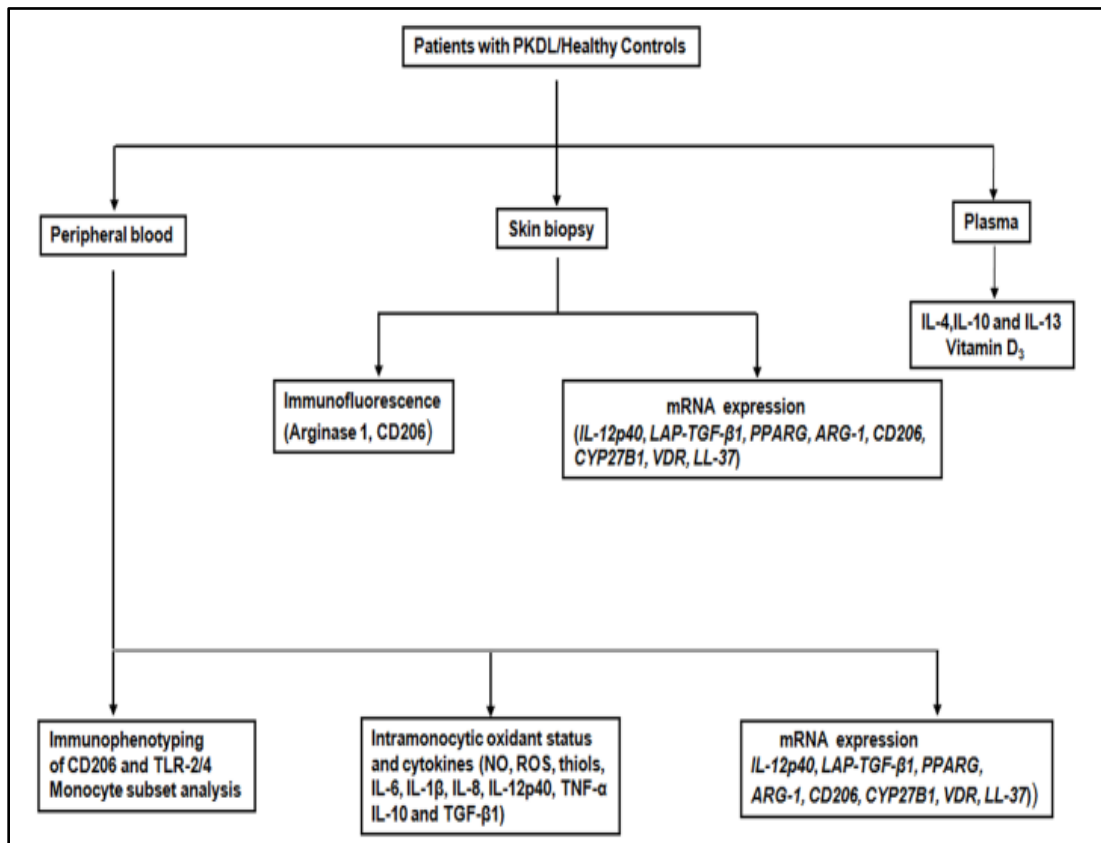


Figure 18. Immunological analysis of PKDL patients (Mukhopadhyay et al., 2015).

Non-invasive serological diagnostic techniques with high sensitivity and specificity, such as DAT, rK39, and Katex (urine dipstick), have made significant progress in leishmaniasis control. Drug combinations have been proven as an essential feature of antimicrobial treatment due to their design or use to (i) increase activity through the use of compounds with synergistic or additive activity, (ii) prevent the emergence of drug resistance, (iii) lower required doses, reducing the risk of toxic side effects and cost, or (iv) broaden the

spectrum of activity, such as the use of an antileishmanial with either an anti-inflammatory or immunomodulator in cutaneous leishmaniasis. Combinational drug therapy such as allopurinol plus sodium stibogluconate and paromomycin plus sodium stibogluconate has been used for the VL treatment.

MAJOR HOST IMMUNOLOGICAL RESPONSES IN LEISHMANIASIS:

We know that the host immune response plays a very significant role during the clinical manifestation of Leishmaniasis (Kumar et al., 2017). After the bite of the infected sandfly, the host triggers the innate immune response, which includes interaction with neutrophils and complement to protect the host in the early stages of infection, after which the parasite invades macrophages/monocytes.

Inside host macrophages, *Leishmania* induces phagocytosis without stimulating oxidative burst, followed by the production of anti-inflammatory cytokines such as IL-4 and IL-10 and transforming growth factor (TGF) confers survival of the parasites (Zijlstra et al., 2016). During *Leishmania* infection macrophages play a dual role. These cells are responsible for destruction as well as providing a safe place for internalized parasites. Therefore, the progression & regression of *Leishmania* infection depends on the host's immune responses and infecting *Leishmania* species. Typically, macrophages are at rest as M0 (naive macrophages). Still, the development of functionally distinct macrophage phenotypes depends on the signals provided by the microenvironment in which these cells are found (Tomiotto et al., 2018). Therefore, the M1(classically activated) macrophage activation depends on the signal provided by the Th1 lymphocyte, which secretes various cytokines, primarily interferon-gamma (IFN- γ) and Tumour Necrosis Factor-alpha (TNF- α), which are responsible for the elimination of intracellular parasites by activating the mechanism of oxidative burst. In contrast, Th2 lymphocytes activation, which produces IL-

4 and IL-13 cytokines, induces the M2 (alternatively activated) macrophages which are characterized by biosynthesis of polyamine via activation of the enzyme Arginase (arg) and production of urea and L-ornithine, which are beneficial for the survival of parasite inside the infected macrophages and progression of disease (Arango et al., 2014; Zanluqui et al., 2015). VL is characterized by a mixed Th1/Th2 immune response of adaptive immunity.

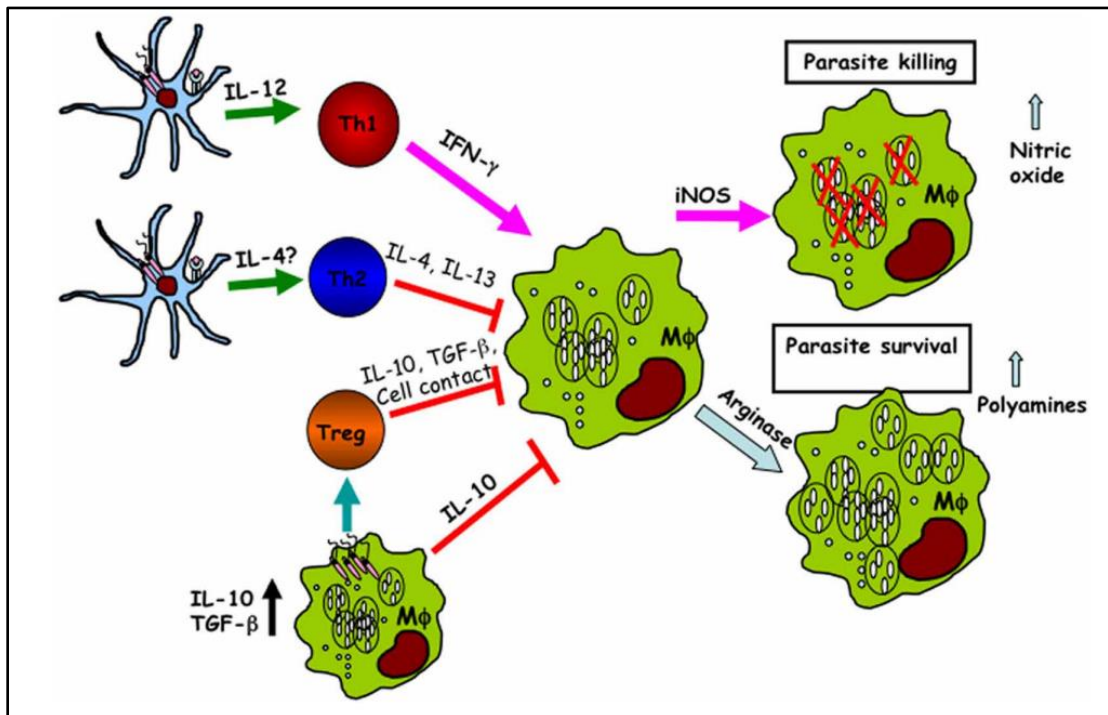


Figure 19. Schematic representation of *Leishmania* infection outcome shown by dendritic cells and macrophages (Liu and Uzonna, 2012).

The parasites change the phenotypic differentiation of antigen-experienced CD4⁺ T cells into the Th2 phenotype. Th2 cells, which release IL-4 and IL-10 cytokines, are responsible for parasite survival and disease progression, whereas disease resistance is conferred by Th1 cytokines like IL-2, IL-12, and interferon-gamma. Recently, IL-17, IL -21, IL -22, and IL -27 have been studied in disease resistance and susceptibility (Ansari et al., 2011; Pitta et al., 2009). In *Leishmania*, macrophages and dendritic cells phagocytose, leading to different functional outcomes. Infected dendritic cells produce IL-12, critical for developing IFN- γ -producing CD4⁺ Th1 cells. Classical activation on infected

macrophages starts with IFN- γ , upregulation of iNOS and production of nitric oxide (NO) and other free radicals that are important for intracellular parasite killing. Intracellular parasite proliferation is favoured by the Th2 cells, which produce IL-4 and IL-13, which leads to upregulation of arginase activity, alternative macrophage activation and the production of polyamines. In addition, some immunoregulatory cytokines, including IL-10 and TGF- β , are produced by naturally occurring regulatory T cells (Treg) and infected macrophages, which further deactivate the infected cells, leading to impaired parasite killing (Kumar et al., 2017; Mukhopadhyay et al., 2015). PKDL, a dermal sequel of VL, is characterized by macular, nodular hyperpigmentation or nodular lesions in the skin. In Indian PKDL, intralesional upregulation of IFN- γ and TNF- α are counterbalanced by TGF- β and IL-10 together with downregulated IFN γ R1 (Ganguly et al., 2010). Recently, in PKDL patients, it was demonstrated that M2 macrophage polarization is induced by an increased mRNA expression of classical M2 markers CD206, ARG1 and PPAR γ in the monocytes and lesion macrophages. PPAR γ is a nuclear hormone receptor family transcription factor that regulates macrophage metabolism by acting on the downstream pathway of STAT6 signalling, and activation of PPAR γ leads to M2 phenotype development (Mukhopadhyay et al., 2015). During *Leishmania* parasites infection, exposure to phosphatidylserine (PS), a major surface characteristic of apoptotic cells and macrophage engulfment of apoptotic cells leads to induction of PPAR γ (Kebir et al., 2010; Wanderley et al., 2006). The parasitized macrophages with activated PPAR γ promote the survival of parasites inside macrophages, whereas the PPAR γ blockade facilitates the removal of the parasites. Thus, harnessing PPAR γ by *Leishmania* parasites helps sustain the alternatively activated (M2) phenotype for survival (Tomiotto et al., 2018).

Peroxisome proliferator-activated receptors (PPARs), TACI (transmembrane activator and calcium modulator and cyclophilin ligand interactor) and mesenchymal stem cells (MSCs) participate in *Leishmania* M1 polarization of macrophage as well as the crotoxin treatment.

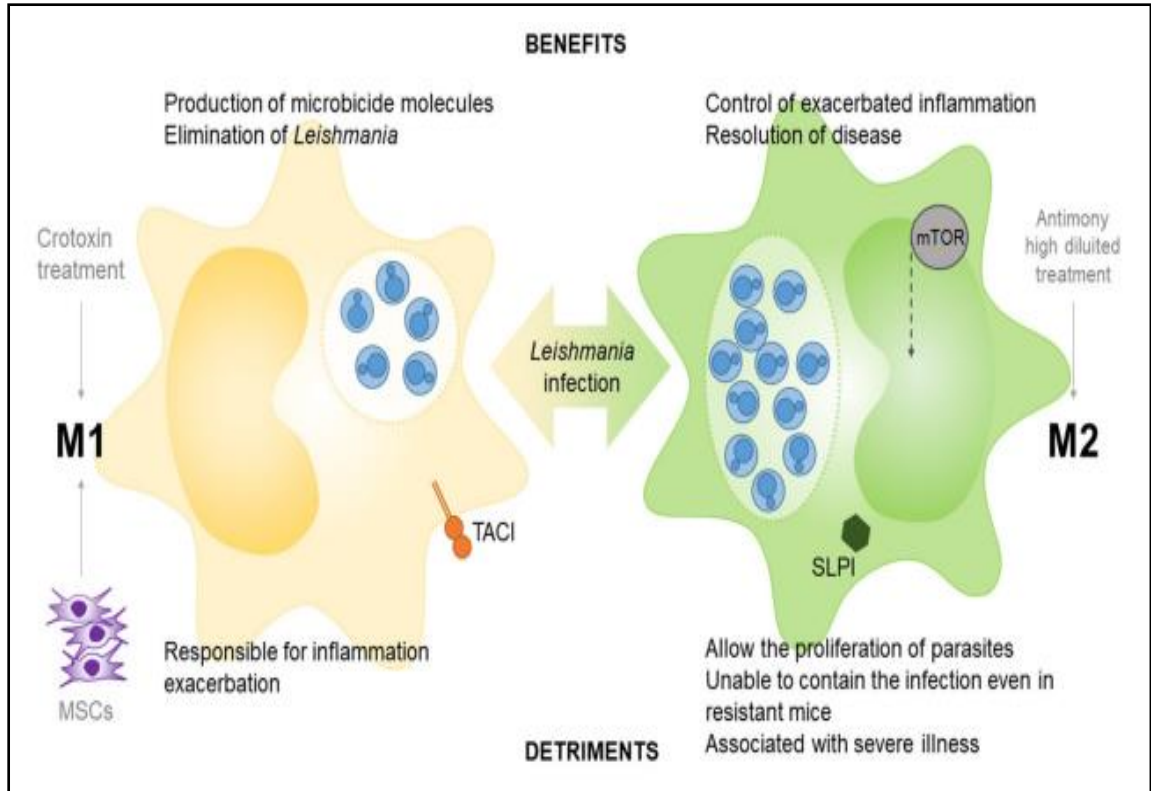


Figure 20. Role of macrophage polarization from M1 to M2 in *Leishmania* infection (Mukhopadhyay et al., 2015).

The participation of the M2 polarization of macrophages in the *Leishmania* model is observed in the treatment with high diluted antimony, the epithelial and myeloid-derived serine protease inhibitor (SLPI), mammalian target of rapamycin (mTOR). In the same geographical region, the association of *L. donovani* with the dermal and visceral area are previously reported (Ranasinghe et al., 2012), but the cause and subsequent development of viscerotropic *L. donovani* parasites to dermatropic PKDL are still unclear. Recently, it was reported that new approaches need to be explored as the current biomarkers in PKDL lesions are unsatisfactory (Zijlstra et al., 2019).

Whole Genome Sequencing (WGS) provides an insight to explore the area more by analysing whole genome Copy Number Variations (CNV) and SNPs change to develop a biomarker for PKDL.

FEEDBACK LOOP FACILITATES PARASITE PERSISTENCE:

- (1) Activated T cells, both CD4 and CD8, produce IFN γ , which in combination with IL-1 β drives IL-27 production by macrophages.
- (2) (2a) IL-27 inhibits the generation of potentially protective Th17 cells and
(2b) promotes together with IL-21 the generation of IL-10-secreting T cells.
- (3) IL-10 causes down-regulation of IL-12, MHC class II and co-stimulatory molecules on macrophages and renders macrophages unresponsive to activation signals such as IFN γ . Together this facilitates parasite replication and disease progression

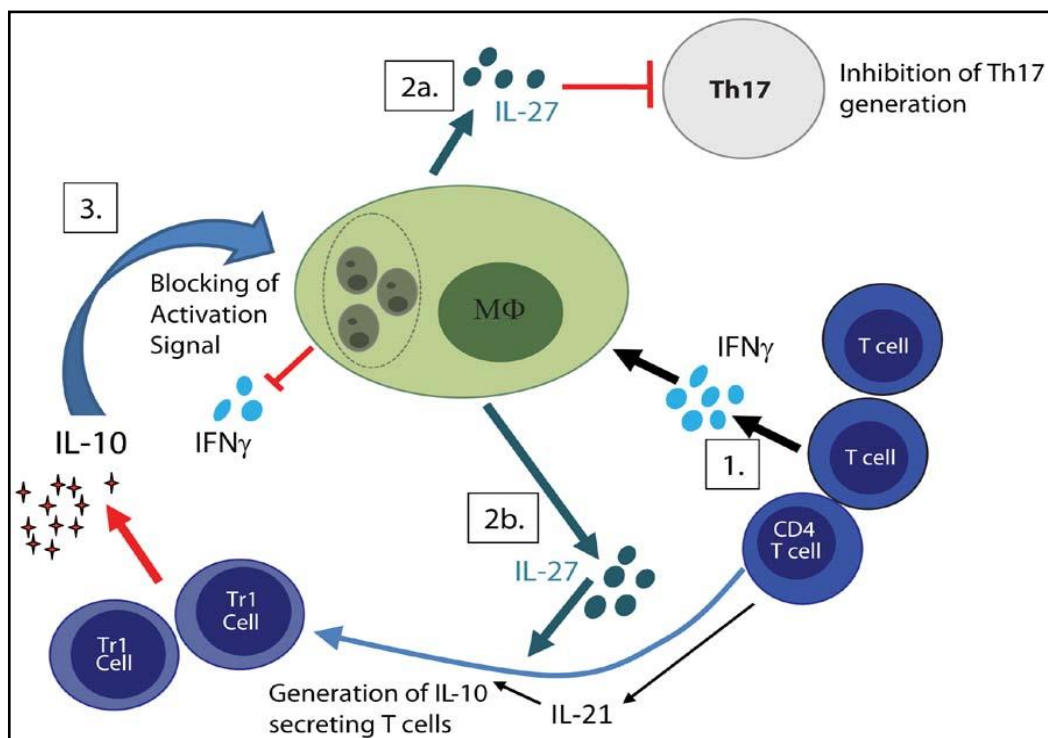


Figure 21. The regulatory feedback loop shows how this immunological loop favours parasite persistence in VL (Kumar et al., 2012).

IMMUNE REGULATION IN HUMAN MACROPHAGES

Human macrophages infected with *Leishmania spp.* are influenced by cytokines found in the skin near the infection site. IFN- γ and TNF, two inflammatory Type 1 cytokines, can cause macrophages to produce ROS synergistically, limiting *Leishmania* replication within the phagolysosome. Inflammatory MCL lesions had an increase in TH17 cells.

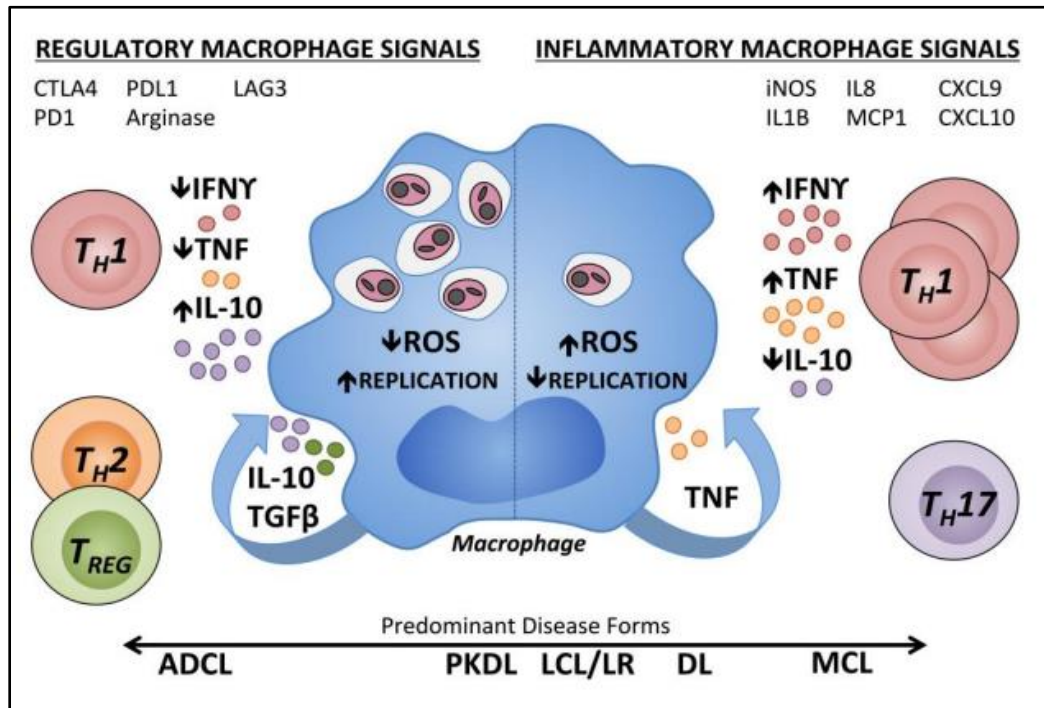


Figure 22. Schematic representation of immune regulation of human macrophages during *Leishmania* infection (Scorza et al., 2017). IFN γ : Interferon γ ; TNF: Tumor necrosis factor, IL10: Interleukin 10, TGF β : Transforming growth factor β , ROS: Reactive oxygen species, ADCL: Anergic diffuse cutaneous leishmaniasis; PKDL: Post Kala-Azar dermal leishmaniasis, LCL: Localized cutaneous leishmaniasis; LR: *Leishmania recidivans*, DL: Disseminated leishmaniasis; MCL: Mucocutaneous leishmaniasis; DTH: Delayed type 1 hypersensitivity; TREG: T regulatory cell; TH: T helper cell; MCP: Monocyte chemoattractant protein, CXCL: Chemokine (C-X-C motif) ligand; CTLA4: Cytotoxic T-lymphocyte associated protein 4; PD1: Programmed cell death protein 1; LAG3: Lymphocyte activation gene 3; PDL1: Programmed death ligand 1.

Transcripts associated with classical (type 1) macrophage activation (iNOS, IL1 β) and chemoattractant transcripts (IL8, MCP1, CXCL9, and CXCL10) have been measured in inflammatory CL lesions. In a regulatory environment (Figure 22 left), there are low levels of type 1 cytokines with a lack of microbicidal effectors. Type 2 cytokines include T-cell or macrophage-derived cytokines. Parasite proliferation results from enhanced polyamines, IL-10 and TGF β , which antagonize the effects of IFN- γ and TNF. The effects of IFN- γ are suppressed by T-REG cells. T cell exhaustion is associated with Inhibitory receptors on CD4 $^+$ or CD8 $^+$ T cells (CTLA4, PD1, and LAG3) and their counter-ligands (CD80, CD86, and PDL1). M2-type non-classical macrophage activation is associated with Arginase activity (Scorza et al., 2017).

PATHOPHYSIOLOGY OF VISCERAL LEISHMANIASIS:

The mice model has a distinct organ-specific pattern of parasite growth during visceral leishmaniasis (VL). Infection of the liver is defined by a rapid increase in parasite burden during the first four weeks of infection, followed by parasite elimination in six to eight weeks. The formation of a Th1-dominated granulomatous response characterised by strong IFN-gamma production by CD4 $^+$ and CD8 $^+$ T cells is endorsed by this self-curing mechanism in the liver. In contrast to the liver, the splenic infection has substantial implications, as seen by increased parasite burden, splenic microarchitecture disturbance, and decreased immune responses, all of which contribute to parasite persistence (Bankoti et al., 2002).

Spleen Infection:

The spleen is a site of chronic inflammation in the experimental VL model, characterised by the presence of a parasite. Chronic splenic infection is associated with splenomegaly and alterations in splenic microarchitecture, both of which affect the immune response (Stanley, 2007).

Dendritic cells (DCs) migrate from the marginal zone (MZ) of the spleen to the T-cell-rich area known as periarteriolar lymphoid sheaths after being exposed to the antigen (PALS). The DCs connect with the T cells in the PALS and this association stimulates an antigen-specific T cell response (Gorak et al., 1998). The splenic MZ is disrupted during chronic *L. donovani* infection due to increased TNF generation by macrophages. DCs (Ato et al., 2002) and T-cells fail to migrate to the PALS due to this disruption, which may result in reduced T-cell priming. The loss of the follicular DC network and the gp38+ fibroblastic reticular cell network is also linked to splenomegaly (Ato et al., 2002).

Liver Infection:

Another primary target organ in the VL experimental model is the liver. The development of granuloma formation, which is one of the main aspects of hepatic resistance, is linked to the progression of disease in the liver (Murray et al., 2001). Establishing a Th1-dominated granulomatous response is associated with the clearance of the infection in the liver. IL-12, which DCs secrete, triggers the Th1 response (Scharton-Kersten et al., 1995). Blocking IL-12 reduced both IFN- production and granuloma formation in the liver of infected mice, indicating that it is an essential cytokine in establishing protective immunity against *L. donovani* (Murray et al., 1997). CD8+ T lymphocytes are also crucial in the parasite's removal from the liver. Indeed, reducing CD8+ T cells prevents the granulomatous response from developing and leads to illness. In VL, hepatic resistance is linked to the production of reactive nitrogen and oxygen intermediates, both of which have been demonstrated to limit parasite growth during the early stages of infection.

MOLECULAR BIOLOGY TOOLS TO STUDY POPULATION GENETICS OF

GENUS: *Leishmania*

Clinical, biochemical, geographical, and epidemiological factors have been used to distinguish *Leishmania sp.* Several methods for differentiating *Leishmania* isolates have been available over the previous decade. For epidemiological purposes, isoenzyme analysis is the most useful taxonomic approach for these goals (Evans et al., 1984). The technique, as well as its theory, remain the gold standard for *Leishmania* taxonomy. The benefit is that it provides a stable marker for each species within a cluster of geographical isolates. Because it evaluates the genotype indirectly, nucleotide substitutions that do not change the amino acid composition and changes in the amino acid composition that do not modify the electrophoretic mobility may go undetected. Monoclonal antibodies are other techniques which are not highly sensitive due to their lack of specificity for isolates from widely dispersed foci (Schonian et al., 2000). Different DNA-based approaches are utilised at various levels to differentiate *Leishmania* isolates.

Multiple methods could be used at the species level, including PCR, which requires species-specific primers, and the study of Random Amplified Polymorphic DNA (RAPD), which uses random oligomers to amplify genomic DNA and so does not require any prior knowledge of the organism's genome. Analysis of kinetoplast and nuclear DNA, including Southern blot hybridisation with particular DNA probes, and analysis of restriction fragment length polymorphism (RFLP), where the substrate can be entire genomic DNA, kDNA, or PCR amplification products are exploited. Molecular karyotyping (including sequencing of species-specific genes) and DNA fingerprinting employing DNA probes are complementary to repetitive DNA sequences. (El Tai et al., 2000).

Previously, a gold standard method, Multilocus Enzyme Electrophoresis (MLEE), based on the MON system, was used for *Leishmania* identification at the species level (Rioux et al., 1990). Currently, to study the genetic diversity and phylogeny analysis of *Leishmania* parasites, the most reproducible technique, Multilocus Sequence Typing (MLST), is used (Zhang et al., 2013; Gelanew et al., 2014; Marlow et al., 2014). MLST approach is also used to detect interspecies and intraspecific variations (Miles et al., 2009, Akhoundi et al., 2016). The emergence of multiple bioinformatics tools and databases provides an effective approach to *Leishmania* clinical research, which has been seen as a magnificent contribution to computational science in the biomedical field. The introduction of PCR technique-based strategy has brought a paradigm shift in the field of biological sciences, especially in species identification, epidemiological surveillance of the disease and evolutionary biology etc. Sanger sequencing using the dideoxy chain termination method has been the gold standard for establishing a DNA sequence and analysing genomic variants for more than 30 years.

Genome-wide sequencing applications are now facilitated by high-throughput next-generation sequencing (NGS) systems, which have shifted genomics research towards DNA sequencing as a key molecular research tool over the last decade. The introduction of high-throughput sequencing has improved virtually every aspect of biotechnology. In terms of the number of base pairs sequenced in a massively parallel manner, these NGS platforms with different sequencing chemistries and instrumentation produce sequencing data several times higher than traditional ways of sequencing; thus, sequencing of a complete genome with desired depth now takes a few hours or days. In the case of microorganisms, massively parallel sequencing has revolutionised the identification of genomic variants, disease-associated markers, and molecular evolutionary studies.

The first generation of sequencing is clearly defined by the Sanger and Maxam–Gilbert procedures, which could sequence a few hundred base pairs (bp) at a time and be utilised for individual gene sequencing. Sanger sequencing would take a long time to sequence 3 billion base pairs of the human genome since it requires around 6 million 500-bp-long DNA fragments. The scientific community was introduced to next-generation sequencing methods, often known as second-generation sequencing. Second-generation sequencers could parallelise the sequencing reaction in a vast way, which was a significant improvement over the first. Roche 454, Illumina Solexa, and ABI SOLiD technologies are examples of these technologies.

NEXT-GENERATION SEQUENCER

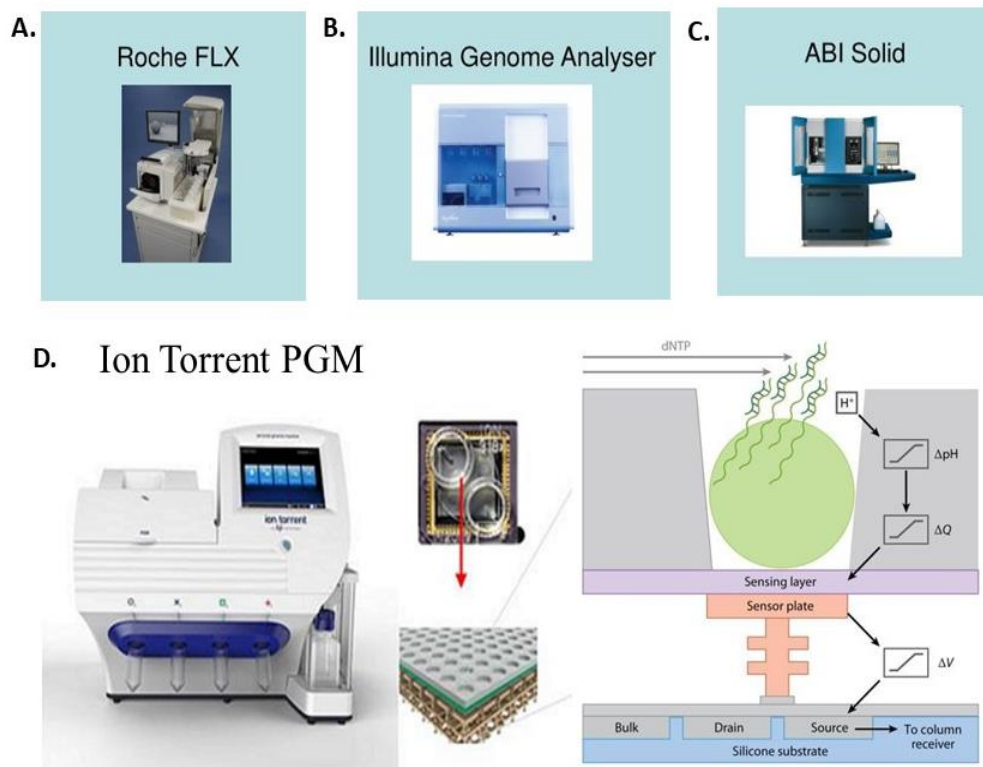
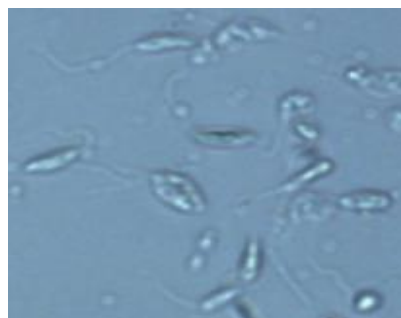


Figure 23. The Next-Generation Sequencing (NGS) systems. A. Roche; B.Illumina Solexa; C. ABI solid; D. Ion TorrentPersonal Genome Machine® (PGM).

Third-generation sequencing or next-next-generation sequencing refers to techniques that examine single molecules of DNA without first amplifying them using PCR, avoiding the biases introduced by PCR amplification and de-phasing. Third-generation sequencers, on the other hand, were created to make sequencing less expensive than second-generation sequencing. PacBio and Helicosbio are examples of third-generation sequencers (Gupta et al., 2014). On the other hand, Ion Torrent could fall in between the second and third generations. Although it is not a single-molecule sequencing approach, it differs from the second generation because it uses "post-light" sequencing rather than fluorescence detection. In the Ion Torrent technology, read lengths of 200 bp are generated, which are utilised to bridge gaps in the assembly created by other technologies. Although the read duration is substantially shorter than Roche-454 and PacBio, Ion Torrent can be a viable option due to lower costs. This technique's quick run time allows for repeated runs to generate more data in the same amount of time (Gupta et al., 2014).



AIM AND OBJECTIVES **OF THE PRESENT STUDY**



INTRODUCTION

A complex disease form leishmaniasis is associated with diverse clinical manifestations ranging from disfiguring cutaneous lesions to fatal visceral leishmaniasis in humans. Visceral Leishmaniasis (VL) or Kala-azar (KA) is a neglected tropical disease caused by vector-borne, obligate intracellular, protozoan parasite *Leishmania donovani* belonging to the genus *Leishmania* in the Old and New World. Recently, the association of *L. tropica* with the disease were also reported (Sacks et al., 1995; Khanra et al., 2011, 2012). Once thought eradicated, the disease VL or KA came back to the Indian subcontinent. The control strategy is besieged with the problem of 60%-70% drug unresponsiveness cases towards Sodium stibogluconate (SSG), the first-line drug. The second line drugs - Amphotericin B & Pentamidine etc. are costly & extremely toxic, making them unsuitable for the long course of treatment (Croft et al., 2006). Miltefosine (MIL) was introduced in 2003 to treat especially in SSG unresponsive KA cases. MIL unresponsive cases were reported within a few years (Arif et al., 2008; Khanra et al., 2017). Hence, the emergence of drug-resistant parasites has complicated the elimination strategies much more challenging. The evolution of parasites in response to chemotherapeutic agents and proliferation within different hosts is known to be influenced by their population genetic structure variation (Downing et al., 2011). In response to stress, *Leishmania* parasite can re-shape their genomes rapidly (Grunebast and Clos, 2020), implying that strategic manipulation of host cellular stress and drug responses is an integral phenomenon of host-parasite interaction. *Leishmania* genome plasticity indicates the importance of genomic variation as a survival strategy acquired by parasites in response to the changing cellular environmental conditions and drug pressure (Alam et al., 2009; Schonian et al., 2008).

Unfortunately, in India, after apparent cure 10% to 20% of VL patients develop Post Kala-azar Dermal Leishmaniasis (PKDL), manifested by symptomatically painless macular, hypopigmented skin patches and/or papulonodular skin lesions (Burza et al., 2018), which may not be very harmful, but the lesions can act as a reservoir in the transmission cycle of the parasites during sand flies bite (Zijlstra et al., 2017). The development of PKDL in VL patients is a gradual transition of the parasites from a visceral organ such as bone marrow, lymph node or spleen to the dermis of patients. These intermediate conditions where active VL is associated with PKDL are referred to as para-Kala-azar Dermal Leishmaniasis (para-KDL) (Zijlstra, 2019; Zijlstra et al., 2000; Zijlstra et al., 2003). Though VL associated with PKDL was rarely found in the Indian subcontinent (Zijlstra et al., 2003), nine such cases have recently been reported from Bihar, India (Kumar et al., 2016). Therefore, an association of VL along with PKDL cases requires further attention to understand the pathology, clinical diagnosis and treatment regime of these emerging incidences of para-KDL. Although several recent studies have been conducted using various molecular biological techniques to explore the population genetics of *Leishmania sp.*, only limited data is available in the clinical context on parasite population diversity and their genome modification during treatment (Schonian et al., 2008). Hence, the present study is undertaken to conduct the molecular biological characterization and genome-wide study of the collected clinical isolates of *Leishmania sp.* from Indian KA and para-KDL patients.

In the background of the above:

The present study has tried to contribute toward the development and evaluation of molecular biological methods to characterize both drug-sensitive and drug-resistant clinical isolates of KA along with para-KDL isolates (clinically diagnosed as PKDL) to establish differentiation, if any, at the genome level.

The aim of the present work is:

- To identify species of the collected clinical isolates from Indian Kala-azar and para-KDL patients.
- To characterize the group of housekeeping genes that globally differentiate the position of Indian clinical isolates of *Leishmania sp.* relative to other reference strains and check whether any intra or interspecies variations are present among Indian *Leishmania* isolates.
- To identify the Sequence Type (ST), that describes the phylodynamics and population genetic diversity of the *Leishmania sp.* Which helps to strengthen the epidemiological surveillance in the country.
- To identify the possible genetic variations and heterogeneity in *L. donovani* clinical isolates of Indian KA and para-KDL by Whole Genome Sequencing (WGS) and determine their potential roles in the observed drug resistance and PKDL manifestation.

The objective of the present work:

- Fresh isolation of clinical isolates from clinically diagnosed VL or Kala-azar (KA) and Para Kala-azar Dermal Leishmaniasis (para-KDL) cases directly from the field.
- Along with WHO reference strains, stabilization, maintenance and cryopreservation of the recent and all other previously collected clinical isolates.
- Characterization of the *Leishmania* parasites isolated from KA and para-KDL patients with the help of PCR-RFLP analysis of ITS, ITS1 and hsp70 amplicons on all the studied clinical isolates.
- Study the drug sensitivities of the clinical isolates for Sodium Stibogluconate (SSG) and Miltefosine (MIL) in the respective screening systems.

- Ascertain intraspecific and interspecific variations, if any, with the help of Multilocus Sequencing Typing (MLST) analysis of the fifteen housekeeping genes *aco*, *alat*, *enol*, *pgm*, *hgppt*, *spdsyn*, *asat*, *gpi*, *nh1*, *icd*, *mpi*, *fh*, *pgd*, *g6pdh* and *pmm* respectively.
- Determination of Nucleotide sequence and genetic diversity indices analysis.
- Determination of Protein sequence, Allele profile and Sequence Type (ST).
- Assessment of haplotype and construction of dendrogram for phylogenetic analysis.
- Estimation of the ratio of non-synonymous (dN) to synonymous (dS) substitution per nucleotide site by graphical representation.
- Clonal population selection of *Leishmania donovani* isolates for Whole Genome Sequencing.
- Genomic DNA isolation and library preparation of the *Leishmania* parasites isolated from KA and para-KDL patients.
- Genome-wide comparison by whole genome sequencing analysis of the *Leishmania* parasites genome of both drug-sensitive and drug-resistant isolates of KA and para-KDL isolates.
- Estimation of some values, gene clustering and pathway analyses by using different software and programming.
- Investigation of the possible association of aneuploidy with the development of visceral to the dermal tropism of the parasite and their observed drug resistance in Indian VL patients.
- Principal Component Analysis (PCA) of the type of mutation among the genes in the individual strains to evaluate the significance of genetic differences across the strains.

- Identification of the unique set of mutated genes in MIL-resistant VL (VL-MIL-R) and SSG-sensitive para-KDL (para-KDL-SSG-S) strains which encode previously known proteins and assessment of chromosome Copy Numbers Variations (CNV).
- Prediction of the secondary structure of the mutated transporter and surface proteins in para-KDL strains followed by an analysis of the mutational domain using different bioinformatic tools.



MATERIALS & METHODS

Reagents

The tissue culture chemicals like Bovine serum albumin (BSA), HEPES, medium M199, medium RPMI-1640, 2-mercaptoethanol (2-ME), Sodium bicarbonate, streptomycin, penicillin and other reagents like 3,3',5,5'-Tetramethylbenzidine (TMB), EDTA, Bromophenol blue, Giemsa, Proteinase K, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide (MTT) were obtained from Sigma Chemical (St. Louis, MO, USA). Tween-20, Cell culture DMSO (Dimethyl Sulfoxide), HPLC grade Methanol, Silver nitrate, Formaldehyde solution and Ethanol were obtained from MERCK, USA. Schneider's Drosophila medium, Fetal Bovine Serum (FBS), was procured from Gibco, USA. Primers were procured from Eurofins Scientific GmbH, Germany, and some gene-specific primers were procured from Sigma chemical. DNA Ladder was obtained from G-Bioscience. Taq DNA polymerase, dNTP, and PCR buffer were obtained from Genet Bio, USA. Agarose was procured from Lonza Rockland, USA. Restriction enzymes (FspI, MnaII, HaeIII, HinfI, BstDEI, Bst2UI, BstSCI, HpaI, FauNDI, MseI and AcoI) and specific buffer were procured from Bioenzyme, USA. Buffer saturated phenol, 5X DNA buffer. DNA library by OneTouch™ 2 System (Thermo-Fisher), Ion Proton Semiconductor Sequencer using Ion PI™ Chip V3 and Ion PI™ Hi-Q™ Sequencing 200 reagents were used to sequence enriched library. Qubit DNA HS kit was used for quantification and their quality check was carried out by High sensitivity DNA reagent kit in Agilent 2100 libraries Bioanalyzer. Ion Proton Semiconductor Sequencer using Ion PI™ Chip V3 and Ion PI™ Hi-Q™ Sequencing 200 reagents used for enriched library sequencing. Miltefosine and Sodium Stibo Gluconate was a kind gift from Prof. Syamal Roy, Professor of Biological Sciences, AcSIR & Chief Scientist, CSIR- Indian Institute of Chemical Biology, Kolkata, India.

1. COLLECTIONS AND MAINTENANCE OF CLINICAL ISOLATES

1.1 Ethical Statement

Collection of Bone marrow or splenic aspirates from Kala-azar and Para kala-azar dermal leishmaniasis patients was approved by the Ethical Committee of the Calcutta National Medical College, Kolkata, India. Written consent was obtained from each patient or guardian (for a minor patient) before the study began.

1.2 Splenic Aspirates

Firstly, palpate the spleen and outline its margins on the patient's abdomen with a pen. The skin was penetrated with the 21-gauge (0.8-mm) needle attached to the 5-ml syringe, midway between the edges of the spleen, 2-4 cm below the costal margin at an angle of 45° to the abdominal wall. Then syringe plunger was pulled back to approximately the 1-ml mark to apply suction. With a quick in-and-out movement, the needle was pushed into the spleen to the full needle depth and then withdrawn completely, maintaining suction throughout. These procedures were conducted by the expert physicians of the hospital/s at the time the clinical isolates were collected. Prof. (Dr) Anjan K Das, Professor, Department of Pathology, Cooch Behar Government Medical College, West Bengal, India, is one of our collaborators who has helped us in collecting the isolates.

1.3 Bone Marrow Aspirates

A particular type of needle was used to puncture the sternum or iliac crest and marrow were aspirated using a syringe containing an anticoagulant such as heparin or EDTA. Only two or three drops were given in each culture tube containing Schneider's *Drosophila* medium and 20% Fetal Bovine Serum (FBS).

These procedures were conducted by the expert physicians of the hospital/s at the time the clinical isolates were collected. Prof. (Dr) Anjan K Das, Professor, Department of Pathology, Cooch Behar Government Medical College, West Bengal, India, is one of our collaborators who has helped us in collecting the isolates.

1.4 Parasite Culture and Maintenance

Giemsa staining procedure was used to confirm the presence of amastigotes in the bone marrow or spleen of the patients. Amastigote to promastigote transformation of the parasites occurred in culture in medium M199 supplemented with 20% heat-inactivated FBS and 100 IU/mL of penicillin and 100 mg/mL of streptomycin in a 22° C. Promastigotes were maintained in media M199 supplemented with 10 % heat-inactivated FBS and 100 IU/mL of penicillin and 100 mg/mL of streptomycin in a 22° C.

1.5 Cryopreservation

Preservation of clinical isolates after initial isolation was necessary for future clinical isolates depositories. *Leishmania* Promastigotes can be stored in low-temperature mechanical freezers (-80°C) or in liquid nitrogen containers (-196°C) in 2 ml cryovial (plastic freezing vial with air-tight screw caps).

During cryopreservation, the cell suspension *Leishmania* parasite culture was pellet down by centrifugation (3000 rpm at room temperature), and then 90% FBS plus 10% DMSO were added to resuspend the pellet. The resuspended cells are transferred into the cryovial after that, kept the vial at -20° C overnight then placed frozen cryovial at -80°C, for long-term preservation, transferred the vial from -80°C to -196°C. To avoid cell cold shock, a gradual slowdown in temperature would allow. Parasites' minimum concentration was 10⁶/ml.

2. ISOLATION OF TOTAL NUCLEIC ACID

Total nucleic acids from all clinical isolates (n=15) and WHO reference strain for *Leishmania donovani* (DD8) and *L. tropica* (K27) were prepared as per standard protocol (Bhattacharya et al. 1993). In short, *Leishmania* promastigotes were suspended in Phosphate Buffered Saline (PBS, 0.02 M, pH 7.4). The cell pellets from different strains were obtained by centrifugation at 5000 rpm for 10 min at 4°C. Pellets were washed in Phosphate Buffered Saline (PBS) and resuspended in 400 µl of NET buffer (50 mM NaCl; 10 mM EDTA; 50 mM Tris- HCl pH 7.4). DNA was prepared from the samples by digestion with 100 µg/ml proteinase K in the presence of 1% SDS overnight at 37°C, followed by phenol-chloroform extraction and ethanol precipitation. The DNA was dissolved in Tris EDTA (TE) buffer and stored at -20°C till use.

3. STRAINS/ ISOLATES USED IN THE PRESENT STUDY

3.1 Reference Strains:

L. donovani strain AG83 (MHOM/IN/83/AG83) was originally isolated from a treated Kala-azar patient at the Calcutta School of Tropical Medicine (Manna et al., 2005) and the World Health Organization (WHO) *Leishmania donovani* (DD8) and *L. tropica* (K27) were used in this study as a reference strains. The annotated *L. donovani* genome for BPK282/0cl4 is also used as a reference genome for Ion Torrent Suite Version 4.2, which is available in the EMBL Nucleotide Sequence Database. Information related to reference strain/isolate is mentioned in Table 6.

Table 6: Source and a brief history of reference strains of *Leishmania sp.* used for the present study.

Strains/Isolate	Origin / Year of collection	Species	Reference No.	Comments
AG83	INDIA/1983	<i>Leishmania donovani</i>	MHOM/IN/83/AG83	Obtained from Dr Syamal Roy IICB Kolkata
DD8	INDIA/1980	<i>Leishmania donovani</i>	MHOM/IN/80/DD8	WHO Reference strain
K27	SUDAN/1974	<i>Leishmania tropica</i>	MHOM/SU/74/K27	WHO Reference strain
BPK282/0cl4	Nepal/-	<i>Leishmania donovani</i>	BPK282A1	EMBL Nucleotide Sequence Database

3.2 Clinical Isolates:

A total of fifteen (n=15) clinical isolates are used in the present study. Information related to the patient's age, sex, clinical status, collection year and country of the isolates from visceral leishmaniasis (VL) or Kala-azar (KA) and para-kala-azar dermal leishmaniasis (para-KDL) patients were described in Table 7.

Table 7: Source and a brief history of clinical isolates of *Leishmania sp.*

Study Code	Age	Sex	Country	Year of Collection	Clinical Status
H1	10 years	Female	India	2014	VL
A2	3 yrs 6 months	Female	India	2014	VL
P2	32 years	Female	India	2013	VL
P6	20 years	Female	India	2014	para-KDL
KK	7 years	Female	India	2016	VL
BK	8 years	Female	India	2016	VL
T2	2 yrs 3 months	Male	India	2010	VL
T8	11 years	Male	India	2012	VL
T9	5 years	Female	India	2012	VL
T10	35 years	Male	India	2014	para-KDL
T3	11 years	Male	India	2010	VL
T4	56 years	Male	India	2010	VL
T7	17 years	Female	India	2010	VL
M1	10 years	Female	India	2013	para-KDL
T5	12 years	Male	India	2010	VL

4. PCR AMPLIFICATION OF THE DNA FROM CLINICAL ISOLATES OF KA, para-KDL AND REFERENCE STRAINS

- The internal transcribed spacer (ITS) & ITS1 regions were individually amplified from DNA (El Tai et al., 2000; Schonian et al., 2003) of all samples with the following primers:

LITSR (5'CTGGATCATTTTCCGATG-3') and

LITSV(5'ACACTCAGGTCTGTAAAC-3') &

LITSR(5'-CTGGATCATTTTCCGATG-3') and

L5.8S (5'-TGATACCACTTATCGCACTT3').

- As previously reported, the hsp70 PCR amplification was performed with the primers (Montalvo et al., 2010).

HSP70sen (5'GACGGTGCCTGCCTACTTCAA-3') and

HSP70ant (5'CCGCCCATGCTCTGGTACATC-3')

4.1 Internal Transcribed Spacer (ITS) And ITS1 PCR Amplification:

PCR reactions were performed in volumes of 50 µl. 200ng DNA samples were added to a PCR reaction mixture containing 200µM of each dNTP, 5 µl 10x PCR buffer, 1-2 units of *Taq* polymerase and 25 pmol of each primer. After initial denaturation of 5 min at 95°C, 35 PCR cycle was carried out using the following cycle parameter: 30 sec at 95°C, 30 sec at 46°C for primer pair LISTR/LITSV or 30 sec at 48°C for primer pair LITSR/L5.8S, 1 min at 72°C. At the end of the PCR cycle, all tubes were kept for 6 minutes at 72°C for an extension to go completion. The products were electrophoresed in 1.2% agarose gel in TAE buffer and analysed with EtBr by the Bio-Rad Gel Doc system.

4.2 Heat Shock Protein (hsp70) PCR Amplification:

The HSP 70 PCR amplification reactions were performed in a 50 µl PCR reaction mixture containing 200ng DNA samples, 200µM of each dNTP, 5 µl 10x PCR buffer, 2 units of Taq polymerase and each primer were added at a concentration of 10 pmol/µl to the final reaction mixture. After the initial denaturation of 5 min at 95°C, PCR was carried out for 33 cycles using the following parameters: 40 sec at 95°C, 40 sec at 52°C, and 1 min at 72°C. At the end of the PCR cycle, all tubes were kept for 8 minutes at 72°C for an extension to go completion. The products were analysed by 1.2% agarose gel in TAE buffer and band visualised with EtBr by Bio-Rad Gel Doc system.

5. Hae III RESTRICTION ANALYSIS OF hsp70 PCR PRODUCT

The hsp70 PCR products ethanol precipitation was followed by reconstituted in 15 µl of autoclaved distilled water. The digestion was performed with 10 µl of amplicon solution in 10 µl of the optimal buffer using 5U HaeIII. Reactions were incubated for 4 hours at 37°C and analysed by electrophoresis in a 3% agarose gel.

6. RESTRICTION ANALYSIS OF ITS AND ITS1 PCR PRODUCT

The amplified ITS1 region was digested using restriction enzymes Hae III, Mse I and the amplified ITS region was digested using restriction enzymes Fsp I, Hinf I, Mse I, BstDE I, BstSC I, Bst2U I, Aco I, FauND I, Hpa I and Mna1. Briefly, the protocol is as follows, for diagnostic samples, 10 µl of the DNA were restricted by the addition of 5 units of enzyme and 2 µl of the corresponding 10x buffer and 2 µl of distilled water and incubated at 37°C for 4 hours. The Restriction products were analyzed by 2% agarose gel electrophoresis in TAE buffer.

7. MARKER SELECTION

Marker selection for coding gene fragments for fifteen metabolic enzymes was based on the presence of variable sequence diversity, as previously reported (Marco et al., 2015; Mauricio et al., 2006; Zemanova et al., 2007). The list of primers, location at chromosome (Chr.) and the name of gene markers used in the MLST study are given in Table 8.

Table 8: List of primers and gene markers used in the MLST scheme

Target gene	Enzyme entry	Chr. location	Gene length (bp)	Amplicon size (bp)	Primer sequence 5'-3'
Aconitase (aco)	EC 2.5.1.16	18	2691	550	Fw: CAAGTTCCTGRCGTCT CTGC Rv: GAGTCCGGGTATAGC AKCCC
Alanine aminotransferase (alat)	EC 2.6.1.21	12	1494	600	Fw: GTGTGCATCAACCCM GGGAA Rv: CGTTCAGCTCCTCGTT CCGC
Hypoxanthine -guanine phosphoglucomutase (hgprt)	EC 2.4.2.8	21	636	412	Fw: GCTCTACCTGCTGTGC GTGC Rv: ATCGCGCAGCTCGCG RTACG
Spermidine synthase (spdsyn)	EC 2.5.1.16	4	903	400	Fw: CGAACCTGTCGCTGA CGTG Rv: GAYTCGCCCTGGTTGC ACAC
Nucleoside hydrolase (nh1)	EC 3.2.2.2	18	1100	1010	Fw: CTTGCTTACGCCGAG ATAC Rv: GAAAAAAAAGACGCT TCACACAAGC

Target gene	Enzyme entry	Chr. location	Gene length (bp)	Amplicon size (bp)	Primer sequence 5'-3'
6Phosphogluconate dehydrogenase (pgd)	EC 1.1.1.44	35	1440	1100	Fw: GAACGAATCCCTTATT CTCYATG Rv: GGAACCGGTTGAGCG GC
Phosphoglucomutase (pgm)	EC 5.4.2.2	21	1770	500	Fw: CAGAGAAGCTGACGT CCCAG Rv: GACGGGTTACGAAG AAGCG
Phosphomannomutase (pmm)	EC 5.4.2.8	35	744	536	Fw: TTCAAGCTTGGCGTC GTCGG Rv: TAATCGTTRCCGCCCT CTGA
Mannose phosphate isomerase (mpi)	EC 5.3.1.8	32	1287	1050	Fw: ATGTCTGAGCTCGTA AAGCT Rv: CTACCTGTCGCTCAAG TC
Glucose-6-phosphate dehydrogenase (g6pdh)	EC 1.1.1.49	34	1686	1100	Fw: ATGTCGGAAGAGCAG TCT Rv: TCACAGCTTATTCGAG GGAA
Enolase (enol)	EC 4.2.1.11	14	1290	450	Fw: GCTGCCGATCCTGAT GGAGG Rv: ACCCGTTCTCCATGCA CAGC
Aspartate aminotransferase (asat)	EC 2.6.1.1	35	1169	>1100	Fw: ACGAGCGCCGTCCGY AA Rv: TTCCYMCATCCACCAA GC

Target gene	Enzyme entry	Chr. location	Gene length (bp)	Amplicon size (bp)	Primer sequence 5'-3'
Fumarate hydratase (fh)	EC 4.2.1.2	29	1707	>1100	Fw: AGCGTCTTGTGTTCC CA Rv: GAGCCCGTGAAGGA GGC
Isocitrate dehydrogenase (icd)	EC 1.1.1.42	10	1278	1100	Fw: ATGTTCCGCCATGTTT CGGC Rv: TTACGCGCTCATCGCC TT
Glucose-6-phosphate isomerase (gpi)	EC 5.3.1.9	12	1818	>1100	Fw: GAATCCCTTTTCAAGA TGAGCGATTAT Rv: CCCCTGAGAGGCAAT CACAG

8. PCR AMPLIFICATION AND SEQUENCING OF THE DNA OF CLINICAL ISOLATES OF KA AND para-KDL WITH MLST PRIMERS

Amplification was performed in a thermal cycler with the protocol: The PCR amplification was performed in a 50 µl reaction mixture containing 0.5µM of each forward and reverse primer, 0.2Mm deoxyribonucleotide triphosphate, 0.5U Taq polymerase and 125ng DNA. Amplification was performed with an initial denaturation at 96°C for 5min, followed by 35 cycles with 96°C for 1 min, annealing varies 55-60°C for 1 min, 72°C for 1 min and final extension at 72°C for 10 min. five microliters of each amplified product were electrophoresed in a 1% metaphor agarose gel (LONZA Bioscience) in TAE buffer and analysed with EtBr and molecular size marker (100 bp ladder, GeNei) to confirm PCR amplicon size and the remaining 45 µl of the amplified product was purified by AxyPrep PCR clean up kit (Lot#:04113KB1).

Purified PCR products were subjected to Sanger sequencing (Applied Biosystems, Eurofins Scientific GmbH, Germany). Sequencing results were analysed with different bioinformatics software to study the genetic diversity parameters, Sequence Typing (ST) and construction of phylogenetic relationships. The list of software and studied areas is given in table 13.

9. NUCLEOTIDE SEQUENCE AND GENETIC DIVERSITY INDICES ANALYSIS

The consensus sequence assembly of each gene was aligned and trimmed, the shape parameter of the gamma distribution was evaluated using eight categories GTR model (Lanave et al., 1984), and uniform rates among sites were observed with the transition/transversion bias (R), using MEGA X (Kumar et al., 2018). The ratio of non-synonymous and synonymous substitution rates (dN/dS) was calculated using the SNAP software database <http://www.hiv.lanl.gov>, a program based on the Nei and Gojobori 1986 method (Korber et al., 2000). DNAsp v6 software (Rozas et al., 2017) was used to calculate the genetic diversity indices (number of polymorphic sites S, total number of mutations Eta, nucleotide diversity Pi, number of haplotype h, haplotype diversity Hd, theta per site) for each marker genes.

10. DETERMINATION OF ALLELE PROFILE AND SEQUENCE TYPE (ST):

Based on the difference in single nucleotide polymorphism (SNP) position for each of the fifteen loci, allele profiles were generated, which are used to assign sequence type (ST) for each isolate. The alleles for each locus and STs for each isolate were assigned using MLSTest software based on genotypic diversity according to the allelic profiles (Tomasini et al., 2013).

11. PROTEIN SEQUENCE ANALYSIS

Molecular mass (m) and overall charge indicated by the inferred protein sequence's theoretical isoelectric point (pI) was resolved through the <http://www.expasy.org> program.

12. ANALYSIS OF HAPLOTYPE

Haplotype analysis was performed by DNAsp v6 software (Rozas et al., 2017) using the PHASE algorithm, which automatically assigns the haplotype number to each unique sequence. Then haplotype data were utilised in DNAsp to calculate haplotype diversity and haplotype networks were constructed by PopART software (Leigh et al., 2015) using the TCS algorithm based on its implementation of statistical parsimony (Crandall et al., 2000).

13. PHYLOGENETIC ANALYSIS

The phylogenetic relationship among concatenated sequences was performed by using Neighbour Joining (NJ) method tested with 1000 bootstrap replicates in MLSTest software (Tomasini et al., 2013). A phylogenetic tree of every single locus was constructed using the maximum likelihood method with 1000 bootstrap replicates based on the HKY model (Hasegawa et al., 1985) with Nearest-Neighbor-Interchange (NNI) algorithm using MEGA X (Kumar et al., 2018).

14. WHOLE GENOME SEQUENCING (WGS):

14.1 The Isolates Used In The Study

Ten clinical isolates from KA and para-KDL patients (Splénomegaly along with hypopigmented skin lesion) of India (Table 4), including six from our previous studies (Khanra et al., 2012; Khanra et al., 2017; Khanra et al., 2016) were analysed in the present work. Details of the isolates used in the present study are given in Table 9.

The occurrence of amastigotes in the bone marrow of the patients was established by Giemsa staining and 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 (Resistance, R; Sensitive, S) 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% CO₂.

14.2 Parasite Cloning

All promastigotes used in the present study were cloned as described elsewhere (Bhattacharyya et al., 2002; Pal et al., 2001). Briefly, exponentially growing cells were serially diluted in medium M-199 supplemented with 20% fetal bovine serum, 50% conditioned medium (Sterilized culture supernatants from exponentially growing promastigotes in M199 media containing 20% FBS), 50 mg/ml gentamycin, 100 U/ml penicillin and 100 mg/ml streptomycin. The cell suspension of 100 µl aliquots was added to each well of 96 well tissue culture plates (Tarson, USA) and incubated at 22°C for 3 weeks.

14.3 Genomic DNA (gDNA) Isolation and Genotyping

The gDNA is isolated from the cloned promastigote strains using a QIAamp DNA isolation kit (Qiagen, cat. no. 19413). Qubit assay protocol was used to measure the concentration of DNA. And then, genotype characterisation is performed by ITS1 sequencing and PCR-RFLP methods (Alam et al., 2009; Khanra et al., 2012; Khanra et al., 2016)

Table 9: Details of the isolates used in the WGS study.

Isolate Study Code	WHO Reference Code	Diseases Type	Method Used for genotyping	Drug Susceptibility R=Resistant S=Sensitive		SSG EC50 ± SD (µg/mL)	MIL EC50 ± SD (µM)
				SSG	MIL		
R1	MHOM/IN/2014/H1	VL	PCR-RFLP & ITS1 Sequencing	R	S	15.66±3.01	2.279±0.24
R2	MHOM/IN/2012/T8	VL	PCR-RFLP	R	S	18.77±4.84	4.29±0.59
R3	MHOM/IN/2013/P2	VL	PCR-RFLP	R	S	27.71±3.01	3.66±0.47
S1	MHOM/IN/2010/T7	VL	PCR-RFLP	S	S	3.79±1.35	2.47±0.13
S2	MHOM/IN/2010/T3	VL	PCR-RFLP	S	S	4.5±0.542	2.39±0.025
S3	MHOM/IN/2014/A2	VL	PCR-RFLP & ITS1 Sequencing	S	S	3.97±0.21	4.57±0.68
K1	MHOM/IN/2013/M1	para-KDL	PCR-RFLP	S	S	3.9±0.24	3.88±0.33
K2	MHOM/IN/2014/P6	para-KDL	PCR-RFLP & ITS1 Sequencing	S	S	4.04±0.48	3.39±0.44
K3	MHOM/IN/2014/T10	para-KDL	PCR-RFLP & ITS1 Sequencing	S	S	3.49±1.14	3.318±0.628
T9	MHOM/IN/2012/T9	VL	PCR-RFLP	S	R	4.69±0.302	13.47±0.872

14.4 Genomic DNA Library Preparation

High-quality RNA free DNA libraries were prepared using Ion Xpress™ Plus Fragment Library Kit (Thermo-Fisher). The protocol is as follows:

14.4.1 gDNA fragmentation (Ion Shear™ Plus Reagents)

Enzymatic fragmentation of gDNA into blunt-ended fragments with Ion Shear™ Plus Reagents. This method is used to construct 200-300-base-read libraries for the Ion Proton™ Sequencer. No end-repair is required and the fragmented gDNA is ready for adapter ligation. gDNA (120 ng) was taken and adjusting the concentration as necessary (For 1 µg input: Prepare 10 µL at 100 ng/ µL in Nuclease-free Water or 10 mM Tris, pH 7.5-8.5). The Ion Shear™ Plus 10X Reaction Buffer and the Ion Shear™ Plus Enzyme Mix II were added to the sample and adjusted to the total reaction volume is 50 µL and then incubated at 37° C for 5 mins.

14.4.2 Purification of the fragmented gDNA

99 µL of Agencourt® AMPure® XP Reagent (1.8X sample volume) was added to the sheared DNA samples, after thoroughly mixing (pulse-spin) the bead suspension with the DNA, and then incubated the mixture at room temperature for 5 minutes. Again pulse-spin and place the tube in the DynaMag™-2 magnetic rack until the solution is clear of brown tint. Then the supernatant was discarded without disturbing the tint.

500 µL of freshly prepared 70% ethanol was added, and after ethanol wash, 25 µL of Low TE was directly added to the pellet to disperse the beads and then placed the tube in the magnetic rack until the solution clears (for at least 1 minute). Finally, the supernatant containing the eluted DNA was transferred to a new PCR tube without disturbing the pellet.

14.4.3 Adapters Ligation and nick-repair

Reaction Setup for Barcoded Libraries	
Component	Volume by Input gDNA (120 ng)
gDNA	~25 μ L
10X Ligase Buffer	10 μ L
Ion P1 Adapter	2 μ L
Ion Xpress™ Barcode X	2 μ L
dNTP Mix	2 μ L
Nuclease-free Water	49 μ L
DNA Ligase	2 μ L
Nick Repair Polymerase	8 μ L
Total	100 μL

In a 0.2-mL PCR tube, the reagents as indicated above were combined. Then, the reaction mixture was incubated at 25°C for 15 min, 72°C for 5 min, and held at 4°C in a thermal cycler for barcoded libraries.

14.4.4 Purification of the adapter-ligated and nick-repaired gDNA

1.2X sample volume of Agencourt® AMPure® XP Reagent was added to the sheared gDNA samples and the bead suspension with the gDNA was mixed thoroughly by pulse-spin followed by 5 minutes incubation at room temperature. Again pulse-spin and place the sample tubes in the DynaMag™-2 magnetic rack until the solution gets cleared of brown tint. Then, the supernatant was discarded without disturbing the tint. 500 μ L of freshly prepared 70% ethanol was added, and after ethanol wash, 25 μ L of Low TE was directly added to the pellet to disperse the beads and then placed the tube in the magnetic rack until

the solution clears (for at least 1 minute). Finally, the eluted gDNA supernatant was transferred to a new PCR tube without disturbing the pellet.

14.4.5 Size-selection of the unamplified 200-base-read libraries (330-bp target peak)

Size-selection of the unamplified libraries was performed by E-Gel® iBase™ unit.

The amber filter was placed over the E-Gel® iBase™ unit. Run size and run time to the reference line were 2% program and 12–14 min in the E-Gel® Size Select™ agarose gels, respectively.

14.4.6 Platinum PCR and purification of the unamplified libraries

The procedure was as follows, total reaction mixtures of 130 µL contained 100 µL Platinum® PCR SuperMix High Fidelity, 5µL Library Amplification Primer Mix and 25 µL Unamplified library. Amplification was performed in a thermal cycler with the protocol: 10 total cycles using the program, 5 min at 95°C, 15 sec at 95°C, and 15 sec at 58°C. At the end of the PCR cycle, all tubes were kept for 1 min at 70°C for an extension to go to completion. The products after the PCR run were purified by a 1.2X sample volume of Agencourt® AMPure® XP Reagent, as mentioned in section **14.4.4**.

14.5 Library Quantification and Sequencing

A Qubit DNA HS kit quantified the gDNA libraries and their quality check was carried out by High sensitivity DNA reagent kit in Agilent 2100 Bioanalyzer. Enriched and template-positive Ion PI™ Ion Sphere™ Particles (ISPs) containing clonally amplified gDNA up to 200 bp were prepared from the qualified DNA library by OneTouch™ 2 System (ThermoFisher). The enriched library was sequenced in Ion Proton Semiconductor Sequencer using Ion PI™ Chip V3 and Ion PI™ Hi-Q™ Sequencing 200 reagents. The mean depth coverage of 117-fold was obtained for each strain in WGS; on average, 97% of the sequence reads could be mapped to the reference genome (Table 10).

Table 10: WGS mapped reads values.

Samples	Total reads	Mapped reads	% of mapped reads	Mean Depth (X)	Mean read length	Total Bases	Q20 bases	% of Q20 bases
K1	19003650	18540772	97.56	92.97	166	3148927553	2731843982	86.75
K2	20477030	19855407	96.96	99.34	166	3395825206	2914555646	85.83
K3	19586381	19259372	98.33	100.8	175	3430176857	2950140559	86.01
R1	26397822	25959501	98.34	131.2	187	4938052645	4210824415	85.27
R2	17832587	17276946	96.88	84.6	163	2901941960	2542944618	87.63
R3	28207892	27719884	98.27	137	163	4607621212	3986382096	86.52
S1	23735281	22247399	93.73	118.6	179	4255056522	3570856846	83.92
S2	31457556	30197681	96.00	152.2	187	5891947240	5003696719	84.92
S3	25040709	24301492	97.05	124.7	170	4256900105	3680517995	86.46
T9	28013270	26962279	96.25	138.5	190	5320238423	4498278254	84.55
Average	23975218	23232073	96.94	118.0	175	4214668772	3609004113	85.92

14.6 WGS Variant Calling

The adapter sequences were removed from the reads. 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 the 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.

14.7 Somy Value Estimation

The chromosomal somy values were estimated by using the method based on the 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 tend to be repetitive telomeric regions and their depths were unreliable (Domagalska et al., 2019). Similarly, certain over-amplified chromosome regions were excluded from median depth calculation (Table 10). The resulting raw median read depth of an individual chromosome was normalised against the median read depths of the 15 chromosomes - 7 neighbouring chromosomes on each side and the chromosome itself (Dumetz et al., 2017).

The chromosome somy value S was finally obtained by dividing the $2 \times$ raw median depth value of the chromosome with its normalised median depth value. The range of monosomy, disomy, trisomy, tetrasomy, and pentasomy was defined as described previously (Dumetz et al., 2017) as $S < 1.5$, $1.5 \leq S < 2.5$, $2.5 \leq S < 3.5$, $3.5 \leq S < 4.5$ and $4.5 \leq S < 5.5$, respectively. The somy values of each chromosome in a sample were plotted to create the clustered heatmap using the Rplot of heatmap2.

Table 11. The range of aneuploidy values in the heatmap.

$S < 0.5$	Monosomy
$1.5 \leq S < 2.5$	Disomy
$2.5 \leq S < 3.5$	Trisomy
$3.5 \leq S < 4.5$	Tetrasomy
$4.5 \leq S < 5.5$	Pentasomy

Table 12: Excluded regions for somy estimation.

Chromosome	Coordinates	Remarks
Ld01	270000 < x	
Ld02	280000 < x < 325000	Mini-exons
Ld05	440000 < x	
Ld10	40000 < x < 275000	gp63
Ld12	380000 < x < 570000	
Ld19	700000 < x	
Ld22	710000 < x	
Ld23	90000 < x < 110000	H-locus
Ld27	1000000 < x < 1050000	rDNA genes
Ld36	2550000 < x < 2580000	M-locus
All Chromosomes	7000 from 5' end & 2000 from 3' end are excluded	

14.8 Bioinformatic Analysis (Gene clustering and pathway):

Principal component analysis (PCA) was performed based on the types of mutations in the coding regions and 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 web tool to calculate and draw custom Venn diagrams at <http://bioinformatics.psb.ugent.be/webtools/Venn/>.

To understand the functional involvement of mutated genes in individual strains the pathway analyses were carried out by mapping the respective sets of the mutated gene of each strain onto the *L. donovani* specific KEGG pathway database (Du et al., 2014). A list of software and programming used in the present study is provided in table 13.

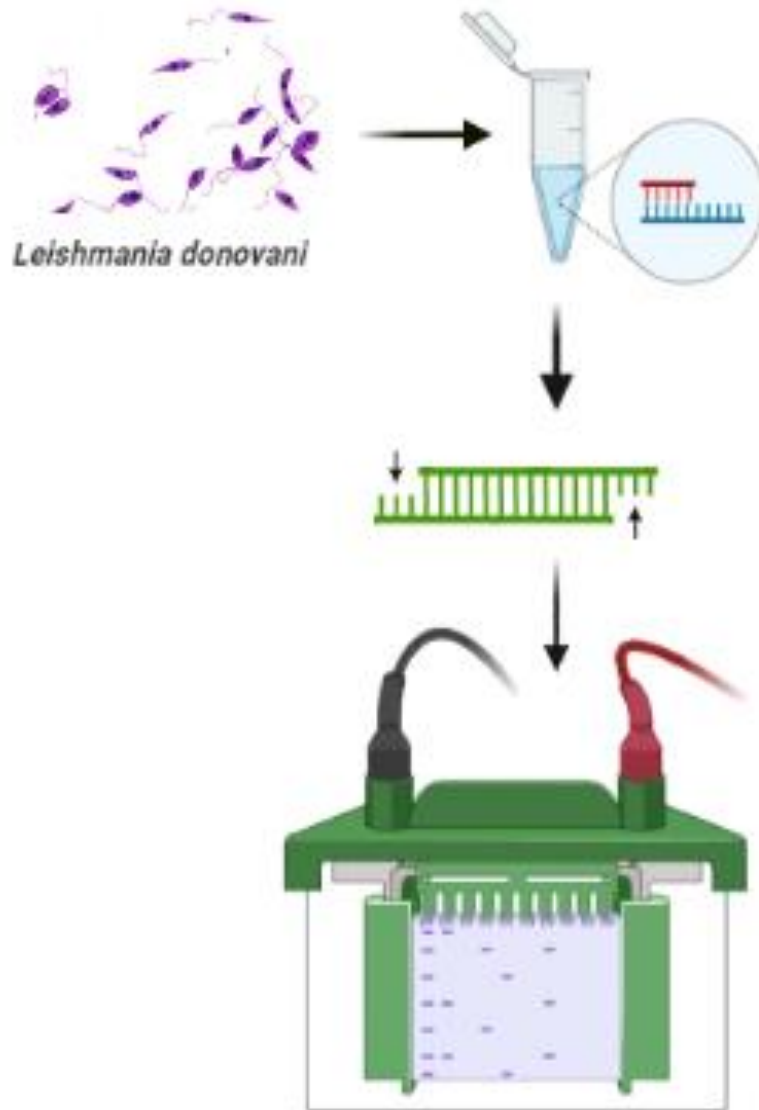
Table 13: List of software and programming used in the study.

Software and Programming	Analyses
Clustal Omega	Multiple Sequence Alignment Analysis
CLC genomics	SNPs analysis
DNAsp v6 software	Genetic diversity indices
EMBOSS Transeq	Sequence translation
GeneDB	An annotation database for pathogens
http://www.expasy.org program	Protein sequence analysis
I-Tasser Online Software	Secondary structure analysis
KEGG online database	Metabolic pathway study
MEGA X	Phylogeny network
MLST software	Sequence typing and allelic polymorphism
NCBI BLAST	Sequence alignment
PopART software	Haplotype analysis
PyMol 2 Software	Protein 3D structure analysis
R plot of heatmap2	Whole genome heat map construction
UniProt	Protein sequence analysis

Gene Bank Submissions:

A total of 176 nucleotide sequences and 10 clinical isolates whole genome sequences are submitted to National Centre for Biotechnology Information (NCBI), U.S. National Library of Medicine USA, by following GenBank Accession numbers:

1. MK720114- MK720121 (ITS1 sequences)
2. MN031875- MN031878, MK732447- MK732448 (ITS sequences)
3. MK764429- MK764435 (HSP70 sequences)
4. MN398901, MN883540-MN883559, MT625163-MT625221, MT634209-
MT634224, MT741860-MT741907, MT748045-MT748055 (MLST housekeeping
genes)
5. Data Accession: The sequence data related to whole genome sequencing of the
studied ten isolates are available under the BioProject Accession number
PRJNA598250.



CHAPTER 1

Restriction Fragment Length Polymorphism Analysis [RFLP] of the Indian Clinical Isolates of Kala-azar and Para Kala-Azar Dermal Leishmaniasis

1.1 Introduction

Visceral leishmaniasis (VL) or Kala-azar (KA) is a vector-borne parasitic disease caused by replication of the parasite *Leishmania sp.* within the macrophage-phagocytic system inside the vertebrate host. According to the WHO, three lakh cases of VL are reported worldwide annually, with forty to fifty thousand deaths. About 95% of the VL incidences occur in India, Brazil, Ethiopia, Kenya, Nepal, Sudan and South Sudan. In India, 90% of the cases occur in Bihar (Burza et al., 2018; Mann et al., 2021). VL can be fatal if left untreated and requires accurate and rapid diagnosis. Parasitological and serological methods are used for laboratory diagnosis of VL but for species identification, microscopy gives no result as under the microscope all parasites are similar morphologically. Multi Locus Enzyme Electrophoresis (MLEE) or Isozyme study of these protozoan parasites is considered as the gold standard by WHO (Rioux et al., 1990, Banuls et al., 2007) but the main drawback of the technique is mass cultivation of the parasites is required to get the extracts for enzyme electrophoresis and standardization of the technique is difficult. On the other hand, the introduction of Polymerase Chain Reaction (PCR) and modification of the basic techniques gave researchers of the field better choice for strain or species identification with great ease. These techniques are more sensitive and specific than other methods (Godoy et al., 2016; Koltas et al., 2016). Among many PCR based characterization methods, RFLP-PCR scores high and the RFLP markers give species specific profiles of the organisms. In some situations, species identification is crucial, especially for patients with *Leishmania* and HIV co-infection (Godoy et al., 2018). VL transmission is anthroponotic in the Indian subcontinent (Alvar et al., 2004), with Post Kala-azar Dermal Leishmaniasis (PKDL). Being a dermal sequel of VL, it is characterized by a macular, maculopapular or nodular rash. After an apparent cure from the VL in 20% to 60% cases, patients may develop PKDL (Zijlstra et al., 2019).

The problem leads to more badly when the arrival of first-line drug antimonial unresponsive cases is reported in 50% to 60% of KA patients (Croft et al., 2006). Moreover, reports of the association of other species and another genus with the disease were coming up (Sacks et al., 1995; Khanra et al., 2011; 2012; Desjeux et al., 2001). This situation again forces molecular biologists to enrich the knowledge of causal agents of disease by species identification for the proper drug approach. PCR-RFLP based methods are rapid tools with high accuracy in diagnosing the species of the causative organism (Schwartz et al., 2006; Coutinho et al., 2011; Kaufer et al., 2019).

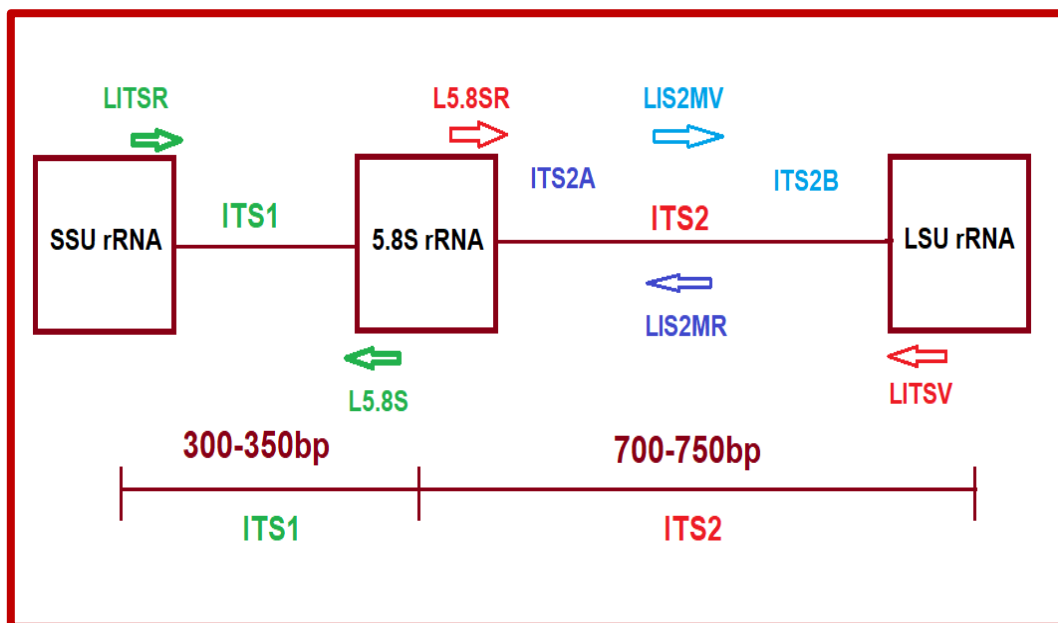


Figure 1.1. (A) The position of the internal transcribed spacer (ITS) in the ribosomal operon amplified with *Leishmania* specific primers.

Polymerase chain reaction based molecular diagnosis of leishmaniasis is typically targeted at the internal transcribed spacer (ITS), a region of the 1000bp ribosomal RNA genes. ITS sequences are composed of highly conserved regions, hence allowing their use in PCR for diagnostic purposes, and they have polymorphic regions (ITS1 & ITS2) that can be used in restriction fragment length polymorphism (RFLP) assays to determine *Leishmania* species status of clinical isolates of Indian KA.

ITS1, 300 to 350bp sequence separating the coding genes for small subunit (SSU) rRNA and 5.8S rRNA and ITS2, 700 to 750bp sequence separating the coding genes for 5.8S rRNA and large subunit (LSU) rRNA (Figure 1.1A).



Figure 1.1. (B) Schematic overview of the HSP70 coding region from *L. major* strain [the hatched boxes indicate primer annealing sites] (Fraga et al., 2010).

Another marker for species identification is RFLP profile of the *hsp70* gene (fragment encoding the 70-kDa heat shock protein, Figure 1.1B). Hsp70 PCR-RFLP with different restriction endonucleases is used to identify *Leishmania* species from the New and Old World (Fraga et al., 2010; Montalvo et al., 2010). HSP70 gene coding sequences are beneficial for species identification and species discrimination and phylogenetic tree analysis (Reimão et al., 2020).

In the present study, our chief goal was the species characterization of the available clinical isolates by PCR- RFLP analysis. We have described the Hsp70-RFLP, ITS-RFLP and ITS1-RFLP profiles of the clinical isolates of KA (n = 12) and para-KDL (n = 3) collected from India (2010–2016). We have also used two WHO reference strains, DD8 (*Leishmania donovani*) and K27 (*L. tropica*), to identify the species status of recently collected isolates. The present study also helped to estimate the species-specific markers relatedness between clinical isolates of *Leishmania* and WHO reference strains for molecular taxonomical purposes.

1.2 Outline of the work

1.2.1 Clinical Isolates of Indian KA & para-KDL and WHO reference strains

In the present study, we carried out the RFLP analysis with fifteen Indian clinical isolates from KA and para-KDL patients, along with two World Health Organization (WHO) reference strains for *Leishmania donovani*, DD8 and *L. tropica*, K27.

1.2.2 Genomic DNA isolation

Genomic DNA isolation from all Indian KA clinical isolates and para-KDL isolates along with two WHO reference strains was done as per the standard protocol (Khanra et al., 2011).

1.2.3 PCR amplification

PCR amplification of Heat Shock Protein 70 (HSP70), Internal Transcribed Spacer (ITS) and ITS1 of all clinical isolates of KA and para-KDL and reference strains were performed. The products were analysed with 1.2% agarose gel in TAE buffer.

1.2.4 Restriction cut analysis of ITS and ITS1 PCR product

Using restriction enzymes Hae III, MseI, Fsp I, Hinf I, Mse I, BstDE I, BstSC I, Bst2U I, Aco I, FauND I, Hpa I and MnaI the amplified ITS and ITS1 region was digested. The Restriction PCR products were analysed by 2% agarose gel electrophoresis in TAE buffer.

1.2.5 Restriction cut analysis of the HSP70 PCR product

The PCR amplicons of the hsp70 gene were digested using the 5U HaeIII restriction enzyme. The resulting products were analysed in a 3% agarose gel electrophoresis.

Detailed materials and methods are available in the Materials and Methods section.

1.3 Results

The PCR-RFLP method for genotyping is based on prior information about the amplified DNA sequence. PCR-RFLP is still widely used today for molecular characterization of *Leishmania sp.* This method identifies small gene changes that create or eliminate sites where a single base substitution can be digested by a particular restriction endonuclease (Banuls et al., 2007). In the present study, fifteen clinical isolates, including nine from our previous studies (lane 9 to lane 17; Khanra et al., 2012) were examined, along with two WHO reference strains, *L. donovani* (DD8) and *L. tropica* (K27). The PCR product of the entire ITS, ITS1 and ITS2 region of the genus *Leishmania* was approximately 1000 bp, 300-350 bp and 700 bp, respectively (Figure 1.2 & 1.3).

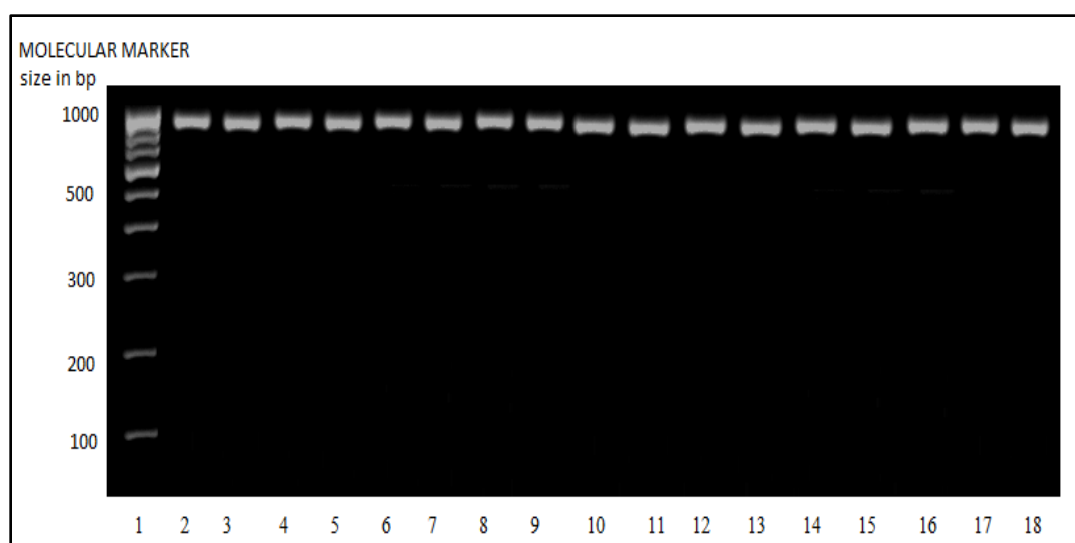
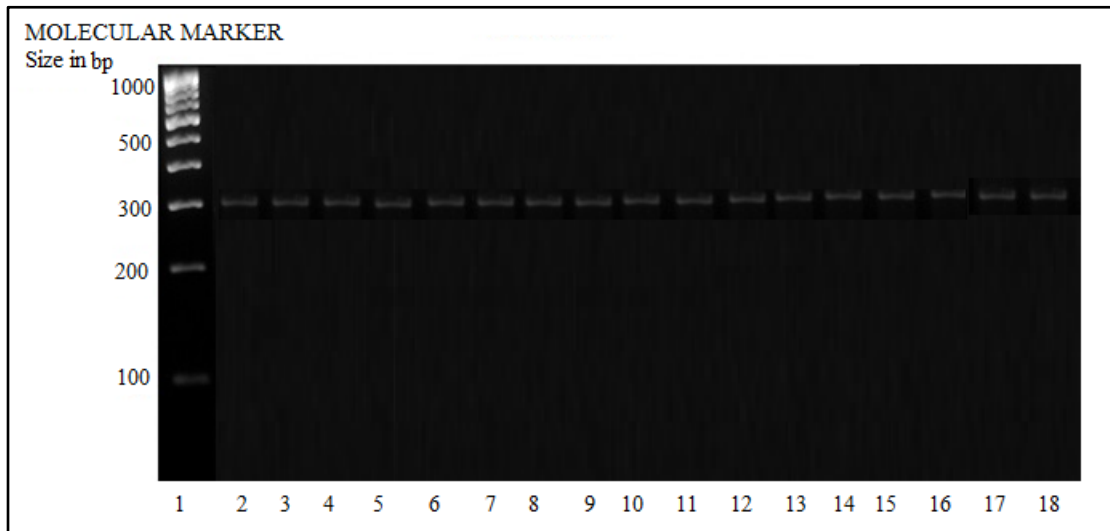


Figure 1.2. The PCR amplified nuclear DNA ITS region of different clinical isolates of KA and para-KDL along with WHO reference strain DD8 and K27 using primer LITSR/LITSV; Lane 1 MWmarker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 P6, lane 6 T10, lane 7 KK, lane 8 BK; lane 9 M1, lane 10 T9, lane 11 T8, lane 12 P2, lane 13 T7, lane 14 T4, lane 15 T3, lane 16 T2, lane 17 T5, lane 18 K27

(A)



(B)

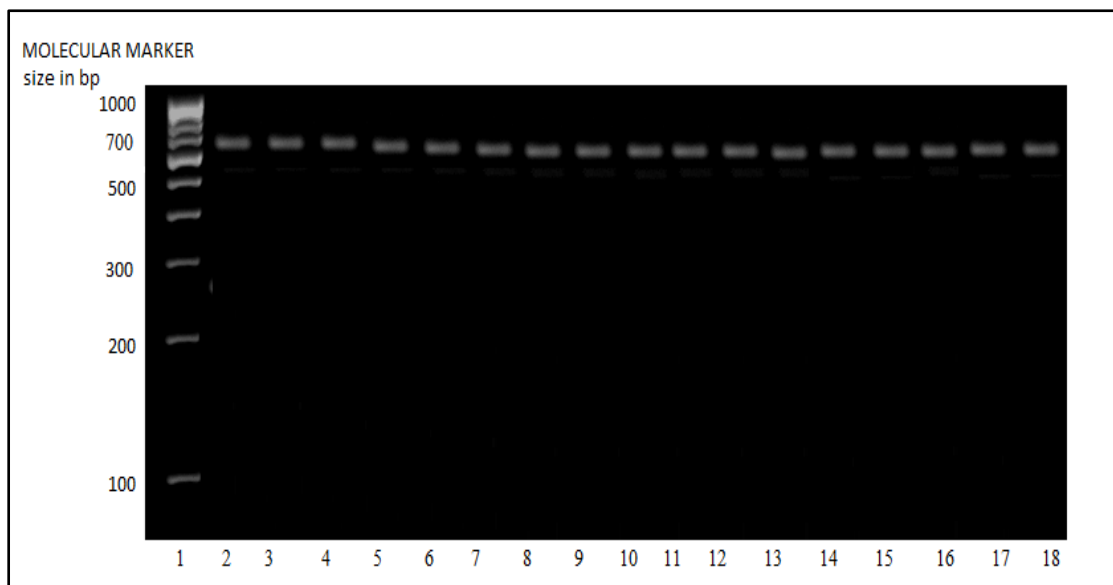


Figure 1.3. The amplified (A) ITS1 and (B) ITS2 region of nuclear DNA of different clinical isolates of KA and para-KDL along with reference strain DD8 and K27 using primers LITSR/L5.8S & L5.8SR/LITSV.; Lane 1 MWmarker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 P6, lane 6 T10, lane 7 KK, lane 8 BK; lane 9 M1, lane 10 T9, lane 11 T8, lane 12 P2, lane 13 T7, lane 14 T4, lane 15 T3, lane 16 T2, lane 17 T5, lane 18 K27

In the present study, restriction digestion products of HSP70, ITS1 and ITS amplicons with the eleven restriction enzymes HaeIII, Fsp I, Hinf I, Mse I, BstDE I, BstSC I, Bst2U I, Aco I, FauND I, Hpa I and MnaI were used to characterize and identify the species status of recently collected clinical samples from Indian KA and para-KDL patients. The HSP70 and ITS1 amplicon restriction fragment length polymorphism is a well-established molecular tool for parasite characterization, including *Leishmania* (Montalvo et al., 2010). We performed RFLP analysis of the HSP70 (Figure 1.4) and ITS1 (Figure 1.5) PCR amplified products from all fifteen clinical isolates (lane 3 to lane 17) of KA and para-KDL and two reference strains (lane 2 and lane 18) were digested with the restriction enzyme HaeIII.

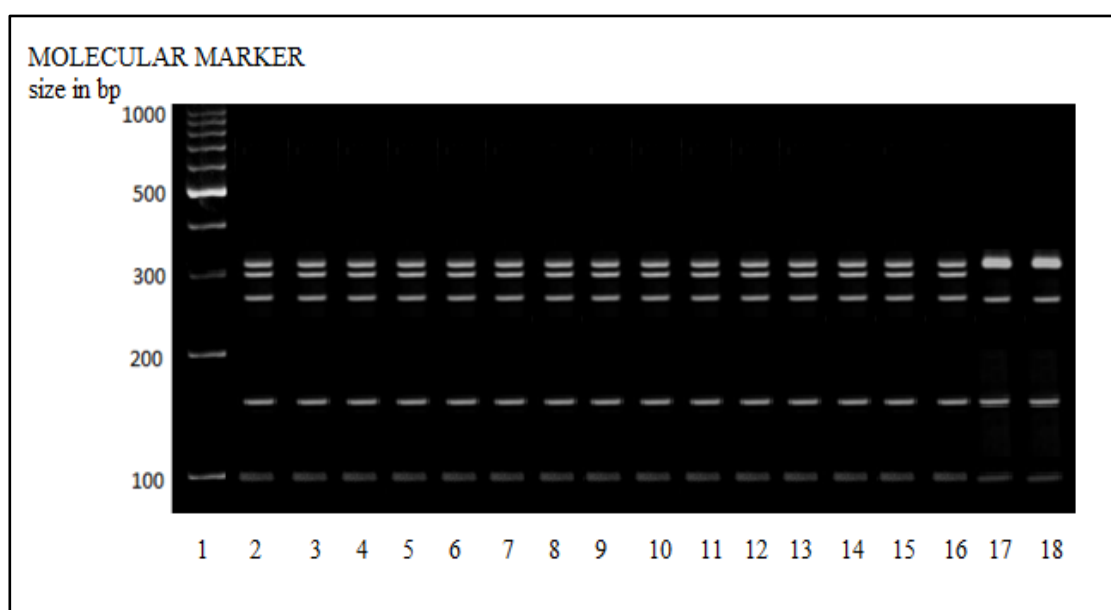


Figure 1.4. The amplified HSP70 coding regions of nuclear DNA of different clinical isolates of KA and para-KDL along with DD8 and K27 were digested using restriction enzyme HaeIII; Lane 1 MWmarker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 P6, lane 6 T10, lane 7 KK, lane 8 BK; lane 9 M1, lane 10 T9, lane 11 T8, lane 12 P2, lane 13 T7, lane 14 T4, lane 15 T3, lane 16 T2, lane 17 T5, lane 18 K27

Interestingly, HaeIII produced different patterns with HSP70 (Figure1.3) and ITS1 amplicon that unambiguously differentiated *L. tropica* from *L. donovani* (Figure1.4). We noticed the similarity of lane 3 to lane 16 with *L. donovani* standard strain DD8 (lane 2) but not with *L. tropica* standard strain K27 (lane 18), suggesting their species identity as *L. donovani* while lane 17 clinical isolate, T5 showed similar banding pattern as that of K27 (*L. tropica*) at lane 18. T5 is a unique clinical isolate which was collected from a confirmed Indian Kala-azar patient and was found to be *L. tropica* (Khanra et al., 2012).

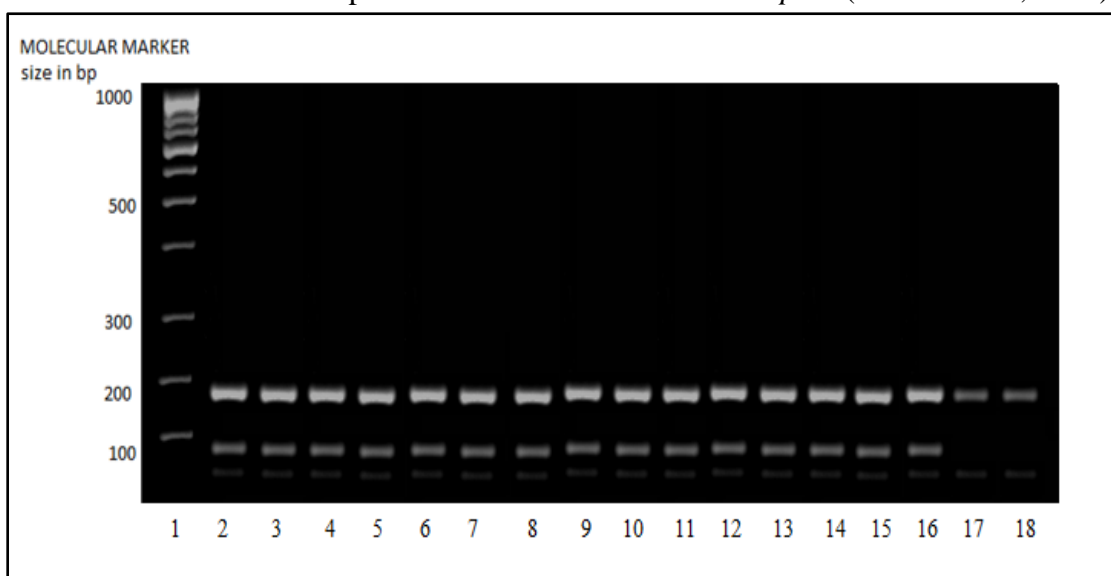
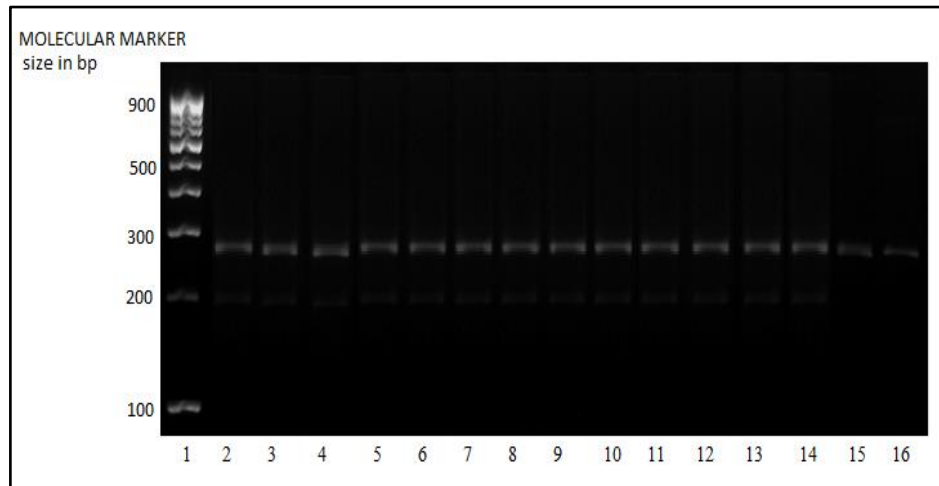


Figure 1.5. The amplified ITS 1 region of nuclear DNA of different clinical isolates of KA and para-KDL along with DD8 and K27 were digested using restriction enzyme Hae III; Lane 1 MW marker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 P6, lane 6 T10, lane 7 KK, lane 8 BK; lane 9 M1, lane 10 T9, lane 11 T8, lane 12 P2, lane 13 T7, lane 14 T4, lane 15 T3, lane 16 T2, lane 17 T5, lane 18 K27

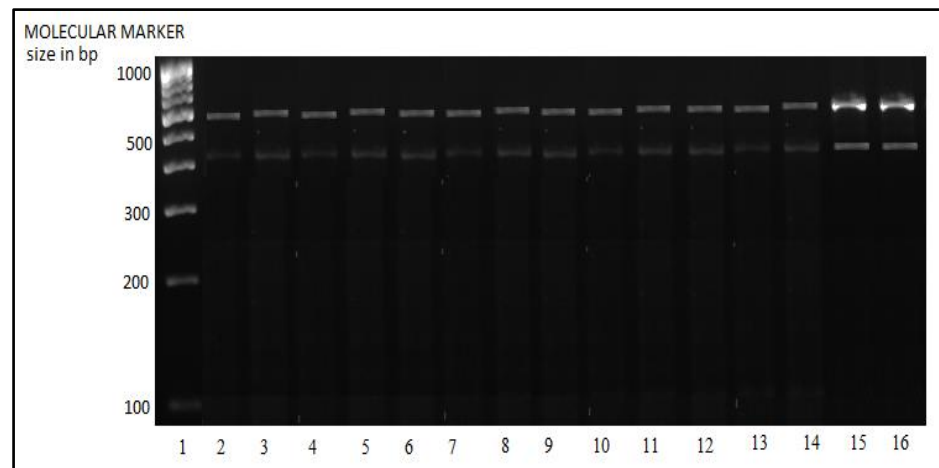
Our recently collected clinical isolates' samples were in lanes 3 to lane 8 which have not yet been identified at the species level but the isolates from lanes 9 to 17 were included in the present study to characterize them with other restriction enzymes which had not been done yet and also for other molecular biological characterization works that had been performed here only.

RFLP analysis of the PCR products of ITS amplicon from the studied clinical samples of KA and para-KDL from India was performed by digesting with six restriction enzymes - Hinf I (Figure 1.6; A), BstDE I (Figure 1.6 B), BstSC I (Figure 1.6; C), Bst2U I (Figure 1.6; D), Mse I (Figure 1.7; A), Fsp I (Figure 1.7; B).

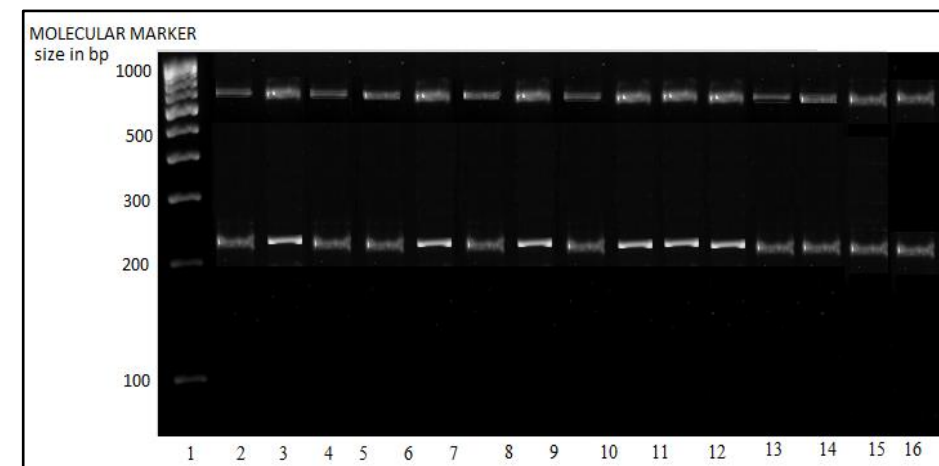
(A)



(B)



(C)



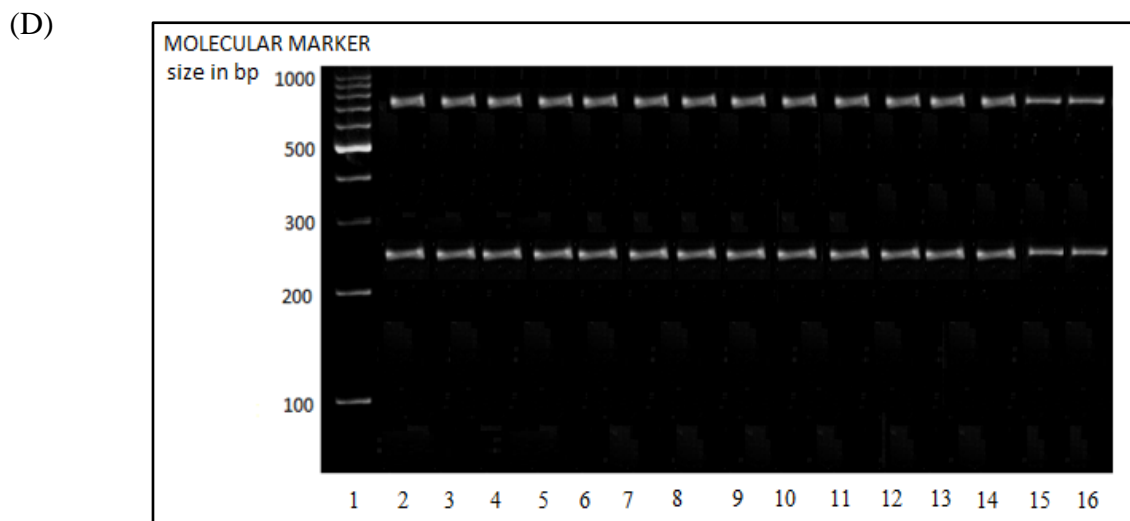
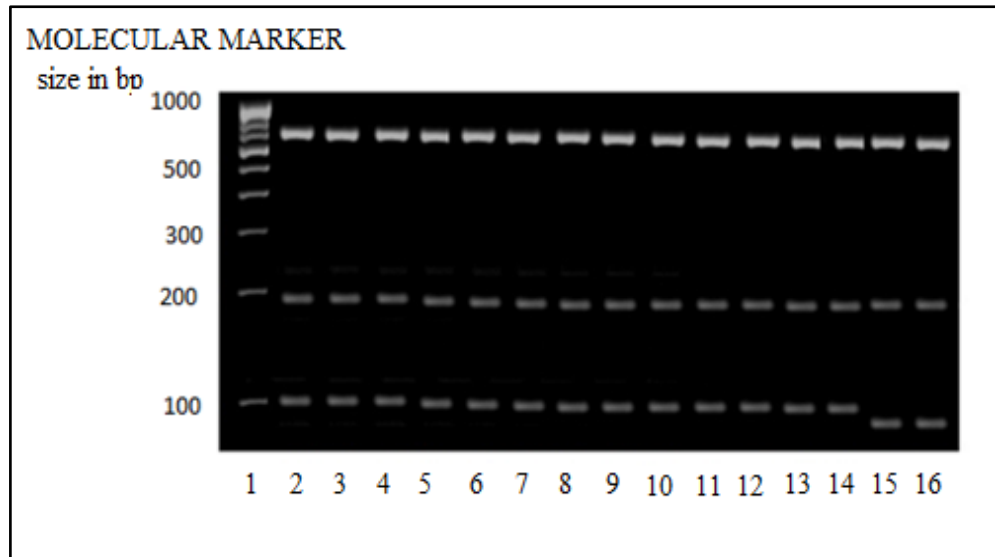


Figure 1.6. The amplified ITS region of nuclear DNA of different clinical isolates of KA and para-KDL along with DD8 and K27 were digested using a restriction enzyme. **A. *Hinf* I, B. *BstDE* I, C. *BstSC* I, D. *Bst2U* I;** Lane 1 MWmarker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 M1, lane 6 P2, lane 7 P6, lane 8 T10; lane 9 T9, lane 10 T8, lane 11 T7, lane 12 T4, lane 13 T3, lane 14 T2, lane 15 T5, lane 16 K27

We have noticed the similarity in the RFLP pattern of four restriction enzymes, *Hinf* I, *BstDE* I, *BstSC* I and *Bst2U* I of *L. donovani* standard strain DD8 (lane 2; Figure 1.6; A, B, C & D) with that of *L. tropica* standard strain K27 (lane 16; Figure 1.6; A, B, C & D) proving the inability of these restriction markers to differentiate these two species. Thus, they were rejected.

Interestingly, two markers *Mse* I (Figure 1.7; A) and *Fsp* I (Figure 1.7; B) produced unique patterns of ITS amplicons for *L. donovani* (lane 2; Figure 1.7; A, B) and for *L. tropica* (lane 16; Figure 1.7; A, B), clearly differentiating them species wise. Our recently collected clinical samples (lane 3 to lane 8) have shown similar restriction digestion patterns as that of DD8 (lane 2). The result again proved that the identifications of recently collected clinical isolates are *Leishmania donovani*.

(A)



(B)

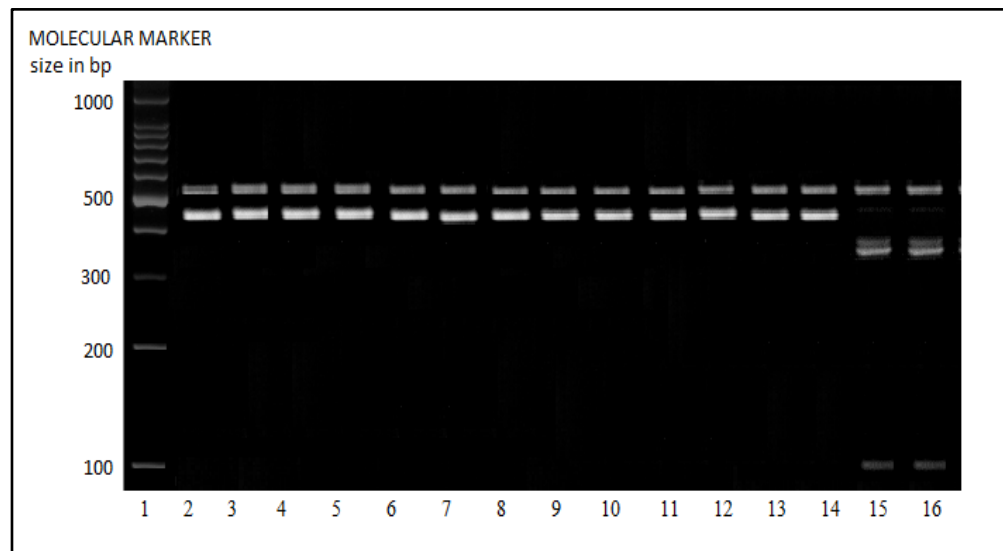


Figure 1.7. The amplified ITS region of nuclear DNA of different clinical isolates of KA and para-KDL along with DD8 and K27 were digested using a restriction enzyme. **A. *Mse I*, B. *Fsp I***; Lane 1 MWmarker (100 bp); lane 2 DD8, lane 3 H1, lane 4 A2, lane 5 M1, lane 6 P2, lane 7 P6, lane 8 T10; lane 9 T9, lane 10 T8, lane 11 T7, lane 12 T4, lane 13 T3, lane 14 T2, lane 15 T5, lane 16 K27

1.4 Discussion

Leishmania donovani has been recognized as the causative agent of KA and PKDL in India for over a century, but the involvement of other species or other genus with Indian KA has recently been reported (Sacks et al., 1995; Khanra et al., 2011; 2012; Srivastava et al., 2010). Therefore, it is of utmost significance to characterise the clinical samples at the species level to identify the responsible species for the concerned disease type. PCR-based molecular techniques have been proven to be more sensitive and powerful tools to detect *Leishmania* parasites directly from the clinical samples and for parasite characterisation (Khanra et al., 2012; Burza et al., 2018). Many such techniques have been developed by researchers in the field: RAPD-PCR (Welsh and McClelland, 1990; Williams et al., 1990; Welsh et al., 1991; 1992), AP-PCR (Welsh and McClelland, 1990;), RFLP-PCR (Rotureau et al., 2006; Banuls et al., 2007) etc. In the case of RAPD-PCR and AP-PCR, knowledge of the DNA sequence is not required because PCR uses only relatively short oligonucleotides of arbitrary sequence. These techniques provide an efficient way to obtain genetic markers in all organisms (Welsh and McClelland, 1990; Williams et al., 1990; Welsh et al., 1991; 1992) but they have their own drawbacks.

RFLP-PCR, on the other hand, has long been used as an alternative to sequencing. In this technique, no prior oligonucleotide synthesis nor previous sequence knowledge is needed. This method identifies minor variations in the nucleotide where a single base substitution either insertion or deletion of a nucleotide site capable of being digested by a specific restriction endonuclease (Banuls et al., 2007).

The analysis of ITS1 and HSP70 amplicon restriction digestion with Hae III restriction enzymes has been carried out here to examine the genome of fifteen clinical isolates, along with two WHO reference strains.

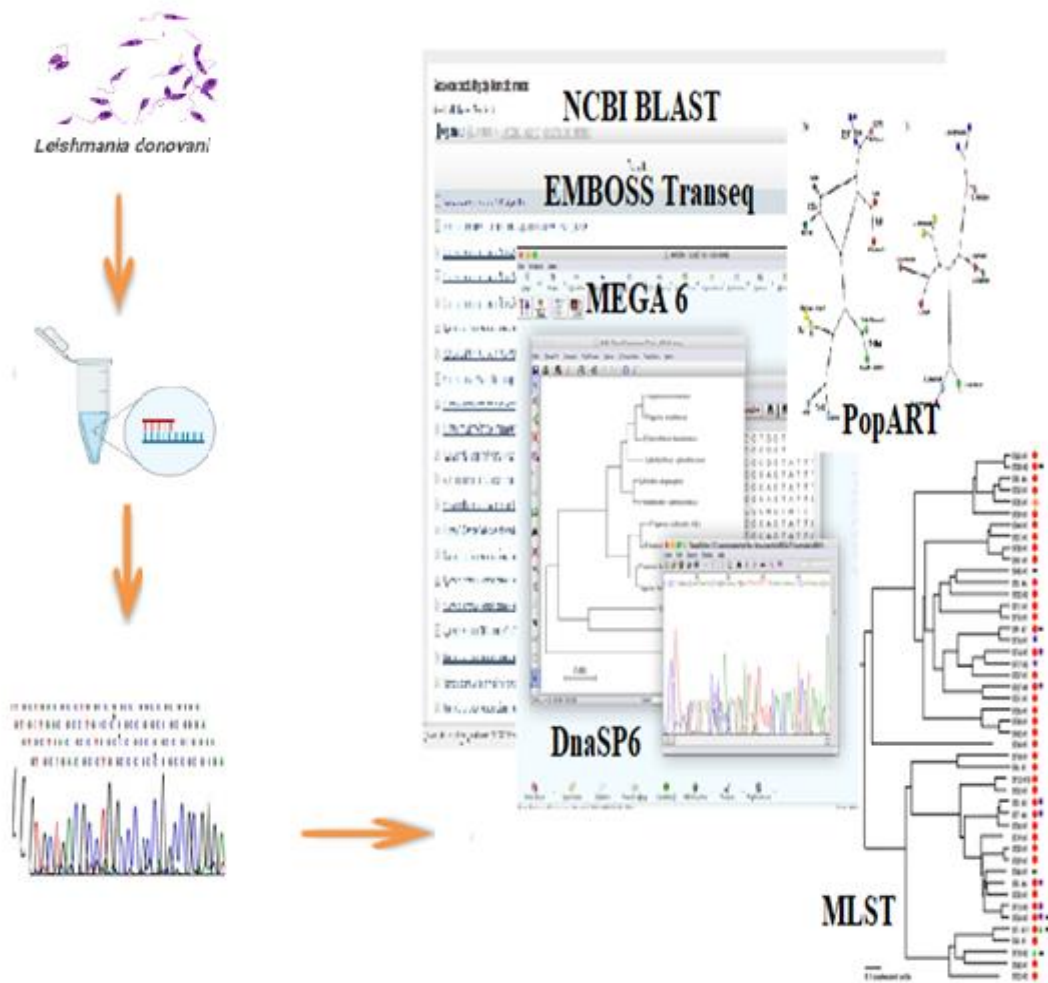
Our study with a limited number of isolates demonstrated that all recently collected clinical isolates from Indian KA and clinically diagnosed PKDL patients were *Leishmania donovani* as they showed PCR-RFLP profiles similar to that of *L. donovani* WHO reference strain, DD8. It should be mentioned here that in the present study, we have one clinical isolate, T5 which was isolated from a confirmed Kala-azar patient but came out as *L. tropica* (Khanra et al., 2012). This isolate is included here for further characterization of the same at the genome level because of its uniqueness. Some of our observations have been published elsewhere (Sarraf et al., 2021).

All para-KDL clinical isolates were found out to be *L. donovani* which is in congruence with an earlier report (Burza et al., 2018) that *L. donovani* is a causative agent PKDL in the Indian subcontinent. This complex association of VL and PKDL infection in the same patients may lead to the manifestation of para-Kala-azar Dermal Leishmaniasis (para-KDL) in the Indian subcontinent.

Despite the practical utility of ITS1 as a target site for identifying pathogenic *Leishmania* species, the main drawback of ITS1 as a diagnostic marker is its inability to distinguish between closely related species within the *Viannia* subgenus (Kaufer et al., 2019). Keeping such drawbacks in mind, the analysis of ten different markers has been carried out in the ITS and ITS1 amplicon in the present study. None of them showed any restriction site in the ITS1 amplicon which is a 300 to 350 bp sequence, proving the inability of these restriction markers to cut the ITS1 amplicon.

Two markers (Mse I & Fsp I) unambiguously have differentiated *L. donovani* from *L. tropica* by the ITS amplicon only.

Four markers (Hinf I, BstDE I, BstSC I and Bst2U I) could not be able to differentiate between *L. donovani* from *L. tropica* by the ITS amplicon. While four restriction enzymes, Aco I, FauND I, Hpa I and Mna1, have not shown any restriction cut during restriction digestion of the ITS amplicon. Accurate species identification in clinical leishmaniasis cases is essential, mainly when dealing with VL cases in the Indian subcontinent, where VL caused by *L. donovani* and *L. tropica* have been reported (Sacks et al., 1995; Khanra et al., 2011). The results highlighted the need for providing clinicians with the tools specific for diagnosing and determining the causative agent.



CHAPTER 2

Multilocus Sequence Typing (MLST) of *Leishmania* sp. Isolated from Indian Kala-azar and Para Kala-Azar Dermal Leishmaniasis Patients

2.1 Introduction

Leishmaniasis is a neglected tropical disease transmitted by the protozoan parasite belonging to the genus *Leishmania*. Disease transmission is carried out by being bitten by an infected female phlebotomine sandfly. Leishmaniasis has three main forms: Visceral Leishmaniasis (VL) or Kala-azar (KA), Cutaneous Leishmaniasis (CL) and Mucocutaneous Leishmaniasis (MCL). Also, the pathological symptoms of all these diseases and their mutations are very different, of all the forms, the most severe form is Kala- Azar (KA) or visceral leishmaniasis (VL), which can be fatal if left untreated. VL is predominant on four continents and is endemic in 88 countries, 72 of which are developing countries (Arif et al., 2008). Globally, more than 90% of all cases of VL occur in India, Bangladesh, Brazil, Iran, Nepal, Ethiopia and South Sudan (Alvar et al., 2012). In India, apparently cured KA patients develop post-kara-Azar dermal leishmaniasis (PKDL) in 10% to 20% of cases, an abnormal skin lesion such as macular, papular, or nodular. In the northwestern states of India (the main regions of Rajasthan and Punjab), CL is caused by *L. tropica* and *L. major*. The most damaged area in Rajasthan is the Bikaner district (WHO, 2017). In addition, co-infection with *Leptomonas* and VL indicates a genetic association between the *Leishmania donovani* isolate in India and the genus *Leptomonas* (Singh et al., 2013). Genetic divergence and recombination were recently reported in a Bangladesh *L. donovani* isolate (Banu et al., 2019). VL was almost wiped out in the Indian subcontinent but reappeared in 1977 with drug resistance, causing several consecutive pandemics (Downing et al., 2012). Even after extensive control measures, VL persistence may be due to genetic variation in *L. donovani* that cannot be treated with current treatment regimens. Historically, *Leishmania donovani* has been known to cause KA, but recent reports have confirmed an association between *L. tropica* with this disease (Sacks et al., 1995; Khanra et al., 2012), a parasite of various genotypes in the country that has been shown to exist.

Hence, the periodic taxonomical characterisation of the clinical isolates is crucial for epidemiological surveillance and proper treatment regime. Many molecular techniques have been implemented to understand the current genetic diversity among the *Leishmania* parasite population. A gold standard method Multilocus Enzyme Electrophoresis (MLEE), based on the MON system, is used to identify the *Leishmania* parasite at the species level (Rioux et al., 1990). Currently, the most reproducible technique, Multilocus sequence typing (MLST), is used for genetic diversity and phylogeny analysis of *Leishmania* parasites (Zhang et al., 2013; Gelanew et al., 2014; Marlow et al., 2014) and also a new approach for detecting interspecies and intraspecific variations of the genus *Leishmania* (Miles et al., 2009; Akhoundi et al., 2016) which revealed that MLST was a highly discriminatory alternative to MLEE (Marco et al., 2015; Mauricio et al., 2006; Zemanova et al., 2007). The emergence of unresponsive drug cases and the use of different hosts and vectors might be associated with genetic structural changes in the parasite population (Banu et al., 2019). Furthermore, HIV and *Leptomonas* co-infection with VL again complexed the situation. Currently, limited information is available about the genetic diversity of *Leishmania sp.* in India. Hence, to address this, we applied the MLST approach to explore various genetic diversity parameters, i.e., nucleotide diversity (P_i), to measure the average number of nucleotide variations per site between any two DNA sequences chosen randomly from the studied sample population, haplotype diversity (H_d) to estimate the number of haplotypes (gene) in studied loci, gamma parameters to understand the variation of substitution rates among sites, Transition/Transversion Ratio (R) to evaluate the changes in the protein sequence, Tajima's D value to distinguish between the nucleotide sequence evolving neutrally or one evolving under a non-random process and phylodynamics to understand the relatedness among the studied Indian *Leishmania spp.* population.

To accomplish the objectives of the present study, here we standardized and analysed the MLST of the total of fifteen clinical isolates of which 12 were from Visceral Leishmaniasis patients and three from clinically diagnosed PKDL patients with VL, i.e., para Kala-azar Dermal Leishmaniasis (para-KDL) taking fifteen housekeeping gene loci.

2.2 Outline of the work

2.2.1 Studied Clinical Isolates and Reference Strains

The list of clinical isolates and reference strains was mentioned in the material and method section (Tables 6 and Table 7). The present study was performed with 12 species and two subgenera, *Leishmania* and *Viannia*. We analysed the MLST of the fifteen clinical isolates from KA and para-KDL patients with 15 marker genes. MLST analyses were classified into three schemes according to the selection of different housekeeping genes and their respective available reference isolates sequences from NCBI-BLAST (Table 2.1).

2.2.2 Marker Selection

In the present study, for the first time, we reported the DNA sequence diversity and sequence typing of Indian clinical isolates for the following fifteen housekeeping genes, namely aconitase (*aco*), alanine aminotransferase (*alat*), enolase (*enol*), phosphoglucomutase (*pgm*), hypoxanthine-guanine phosphoribosyltransferase (*hgprt*), spermidine synthase (*spdsyn*), aspartate aminotransferase (*asat*), glucose-6-phosphate isomerase (*gpi*), nucleoside hydrolase 1 (*nh1*), isocitrate dehydrogenase (*icd*), mannose phosphate isomerase (*mpi*), fumarate hydratase (*fh*), 6-phosphogluconate dehydrogenase (*pgd*), glucose-6-phosphate dehydrogenase (*g6pdh*) and phosphomannomutase (*pmm*). The study of gene locus phosphomannomutase (*pmm*) was separately placed for intraspecies and interspecies MLST analysis due to a smaller number of available reference sequences. For locus *pmm*, intraspecies analysis was performed with nine Indian *Leishmania donovani*

isolates, including one (AG83) reference strain. Interspecies analysis was performed with fourteen *Leishmania sp.* reference strains and eight Indian *L. donovani* isolates.

Amplified *pmm* locus sequences of eight Indian clinical isolates were submitted in the GenBank with accession number MT634217-MT634224. The Materials and Methods section would give detailed information about the gene primers and loci amplicon size with GenBank accession number.

Table 2.1 List of studied housekeeping genes, sample size and species.

Species Level	Housekeeping Genes Set	Sample Size (n)	Species
Intra-species	Scheme A (<i>aco, alat, enol, hgprt, pgm, spdsyn</i>)	16	<i>Leishmania donovani</i>
	Scheme B (<i>pgd, asat, nh1, gpi</i>)	18	
	Scheme C (<i>icd, fh, mpi, g6pdh</i>)	15	
Inter-species	Scheme A (<i>aco, alat, enol, hgprt, pgm, spdsyn</i>)	40	<i>L. donovani</i>
	Scheme B (<i>pgd, asat, nh1, gpi</i>)	50	<i>L. infantum</i>
	Scheme C (<i>icd, fh, mpi, g6pdh</i>)	40	<i>L. mexicana</i>
			<i>L. braziliensis</i>
			<i>L. peruviana</i>
			<i>L. guyanensis</i>
			<i>L. panamensi</i>
			<i>L. tropica</i>
			<i>L. lainsoni</i>
			<i>L. major</i>
			<i>L. gerbilli</i>

2.2.3 Genomic DNA isolation and PCR amplification

- Genomic DNA isolation from all clinical isolates and two WHO reference strains was performed as per standard protocol (Bhattacharyya et al., 1993). PCR amplification was performed by 35 cycles with denaturation at 96⁰C followed by 55-60⁰C annealing with 72⁰C extension.
- PCR purification and sequencing were done.
- Data analysis and interpretation (genetic diversity, allelic profile, sequence type, haplotype and phylogeny) were performed using various bioinformatics tools.

The detailed protocol of the present study was explained in the Materials and Methods section.

2.3 Results

2.3.1 Analysis of Intraspecies Genetic Diversity Indices

- To study the evolutionary rate differences among sites, a distinct gamma distribution was estimated for each studied nucleotide sequence. In the studied isolates population, gamma distribution's shape parameter has varied. The highest gamma distribution was observed in *enol*, *pgm*, *pgd*, *icd* and *spdsyn* (200), representing the very small variation of substitution rates, but very low for *aco*, *asat*, *nh1*, *gpi* and *mpi* (0.050), indicating a high variation of substitution rates among sites (Table 2.2).
- The negative value of Tajima's D specifies an excess of rare alleles. Here, phosphomannomutase (*pmm*) locus showed, at $P < 0.05$, a statistically significant negative value of Tajima's D (Table 2.2).

- In intraspecies genetic diversity indices, the highest number of mutations and variable sites (358,389) were observed in the fumarate hydrogenase (*fh*) gene. No intraspecies polymorphism was observed in the Spermidine synthase (*spdsyn*) locus. Hence, this is the conserved locus in the Indian *Leishmania donovani* (Table 2.2).
- Nucleotide diversity (Pi) and haplotype diversity (Hd) indices are explained with the help of a bar diagram (Figure 2.1), indicating the highest Pi in *fh* followed by *enol* and *g6pdh* and the highest Hd in *asat* locus.
- Haplotype networks in the highest nucleotide diversity (Pi) loci represent the total number of 8,3 and 5 haplotypes in *fh*, *enol* and *asat* respectively (Figure 2.2).
- Intraspecies allelic profiles of each eleven *L. donovani* isolates (H1, A2, KK, BK, P2, P6, T2, T8, T9, T10) along with one WHO references strain (DD8) for the seven housekeeping genes *aco*, *alat*, *asat*, *enol*, *gpi*, *nhl* and *pgm* are indicating the respective sequence type (ST) of each strain with Discriminatory Power (DP) value 1.0. A DP value of 1.0 would indicate that the sequence typing method could differentiate each member of the *Leishmania* population from all other members of that population (Figure 2.3 A).
- Intraspecies phylogenetic analysis based on the concatenated sequence of seven housekeeping genes *aco*, *alat*, *asat*, *enol*, *gpi*, *nhl* and *pgm* respectively, of the *L. donovani* isolates of Indian VL and para-KDL patients, along with WHO reference strain of *L. donovani* (DD8), revealed four distinct clusters in the Indian clinical isolates. This indicated that intraspecies diversity is present among the Indian *L. donovani* clinical isolates of India (Figure 2.3 B).

➤ Table 2.2 Intraspecies Genetic Diversity Indices.

Gene	S	Eta	h	Hd	Pi Nucleotide diversity (per site)	Gamma Distribution	Tajima's D Value	P-Value
<i>aco</i>	18	20	10	0.955	0.01461	0.0500	0.47127	> 0.10
<i>alat</i>	1	1	3	0.167	0.00033	7.2386	-1.14053	> 0.10
<i>enol</i>	41	41	3	0.425	0.03981	200	0.84819	> 0.10
<i>hgprt</i>	1	1	2	0.182	0.00053	0.1075	-1.12850	> 0.10
<i>pgm</i>	6	6	4	0.673	0.00296	200	-1.34209	> 0.10
<i>spdsyn</i>	0	0	0	0	0	200	0	0
<i>pgd</i>	4	4	4	0.455	0.00071	200	-1.74687	> 0.05
<i>asat</i>	30	36	12	0.971	0.00679	0.0500	-1.23337	> 0.10
<i>nh1</i>	21	21	11	0.856	0.00469	0.0500	-1.09459	> 0.10
<i>gpi</i>	11	13	10	0.923	0.00324	0.0500	-0.55150	> 0.10
<i>icd</i>	2	2	3	0.345	0.00069	200	-1.42961	> 0.10
<i>fh</i>	358	389	8	0.909	0.11236	0.4133	-0.05530	> 0.10
<i>mpi</i>	5	5	5	0.810	0.00220	0.0500	0.47090	> 0.10
<i>g6pdh</i>	81	81	6	0.833	0.02426	0.1877	-1.05025	> 0.10
<i>pmm</i>	50	50	6	0.889	0.02433	4.6127	-1.78812	*< 0.05

S: Number of variable (polymorphic) sites

Eta: Total number of mutations

h: Number of Haplotypes

Hd: Haplotype (gene) diversity

Pi: An index of nucleotide diversity

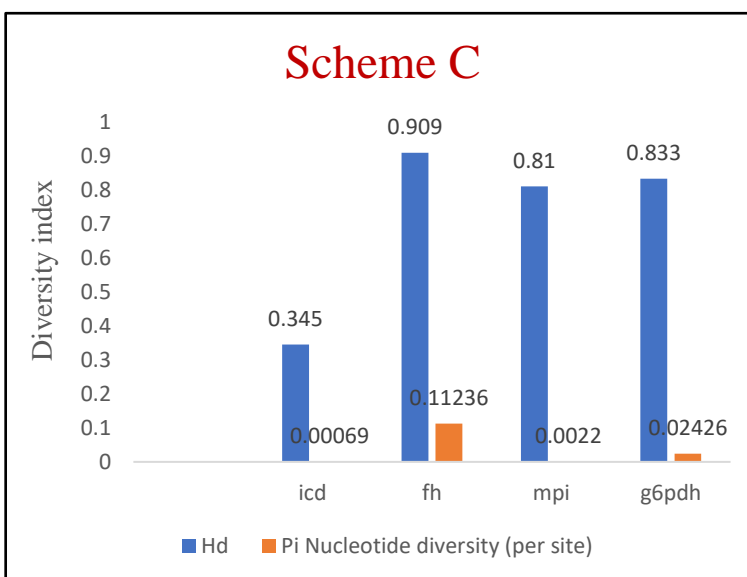
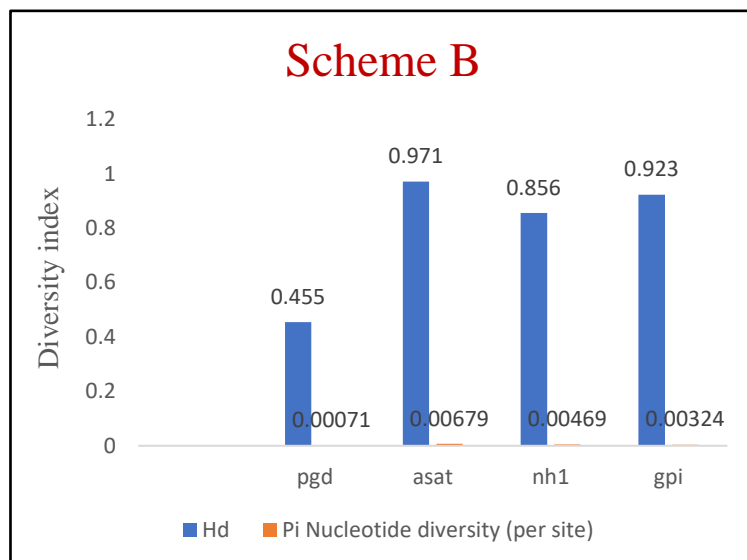
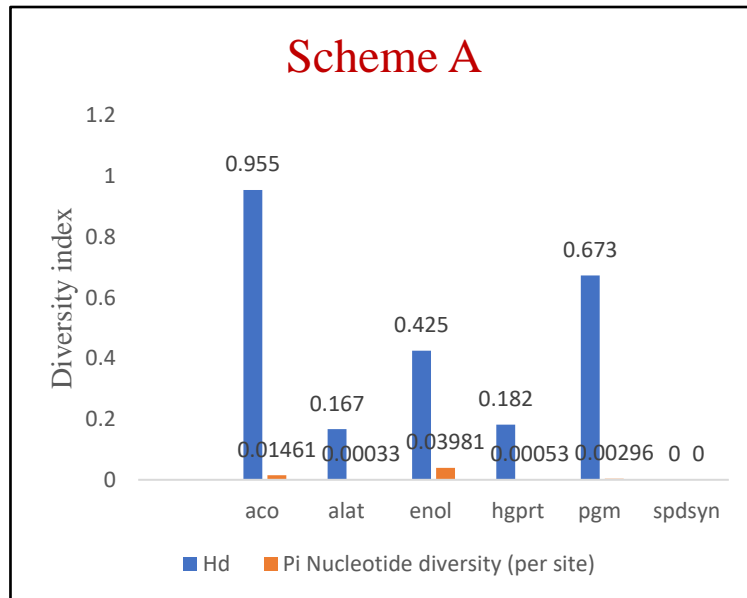


Figure 2.1. Comparison of intraspecies haplotype diversity (Hd) and nucleotide diversity (pi) indexes for each marker in the studied MLST scheme. Except one gene *spdsyn* which is conserved in Indian *Leishmania donovani* isolates (Scheme A). The *fh*, *enol* and *g6pdh* genes in Indian *Leishmania donovani* isolates showed the highest value of nucleotide diversity (Orange bar), followed by *aco*, *alat*, *hgprt*, *pgm*, *icd*, *mpi*, *pgd*, *ast*, *nh1* and *gpi* while the *asat*, *aco*, *gpi* and *fh* genes in Indian *Leishmania donovani* isolates showed the highest values of haplotype diversity (Blue bars), followed by *nh1*, *g6pdh*, *mpi*, *pgm*, *pgd*, *enol*, *icd*, *hgprt* and *alat* genes.

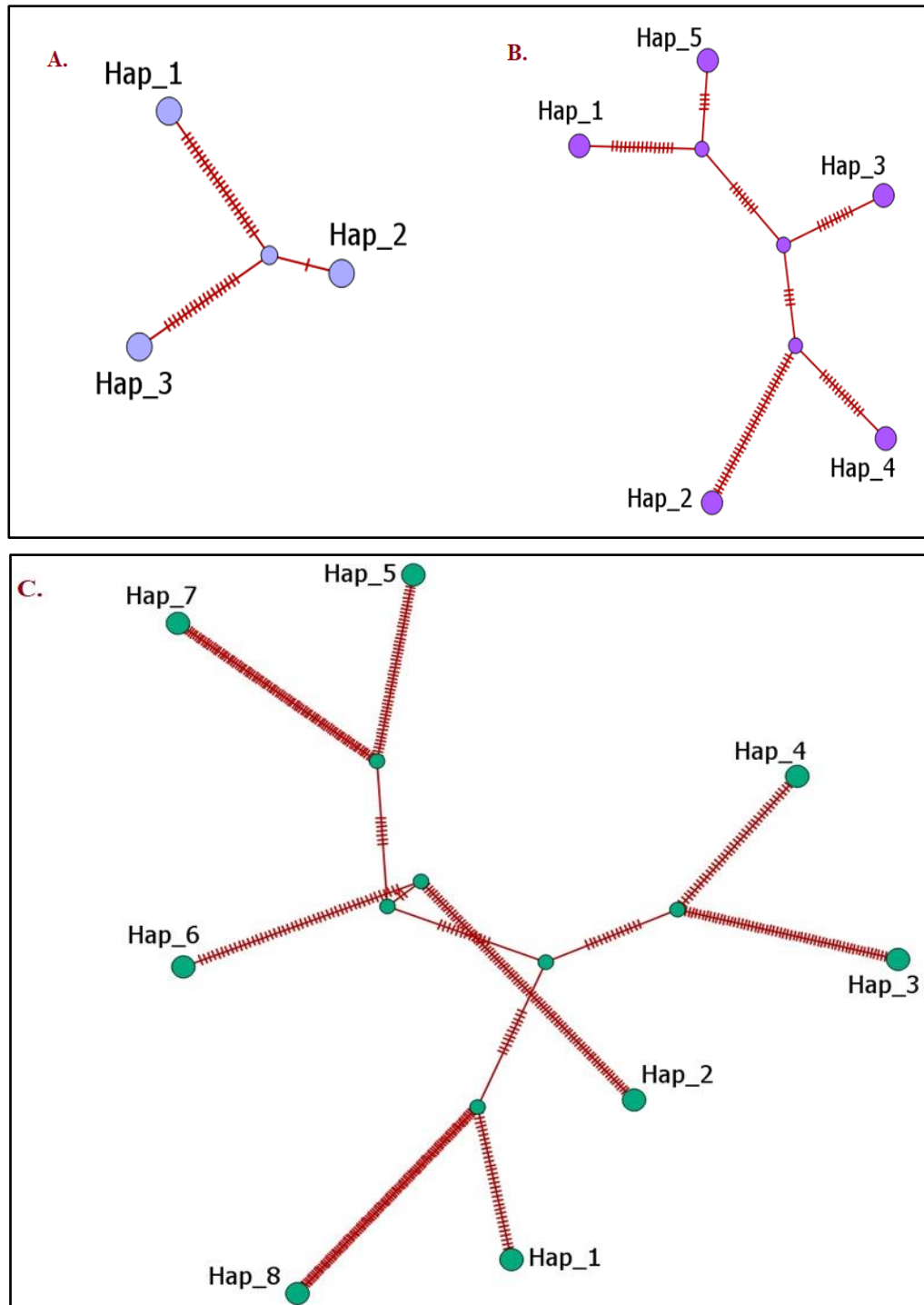


Figure 2.2. Haplotype (Hap) networks in the highest nucleotide diversity (π) loci of (A) *enol*, (B) *g6pdh* and (C) *fh* were built by a statistical parsimony estimation approach based on haplotype reconstruction by TCS algorithm using PopART software. Hypothetical ancestors were represented by small circles and mutational steps between the individual haplotypes were indicated by hatch marks.

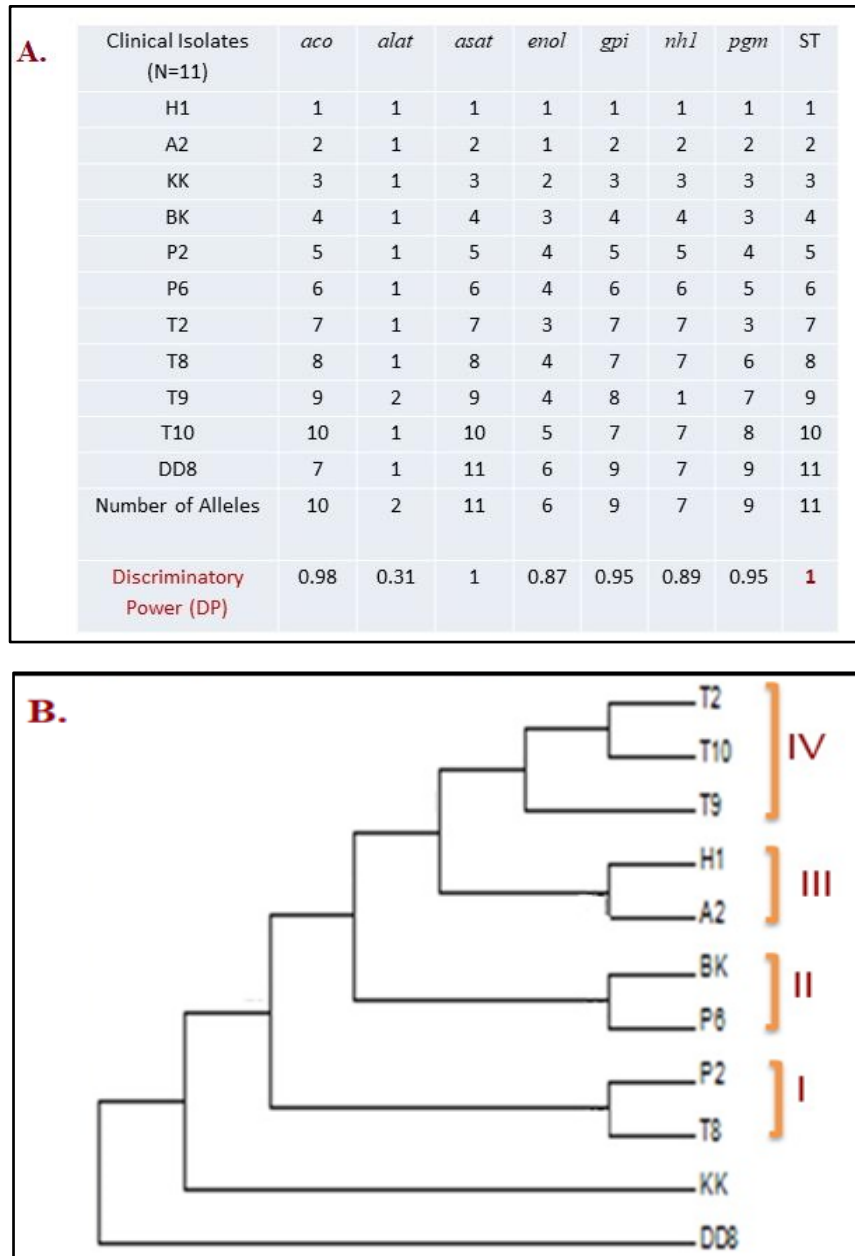


Figure 2.3. (A) Intraspecies allelic profiles show each eleven isolates' sequence type (ST). A Discriminatory Power (DP) value of 1.0 would indicate that the sequence typing method could differentiate each member of the *Leishmania* population from all other members of that population. (B) Intraspecies phylogenetic analysis based on the concatenated sequence of seven housekeeping genes (*aco*, *alat*, *asat*, *enol*, *gpi*, *nhl* and *pgm*) formed four distinct clusters in the Indian clinical isolates of *Leishmania donovani*.

2.3.2 Analysis of Interspecies Genetic Diversity Indices

- A discrete gamma distribution was used to study evolutionary rate differences among sites of the housekeeping gene loci. The studied population shape parameter of gamma distribution has varied. The high value of the gamma parameter indicates a very low variation of substitution rates among sites and vice-versa. Here, it was particularly high (200) for *pgd*, but very low (0.0854) for *nh1* (Table 2.3 A).
- The housekeeping gene fragments had 18-385 polymorphic sites among the studied group of *Leishmania sp.* with *hgprt*, *icd* and *nh1* having the least and *aco*, *asat* and *fh* being the most segregating regions. The average G+C content was above 50% for all genes (Table 2.3 B).
- The estimated Transition/Transversion bias (R) was mostly above one for all genes except *fh* & *g6pdh* ($1 > R > 0.5$), implying that transition and transversion did not occur with the same rate and highest R-value in *icd* indicating the substitution pattern and rates of transition was greater than transversion with the 25% nucleotide frequencies for adenine (A), cytosine (C), guanine (G) and the ratio of thymine (T) & uracil (U) (Table 2.3 B).
- The ratio of nonsynonymous/synonymous rate (dN/dS) also varied and it was below one in six and above one in nine gene loci. These nine loci are *aco*, *enol*, *asat*, *nh1*, *gpi*, *fh*, *mpi*, *g6pdh* and *pmm*, indicating positive selection with more non-synonymous substitutions (Table 2.3 A).
- The significant Tajima's D Value at $P < 0.001$ was observed for variable regions of *pgd*, *asat*, and at $P < 0.05$, the significant Tajima's D Value was observed for variable regions of *gpi*, *mpi* and *g6pdh* gene loci, signifying the more number of rare alleles (Table 2.3 A).

Table 2.3 (A) Intraspecies Genetic Diversity Indices.

Gene	S	Eta	h	Hd	Pi Nucleotide diversity (per site)	Gamma Distribution	Tajima's D Value	P -Value
<i>aco</i>	124	142	23	0.94	0.08139	0.5438	0.82666	> 0.10
<i>alat</i>	115	128	14	0.82	0.07782	0.6416	1.20872	> 0.10
<i>enol</i>	74	85	15	0.84	0.05657	0.1926	0.30619	> 0.10
<i>hgprt</i>	52	53	9	0.70	0.05628	2.7227	1.83899	> 0.05
<i>pgm</i>	95	99	18	0.89	0.07429	1.2447	1.22444	> 0.10
<i>spdsyn</i>	66	67	9	0.74	0.06816	1.3249	1.31113	> 0.10
<i>pgd</i>	79	79	10	0.65	0.00406	200	-2.76241	*** < 0.001
<i>asat</i>	89	97	16	0.76	0.00528	0.0979	-2.53709	*** < 0.001
<i>nh1</i>	27	27	17	0.87	0.00339	0.0854	-1.53608	> 0.10
<i>gpi</i>	83	85	16	0.77	0.006470	0.1523	-2.17390	* < 0.05
<i>icd</i>	18	18	10	0.77	0.00443	0.1219	-1.49334	> 0.10
<i>fh</i>	385	418	14	0.62	0.04866	0.5457	-1.66475	> 0.05
<i>mpi</i>	63	64	12	0.79	0.00755	3.8885	-2.11703	* < 0.05
<i>g6pdh</i>	133	143	12	0.83	0.01496	0.4408	-2.13873	* < 0.05
<i>pmm</i>	146	162	12	0.93	0.08677	1.65	-0.16783	> 0.10

S: Number of variable (polymorphic) sites

Eta: Total number of mutations

h: Number of Haplotypes

Hd: Haplotype (gene) diversity

Pi: An index of nucleotide diversity

Table 2.3 (B) Intraspecies Genetic Diversity Indices.

Gene	Number of alleles	DP	Theta (θ)	dN/dS	R	G+C Content
<i>aco</i>	23	0.943	0.0665	2.36	1.82	0.575
<i>alat</i>	14	0.833	0.0585	0.18	2.26	0.619
<i>enol</i>	21	0.915	0.0521	3.91	1.56	0.601
<i>hgprt</i>	9	0.721	0.0374	0.169	4.15	0.605
<i>pgm</i>	18	0.891	0.0556	0.2233	2.85	0.600
<i>spdsyn</i>	9	0.751	0.0502	0.2338	3.41	0.632
<i>pgd</i>	10	0.66	0.0187	0.2344	1.65	0.616
<i>asat</i>	16	0.771	0.0186	1.77	1.07	0.619
<i>nh1</i>	17	0.88	0.0064	1.18	2.76	0.601
<i>gpi</i>	16	0.779	0.017	1.49	2.21	0.550
<i>icd</i>	10	0.781	0.008	0.049	19.1	0.585
<i>fh</i>	14	0.642	0.0874	1.62	0.58	0.626
<i>mpi</i>	12	0.807	0.0185	1.10	2.85	0.604
<i>g6pdh</i>	12	0.844	0.0350	1.42	0.93	0.558
<i>pmm</i>	12	0.933	0.0905	2.77	1.47	0.520

DP: Discriminatory power

θ index (per site): The theta index is defined as an indicator of mutation rate per nucleotide site per generation.

dN/dS: Ratio of non-synonymous (dN) to synonymous (dS) substitution per nucleotide site.

Transition/Transversion Ratio (R) = This is the ratio of the number of transitions to the number of transversions for a pair of sequences.

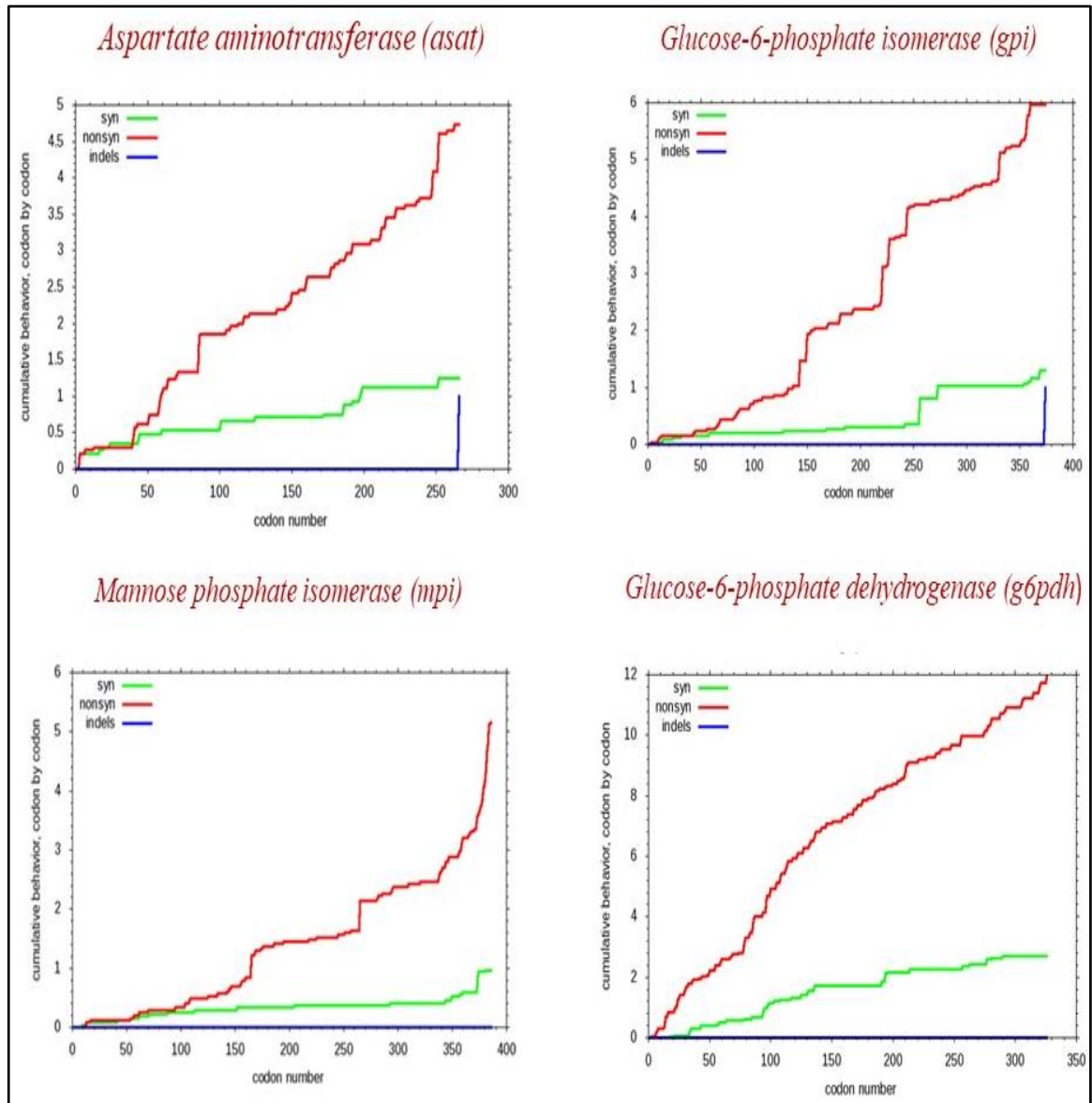


Figure 2.4. Cumulative dN/dS graph with InDels: Graphical representation of interspecies non-synonymous (nonsyn- red line) substitutions loci with a significant number of rare alleles based on the Tajima's D Value which ($P < 0.001$) is observed for variable regions of *asat*, and ($P < 0.05$) for *gpi*, *mpi* and *g6pdh* gene loci respectively.

- Eight out of 15 gene markers showed ambiguous sites at *aco* (8 sites), *g6pdh* (1 site), *pgm* (5 sites) and *pgd* (2 sites) with non-synonymous change and *asat* (1 site), *icd* (4 sites), *mpi* (4 sites) and *alat* (1 site) with synonymous change (Table 2.4).

Table 2.4: Ambiguous sites within *Leishmania sp.* for eight studied MLST genes.

Gene	Strain	Nucleotide position/nucleotide								Amino acid/ position						pl	Mw		
aco		26	27	99	109	330	351	352	439	9	33	37	110	117	118	147			
	M6426	G/T	A	C/T	A	A	A	A	C	G	L	E	D	Q	N	P	5.57	18823.82	
	K27	A	A/G	C	A/G	G	A/G	A/G	C	S	S	E	G	R	N	P	5.36	18756.54	
	CRO9	T	A	C	A	A/C	A	A	C	C	S	R	D	Q	N	P	5.51	18810.7	
	NNO11,RDO12a,RDO12a,RT014	T	A	C	A	A	A	A	C/T	C	S	R	D	Q	N	P	5.51	18810.7	
	H1,P6,BK,KK,A2,P2,T2,T8,T9,T10	A	A	C	A	G	A	A	C	N	S	R	G	Q	N	P	5.80	18832.56	
alat		297								99									
	NNO11,DSO7,MSO6,RDO12a,RDO12b,RT014	C/T								N						6.14	19181.37		
	LC39	T								N						6.14	19181.37		
	H1,A2,P2,P6,KK,BK,T2,T8,AG83,BPK282A1,LdCL,DD8,,T10	C								N						6.31	19308.62		
	T9	C								N						6.31	19260.58		
	T5	C								N						7.48	19241.62		
	HU3,MDP1,NNP2,JPCM5,TR01	C								N						6.31	19338.65		
5ASKH,Friedlin	C								N						6.32	19393.69			
pgm		81	116	156	247	393							27	39	52	83	131		
	CRO9	A/G	A	C	G	C							V	F	L	V	V	4.91	15317.4
	K27	G	A	C/T	C/G	G/T							V	Y	L	L	M	4.82	15386.47
	UA301,U1103	G	A	C	G	G							V	Y	L	V	V	4.72	15304.3
	NNO11,AFO15,OOO8,MSO6,RDO12a,RDO12b,RT014	G	A/T	C	G	C							V	F	L	V	V	4.91	15317.4
	NNO10,DSO7,	G	T	C	G	C							V	F	L	V	V	4.91	15317.4
	H1,A2,P2,P6,KK,BK,T2,T8,T9,T10,M1	G	A	C	G	G							V	Y	L	V	I	4.82	15339.45
Gerbilli	G	A	C	G	G							V	Y	L	V	M	4.82	15244.43	
pgd		167				720				56				240					
	LEM3946	C/T				C/T				X				G				6.95	33720.5
	HU3,3S,LEM3429,LEM3472,LEM3463	T				T				A				G				7.06	67441.61
	H1,A2,P6,KK,BK,T9,T10,AG83	C				C				A				G				6.95	33680.25
asat		187								63									
	Francesca	C/T								Y								7.30	43268.54
	H1,A2,P2,P6,KK,BK,T2,T9,T10,M1	C								Y								7.27	43159.45
icd		24	102	237	321							8	34	79	107				
	LEM3429	C/T	A/G	C/T	C							V	S	N	G	9.21	19846.16		
	HUSSEN,LEM3463	T	G	C	C							V	S	N	G	9.21	19846.16		
	LEM3946	T	A/G	C/T	C							V	S	N	G	9.21	19846.16		
	LEM3472,3S,GILANI	C	A	T	C							V	S	N	G	9.21	19846.16		
	GEBRE1	C	A	T	T							V	S	N	G	9.21	19846.16		
	DEVI	C	G	T	C/T							V	S	N	G	9.21	19846.16		
	JPCM5,LEM75,BCN16,PM1,LPN114,LSL29,JMT260,LEM2298,LLM373,LLM175,RRR-B,H1,A2,P6,KK,BK,T9,DD8	C	G	T	T							V	S	N	G	9.21	19846.16		
mpi		257	349	594	755							86	117	198	252				
	RRR-B	C/T	A	A	C/T							T	P	K	L	5.69	29604.01		
	ISS510	C	A	A/G	C							T	P	E	L	5.43	29604.96		
	ISS800	C	A	G	C							T	P	E	L	5.43	29604.96		
	LEM3429,LEM3463	T	A/G	A	T							T	P	K	L	5.69	29632.03		
	H1,A2,P2,P6,KK,BK,T8,T9,T10,M1,DD8	T	A	A	T							T	P	K	L	5.55	29603.97		
g6pdh		300								100									
	LSL29,LEM2298,LEM3249,LLM373,LEM935	A/T								M								6.38	36810.32
	Friedlin	G								R								6.3	36825.19
H1,A,P6,KK,DD8,AG83	A								K								6.74	36937.36	

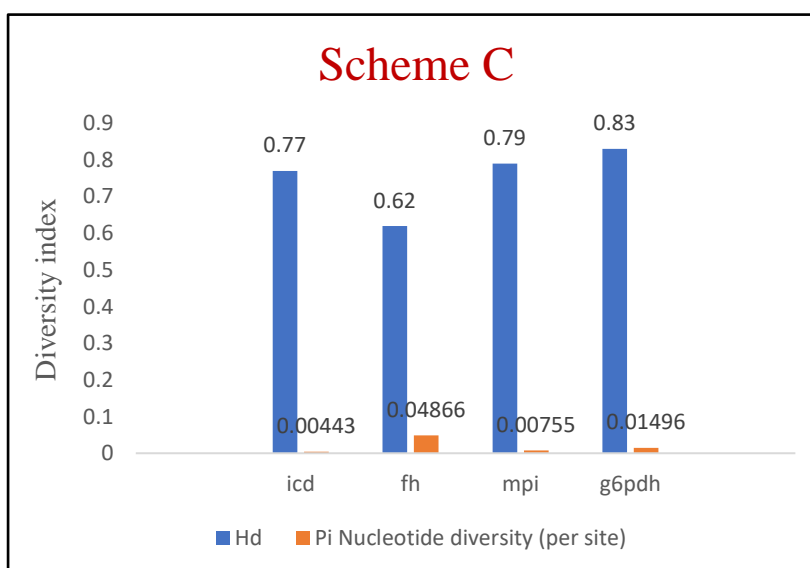
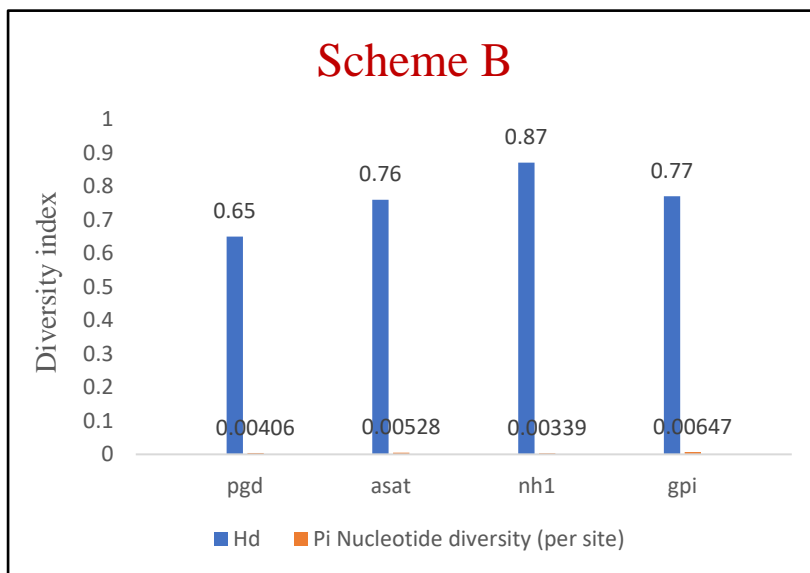
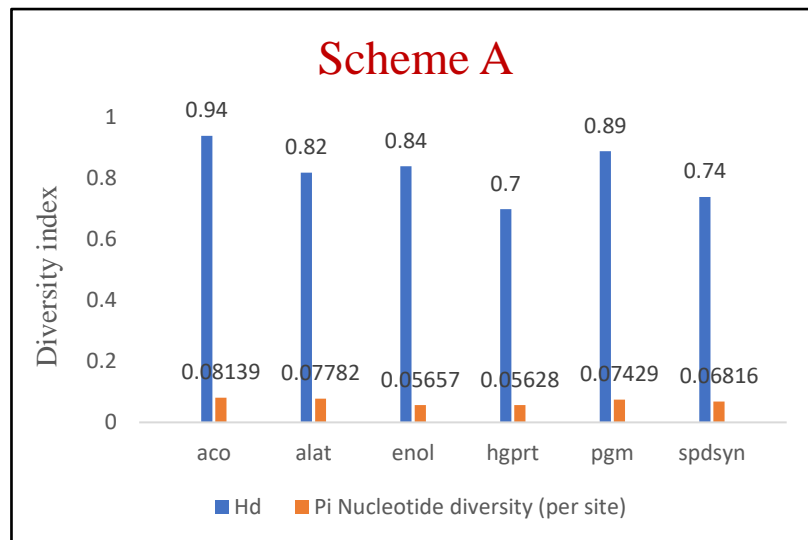


Figure 2.5. Comparison of interspecies haplotype diversity (Hd) and nucleotide diversity (pi) indexes for the each housekeeping genes in the studied MLST scheme. The highest value of nucleotide diversity (Orange bar), are shown by Scheme A housekeeping genes (*aco*, *alat*, *enol*, *hgprrt*, *pgm* and *spdsyn*) while the highest values of haplotype diversity (Blue bars), shown by *aco* gene (Scheme A) and lowest by *pgd* gene (Scheme B) and *fh* gene (Scheme C).

2.3.3. Analysis of Sequence Type (ST) and Haplotype distribution

- The sequence typing was analysed by determining the number of alleles per locus. The SNPs generated 23,14,21,9,18,9,10,16,17,16,10,14,12,12 and 12 alleles in *aco*, *alat*, *enol*, *hgprt*, *pgm*, *spdsyn*, *pgd*, *asat*, *nhl*, *gpi*, *icd*, *fh*, *mpi*, *g6pdh* and *pmm* loci respectively (Table 2.3 B).
- Using the concatenated dataset, the number of alleles varied among the locus and described 23, 26 & 23 ST with a DP value of 0.953,0.958 & 0.951 in each MLST Scheme A, B and C respectively. Some Indian isolates represented unique STs in each concatenated dataset, ST 1-9 in Scheme A, ST 1-8 in Scheme B and in Scheme C ST1-4 & ST 7. Among all loci, *spdsyn*, *pgd* and *icd* with 7,9,8 alleles showed the least and *aco*, *nhl* and *g6pdh* with 19,11 & 14 alleles indicated the maximum polymorphism within each group of the entire data set (Table 2.5 A, B & C).
- Interspecies concatenated haplotype number and haplotype diversity showed *hgprt* and *spdsyn* as the least and *aco* as the most variable loci in the complete data set (Figure 2.6 A). Haplotype (Hap) networks based on the highest interspecies nucleotide diversity (Π) loci were shown in (Figure 2.6 B).

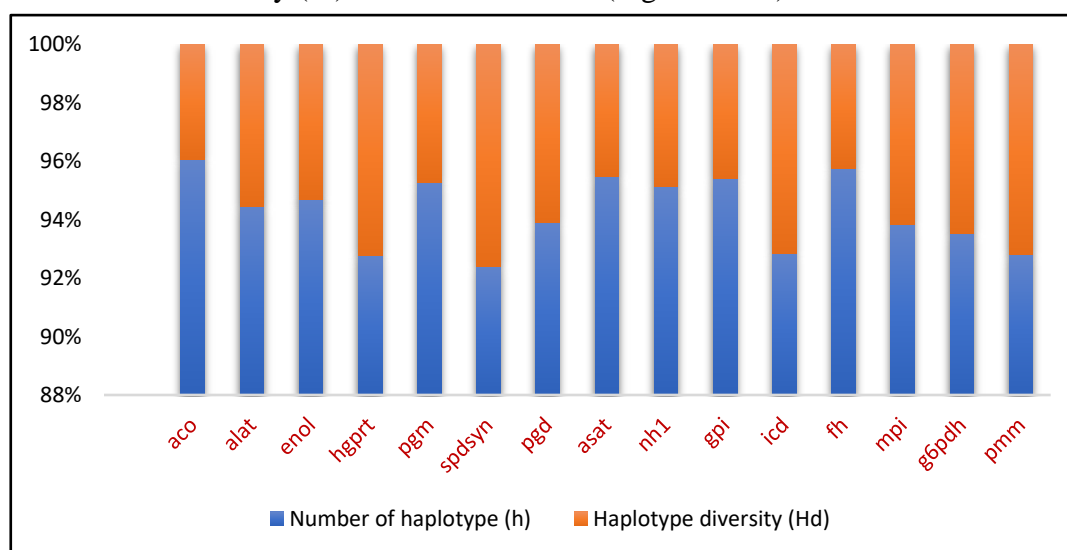


Figure 2.6. (A) Interspecies concatenated Haplotype diversity indices.

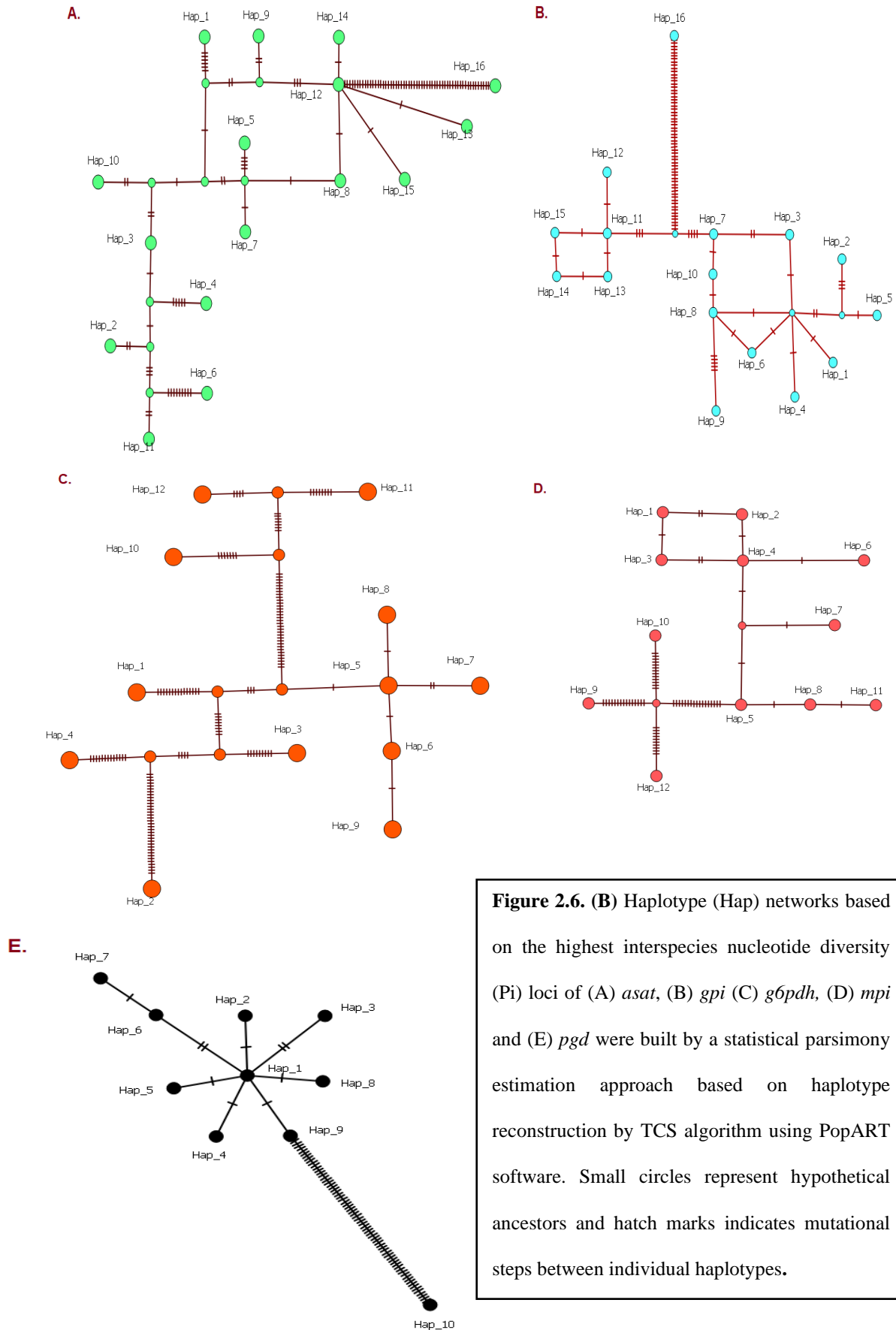


Figure 2.6. (B) Haplotype (Hap) networks based on the highest interspecies nucleotide diversity (P_i) loci of (A) *asat*, (B) *gpi* (C) *g6pdh*, (D) *mpi* and (E) *pgd* were built by a statistical parsimony estimation approach based on haplotype reconstruction by TCS algorithm using PopART software. Small circles represent hypothetical ancestors and hatch marks indicates mutational steps between individual haplotypes.

Table 2.5 (A) Concatenated sequence type (ST) of the MLST Scheme A.

Isolates	pgm	spdsyn	aco	alat	enol	hgprt	ST
H1	1	1	1	1	1	1	1
A2	2	1	2	1	1	1	2
P2	3	1	3	1	1	1	3
P6	3	1	4	1	1	1	4
KK	3	1	5	1	1	2	5
BK	3	1	6	1	1	1	6
T8	4	1	7	2	2	1	7
T9	1	1	8	1	1	1	8
T10	1	1	9	1	1	1	9
AG83	3	1	10	1	1	1	10
DD8	3	1	10	1	1	1	10
BPK282A1	3	1	10	1	1	1	10
LdCL	3	1	11	1	1	1	11
MDP1	5	2	12	3	3	1	12
NNP2	5	2	12	3	3	1	12
JPCM5	5	2	12	3	3	1	12
TR01	5	2	12	3	3	1	12
AFO15	6	3	13	4	4	3	13
CRO9	7	3	13	4	4	3	14
DSO7	6	3	13	4	4	3	13
MSO6	6	3	13	4	4	3	13
NNO10	6	3	13	4	4	3	13
NNO11	6	3	14	4	4	3	15
OOO8	6	3	13	4	4	3	13
RDO12a	6	3	14	4	4	3	15
RDO12b	6	3	14	4	4	3	15
RTO14	6	3	14	4	4	3	15
M2904	8	3	13	4	5	3	16
5ASKH	9	4	15	5	6	4	17
Friedlin	10	5	15	5	7	5	18
LC39	11	3	13	6	4	6	19
UA301	12	6	16	7	8	7	20
U1103	13	7	17	8	9	8	21
M4147	14	3	18	9	10	9	22
PSC-1	15	3	19	10	10	9	23
Number of Alleles	15	7	19	10	10	9	23
Number of Polymorphisms	85	59	107	101	64	52	468
Discriminatory Power	0.884	0.722	0.93	0.797	0.797	0.721	0.958

Table 2.5 (B) Concatenated sequence (ST) type of the MLST Scheme B.

Isolates	<i>asat</i>	<i>gpi</i>	<i>nhl</i>	<i>pgd</i>	ST
H1	1	1	1	1	1
A2	2	2	2	2	2
P6	3	3	3	1	3
KK	4	4	4	1	4
BK	5	5	5	1	5
T9	6	6	1	1	6
T10	7	7	6	3	7
DD8	8	8	7	4	8
DEVI	9	7	6	1	9
THAK35	9	7	6	1	9
SC23	10	9	8	1	10
LRC-L53	10	9	8	1	10
MRC74	10	9	9	5	11
LRC-L57	10	10	9	1	12
Mutinga H9	10	9	9	1	13
HU3	10	11	10	6	14
GILANI	11	11	10	1	15
LEM3463	11	11	6	7	16
GEBRE 1	11	11	10	1	15
LEM3946	10	12	6	6	17
Wangjie-1	10	13	11	8	18
MARZ-KRIM	10	9	12	1	19
HUSSEN	10	13	6	1	20
SUKKAR 2	10	13	6	1	20
BUMM3	10	13	6	1	20
LEM75	11	9	13	9	21
BCN16	11	9	13	9	21
PM1	11	9	13	9	21
LPN114	11	9	13	9	21
LSL29	11	9	13	9	21
IMT260	11	9	13	9	21
IPT-1	11	9	13	9	21
LEM2298	11	9	13	1	22
Strain A	11	9	13	1	22
LEM3249	11	9	14	1	23
LLM373	11	9	14	1	23
LEM189	11	9	14	1	23
BUCK	11	9	14	1	23
LLM175	11	9	14	9	24
ISS1036	11	9	10	1	25
Francesca	11	9	14	9	24
AG83	9	7	6	1	9
Number of Alleles	12	13	14	9	26
Number of Polymorphisms	29	18	21	10	78
Discriminatory Power	0.747	0.751	0.878	0.616	0.953

Table 2.5 (C) Concatenated sequence type (ST) of the MLST Scheme C.

Isolates	<i>fh</i>	<i>g6pdh</i>	<i>icd</i>	<i>mpi</i>	ST
H1	1	1	1	1	1
A2	2	2	1	2	2
P6	3	3	1	2	3
KK	4	4	1	2	4
AG83	5	5	1	3	5
DD8	5	5	1	3	5
JPCM5	5	6	2	4	6
DEVI	5	5	2	3	7
THAK35	5	5	1	3	5
GILANI	6	7	3	4	8
HUSSEN	5	7	4	3	9
LEM3429	5	7	5	5	10
GEBRE1	6	7	6	4	11
LEM3946	5	7	5	3	12
LEM3463	5	7	4	5	13
SC23	7	8	1	6	14
3S	6	7	3	4	8
LEM75	5	6	2	4	6
BCN16	5	6	2	4	6
PM1	5	6	2	4	6
LPN114	5	6	2	4	6
LSL29	5	9	2	4	15
IMT260	5	6	2	4	6
LEM2298	5	9	2	4	15
LEM189	5	6	1	4	16
LEM3249	5	9	1	4	17
BUCK	8	6	1	4	18
LLM175	5	6	2	4	6
ISS800	5	6	1	7	19
ISS1036	5	6	1	4	16
LLM373	5	9	2	4	15
LEM3472	5	7	3	3	20
K27	9	10	7	8	21
Friedlin	10	11	8	9	22
RRR-B	5	6	2	3	23
Number of Alleles	10	11	8	9	23
Number of Polymorphisms	324	130	14	52	520
Discriminatory Power	0.557	0.832	0.763	0.733	0.951

- Out of 15 studied loci only two *alat* and *enol* loci get amplified in the Indian *L.tropica* (T5) isolate, where the T5 isolate had unique ST3 for *alat* locus (Figure 2.7A) and for *enol* locus, the same ST6 shared by two *L. donovani* (T3, T7) isolates (Figure 2.7 B).

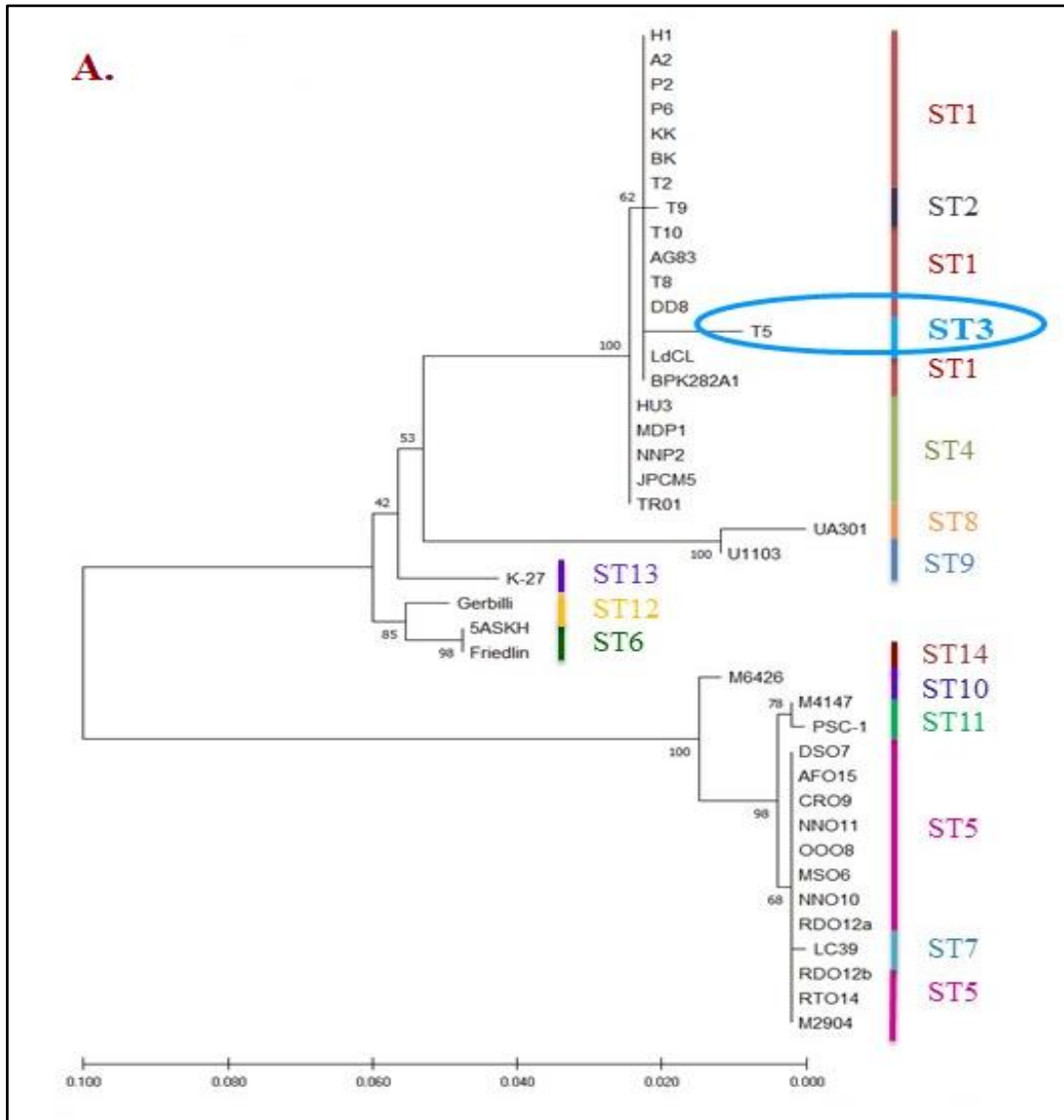


Figure 2.7. (A) Rooted phylogenetic tree of alanine aminotransferase (*alat*) locus from 41 isolates of the *Leishmania sp.* was constructed using the maximum likelihood method with 1000 bootstrap replicates. The colourful line in front of isolates indicates their respective sequence type (ST). Blue circled **ST3** uniquely present in the **Indian *L. tropica*** (T5) isolate of **VL patients**.

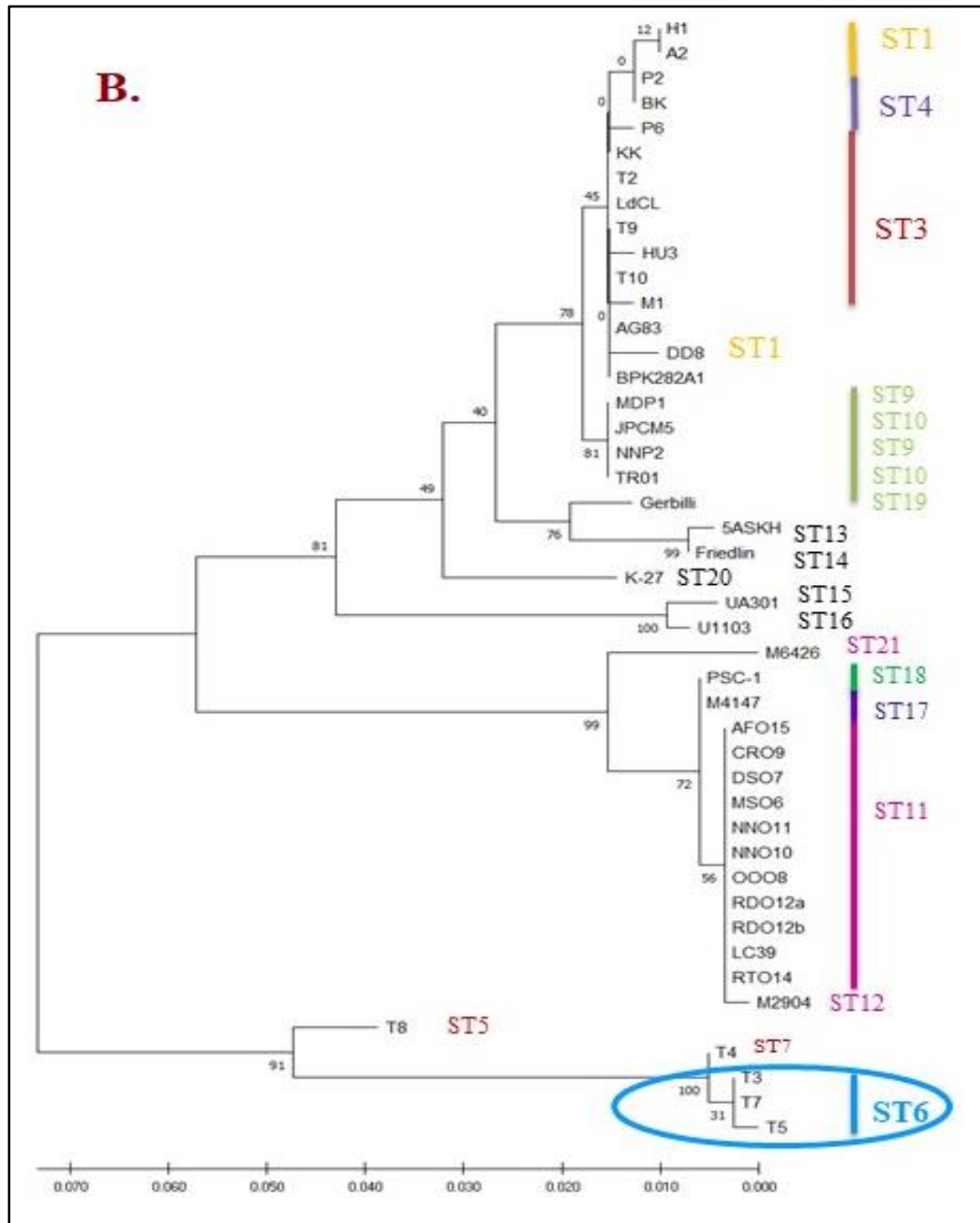


Figure 2.7. (B) Rooted phylogenetic tree of enolase (*enol*) locus from 44 isolates of the *Leishmania* sp. was constructed using the maximum likelihood method with 1000 bootstrap replicates. The colourful line in front of isolates indicates their respective sequence type (ST). Blue circled ST6 is notably shared by two Indian *L. donovani* isolates along with the Indian *L. tropica* (T5) isolate of VL patients.

2.3.4 Phylogenetic analysis

Phylogenetic analysis of the Scheme A concatenated sequence of six genes (*aco*, *alat*, *enol*, *hgprt*, *pgm* and *spdsyn*) could clearly differentiate two species within *Leishmania* species complexes and species level organisation with high bootstrap support. All branches reflected the geographical origin of respective *Leishmania* strains. Interestingly, one common cluster was formed by one Nepal (BPK282A1) and one Canada (LdCL) *L. donovani* strain (Figure 2.8). No statistically significant incongruence was detected among loci (ILDp = 0.009) and distance matrices had high and statistically significant Kendall's $W = 0.8970$ (p value= 0.01), supporting the concatenating loci dendrogram of MLST Scheme A. However, among all fifteen gene loci, the phylogenetic analysis of the individual locus isocitrate dehydrogenase (*icd*) sequence could clearly distinguish the geographical origin of *Leishmania* strains with maximum bootstrap support (Figure 2.9 A) and the locus phosphomannomutase (*pmm*) sequence clearly differentiates the *Leishmania* species complexes (Figure 2.10).

While phylogenetic analysis of Scheme C concatenated sequence of four genes (*icd*, *mpi*, *g6pdh* and *fh*) could clearly indicate the species-specific grouping of Indian *L. donovani*. At the same time, the other African *L. infantum* (3S) grouped with Ethiopian *L. donovani* (GILANI) and Sudanese *L. archibaldi* (GEBRE 1). Malta strain BUCK is intermediate between European strains (Figure 2.9 B). In the rooted maximum likelihood (log - 855.1867) phylogenetic tree of individual locus isocitrate dehydrogenase (*icd*), the position of one Indian *L. donovani* strain DEVI which takes a middle position in the European strain (Figure 2.9 A) which was resolved during their concatenated phylogeny analysis four gene *icd*, *mpi*, *g6pdh* and *fh* (Figure 2.9 B).

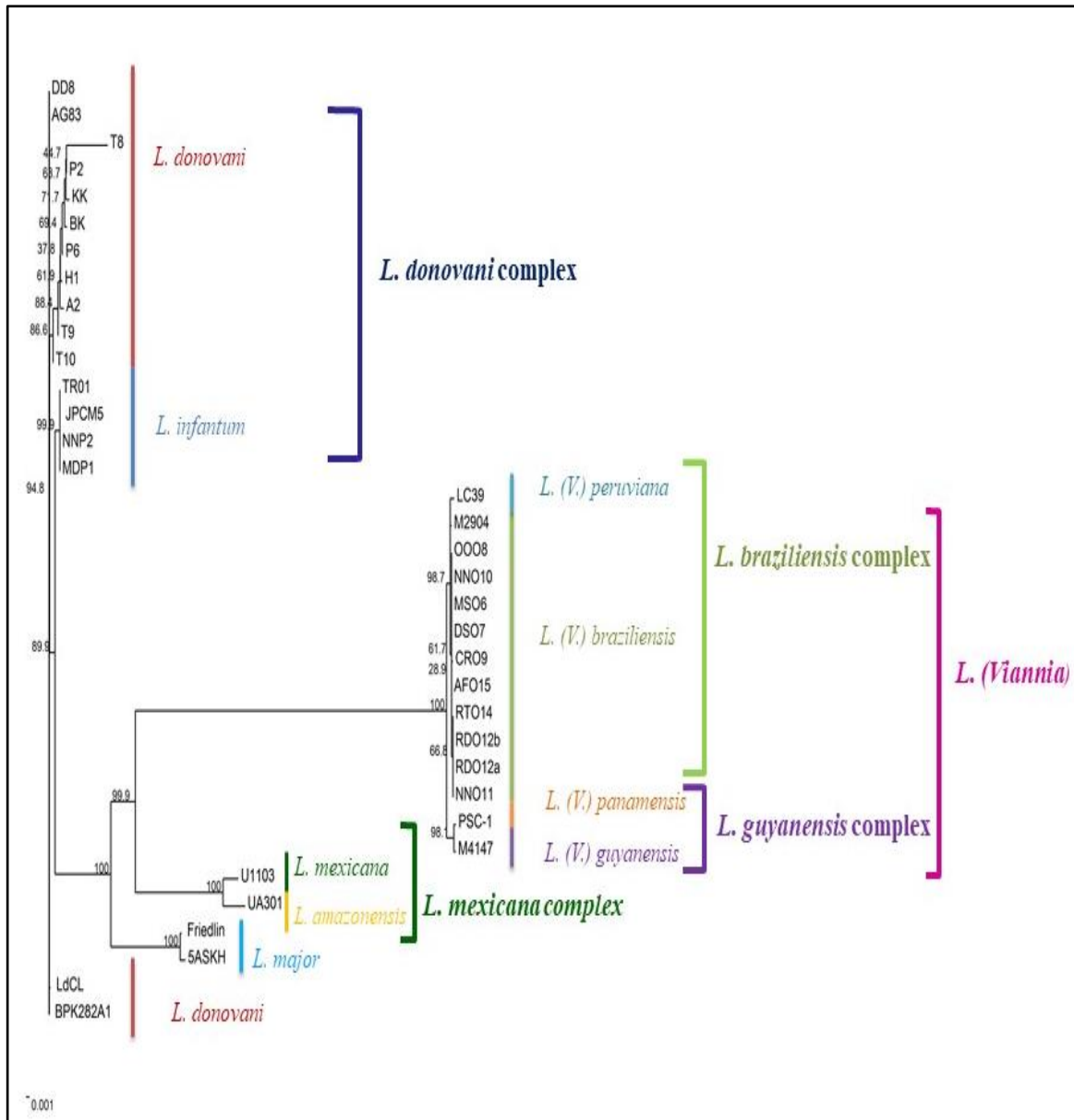


Figure 2.8. The phylogenetic analysis of six concatenated sequences loci of *aco*, *alat*, *enol*, *hgpri*, *pgm* and *spdsyn* could clearly differentiate the two species within the *Leishmania* species complexes and *L. (Viannia)* subgenus level with high bootstrap support by NJ method, 1000 bootstrap replicates using MLSTest software. The colourful line in front of isolates indicates their respective species status, species complex and subgenera.

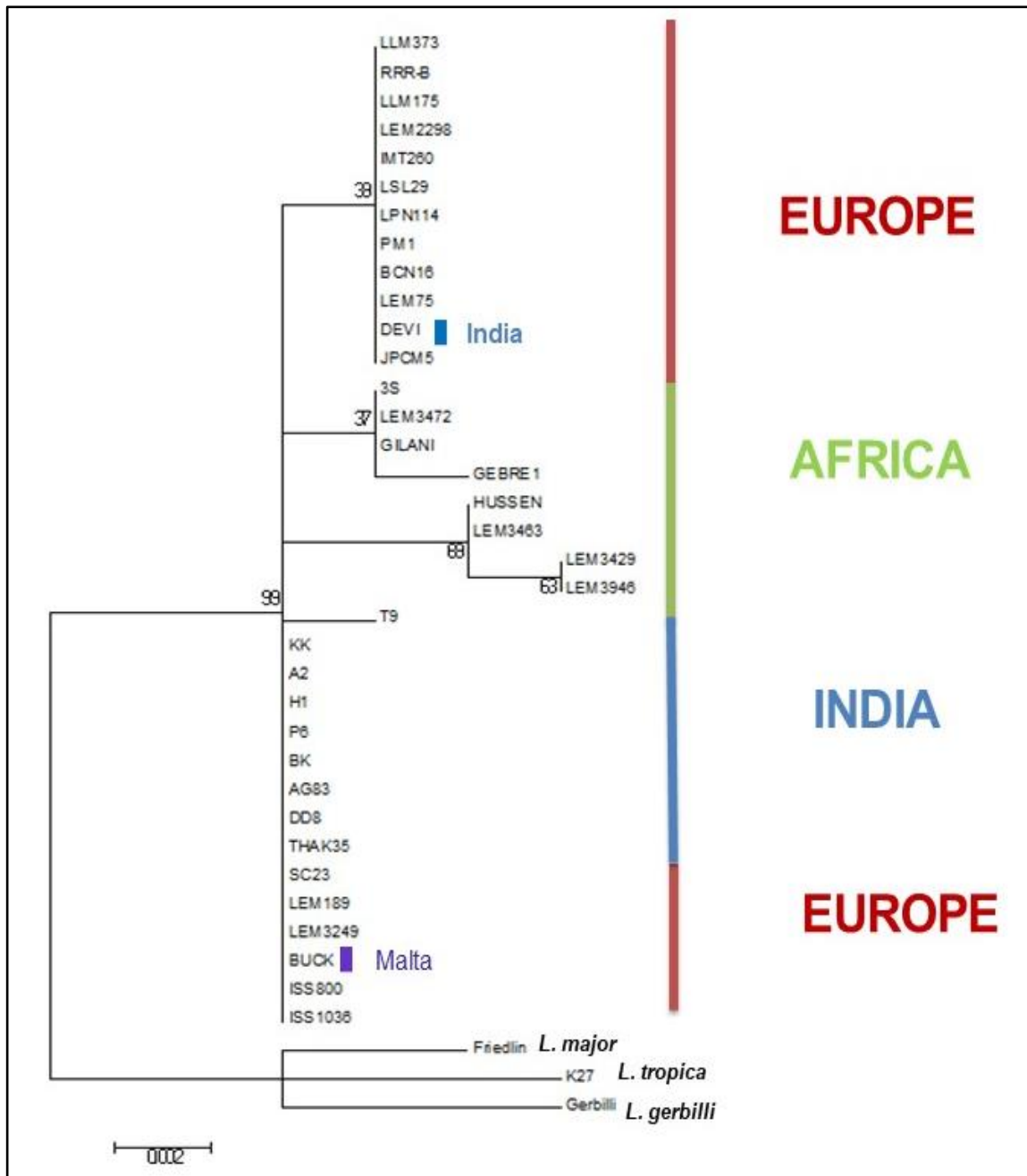


Figure 2.9. (A) Rooted maximum likelihood (log -855.1867) phylogenetic tree of isocitrate dehydrogenase (*icd*) locus from 38 isolates of the *Leishmania sp.* was constructed using the maximum likelihood method with 1000 bootstrap replicates. The colourful line in front of isolates indicates their respective geographical origin.

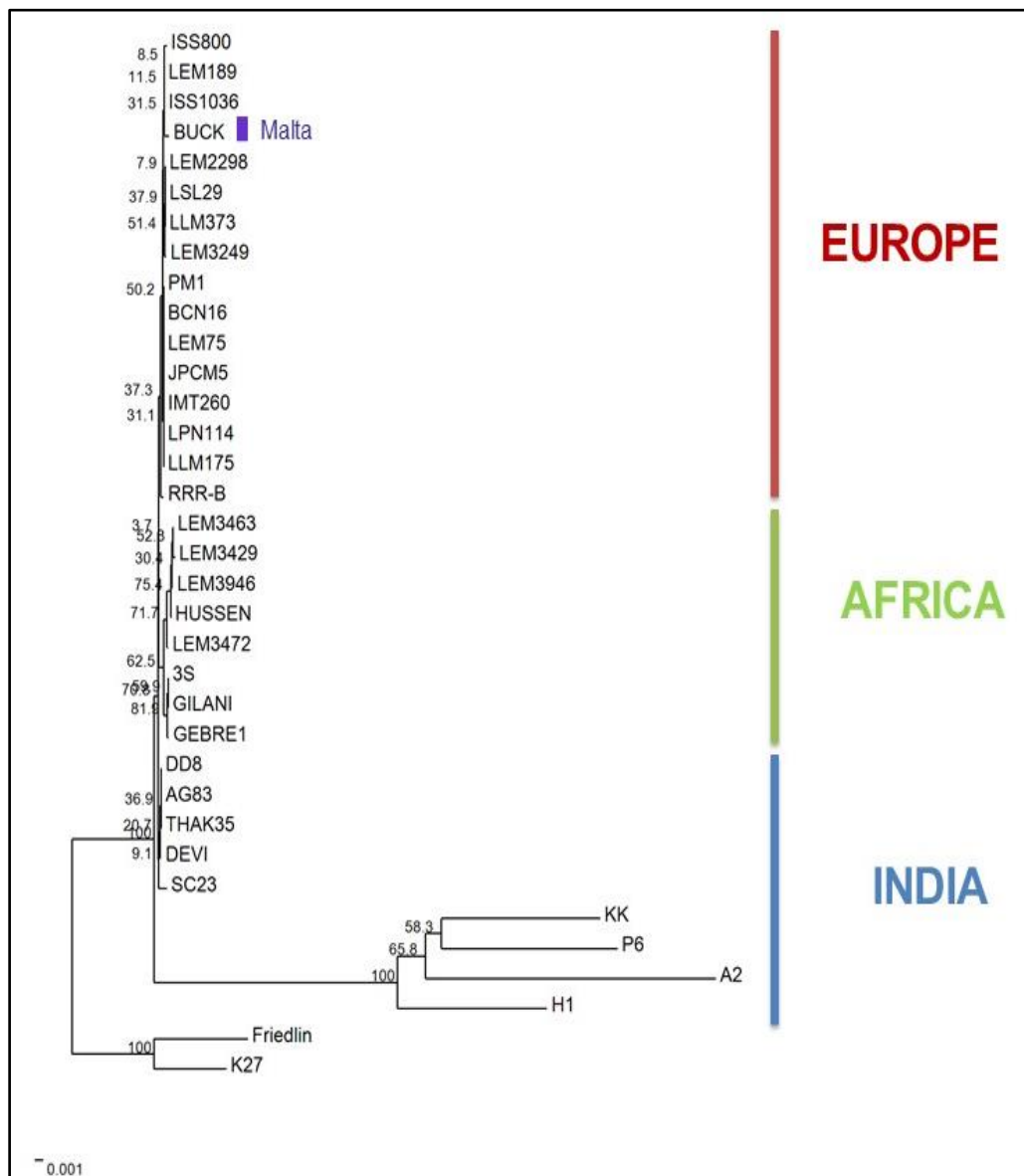


Figure 2.9. (B) The phylogenetic analysis of four concatenated sequences loci of *icd*, *mpi*, *g6pdg* & *fh* could clearly differentiate the *Leishmania* isolates according to their respective sub-continental origin with high bootstrap support by NJ method, 1000 bootstrap replicates using MLSTest software. The colourful line in front of isolates indicates their respective geographical origin. (ILDp = 0.07902; Kendall's W = 0.6075, p value = 0.01).

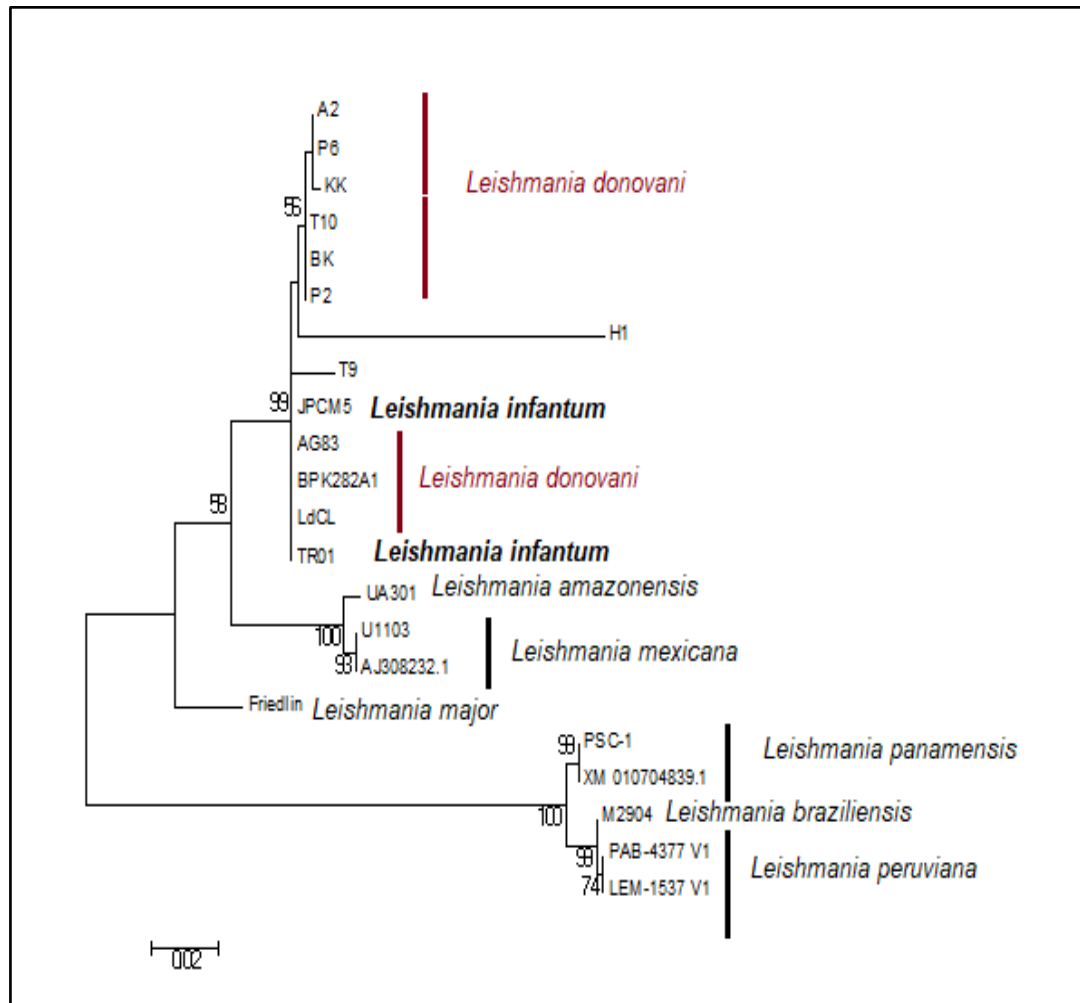


Figure 2.10. Rooted phylogenetic tree of phosphomannomutase (*pmm*) locus from 22 isolates of the *Leishmania sp.* was constructed using the maximum likelihood method with the highest log likelihood (-1520.9094) of 1000 bootstrap

2.4 Discussion

In the Indian subcontinent, *Leishmania donovani* was previously detected as a genetically homogeneous population (Alam et al., 2009) though some genetic variations in the Indian *Leishmania* population have recently been reported (Downing et al., 2011, 2012; Schonian et al., 2011; Imamura et al., 2016).

Therefore, a better and higher discriminatory method is necessary to identify diverse genotypes within closely related *Leishmania* strains from the Indian subcontinent because distinguishing genotypes within the closely related strains help to understand the clinical and epidemiological aspects of VL management. Several methods have been used to distinguish the *Leishmania donovani* complex. MLEE, is still the ‘Gold standard’ for parasite typing, but this method is less efficient in evaluating nucleotide changes and against poor discriminating efficacy to closely related strains (Ochsenreither et al., 2006; Schonian et al., 2011). In addition, Multilocus Microsatellite Typing (MLMT) has less competence for differentiating genome-wide SNP analysis among the Indian and Nepal isolates of *Leishmania donovani* (Downing et al., 2012). Hence, MLMT is unsuitable for explaining phylogeny (Schonian et al., 2011). Recently, for typing *Leishmania donovani* complex a PCR sequencing based powerful genotyping approach, Multilocus Sequence Typing (MLST) method is highly reproducible with results (Pérez-Losada et al., 2017). This approach is beneficial for the interspecies and intraspecies analysis of the microorganism, including the genus *Leishmania* (Akhoundi et al., 2016). MLST data analysis provides a more effective and easier approach for typing the microorganism population for a developing country like India. Moreover, to study population genetic structural analysis in a small population size, MLST is a better option with a low budget (Lytsy et al., 2017; Perez-Losada et al., 2017). Thus, the present study is focused on examining the genetic diversity and phylogenetic relationship among the *Leishmania* population of the Indian clinical isolates at the intraspecies and interspecies level by using fifteen enzyme-coding housekeeping genes (*aco*, *alat*, *enol*, *hgpri*, *pgm*, *spdsyn*, *icd*, *fh*, *mpi*, *g6pdh*, *pgd*, *asat*, *nh1*, *gpi*, *pmm*) that were previously identified as an MLST target for the genus *Leishmania*.

In the present study, we grouped fifteen genes in the three MLST schemes -Scheme A, (*aco*, *alat*, *pgm*, *spdsyn*, *enol* and *hgprt*), Scheme B (*asat*, *gpi*, *pgd* and *nh1*) and Scheme C (*icd*, *mpi* *g6pdh* and *fh*) respectively. At the same time, the phosphomannomutase (*pmm*) locus was examined separately, which was previously used for typing Argentina VL isolates and showed a 100% identity with genomic sequences of Argentina isolates (Marco et al., 2015). The *pmm* locus was examined in 8 Indian clinical isolates and the 14 reference strains. At the intraspecies level, the *pmm* locus indicated a statistically significant ($P < 0.05$) Tajima's D value with six haplotypes. On the other hand, at the interspecies level, the situation is different. We observed a statistically non-significant ($P > 0.10$) Tajima's D value with twelve haplotypes.

The intraspecies and interspecies nucleotide polymorphism of the Indian clinical isolates (n=15) was studied in the context of reference strains listed in Table 2.5 A, B & C for the coding sequence of the fifteen housekeeping genes of the two subgenera and 12 species Table 2.1. A discrete gamma distribution was observed in the intraspecies genetic diversity analysis of the Indian *Leishmania donovani*. It was particularly high for *enol*, *pgm*, *icd* and *spdsyn* loci indicating low variation of substitution rates among sites, but very low for *aco*, *asat*, *nh1*, *gpi*, and *mpi* indicating high variation of substitution rates respectively. As the spermidine synthase (*spdsyn*) gene showed no intraspecies polymorphism, this locus is denoted as the conserved locus in the Indian *Leishmania donovani* population. The highest nucleotide diversity (P_i) was observed in three locus *fh* followed by *enol* & *g6pdh* with 8,3 and 5 haplotypes, respectively. The intraspecies allelic profiles of each ten *L. donovani* isolates (H1, A2, KK, BK, P2, P6, T2, T8, T9, T10) along with one WHO references strain (DD8) for the seven housekeeping genes *aco*, *alat*, *asat*, *enol*, *gpi*, *nh1* and *pgm* respectively, were observed with Discriminatory Power (DP) value of 1.0, which indicated that the typing method was able to distinguish each member of a strain population from all

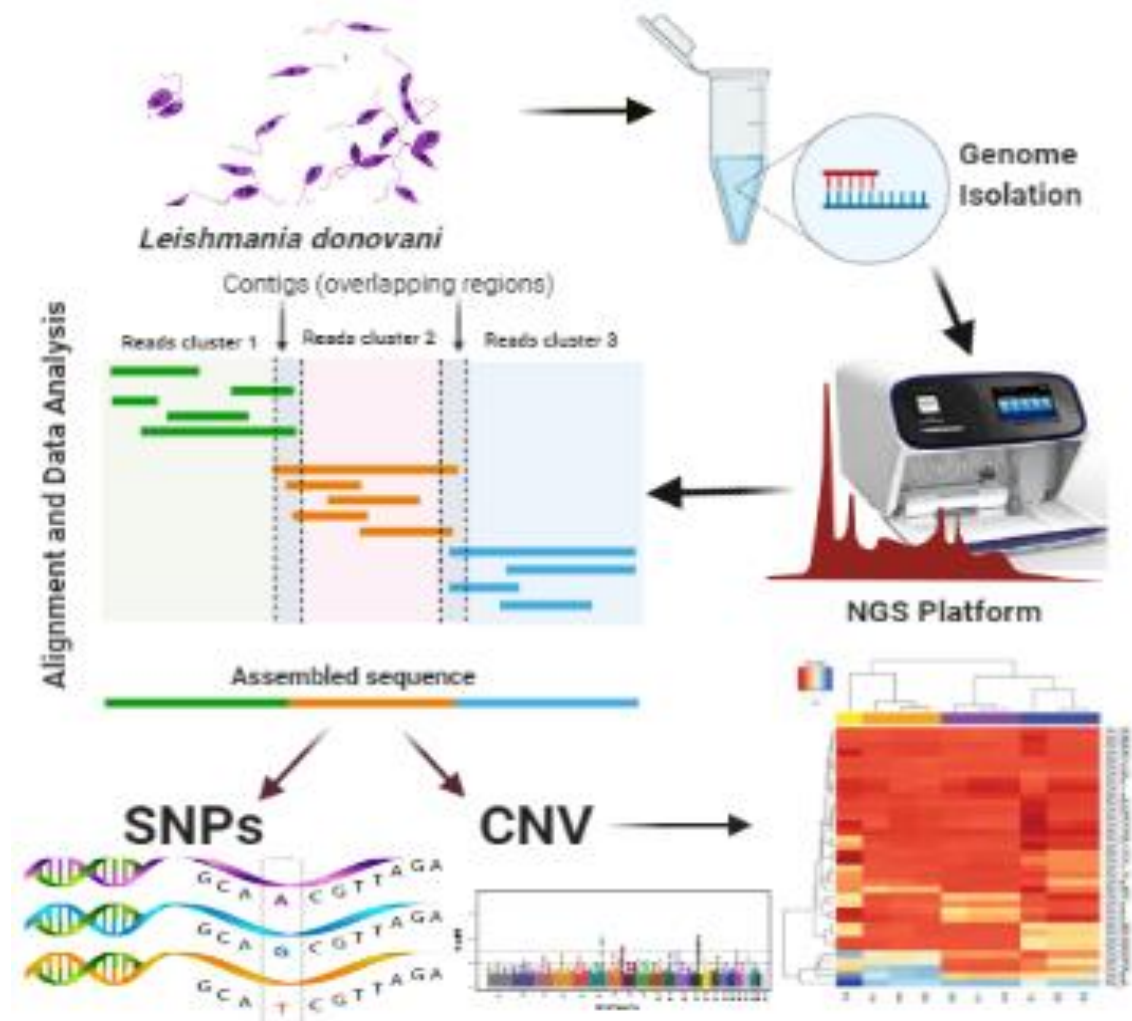
other members of that population. Phylogenetic analysis based on the concatenated sequence of the above seven housekeeping genes (*aco*, *alat*, *asat*, *enol*, *gpi*, *nh1* and *pgm*) revealed four distinct clusters indicating the existence of intraspecies diversity among the Indian VL and para-KDL clinical isolates of *L. donovani*.

In the interspecific genetic diversity parameters analysis, we observed one neutrally evolved enzyme 6-phosphogluconate dehydrogenase (*pgd*), which has a significantly high value (200) of the gamma parameter, implying a very low variation of substitution rates in the coding sequence. While the shape parameter of gamma distribution was particularly very low for *nh1* (0.0854) indicating a very high variation of substitution rates among sites. The analysed sequence type (ST) of each fifteen housekeeping gene indicates that the *aco* gene has the highest ST, i.e., 23 and the *hgprt* gene has the lowest ST, i.e., 9 in the studied *Leishmania* strains. The housekeeping gene loci have 18-385 polymorphic sites among the studied group of *Leishmania sp.*, with *hgprt*, *icd* & *nh1* having the least and *aco*, *asat* & *fh* the most segregating regions. The highest interspecies nucleotide diversity (P_i) was shown by five gene loci *asat*, *gpi*, *g6pdh*, *mpi* and *pgd* respectively, which is explained by Haplotype (Hap) networks in Figure 2.6. B. The average G+C content was higher than 50% for all fifteen genes. The estimated Transition/Transversion bias (R) was mainly above one for all genes except *fh* & *g6pdh* ($1 > R > 0.5$), implying that transition and transversion do not occur at the same rate, and the highest R -value in *icd* indicates the substitution pattern, i.e., the rates of transition are greater than transversion. The varied ratio of nonsynonymous/synonymous rate (dN/dS) was also observed in the studied genes. The significant Tajima's D Value at $P < 0.001$ is observed for variable regions of *pgd*, *asat*, and at $P < 0.05$, the significant Tajima's D Value is observed for variable regions of *gpi*, *mpi* and *g6pdh* gene loci, signifying the more significant number of rare alleles (Table 2.3 A).

Further, analysis of *Leishmania* isolates denoted ambiguous sites in eight out of fifteen genes namely *aco* (8 sites), *g6pdh* (1 site), *pgm* (5 sites) and *pgd* (2 sites) loci with non-synonymous changes and *asat* (1 site), *icd* (4 sites), *mpi* (4 sites) and *alat* (1 site) loci respectively with synonymous changes which validated the result of Tajima's D (Table 2.4). Thus, ambiguous sites in the studied genes revealed the genetic variants of the studied *Leishmania* population. Using the concatenated dataset, the number of alleles has been observed varied among the locus and is described as 26, 23 & 23 ST with a DP value of 0.953, 0.958 & 0.951 in each MLST Scheme A, B and C respectively. Here, some of the Indian isolates showed the unique STs in each concatenated dataset of interspecies analysis, ST 1-9 in scheme A, ST 1-8 in scheme B & scheme C ST1-4 & ST 7. Meanwhile, *pmm* locus ST2 and ST3 were shared by two isolates A2 (VL), P6 (para-KDL) and BK, P2 (VL) respectively. Interspecies concatenated haplotype number (h) and haplotype diversity (Hd) showed *hgprt* and *spdsyn* as the least and *aco* as the most variable loci in the complete data set. Our study on the interspecies phylogenetic analysis of the concatenated sequence of Scheme A loci (*aco*, *alat*, *pgm*, *spdsyn*, *enol* and *hgprt*), differentiated between the two species within *Leishmania* species complexes and subgenus level organization with high bootstrap support. Interestingly, one common cluster was formed by one Nepal (BPK282A1) and one Canada (LdCL) *L. donovani* strain indicating the need for a wider range of sample size to explain the global epidemiology of the *Leishmania* species population. While phylogenetic analysis of the Scheme B concatenated sequence could clearly indicate Indian *L. donovani* species-specific grouping corroborating the previous report (Zemanova et al., 2007) that the African *L. infantum* (3S) grouped with Ethiopian *L. donovani* (GILANI) and Sudanese *L. archibaldi* (GEBRE 1) and the position of Malta strain BUCK as an intermediate position between European strains.

Although we observed four ambiguous sites in the *icd* locus, but their theoretical molecular mass and *pI* remain unaltered (Table 2.4), which supports the previously reported monomorphic nature of the ICD gene (Zemanova et al., 2007). Previously, no interspecific variation was observed in the *pmm* locus of the Argentina VL isolates (Marco et al., 2015), while we observed 146 polymorphic sites and 12 haplotypes in the *pmm* locus and single-locus phylogeny with high bootstrap support, which explains the species status of 22 *Leishmania* isolates (figure 2.10). Still, for the locus *pmm* the status of two Indian *L. donovani* isolates (H1 and T9) was unresolved. The nucleotide sequence amplified herein to be employed in the MLST approach is submitted to (NCBI), U.S. National Library of Medicine USA, by following GenBank with the following accession numbers for; MN398901, MN883540-MN883559, MT625163-MT625221, MT634209-MT634224 & MT741860-MT741907, MT748045-MT748055.

The present study revealed four distinct intraspecies phylogenetic clusters and unique Sequence Type (ST) of Indian *Leishmania* isolates with a discriminatory power value of 1.0 in the coding region of the studied housekeeping genes. Although the observed SNP variations are low in our studied group, our data indicated a significant intraspecies genetic diversity in the concatenated sequences of the studied housekeeping genes. The interspecies phylogenetic analysis of the Indian clinical isolates for the individual locus alanine aminotransferase (*alat*) would be able to distinguish Indian VL isolates of *L. donovani* from *L. tropica* isolates by a unique sequence type, ST3 as it was seen for the T5 isolate (Figure 2.7 A). Thus, the MLST approach is a strong tool for differentiating intraspecies and interspecies variation.



CHAPTER 3

Whole Genome Sequence Analysis of the *Leishmania donovani* Strains from Indian VL and para-KDL Clinical Isolates

3.1 Introduction

Leishmaniases are caused by at least 20 species of the genus *Leishmania* and spread across the mammalian hosts by the female phlebotomine flies, ranging from self-healing cutaneous lesions to potentially fatal visceral forms (Burza et al., 2018). The group of neglected tropical diseases is endemic in almost 100 countries, with about 200,000 reported cases of visceral leishmaniasis (VL) or Kala-azar (KA) annually, leading to as many as 50,000 deaths per year (Burza et al., 2018; den Boer et al., 2011; Lozano et al., 2012). Though VL, primarily caused by *Leishmania donovani*, is prevailing in 100 countries on four continents, 90% of the incidences occur in India, Bangladesh, Nepal, Iran, Brazil, Ethiopia and South Sudan (Burza et al., 2018; Croft et al., 2006). In India, the eastern part, particularly the state of Bihar, is the endemic focus of VL infection, where periodic infection is common. As the *Leishmania* parasites are opportunistic in nature, incidences of VL and HIV co-infection have also been reported. Moreover, clinical isolates of *L. donovani* from India show evidence of *Leptomonas* co-infection (Singh et al., 2013). Even though *L. donovani* is known to be the major causal agent, some recent reports indicate the association of *L. tropica*, *L. amazonensis* and *L. major* with VL manifestation (Khanra et al., 2012; Sacks et al., 1995; Thakur et al., 2018), suggesting complex and possible atypical nature of the diseases. VL was first reported in the Indian sub-continent in the 1820s (Gibson, 1983), affecting the poorest communities and causing epidemics in several regions (Boelaert et al., 2009). Although VL was nearly eradicated from the Indian sub-continent in the 1960s (Thakur, 2007), it resurfaced in 1977 and caused several successive major epidemics (Dye and Wolpert, 1988). One major reason for such incidences could be the increasing number of cases unresponsive to sodium stibogluconate (SSG), the first line of drugs for Indian KA (Jackson et al., 1990; Sundar, 2001; Sundar et al., 2006).

Unfortunately, most of the alternative drugs such as Amphotericin B and Pentamidine have similar limitations and have proven to be very toxic (Amato et al., 1998; den Boer et al., 2011). Miltefosine (MIL), primarily an oral cancer drug, was introduced in 2003 to treat KA patients, especially for the SSG unresponsive cases (Jha, 2006). However, within few years, MIL resistant cases have started appearing as well (Croft et al., 2006). As *L. donovani* causes a major part of the leishmaniasis burden both in terms of morbidity and mortality, particularly in East Africa and the Indian sub-continent (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. Interestingly, among the apparently cured KA 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 lesions (Burza et al., 2018). 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 patients are appropriately referred to as para-Kala-azar Dermal Leishmaniasis (para-KDL). Though VL associated with PKDL was rarely found in the Indian subcontinent previously (Zijlstra et al., 2003), more recently nine such cases have been reported from Bihar, India (Kumar et al., 2016). The remaining load of active VL in increasing number of PKDL cases requires more attention to understand the pathology, diagnosis and treatment regime of these emerging cases of (para-KDL). Though the PKDL symptom might not seem very harmful, the lesions on the patients harbour the parasites which could 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 as their reservoirs and are considered to be a major concern in the elimination programmes of KA.

Hence, it is important to elucidate the mechanism that assists parasite to become dermatotropic from viscerotropic. *Leishmania* parasites can re-shape their genomes rapidly in response to stress (Leprohon et al., 2009b), 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 leading to unfavourable responses to the treatment regimes. Keeping such objectives in mind, we report here the genome comparison analysis of 10 *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 the observed immune evasion during VL treatment leading to PKDL in apparently cured patients.

3.2 Outline of the work

3.2.1. Ethics statements

Collection of Bone marrow or splenic aspirates from KA 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.

3.2.2. Parasite culture

Ten clinical isolates from KA and para-KDL patients (clinically diagnosed as PKDL) of India. Detailed protocols of the parasite culture were explained in the Materials and Methods section.

3.2.3. Whole genome sequencing and data analysis

Isolation of genomic DNA from the cloned promastigote strains followed by high quality RNA free DNA libraries preparation. The enriched library was sequenced in Ion Proton Semiconductor Sequencer using Ion PI™ Chip V3 and Ion PI™ Hi-Q™ Sequencing 200 reagents. The SNPs analysis was carried out using low frequency variant detection option in CLC genomics workbench 12 (Qiagen). Detailed protocols of the whole genome sequencing and data analysis were explained in the Materials and Methods section.

3.2.4. Somy estimation

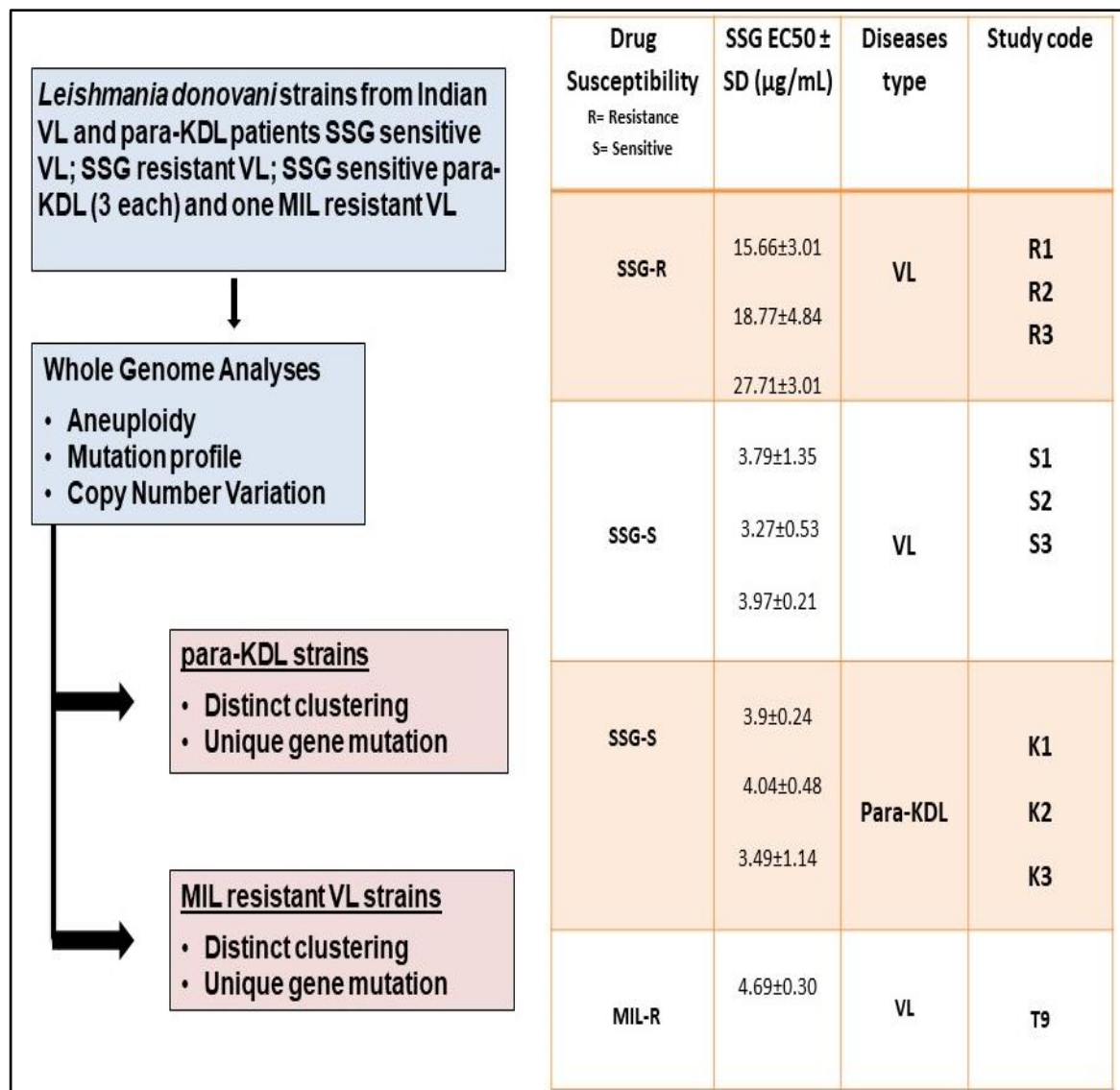
The chromosomal somy values were estimated by using the method based on variability of sequence depth and median read depth of the sample (Downing et al., 2011; Dumetz et al., 2017; Imamura et al., 2016). The somy values of each chromosome in a sample were plotted to create the clustered heatmap using the Rplot of heatmap2. Detailed protocols of the somy estimation were explained in the Materials and Methods section.

3.2.5. Gene clustering and pathway analyses

Principal component analysis (PCA) was performed based on the types of mutations 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 and the pathway analyses were 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). Detailed protocols of the gene clustering and pathway analyses were explained in the Materials and Methods section.

3.2.6. Graphical representation of the present study



3.3 Results

3.3.1. Whole genome sequencing

To investigate the genetic variation that could be responsible for manifestation of PKDL and development of drug resistance in Indian VL patients, whole genome sequencing (WGS) of 10 *L. donovani* strains from the clinical isolates 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) and one strain of Miltefosine resistant VL (VL-MIL-R) were cloned from the bone marrow or splenic aspirates from the patients of Bihar-West Bengal area of India during 2010-2014 (Table 7 in Material and Methods section). The strains were genotyped by PCR-RFLP or PCR-ITS1 sequencing methods and their drug sensitivities were 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 6 in Material and Methods section).

3.3.2. Aneuploidy

In order to analyse possible association of aneuploidy with the development of para-KDL 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). The estimated copy values (Table 3.1) showed that on an average 30 chromosomes were disomic in all strains except in VL-MIL-R strain, where the number was only 23. 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.

Table 3.1: Average chromosome ploidy of different clinical isolates derived based on median read depth analysis. (K: para-KDL-S; R; VL-SSG-R; S: VL-SSG-S; T9: MIL-R)

	Mean_K	Mean_R	Mean_S	T9
Ch01	1.72 (± 0.07)	1.93 (± 0.11)	1.82 (± 0.17)	1.74
Ch02	2.33 (± 0.04)	2.10 (± 0.07)	2.01 (± 0.03)	2.94
Ch03	1.80 (± 0.09)	1.98 (± 0.10)	1.93 (± 0.12)	1.78
Ch04	1.83 (± 0.05)	1.90 (± 0.04)	1.86 (± 0.03)	1.82
Ch05	2.64 (± 0.12)	1.93 (± 0.06)	1.81 (± 0.13)	1.71
Ch06	2.85 (± 0.15)	2.06 (± 0.07)	2.07 (± 0.11)	1.60
Ch07	1.90 (± 0.03)	2.01 (± 0.01)	2.25 (± 0.22)	2.76
Ch08	2.79 (± 0.08)	3.83 (± 0.23)	3.06 (± 0.04)	4.42
Ch09	1.92 (± 0.03)	2.13 (± 0.09)	2.47 (± 0.39)	2.40
Ch10	2.04 (± 0.02)	1.96 (± 0.11)	1.70 (± 0.22)	1.68
Ch11	2.03 (± 0.01)	1.98 (± 0.05)	1.69 (± 0.21)	2.51
Ch12	1.93 (± 0.02)	2.00 (± 0.03)	2.10 (± 0.19)	2.45
Ch13	1.96 (± 0.01)	1.99 (± 0.03)	2.26 (± 0.23)	1.60
Ch14	1.95 (± 0.03)	1.97 (± 0.03)	2.35 (± 0.04)	1.38
Ch15	2.81 (± 0.05)	2.00 (± 0.03)	1.84 (± 0.15)	3.03
Ch16	2.08 (± 0.02)	3.15 (± 0.09)	2.81 (± 0.43)	3.02
Ch17	2.02 (± 0.03)	2.02 (± 0.05)	1.95 (± 0.03)	1.98
Ch18	1.98 (± 0.03)	1.91 (± 0.08)	1.79 (± 0.29)	2.01
Ch19	1.87 (± 0.03)	1.86 (± 0.02)	1.63 (± 0.06)	1.85
Ch20	2.00 (± 0.01)	2.05 (± 0.07)	2.79 (± 0.14)	2.97
Ch21	1.98 (± 0.00)	2.00 (± 0.04)	1.99 (± 0.02)	1.95
Ch22	1.94 (± 0.02)	1.91 (± 0.03)	1.83 (± 0.17)	1.92
Ch23	2.46 (± 0.03)	3.16 (± 0.08)	2.24 (± 0.21)	4.08
Ch24	2.02 (± 0.01)	2.03 (± 0.02)	2.99 (± 0.07)	1.98
Ch25	2.01 (± 0.01)	2.02 (± 0.03)	2.94 (± 0.11)	1.93
Ch26	2.02 (± 0.01)	3.96 (± 0.10)	2.41 (± 0.11)	2.85
Ch27	2.00 (± 0.01)	1.97 (± 0.02)	1.79 (± 0.13)	1.92
Ch28	2.04 (± 0.02)	1.99 (± 0.04)	1.90 (± 0.16)	1.99
Ch29	1.95 (± 0.01)	1.90 (± 0.01)	1.82 (± 0.05)	2.86
Ch30	2.05 (± 0.02)	1.99 (± 0.01)	1.92 (± 0.05)	2.00
Ch31	4.03 (± 0.03)	3.82 (± 0.14)	3.69 (± 0.34)	4.97
Ch32	2.05 (± 0.02)	2.00 (± 0.01)	2.12 (± 0.05)	2.00
Ch33	1.98 (± 0.01)	2.01 (± 0.04)	2.96 (± 0.07)	2.91
Ch34	1.98 (± 0.02)	1.91 (± 0.01)	1.91 (± 0.09)	1.94
Ch35	1.93 (± 0.02)	2.74 (± 0.13)	2.05 (± 0.37)	2.35
Ch36	2.02 (± 0.02)	1.93 (± 0.05)	1.90 (± 0.09)	1.62

S < 0.5; Monosomy, 1.5 ≤ S < 2.5; Disomy, 2.5 ≤ S < 3.5; Trisomy
3.5 ≤ S < 4.5; Tetrasomy, 4.5 ≤ S < 5.5; Pentasomy (Sarraf et al., 2021)

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. 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).

➤ **Aneuploidy profile in MIL resistant VL strain**

Notably, the proportion of tetrasomy was estimated to be more in VL-SSG-R strains. Strikingly, higher proportion of aneuploidy were observed in the VL-MIL-R strain with the presence of monosomic and pentasomic chromosomes (Table 3.1), which could be significant in development of the resistance toward MIL. Comparative aneuploidy profiles of VL-SSG-S, VL-SSR-R and para-KDL-SSG-S with VL-MIL-R are explained in figure 3.1, where we observed that only in VL-MIL-R chromosome 31 showed pentasomy. A greater variability in aneuploidy was observed in VL-MIL-R strains, which could be important in development of the particular resistance (Figure 3.1).

➤ **Aneuploidy profile in SSG sensitive VL & para-KDL strains and SSG resistant VL strain**

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 SSG resistant VL strains were broadly clustered distinctly (Figure 3.2 and Table 3.1). 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).

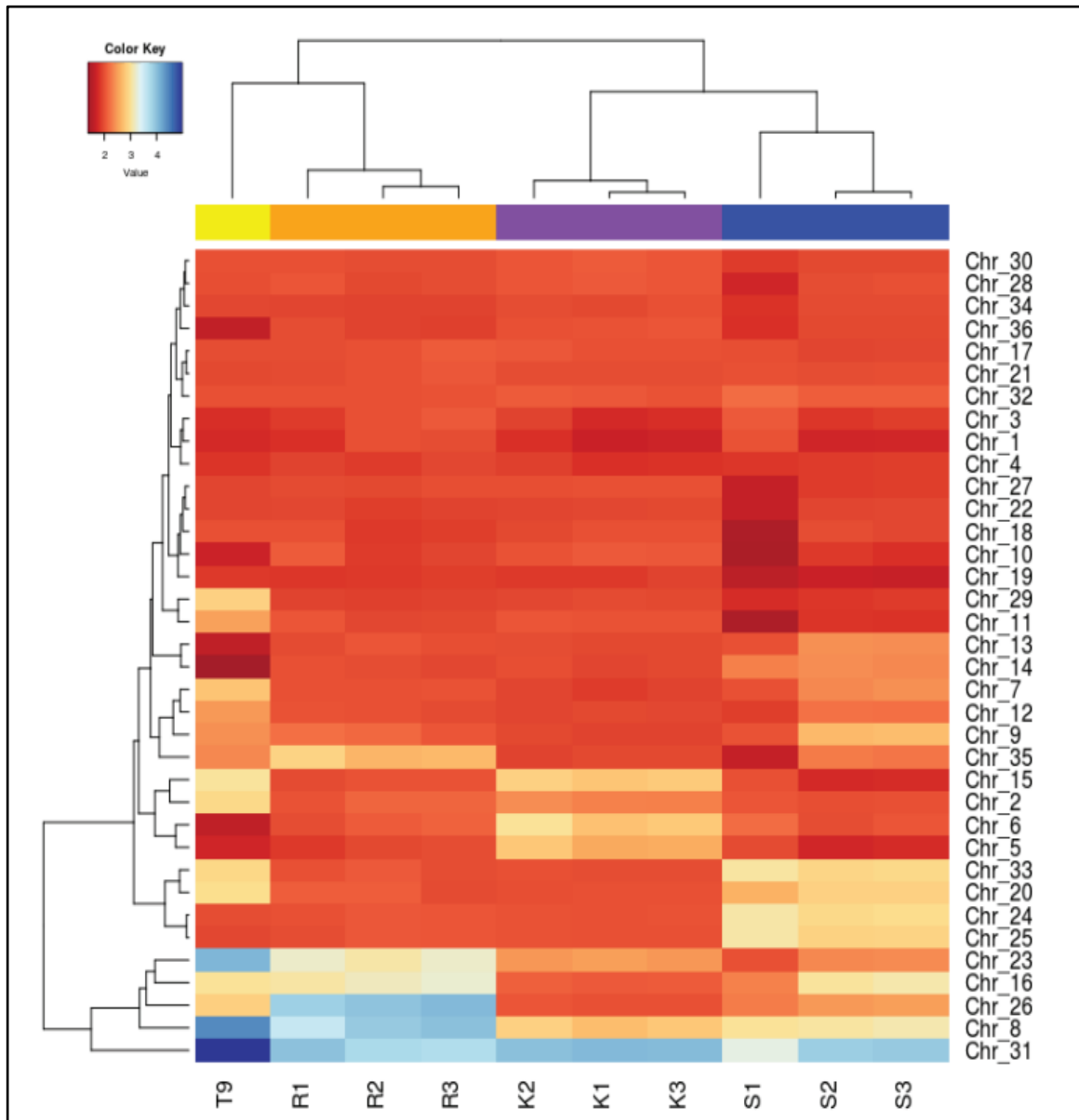


Figure 3.1. Comparison of aneuploidy profiles of the *L. donovani* strains obtained from VL-SSG-S, VL-SSG-R, VL-MIL-R and para-KDL-SSG-S clinical isolates. Chromosome numbers are indicated on the right side of the heat map and the study codes of the isolates are mentioned across the bottom of the map. The dendrogram on the top indicates the clustering of the strains based on the similarity of aneuploidy profile. Similarly, the dendrogram on the left represents the clustering of the chromosomes. The insert shows the colour code that is used to indicate the ploidy of chromosomes.

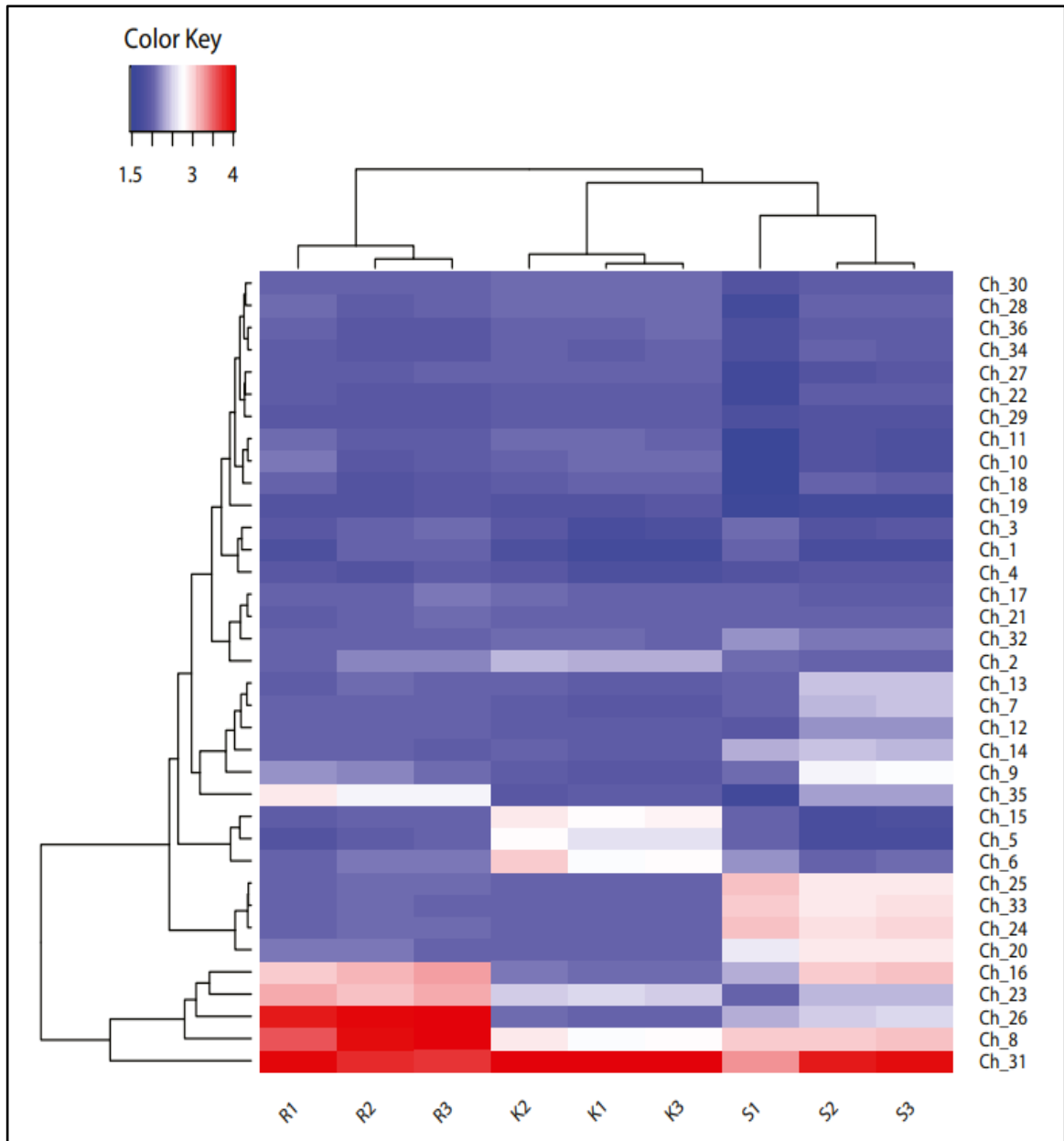


Figure 3.2. Comparison of aneuploidy profiles of the *L. donovani* strains obtained from the VL-SSG-S, VL-SSG-R and para-KDL-SSG-S clinical isolates. Chromosome numbers are indicated on the right side of the heat map and the study codes of the isolates are mentioned across the bottom of the map. The dendrogram on the top indicates the clustering of the strains based on the similarity of aneuploidy profile. Similarly, the dendrogram on the left represents the clustering of the chromosomes. The insert shows the colour code that is used to indicate the ploidy of chromosomes.

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 Figure 2, 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 3.1.

Table 3.2. Distinction between SSG sensitive para-KDL and VL strains based on average chromosome ploidy

Chromosome	Mean_K*	Mean_S*	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

* K, SSG sensitive para-KDL strains. S, SSG sensitive VL strains

Also, VL-SSG-S and para-KDL-S strains could be distinguished based on significant change of average chromosome some 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 (Figure 2 and Table 3.2). 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 some values (Figure 2 and Table 3.3). Strikingly, for certain chromosomes, unique variations were noted when some values of para-KDL-S and VL-SSG-R strains were compared with that of VL-SSG-S strains (Table 3.4). As shown in Table 3, for chromosomes 5, 6, 15 and 16, overall some changes were observed only between VL-SSG-S and para-KDL-S strains. On the other hand, some variation was found in chromosomes 8, 23, 26 and 35 only between VL-SSG-S and VL-SSG-R strains (Table 3.4). Thus, differential some changes in SSG sensitive strains could be important for transition to PKDL or development of the drug resistance.

Table 3.3. The distinction between SSG resistant and sensitive VL strains based on average chromosome ploidy

Chromosome	Mean_R*	Mean_S*	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

* R, SSG resistant VL strains. S, SSG sensitive VL strains

Table 3.4. Distinct change of chromosome ploidy among SSG sensitive VL and para-KDL strains and SSG resistant VL strains

Chromosome	Mean_S*	Mean_K*	Mean_R*
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

*S, SSG sensitive VL strains; K, SSG sensitive para-KDL strains, R, SSG resistant VL strains

3.3.3. Comparative mutational analysis between the coding sequences of drug sensitive and drug resistant VL and para-KDL strains

To estimate the contribution of genome-wide mutation profile of the protein coding genes in the development of para-KDL and drug resistance, the analyses of non-synonymous SNVs and InDels were particularly carried out in the present study. In the non-coding regions, on an average 6700 mutations were identified in each strain, ranging from only 2859 mutations in VL-SSG-R strain R2 to 11600 mutations VL-MIL-R strain T9 (Table 3.5 A). On the other hand, 3837 unique mutations (InDels, MNVs and SNVs) were identified in the protein coding genes across the strains (Table 3.5 B), which were non-synonymous and further analysed in more detail. Here also, the mutation burden was found to be the least in R2 and the most in T9. Overall, on an average only 2.8% mutations were homozygous (Table 3.5 C).

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). Interestingly, homozygous mutations in four genes were detected in the three para-KDL-S strains only (Table 3.6), suggesting their possible relation with the development of para-KDL symptoms.

Table 3.5 (A) Summary of mutation analysis in the non-coding genes of the isolates.

	<i>Deletion</i>	<i>Insertion</i>	<i>MNV</i>	<i>Replacement</i>	<i>SNV</i>	<i>Total</i>
<i>K1</i>	3206	111	149	16	1424	4906
<i>K2</i>	2664	276	168	19	1619	4746
<i>K3</i>	4070	153	144	18	1535	5920
<i>R1</i>	6068	385	173	19	1682	8327
<i>R2</i>	1080	130	157	16	1476	2859
<i>R3</i>	6375	304	181	17	1755	8632
<i>S1</i>	2511	197	218	25	1867	4818
<i>S2</i>	6628	867	193	19	1755	9462
<i>S3</i>	3201	580	185	17	1674	5657
<i>T9</i>	9575	158	182	15	1661	11600
<i>Average</i>	4537.8	316.1	175	18.1	1644.8	6691.8

(B) Summary of mutation analysis in the protein coding genes of the isolates.

para-KDL Strains

	K1		K2		K3	
	Heterozygous	Homozygous	Heterozygous	Homozygous	Heterozygous	Homozygous
Deletion	731	4	576	3	795	5
Insertion	23	0	91	0	39	0
MNV	49	2	63	6	45	4
Replacement	8	0	10	1	8	0
SNV	215	12	242	16	206	15
Total	1026	18	982	26	1093	24

VL-SSG-R Strains

	R1		R2		R3	
	Heterozygous	Homozygous	Heterozygous	Homozygous	Heterozygous	Homozygous
Deletion	557	5	184	2	932	5
Insertion	41	0	25	0	65	0
MNV	47	11	48	0	57	8
Replacement	9	1	8	0	9	1
SNV	201	24	192	11	227	28
Total	855	41	457	13	1290	42

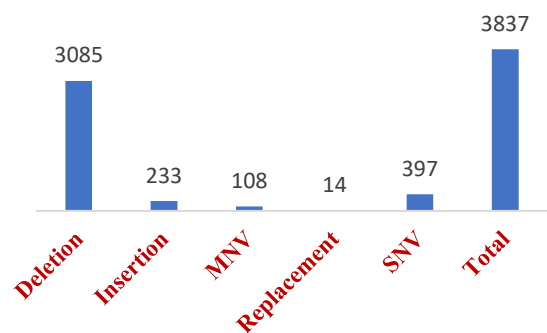
VL-SSG-S Strains

	S1		S2		S3	
	Heterozygous	Homozygous	Heterozygous	Homozygous	Heterozygous	Homozygous
Deletion	704	2	537	3	365	1
Insertion	37	0	92	0	77	0
MNV	60	13	62	8	56	11
Replacement	11	1	9	1	10	1
SNV	239	27	222	24	223	24
Total	1051	43	922	36	731	37

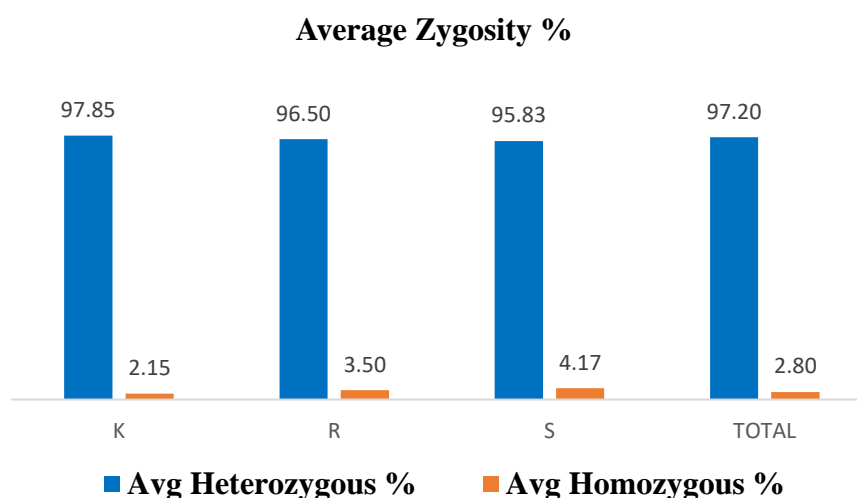
VL-MIL-R Strain

	T9	
	Heterozygous	Homozygous
Deletion	1068	6
Insertion	22	0
MNV	56	8
Replacement	9	1
SNV	206	25
Total	1361	40

Total Unique Mutation



	K Vs S	R Vs S	K+S Vs R
Odds ratio	1.946	1.1728	0.8677
95 % CI:	1.15356 to 3.2829	0.7281 to 1.8890	0.5250 to 1.4339
Z statistic	2.495	0.655	0.554
Significance level	P = 0.0126	P = 0.5123	P = 0.5797

(C) Summary of zygosity in the protein coding genes of the isolates.**Table 3.6 Summary of four homozygous mutations in para-KDL strains.**

GeneDB ID	Amino acid	Encoded Protein
LDBPK_130390	Ile1349Thr	RNA helicase, putative
LDBPK_210700	Ala562Val	phosphoglucomutase, putative
LDBPK_332650	Ala419Thr	UDP-N-acetylglucosamine pyrophosphorylase, putative
Unknown	Gln134Glu	Unknown

The extent of similarities among the different strains based on the pair wise analysis of whole genome non-synonymous mutation profiles in the coding sequences and profiles of the affected genes were shown in Figure 3.3. 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 and VL-MIL-R 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 Figure 3.3 A and B).

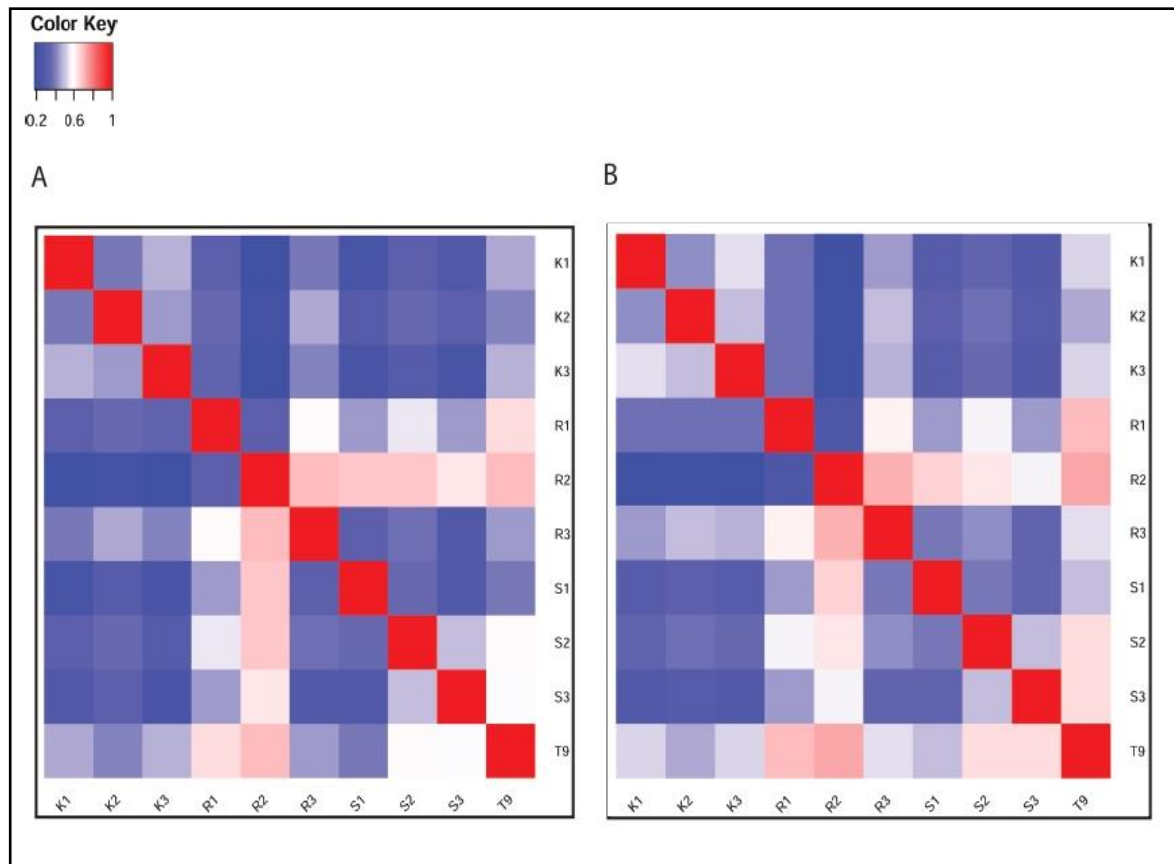


Figure 3.3. Comparison between mutation profiles of the *L. donovani* strains obtained from clinical isolates. The heat map visualizes the similarities between all mutations in the coding sequences of the sequenced strains. Each coloured square in the map represents the fraction of similarity between the 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

The para-KDL-S strains also showed fewer similarities with the other strains (Figure 3.3). Similarly, VL-SSG-S strain S1 and VL-SSG-R strain R3 showed less resemblance with most of the other strains. The VL-MIL-R strain T9, which showed highest mutation burden, displayed a distinct profile with mixed similarity pattern (Figure 3.3).

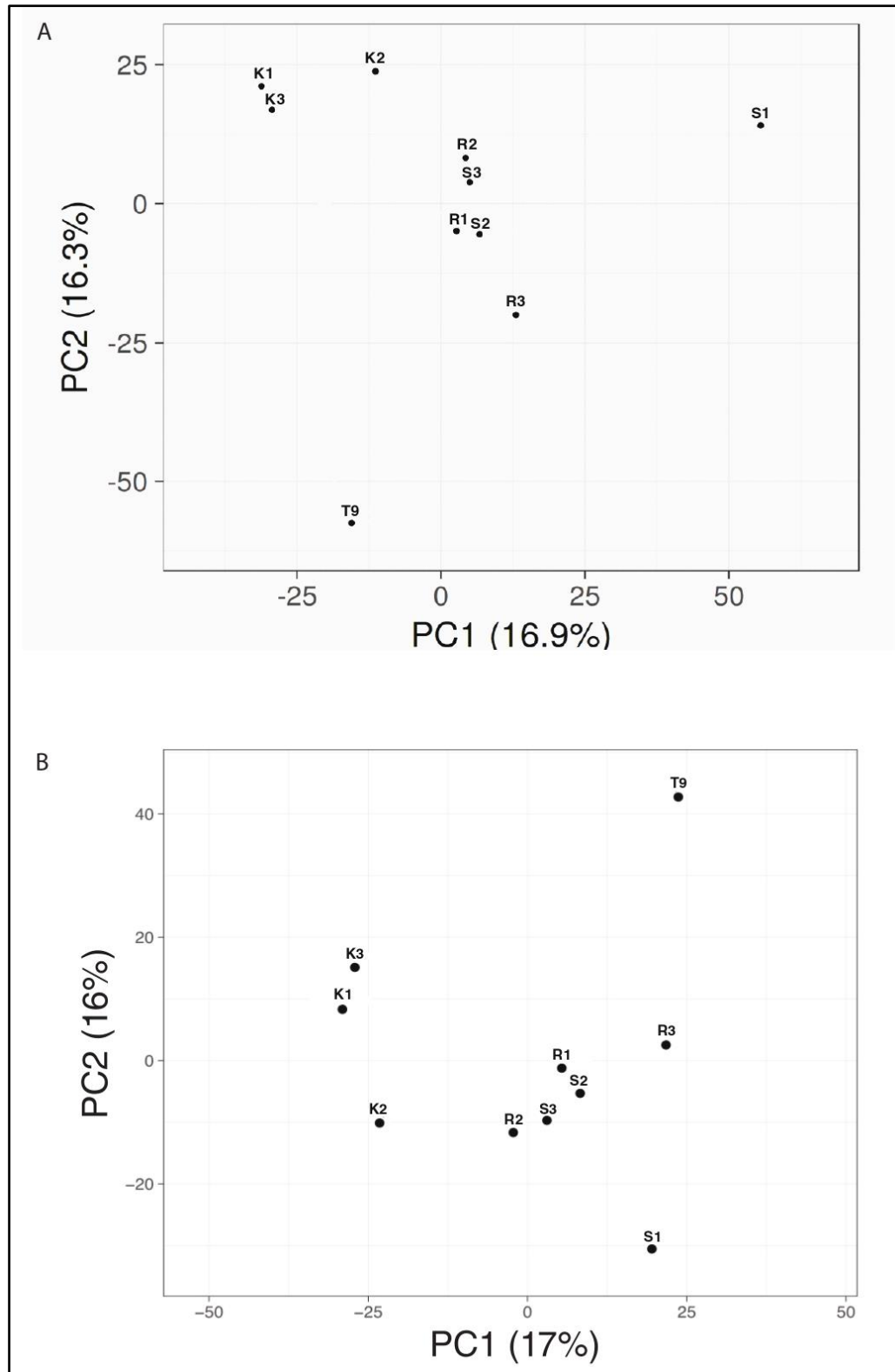


Figure 3.4. Principal component analysis (PCA) is based on (A) the mutational data types in protein-coding genes and (B) the mutated genes in the *L. donovani* strains from the ten clinical isolates.

In order to validate the clustering analysis performed with the types of mutation, the mutated genes were further considered for the principal component analysis. As represented in Figure 3.4 A & B, the clustering analysis using the mutated gene sets highly correlated with that using the types of mutation datasets. In both the clustering analyses in the figure 3.4, the two most significant PCs resulted in the formation of two distinct groups – one consisting of para-KDL-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, VL-SSG-R strain R3 and VL-MIL-R strain T9 were independently and distantly located from the two clusters in the PCA plot. 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 type of mutation among the genes in the individual strains. Further, clustering analysis was performed with the sets of the mutated genes which were likely to get mutated in the respective stains. As shown in Figure 3.4A, two most significant principal components resulted in distinct clustering 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 3.5 B and Figure 3.5 A and D) justified such a separate clustering. Figure 3 also depicted that VL-SSG-S strains S2 and S3 formed a co-cluster with VL-SSG-R strains R1 and R2. Strikingly, VL-SSG-S strain S1, VL-SSG-R strain R3 and VL-MIL-R strain T9 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 (Figure 3.5 B). These data supported the evidence of R3 being far distant strain in comparison to R1 and R2 strains (Figure 3.4). Similarly, the mutational information among the S strains showed comparatively higher mutation burden of S1 strain (Figure 3.5 C),

which was clearly reflected in the PCA analysis plot (Figure 3.4). The occurrence of the highest number of mutations in a single strain of VL-MIL-R strain T9, as shown in Venn diagram (Figure 3.5 D) was also reflected in its distinct position in the PCA plot (Figure 3.4).

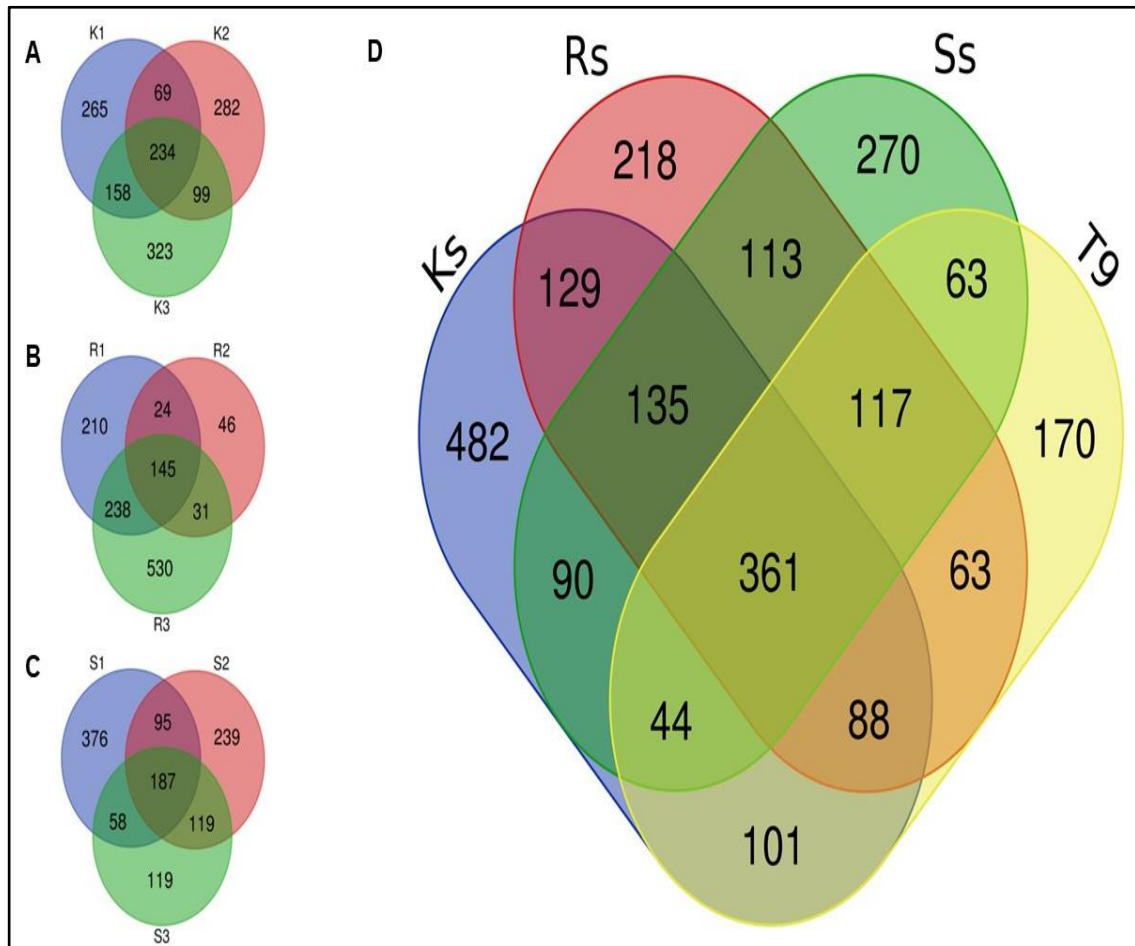


Figure 3.5. Venn diagram of intra- and inter-strain specific clustering analysis.

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, S's and T9 strains simultaneously.

The mutational analysis across the different strains at the genomic level revealed the presence of 61 gene sets which were found to get mutated across the all strains studied. Further, the pathway analyses revealed that most of these 61 genes could be involved in different metabolic processes (Figure 3.6).

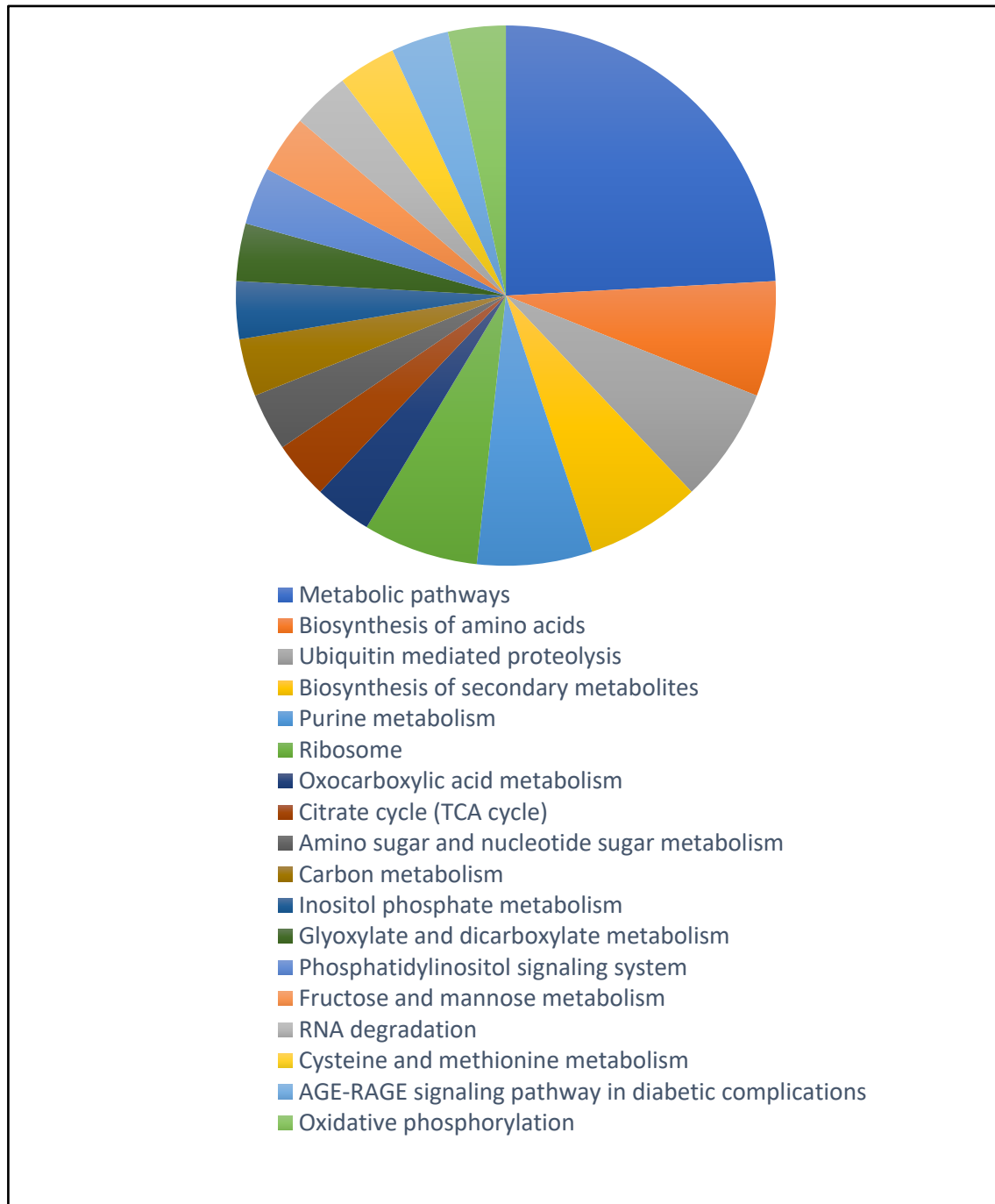


Figure 3.6. The pathway involvements of 61 genes mutated across all the strains.

Table 3.7 List of Genes Mutated Only In para-KDL Strains.

GeneDB ID	Changes in coding sequence		Encoded protein	Cellular Process
	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

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 3.7). As shown in Table 3.7, various cellular processes including metabolism, ion transport, repair and translation could be affected due to the mutations, possibly contributing significantly in the manifestation of PKDL.

The study identified unique mutations in 170 genes of VL-MIL-R strain, 56 of them encoding proteins with previously known function in various cellular processes (Table 3.8). Moreover, a majority of these 56 genes showed frame shift mutations, which might affect the corresponding cellular processes and be linked to appearance of MIL resistance.

Table 3.8 List of Genes Mutated Only In VL-MIL-R Strains.

GeneDB ID	Changes in coding sequence		Encoded protein	Cellular Process
	Nucleotide	Amino Acid		
LDBPK_010330	397delA	Thr133fs	poly(A) export protein, putative	RNA transport
LDBPK_020260	344delG	Gly115fs	protein kinase, putative	Kinase activity
LDBPK_030880	1delA	Met1?	peter pan protein, putative	Ribosome biogenesis
LDBPK_041250	1delA	Met1?	actin	Signaling
LDBPK_050400	2676delA	Asp893fs	structural maintenance of chromosome (SMC), putative	Genetic information processing
LDBPK_060370	191delT	Phe64fs	glutamine synthetase, putative	Metabolism
LDBPK_070880	904delC	Arg302fs	DNAJ-domain transmembrane-like protein	Stress response
LDBPK_071000	510delC	Asp171fs	ubiquitin-conjugating enzyme e2, putative	Transferase activity
LDBPK_071210	1852delG	Val618fs	cation-transporting ATPase, putative	P-type ATPase transporter
LDBPK_081040	416delT	Phe139fs	phosphoribosylpyrophosphate synthetase	Metabolic pathways, PRPP biosynthesis
LDBPK_130300	1104delA	Lys368fs	long-chain-fatty acid-CoA ligase protein, putative	Fatty acid biosynthesis
LDBPK_141420	917delT	Phe306fs	delta-6 fatty acid desaturase, putative	Lipid biosynthesis
LDBPK_150980	976delG	Ala326fs	protein phosphatase 2A regulatory subunit, putative	Phosphatase
LDBPK_181350	1757delT	Phe586fs	heat shock protein, putative	Stress response
LDBPK_190540	229delA	Ile77fs	methionine aminopeptidase, putative	Metabolism
LDBPK_200720	684delA	Asp229fs	serine/threonine protein phosphatase 2A regulatory subunit, putative	Phosphatase
LDBPK_211690	1821delT	Phe607fs	methyltransferase-like protein	Spliceosome Pathway
LDBPK_221450	2131delC	Leu711fs	ser/thr protein phosphatase, putative	Phosphatase

LDBPK_230080	230delT	Phe77fs	agmatinase-like protein	Arginine and proline metabolism
LDBPK_230620	647delT	Leu216fs	oxidoreductase-like protein	Oxidoreductase activity
LDBPK_241410	1436delT	Phe479fs	histone deacetylase, putative	Genetic information processing
LDBPK_242370	1910delC	Pro637fs	TFIIF basal transcription factor complex helicase subunit, putative	DNA excision repair
LDBPK_251620	500delC	Pro167fs	protein kinase, putative	Phosphotransferase
LDBPK_270260	267delT	Phe89fs	flap endonuclease-1 (FEN-1), putative	DNA replication, Base excision repair, non-homologous end-joining
	1173delA	Ala392fs		
LDBPK_271650	6794delA	Lys2265fs	dynein heavy chain, putative	Transport
LDBPK_271730	872delT	Phe291fs	GTP binding protein, putative	Signalling
LDBPK_290610	2166delT	Gln723fs	DNA repair helicase, putative	Repair
LDBPK_291220	798delG	Ser267fs	tryparedoxin-like protein	Metabolism
LDBPK_291230	267delT	Gln90fs	tryparedoxin-like protein	Metabolism
LDBPK_291680	1135delA	Ser379fs	glutamamyl carboxypeptidase, putative	Amino acid metabolism
LDBPK_292350	2256delA	Lys752fs	aminopeptidase, putative	Metabolism
LDBPK_292410	2231delT	Phe744fs	ubiquitin hydrolase, putative	Metabolism
LDBPK_303430	710delT	Phe237fs	PAS-domain containing phosphoglycerate kinase, putative	Metabolism
LDBPK_311300	81delT	His28fs	ATP-binding cassette protein subfamily C, member 5, putative, partial	Transport
LDBPK_311310	207delT	His70fs	ATP-binding cassette protein subfamily C, member 6, putative, partial	Transport
LDBPK_320270	16delA	Arg6fs	protein kinase, putative	Phosphotransferase
LDBPK_321120	1535delT	Leu512fs	dynein, putative	Signaling
LDBPK_321170	437delT	Phe146fs	mitochondrial carrier protein, putative	Transmembrane transporter activity
LDBPK_321260	415delA	Ile139fs	proteasome regulatory non-ATP-ase subunit, putative	Proteasome
	755delA	Lys252fs		pathway
LDBPK_322800	259delC	His87fs	amino acid transporter, putative	Transport
LDBPK_323080	1329delA	Val444fs	tubulin-tyrosine ligase-like protein	Protein modification

LDBPK_332090	604delA	Thr202fs	RNA editing complex protein MP46	Mitochondrial RNA editing
LDBPK_340030	748delT	Tyr250fs	protein kinase, putative	Phosphotransferase
LDBPK_341800	1778delC	Pro593fs	ATP-dependent RNA helicase, putative	Ribosome Biogenesis
LDBPK_342450	1566delT	Phe522fs	RNA helicase, putative	Mitochondrial biogenesis
LDBPK_344120	310delT	Ser104fs	nucleolar protein family a, putative, partial	Ribosome Biogenesis
LDBPK_351720	725delA	Lys242fs	casein kinase II, putative	Ribosome Biogenesis
LDBPK_352460	49delT	Ser17fs	ubiquitin hydrolase, putative	Catalytic activity
	665delA	Lys222fs		
LDBPK_353220	1257delG	His420fs	protein kinase, putative	Phosphotransferase
	4250delA	Lys1417fs		
LDBPK_354220	501delT	Pro168fs	elongation factor 1-alpha, putative	Translation
LDBPK_360590	870delT	Asn292fs	ubiquitin-like protein, putative	Protein processing in endoplasmic reticulum
LDBPK_360980	1642delG	Ala548fs	RNA editing complex protein MP67	Ribonuclease
LDBPK_361700	904delT	Ser302fs	clathrin heavy chain, putative	Transport, endocytosis
LDBPK_363190	128delT	Leu43fs	pre-mRNA branch site protein p14, putative	Spliceosome
LDBPK_364070	496delT	Lys48fs	eukaryotic translation initiation factor 3 subunit, putative	Translation
LDBPK_366510	143delA	Lys48fs	small G-protein, putative	Signaling

3.3.4. Comparative mutational analysis between the coding sequences of SSG sensitive VL & para-KDL and SSG resistant VL strains.

To explain the similarities among the different SSG sensitive VL & para-KDL with SSG resistant VL strains we drawn a heat-map which is based on the pair wise analysis of non-synonymous mutation profiles in the coding sequences and profiles of the affected genes (Figure 3.7). 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.

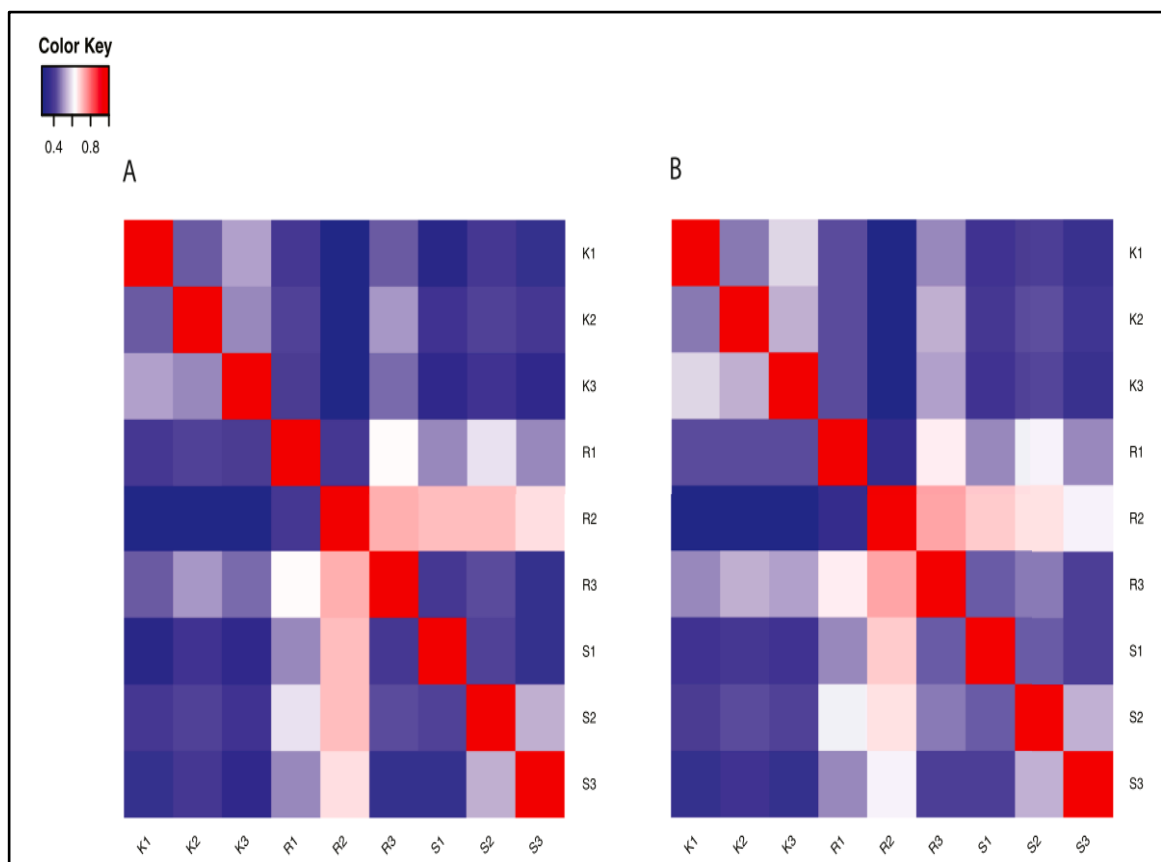
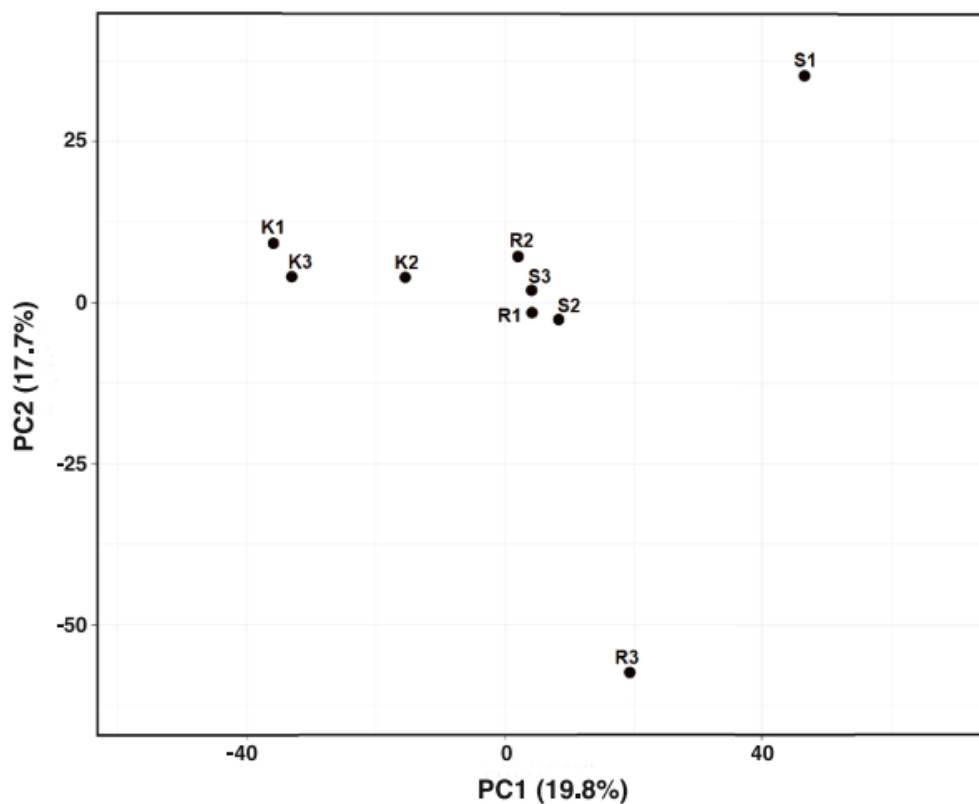


Figure 3.7. Comparison between mutation profiles of the *L. donovani* VL-SSG-S, VL-SSG-R and para-KDL-SSG-S strains. A. Comparison based on the complete mutation profiles. B. Comparison based on the profiles of genes that were mutated

Here also, we observed that the para-KDL-S strains 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). While with the other strains para-KDL-SSG-S strains also showed fewer similarities (mainly lower similarity squares in 6×3 rectangles of R1-S3 \times 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. Here also, the PCA plot analysis justified the previous observation about the separate clustering of VL-SSG-S strain S1 and VL-SSG-R strain R1. In the PCA plot slight deviation of K2 from K1 and K3 in the para-KDL strain was also observed (figure 3.8). The mutational burden was also been explained by the Venn diagram (figure 3.9)

A.



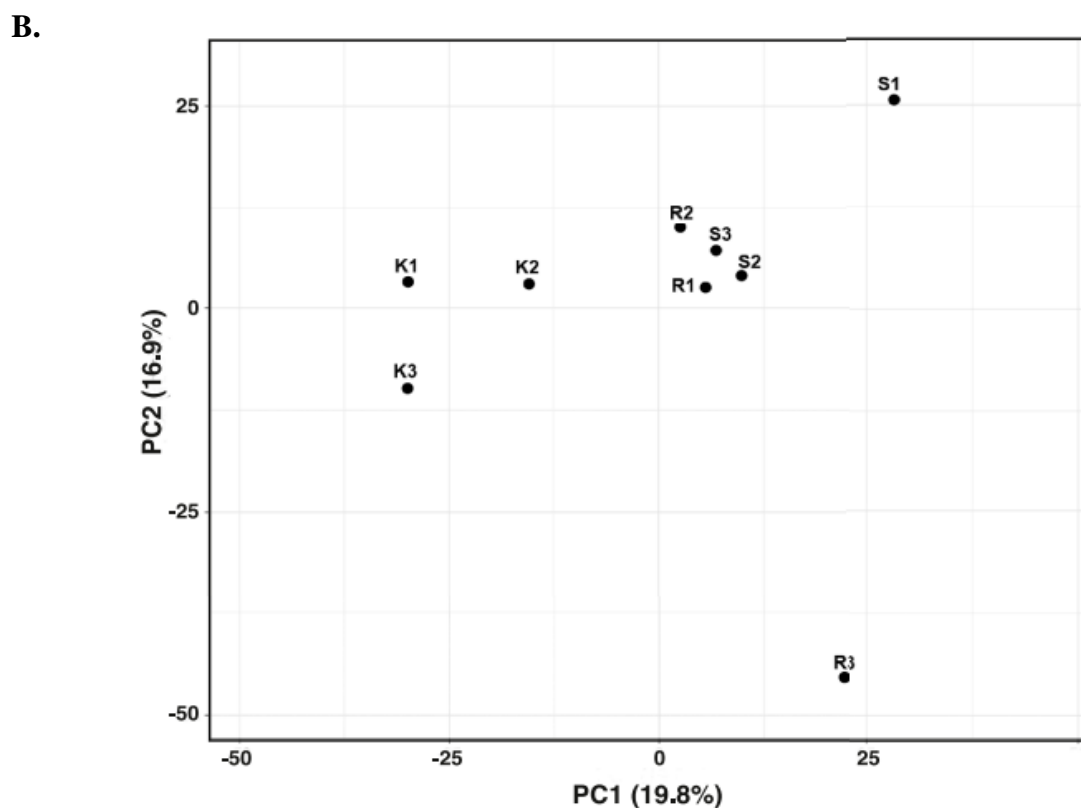


Figure 3.8. Principal component analysis (PCA) based on (A) the mutational data types in protein coding genes and (B) the mutated genes in the *L. donovani* strains from the nine clinical isolates. In both the clustering analyses, two most significant PCs resulted in formation two clusters – one consisting 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.

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 (Figure 3.8, 3.9). However, the strains are obtained from the isolates of the patients having close geographical locations, implying a possible role of different host immune responses for the development of the unique mutation profiles and aneuploidy patterns in the sequenced strains.

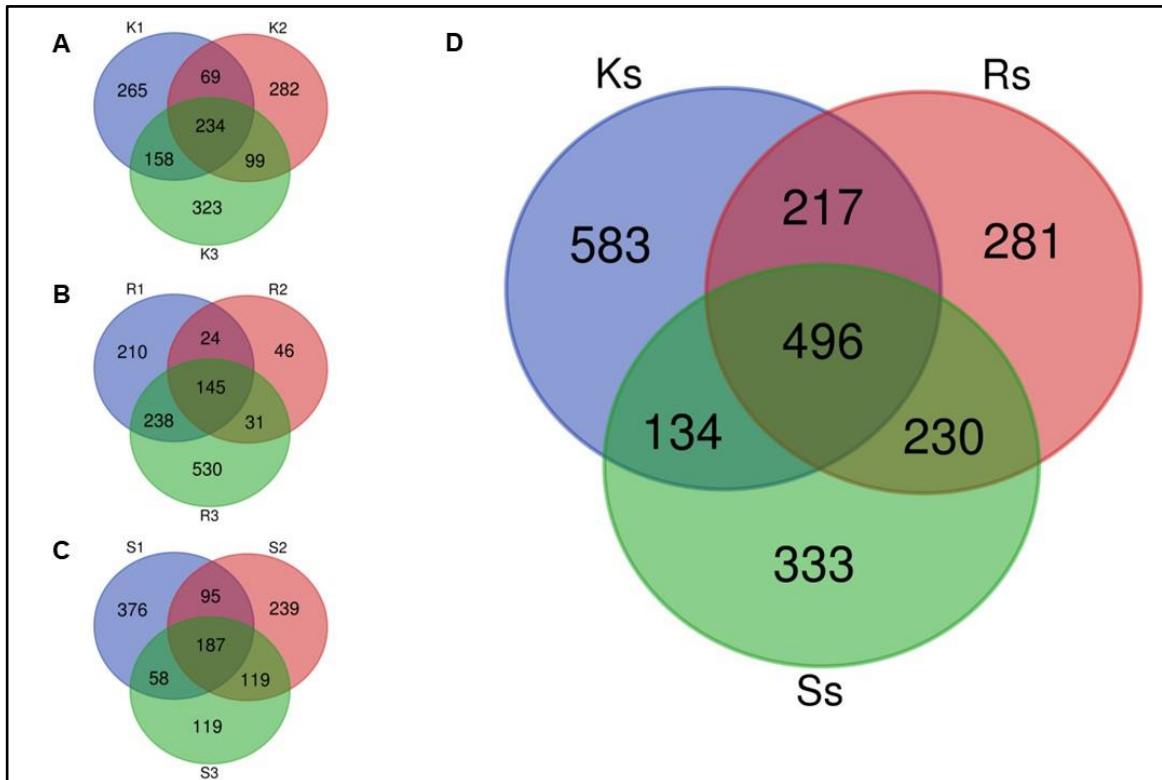


Figure 3.9. Venn diagram of intra- and inter-strain specific clustering analysis.

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.

The association of *L. donovani* with viscerotropism as well as dermatropism in Sri Lanka was reported (Ranasinghe et al., 2012) with differential copy number variation (CNV) of some of the genes between the strains causing visceral and cutaneous (CL) manifestation as a possible contributing factor (Zhang et al., 2014). Significant CNVs were also observed for some of these gene loci in the Indian strains studied here (Tables 3.9 and 3.10). However, no differential CNV was observed among the drug sensitive and resistant VL strains and para-KDL strains except for one hypothetical protein in chromosome 29 (LdBPK_291620), which was found to be absent in VL-SSG-R strains.

Table 3.9. List of genes with CNVs.

GeneDB ID	Encoded protein	Indian Strains			Sri Lankan Strains	
		K ^a	R ^a	S ^a	VL_SL ^b	CL_SL ^b
LdBPK_010790	eukaryotic initiation factor 4A-1	↑	↑	↑	↓	↑
LdBPK_111220	ABC transporter like protein, fragment	-	-	-	↑	↓
LdBPK_200120	phosphoglycerate kinase B, cytosolic	-	-	-	↑	↓
LdBPK_220670	A2 and A2rel repeat cluster	↑	↑	↑	↓	↓
LdBPK_230280	terbinafine resistance locus protein (YIP1)	↑	↑	↑	↓	↑
LdBPK_230290	ABC-thiol transporter	↑	↑	↑	↓	↑
LdBPK_291620	Major facilitator superfamily protein	↓	-	↓	↓	↑
LdBPK_291630	hypothetical protein	↓	↓	↓	↓	↑

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

b upward arrow head indicates higher copy number; downward arrow head indicates lower copy number; -, gene deletion/not found

The absence of ABC transporter like protein (LdBPK_111220) and phosphoglycerate kinase B (LdBPK_200120) was the other notable feature found among the Indian strains. Moreover, similar average copy number of A2 and A2rel repeat cluster (LdBPK_220670) was found in all the Indian strains (Table 3.9), though its lower copy number in CL causing *L. donovani* strains of Sri Lanka was shown to be associated with the impaired survival of the parasite in visceral organs (Zhang et al., 2014). Among the genes that were found to be mutated in CL causing *L. donovani* strains of Sri Lanka (Lypaczewski et al., 2018), two were found to be mutated in all the Indian strains, though the mutation profiles were

different (Table 3.9). On the other hand, 24 genes were found to be 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 in India and CL in Sri Lanka, even though both are skin manifestations caused by the same species of the parasite.

Table 3.10. List of Sri Lankan CL strain genes and their status in sequenced Indian strains.

Mutated In CL_ SL Strains	Start	End	Status In Indian Strains
LdBPK_070700.1	285603	288413	Mutated, but different
LdBPK_120275.1	125277	134975	Mutated, but different
LdBPK_141190.1	493765	494421	No mutation
LdBPK_220840.1	388946	389128	No mutation
LdBPK_250620.1	204904	206916	Mutated, but different
LdBPK_270830.1	349752	350726	No mutation
LdBPK_301640.1	578031	580121	No mutation
LdBPK_311390.1	602799	607259	No mutation
LdBPK_311470.1	670326	670781	Mutated, but different

3.4 Discussion

In the present study, ten *L. donovani* strains including three SSG sensitive VL strains, three SSG resistance VL strains, one MIL resistance VL strain and, most importantly, three SSG sensitive para-KDL strains derived from the patients isolates of Bihar and West Bengal region of India were subjected to whole genome sequencing. The sequence data were analysed to reveal variable aneuploidy and differential mutation burdens across the strains,

which might contribute significantly to modify drug sensitivity and development of PKDL manifestations. All the strains used in the study were promptly derived from the patients' isolates and cultured under identical conditions for similar number of passages before and after cloning 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 could be observed in our study. Therefore, the karyotypes of the sequenced parasites could 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 was observed across the strains with distinct changes in VL-SSG-R and para-KDL-SSG-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 in maintaining a differential pool of messages for prompt survival response to stressful and changing environments in different hosts (Dumetz et al., 2017). It has been reported that the mean gene expression of the trisomic chromosomes is significantly higher than that of disomic chromosomes (Iantorno et al., 2017; Ubeda et al., 2008). Interestingly, trisomy was noticed for chromosomes 5 and 6 exclusively in the para-KDL strains. During depth analysis we observed that the several essential protein-coding genes are located on chromosomes 5 and 6, whose products are involved in various metabolic activity (Manzano et al., 2017; Tovar et al., 1998; Docampo et al., 2011), including three ABCG subfamily of transporters – LABC1 (LdBPK_060080), LABC2 (LdBPK_060090) and LABC3 (LdBPK_060100) which is located on chromosome 6. Recent studies show that LABC2 (ATP-binding cassette G subfamily) transporter is required for the externalization of phosphatidylserine (PS), leading to induction of PPAR γ in parasitized macrophages and activation of M2 phenotype (Campos-Salinas et al., 2013).

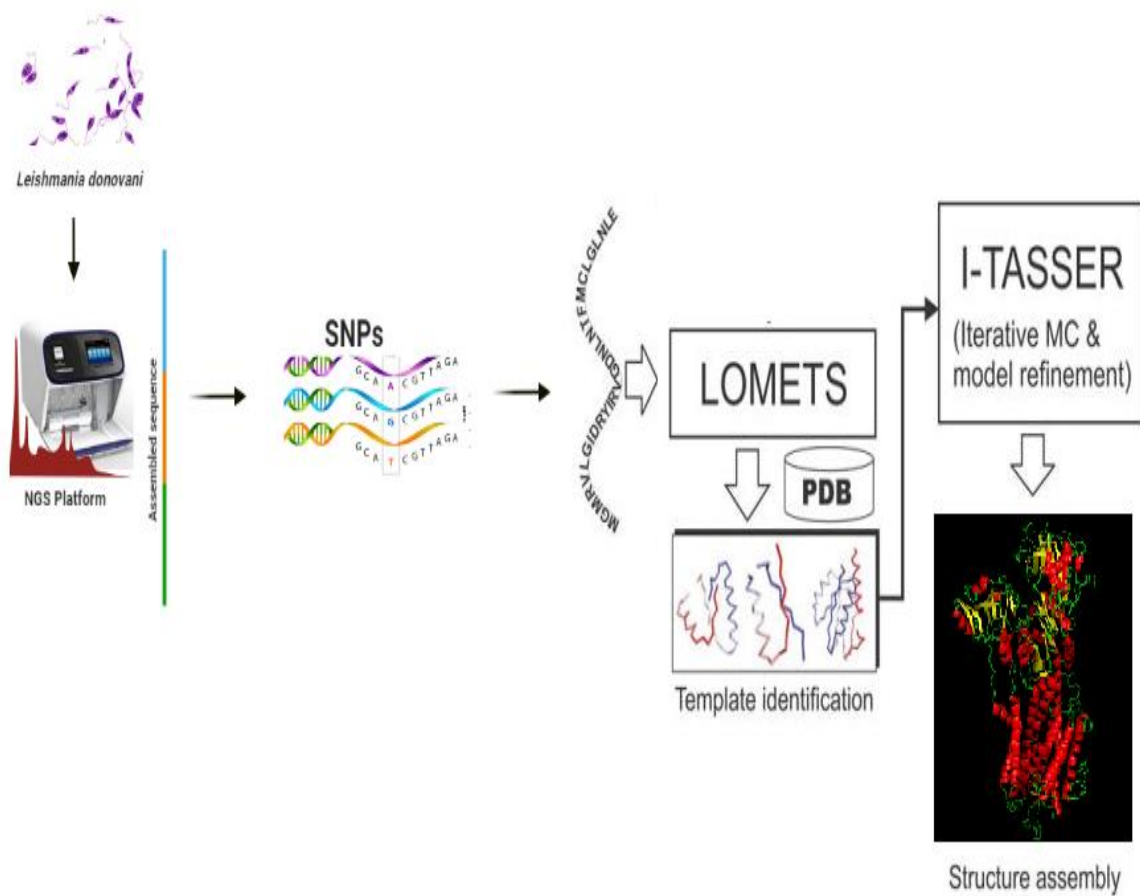
Such polarization to M2 phenotype has been shown to promote *Leishmania* survival and PKDL manifestation (Mukhopadhyay et al., 2015). But the three ABCG subfamily of transporters on chromosome 6 have been found to be deleted in the para-KDL strains studied here. However, three more copies of the similar 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 expression of LABCG4 possibly leading to the survival of the parasite. In our study, distinct clustering of para-KDL strains in pairwise analyses as well as principal component analyses based on non-synonymous mutation profiles in the coding sequences and sets of mutated genes in the individual strains, strongly suggest the covariance of the unique sets of mutations that could be significant for the manifestation of the phenotype. Further identification of 24 genes, that were found to be mutated only in the three para-KDL strains, also supports the commonalities in mutation burden possibly leading to the disease manifestation. Non-synonymous SNPs present in the coding sequences could be particularly crucial, because they could change the structure-function properties of the encoded proteins with possible modification in expression level. Moreover, mutation in the four out of these 24 genes is homozygous in nature (Table 3.6), 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, could provide survival benefits to the parasites even after apparent successful VL treatment and significantly contribute in the development of the symptom. Notably, the clustering and the unique SNPs are identified in the para-KDL strains derived from the isolates of the patients having a wide age range, further

emphasizing the importance of the genomic changes in the development of the phenotype, as they have appeared irrespective of the patient's age. Interestingly, distinct locations in the PCA plots have also been observed for VL-SSG-S strain S1, VL-SSG-R strain R3 and VL-MIL-R strain T9, implying unique mutation burden for each of these strains. Also, the aneuploidy profiles of S1 and T9 strains have been found to be distinct indicating unique genetic nature of each of them. However, the strains have been derived from the isolates of the patients having close geographical locations, but with different genders and diverse ages. Therefore, the unique mutation profiles and aneuploidy patterns in the strains could result from divergent host immune response, as host specific selective processes have been shown to influence parasite genomes. It is known that the translocation machinery comprising LdMT and LdRos3 defines miltefosine sensitivity in *Leishmania* (Khanra et al., 2017; Mondelaers et al., 2016; Perez-Victoria et al., 2006) and previously in the studied miltefosine resistant field isolate, SNPs have been identified in LdMT, not in LdRos (Srivastava et al., 2017). However, no significant mutation and copy number change could be detected in any of the genes in the VL-MIL-R strains studied here, though our previous study showed down-regulation of LdMT and LdRos3 by RT-PCR (Khanra et al., 2017), suggesting possible role of post-transcriptional mechanism in regulation of the expression level. The down-regulation of Hsp70 and the involvement of some lipid biochemical pathways have been shown to be associated with miltefosine resistance (Rakotomanga et al., 2005; Requena et al., 2015; Vacchina et al., 2016). Interestingly, frame shift mutations have been found in genes encoding heat shock proteins (LDBPK_070880 and LDBPK_181350) and proteins involved in fatty acid and lipid biosynthesis (LDBPK_130300 and LDBP1420). Also, ATP-binding cassette (ABC) proteins have been reported to be associated with the appearance of drug resistance in *Leishmania* (Leprohon et al., 2006; Leprohon et al., 2009a), and in the T9 strain, mutations have been found in

genes encoding ABC proteins (LDBPK_311300 and LDBPK_311310). Furthermore, T9 shows tetrasomy at chromosome 23 compared to disomy in VL-SSG-S strains. Strikingly, chromosome 23 harbours the gene encoding multi-drug resistant protein A (MRPA, LDBPK_230290) and 3.4 times higher expression of the gene was observed in T9 compared to that in VL-SSG-S strain in our previous study (Khanra et al., 2017). Higher somy status of chromosome 23 in T9 identified in the present study further corroborate the experimental data. In the present study we identified unique mutations in the 170 genes of VL-MIL-R strain, 56 of them encoding proteins with previously known function in various cellular processes (Table 3.8). Moreover, a majority of these 56 genes showed frame shift mutations suggestions possible role of the mutations in the observed genotype. Immune response in VL – before and after successful treatment and in para-KDL is characterized by changes in cytokine profile explained partly by Th1/Th2 dichotomy of adaptive immunity (Zijlstra, 2016). In VL, Th2 type response is primarily observed resulting in the release of IL-4 and IL-10 cytokines, which are responsible for parasite survival and disease progression. Successful treatment of patient is characterized by resistance to the disease predominantly by Th1 cytokines like IL-2, IL-12 and IFN- γ . Interestingly, para-KDL, which is characterized by the appearance of macular hyperpigmentation or nodular lesion in skin with no sign of systematic illness post-VL treatment, shows Th2 response in skin with persistence presence of IL10, but Th1 response systematically with IFN- γ production. The underlying mechanism of such a dissociation of the host immune response between the skin and the viscera, allowing the persistence of the parasites in the skin to generate PKDL remain largely unknown. Our whole genome sequencing data reveal distinguishing genomic changes among the para-KDL strains irrespective of gender and age of the patients, which could contribute significantly for survival of the persistent parasites in the skin even after apparently successful systematic cure from KA.

Our study also provides further data, which helps to understand the underline mechanisms of associated genes in the arrival of the MIL resistant filed isolates. 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 data of *Leishmania donovani* strains derived from the patient isolates in the present study reveal a strong correlation of development of para-KDL with chromosome aneuploidy and specific non-synonymous genetic variation in the coding sequences.

Data Accession: The sequence data related to the study are available under the BioProject Accession number PRJNA598250.



CHAPTER 4

***In Silico* Structural Analysis of the Mutated Transporter and Surface Proteins of Indian para-KDL Isolates**

4.1 Introduction

Post-kala-azar dermal leishmaniasis (PKDL) is a commonly neglected complication of Visceral Leishmaniasis (VL) or Kala-azar (KA). It is characterized by macular, papular, and nodular lesions and sometimes an asymptomatic skin rash on the face and extremities (Zijlstra et al., 2003). Usually, symptoms start around the mouth and depending upon the severity spread to other parts of the body (Zijlstra et al., 2017). PKDL cases have been reported mainly among the apparently cured VL patients. In Sudan 50 to 60% and in India 10 to 20% of cases have been reported to develop Post Kala-azar Dermal Leishmaniasis which is basically limited to areas where diseases are primarily caused by *L. donovani*. Thus, in absence of a report of any animal reservoir in India, PKDL patients are thought to act as a parasite reservoir and may play an important role during the parasite transmission cycle. This is considered a major issue in the VL elimination program (Burza et al., 2018). Nowadays, the onset of PKDL manifestation within VL patients has also been reported. These cases gradually take an intermediate transition of the disease where parasites invade both the dermal and visceral organs of the patients (Zijlstra, 2019; Zijlstra et al., 2003), known as Para Kara-Azar Dermal Leishmaniasis (para-KDL). Recently, nine such cases were reported from Bihar, India where VL is associated with PKDL (Kumar et al., 2016). The increasing load of active VL in PKDL cases are required more attention to understanding the clinicopathological, diagnosis and treatment regime of these emerging cases of the para-KDL. Hence, it is important to explain the mechanisms that assist parasite survival and transition from the visceral to the dermal area after an apparent cure from VL. Currently, extensive studies have been carried out to understand the underline mechanism of parasites' intracellular survival and the arrival of their drug resistance. Some clinical data reveal that in response to environmental stress and drug pressure *Leishmania* parasites can easily reorganize their genomes rapidly, suggesting the importance of genomic

variation which they can adapt for their survival mechanism (Leprohon et al., 2009b). Previously, to explain the population genetics of the *Leishmania* parasites, many studies have been carried out by using different molecular biological methods. But limited information is available on the genomic modifications of the *Leishmania* parasites in the clinical context (Alam et al., 2009). Recent report suggested that in the Indian sub-continent *L. donovani* population exhibits little genetic variation which can be identified by using existing genetic markers (Imamura et al., 2016) or by more advanced third generation molecular technique such as, Whole Genome Sequencing (WGS) that provides a multidisciplinary field to understand the parasite biology. WGS of different clinical isolates and their comparative study helps to identify significant changes at the genome level in parasite populations that lead to adverse responses to therapeutic regimens. Therefore, as previously observed, it can be safely said that the sequenced parasite represents the parasite of the corresponding original patient isolates (Downing et al., 2011). Keeping such ideas in mind, we performed whole genome sequencing of the *Leishmania donovani* genome of three SSG sensitive para-KDL cloned strains, three SSG sensitive VL cloned strains, three SSG resistant VL cloned strains and one MIL resistant VL cloned strain. After data annotation, we compared the ten *L. donovani* parasites genome and surprisingly, we observed and reported the 24 novel mutations which were only present in para-KDL strains (Sarraf et al., 2021), with thirteen of them being homologous to genes which were associated with previously known cellular functions such as metabolism, transport and signalling. In this chapter we have mainly used computational methodology-based analysis to discuss about the secondary and tertiary structure of the surface and transporter genes which were only get mutated in the para-KDL strains. Here, specifically focus was given to surface protein and transporter encoded genes because it has been observed that in the kinetoplastid parasites, surface proteins and transporters play a vital role during virulence,

intracellular survival and drug metabolism (Maria Valente et al., 2019). Hence, the present study highlighted the *in silico* structural analyses of the five novel mutated genes (out of 13 found) in the para-KDL strains, whose function is homologous to genes which previously been observed as a surface protein and transporter.

4.2 Outline of the work

4.2.1 *Leishmania* isolates

Here, we have taken only para-KDL specific mutated genes. The clinical history of three Sodium Stibogluconate Sensitive (SSG-S) and Miltefosine Sensitive (MIL-S) para-KDL clinical isolates of *L. donovani* are given below in Table 4.1.

Table 4.1 Clinical history of the para-KDL isolates (Khanra et al., 2016; Sarraf et al., 2021)

Study Code	WHO Reference Code	Age (Year)	Sex	Treatment History	Fever	Spleen Size (cm)	Liver Size (cm)	Lesion Type
K1	MHOM/IN/2013/M1	10	Female	Not Received	YES	7	3	Macular
K2	MHOM/IN/2014/P6	20	Female	Inadequate treatment with MIL	YES	8	3	Macular
K3	MHOM/IN/2014/T10	35	Male	MIL	YES	6	3	Macular

4.2.2 Mutation analysis:

The whole genome mutational profile of the protein coding genes among the VL (SSG-S/MIL-S) strains was compared with para-KDL (SSG-S/MIL-S) strains. 24 novel gene mutations are observed in the para-KDL strains only, among them 13 are homologous gene mutations that are encoding previously known cellular functions. Here, for the *in silico* analysis, we selected five homologous genes which act as surface and transporter protein (Table 4.2).

A list of the mutated transporters and surface protein genes with their observed mutation type and nucleotide position in para-KDL strains are given in Table 4.2.

Table 4.2 List of the mutated genes in para-KDL isolate details.

Gene DB ID	Mutation Type	Nucleotide
LDBPK_040010	deletion	2944delA
LDBPK_231970	deletion	760delC
LDBPK_250010	deletion	112delT
LDBPK_310600	deletion	810delT
LDBPK_311290	SNV	3539G>C

4.2.3 Primary structure analysis and functional annotation

UniProtKB online search tool was used for primary structure and functional annotation of the transporter and surface proteins in para-KDL strains.

4.2.4 Protein structure prediction

The secondary and tertiary structure is predicted with I-TASSER. I-TASSER used the PSSpred algorithm for secondary structure analysis (Yang et al., 2015) and the structure templates for modelling is identified by LOMETS (Wu and Zhang, 2017) from the PDB library. Further mutant residue analysis carried out with the help of PyMOL.

4.2.5 Protein motif

L. donovani specific KEGG databases are used to examine the gene-specific metabolic pathway, different protein motifs and their position on the genome (Kanehisa and Goto, 2000).

4.2.6 Normalized B-factor prediction

Using I-TASSER the normalized B-factor is predicted by ResQ programming to indicate the extent of the inherent thermal mobility of residues in proteins (Yang et al., 2015; Wu and Zhang, 2017).

4.2.7 Structure validation

SWISS-MODEL online homology-modelling server is used to validate the 3D structure of the target proteins with the help of Ramachandran plot (Studer et al., 2020).

4.3 Results

In the whole genome sequencing data of para-KDL strains we observed 24 novel genes that were mutated in the coding sequence. In the observed 24 novel genes, 13 of them being homologous to the genes with previously known different cellular functions. Among them two are transporter and three are surface proteins (Table 4.3). As shown in Table 4.3, various cellular processes including metabolism and ion transport could be affected due to the mutations and possibly contribute a significant role towards the development of PKDL.

Table 4.3 List of genes encoding previously known transporters and surface proteins.

GeneDB ID	Nucleotide	Amino acid	Encoded Protein	Cellular Process	Protein Type
LDBPK_040010	2944delA	Met982fs	calcium-translocating P-type ATPase	Ion Transport	Transporter
LDBPK_231970	760delC	Leu254fs	neutral sphingomyelinase activation associated factor-like protein	Metabolism	Surface protein
LDBPK_250010	112delT	Tyr38fs	beta galactofuranosyl transferase	GPI-anchor biosynthesis, Glycosyl-transferases	Surface protein
LDBPK_310600	810delT	Phe270fs	amino acid transporter aATP11, putative	Transport amino acid (Arginine)	Transporter
LDBPK_311290	3539G>C	Ser1180Thr	p-glycoprotein e, partial	Metabolism	Surface protein

To identifying the effect of reported mutations in the secondary and tertiary structure of the functional surface and transporter proteins, we performed *in silico* structural analysis of the respective proteins. UniProtKB online data base are used to predict the primary structure (amino acids sequence) of the respective genes, later online software I-TASSER is used to predict their secondary and tertiary structure which is further analysed by PyMOL and validate by SWISS-MODEL. Predicted secondary structure analysis of transporters and surface proteins explained the effect of the mutation on contacts made by the residue in their predicted tertiary structure. Detailed study of the mutated genes which encoding previously known proteins is as follows:

4.3.1 Calcium-translocating P-type ATPase

Leishmania donovani gene LDBPK_040010 encoding protein calcium-translocating P-type ATPase. The length of the functional protein is 1023 amino acids with 111,495 Da molecular mass. The predicted secondary structure of calcium-Translocating P-type ATPase has seven transmembrane domains (Figure 4.1, Table 4.4) and eight motifs (Table 4.4). The observed nucleotide deletion (2944delA) leads to frameshift mutation (Met982fs) which is present in the 8th motif of the protein that results a truncated protein i.e., 985 amino acids due to insertion of a stop codon at amino acid position 986.

Table.4.4 List of motifs and transmembrane domains in the Calcium-translocating P-type ATPase.

Motif id	From	To
Cation_ATPase_N	18	78
E1-E2_ATPase	126	335
Hydrolase	353	716
Cation_ATPase	424	530
CreA	487	540
HAD	603	712
Hydrolase_3	699	747
Cation_ATPaseC	785	1000

Feature	Position(s)	Length
Transmembrane	66 – 85	20
Transmembrane	91 – 110	20
Transmembrane	255 – 279	25
Transmembrane	304 – 328	25
Transmembrane	761 – 782	22
Transmembrane	834 – 858	25
Transmembrane	926 – 946	21

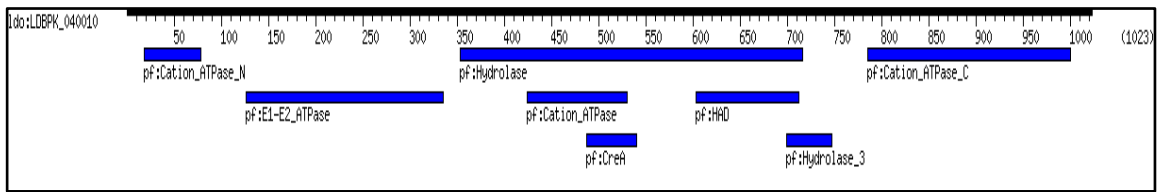
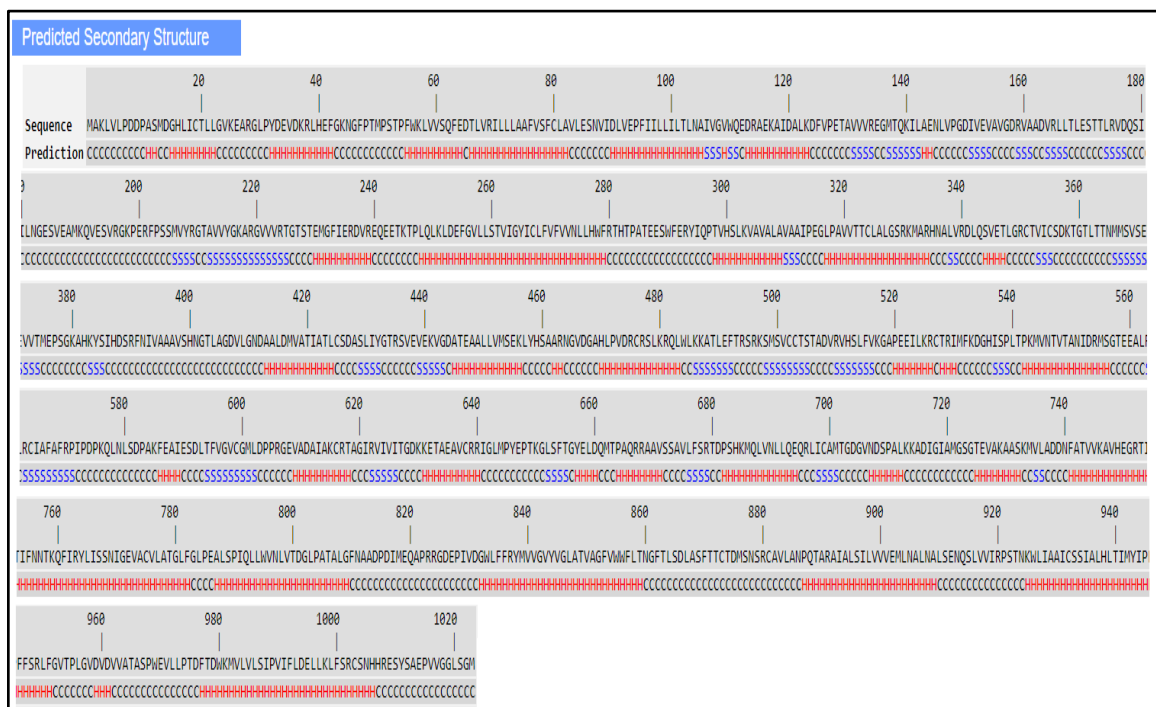


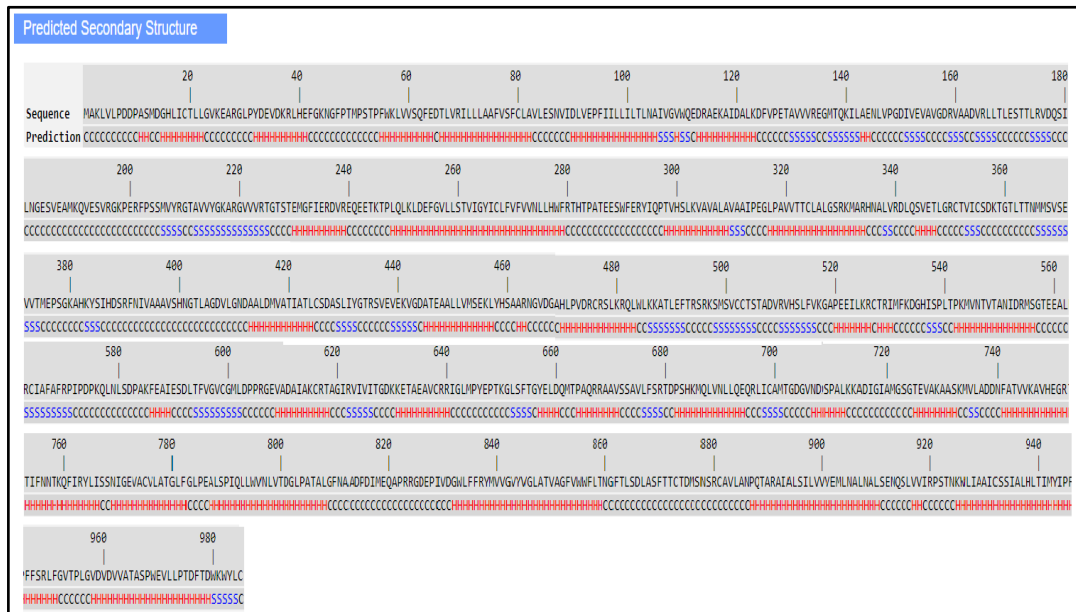
Figure.4.1. *L. donovani* specific KEGG database illustrating eight motifs of the Calcium-translocating P-type ATPase protein.

The eighth motif of protein i.e., cation transporting ATPase C-terminus functionally acts as a vital component of *Leishmania* parasites plasma membrane and plays important role during host and parasite interface (John et al., 2002). The other molecular functions of the encoded gene LDBPK_040010 is associated with the ATP binding, P- type calcium transporter activity and other biological process is associated with the calcium efflux ATPase. The predicted secondary structure, predicted solvent accessibility and normalized B-factor of the wild type calcium-translocating P-type ATPase protein is explained in the Figure 4.2 A, B and C.

A.



A.



B.

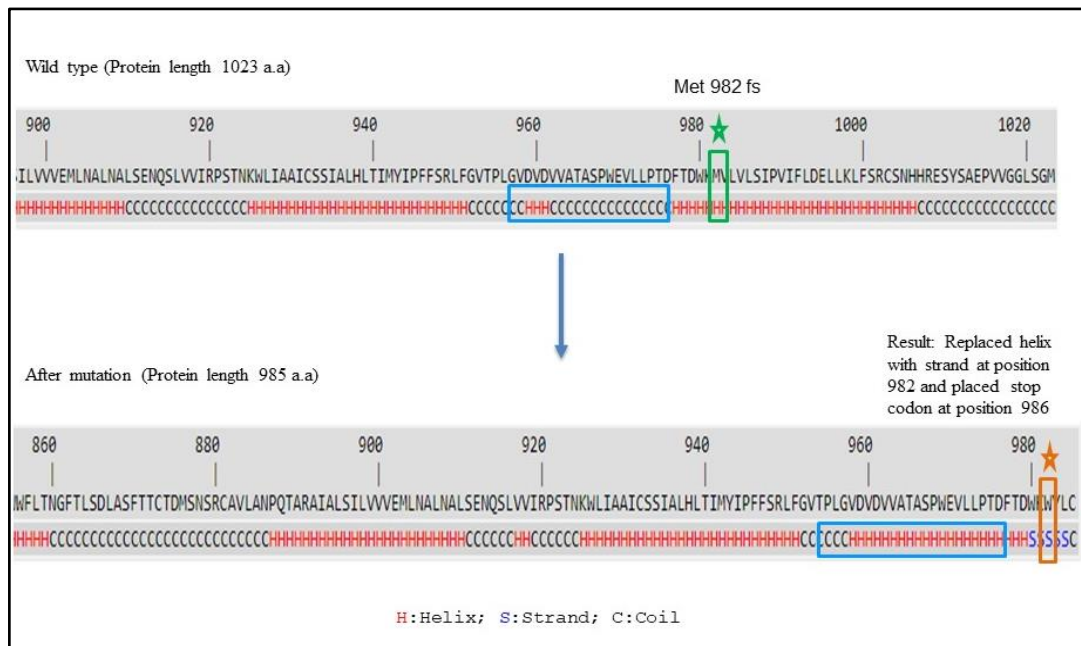


Figure 4.3. I- TASSER results illustrating, A. The predicted secondary structure of the mutant Calcium-translocating P-type ATPase protein with 985 amino acids where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others), B. Comparison between predicted secondary structure of wild type and mutant calcium-translocating P-type ATPase protein. Highlighted green box indicates the site of point mutation, highlighted orange box indicates the site of frameshift where at amino acid position 982, helix (WT) is replaced with a strand (mutant) and highlighted blue box indicating the changes in secondary structure.

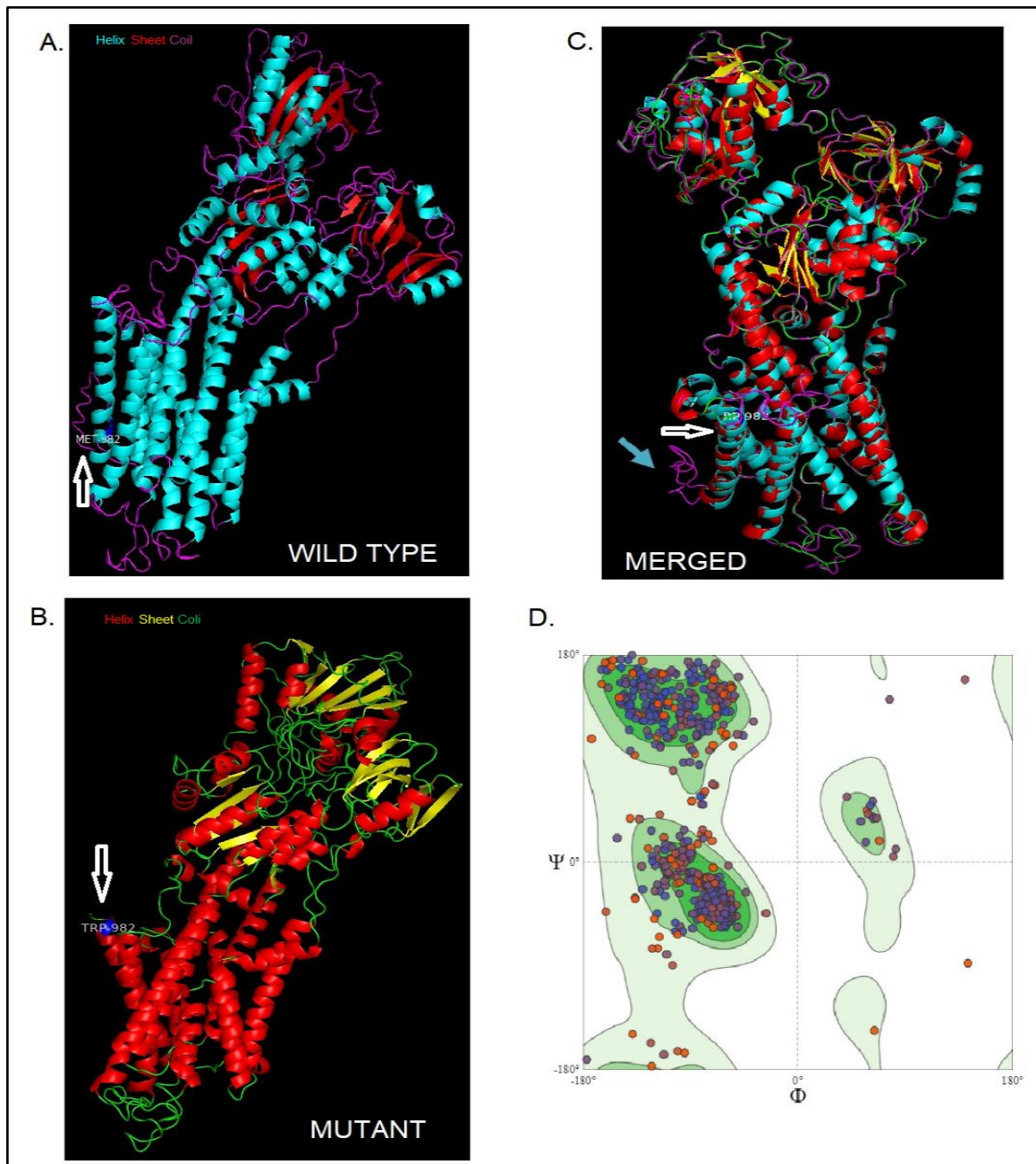


Figure 4.4. I- TASSER homology modelling for predicted secondary structure of the Calcium-translocating P-type ATPase protein, **(A)** Predicted tri-dimensional calcium-translocating P-type ATPase structure **(Wild Type)**, **(B)** Predicted tri-dimensional calcium-translocating P-type ATPase structure **(Mutant)**. The position of the point mutation in the wild type and mutant are shown by arrow head, **(C)** Predicted tri-dimensional overlay structure of the wild type and mutant protein **(Merged)**. White arrow head are used to locate the mutated positions in the tri-dimensional structure of the protein, **(D) Ramachandran plot**, general case for the predicted Calcium-translocating P-type ATPase structure (Wild type) by SWISS-MODEL.

The wild type, mutant and the aligned tertiary (3D) structure of the calcium-translocating P-type ATPase protein along with their mutational residues are shown in the Figure 4.4 A, B & C. A stereochemical assessment was undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral angles i.e Phi and Psi distribution. The Ramachandran plot statistics of the modelled tri-dimensional wild type calcium-translocating P-type ATPase protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (96.9%) and the additional allowed (2.3%) energy regions, meanwhile only the 0.8% were at the disallowed regions (Figure 4.4D). This result suggests that the quality of predicted calcium-translocating P-type ATPase structure is reliable. In Figure 4.4A white arrow head indicates the position (982th amino acid) of mutation where methionine (Met) residue is placed but after the frameshift mutation (figure 4.3,4.4 B) methionine is replaced with tryptophan (Trp) at position 982 in the protein. In the Figure 4.4C the 3D aligned protein structure indicates the variations in the coil region of the protein after the frameshift mutation. The frameshift mutation that leads to insertion of stop codon at amino acid position 986 (Figure 4.3B), results reduction in the functional length of the protein that could be responsible for interrupting molecular function of the calcium-translocating P-type ATPase protein.

4.3.2 Neutral Sphingomyelinase Activation Associated Factor-Like Protein

Leishmania donovani gene LDBPK_231970 encoding neutral sphingomyelinase activation associated factor-like protein, which is a type of surface protein with no transmembrane domain. The length of the functional protein is 925 amino acids with 102,013 Da molecular mass. Predicted secondary structure of neutral sphingomyelinase activation associated factor-like protein has single motif name Beige/BEACH domain from amino acid position

269 to 542 (Figure 4.5). The observed nucleotide deletion (760delC) leads to frameshift mutation (Leu254fs) resulting in truncated protein i.e., 283 amino acids due to the insertion of a stop codon at amino acid position 284. The predicted secondary structure, normalized B-factor and predicted solvent accessibility of the wild type neutral sphingomyelinase activation associated factor-like protein is explained in the Figure 4.6 A, B and C.

A.

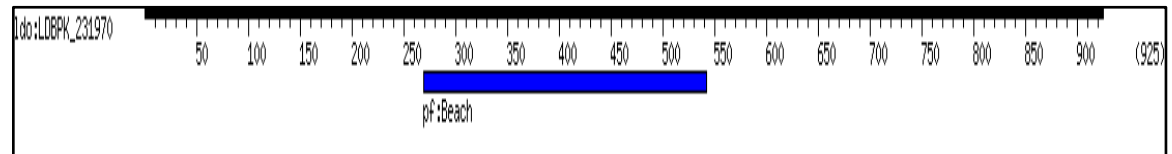
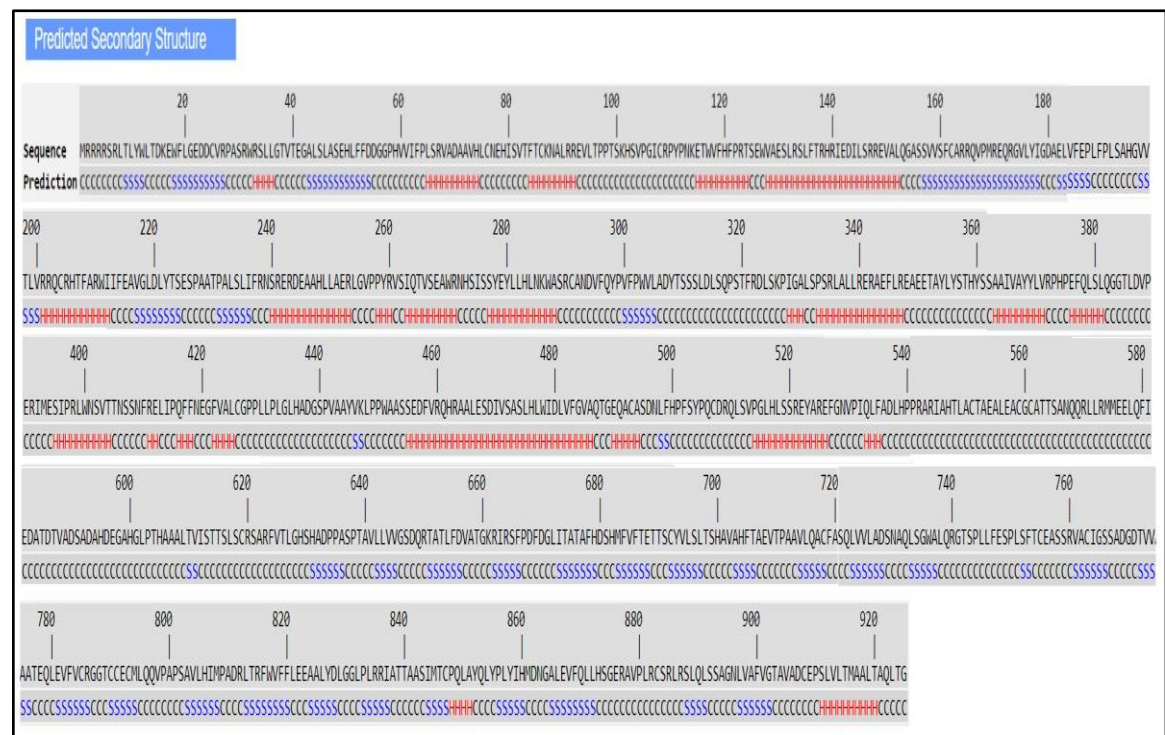
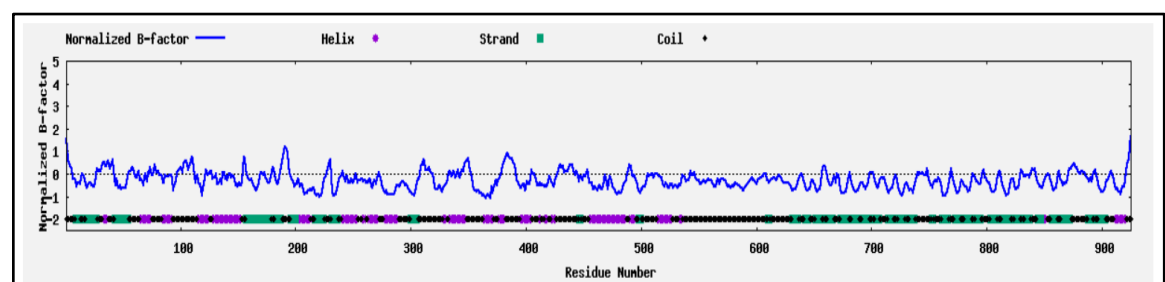


Figure.4.5. *L. donovani* specific KEGG database illustrating single motifs of Neutral sphingomyelinase activation associated factor-like protein.

A.



B.



C.

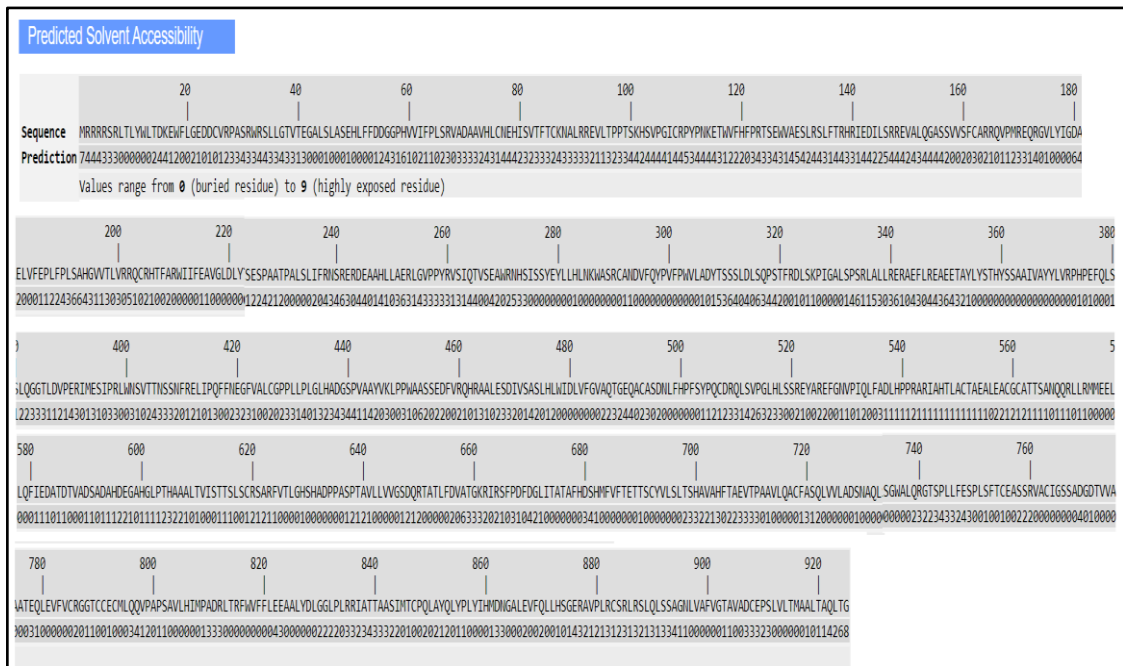


Figure 4.6. I- TASSER results illustrating the 925 amino acids in wild type Neutral sphingomyelinase activation associated factor (NSMAF) like protein, A. Predicted secondary structure where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others), **B. Predicted normalized B-factor** of the wild type protein, **C. Predicted solvent accessibility** are encoded in numerical value states where 0 indicates buried and 9 means highly exposed residue.

The predicted solvent accessibility showing the highly exposed and buried residues (Figure 4.6 C). There is little information known about the NSMAF gene regulation during leishmania infection, previously it has been reported that the Beige/BEACH domain of protein is associated with the tumour necrosis factor (TNF)- α signalling pathway (Montfort et al., 2010). During *Leishmania* infection a pro-inflammatory cytokine TNF- α plays important role in the IFN- γ production whose function is better known in infected macrophage for classical activation i.e., iNOS upregulation, production of nitric oxide (NO) and other free radicals which restricts parasites intracellular survival and promotes intracellular parasite killing (Liu et al., 2012). Hence, structural changes in the protein due to frameshift mutation may alter the biological function of the protein.

secondary structure of neutral sphingomyelinase activation associated factor-like protein indicates the coil (wild type) to helix (mutant) change at mutant position 254 and insertion of stop codon at 284th position. Some secondary structural changes i.e., helix to coil is also observed which is shown by blue box in the Figure 4.7B. The wild type, mutant and the aligned tertiary (3D) structure of the neutral sphingomyelinase activation associated factor protein along with their mutational residues are shown in the Figure 4.8 A, B & C. A stereochemical assessment was undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral angles i.e. Phi and Psi distribution. The Ramachandran plot statistics of the modelled tri-dimensional wild type neutral sphingomyelinase activation associated factor protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (97.0%) and the additional allowed (2.9%) energy regions, meanwhile only the 0.1% were at the disallowed regions (Figure 4.8 D). This result suggests that the quality of predicted neutral sphingomyelinase activation associated factor protein structure is reliable. In Figure 4.8 A white arrow head indicates the position (254th amino acid) of mutation where leucine (Leu) residue is placed but after the frameshift mutation leucine is replaced with serine (Ser) at position 254 in the protein (figure 4.7,4.8 B). In the Figure 4.8C, aligned protein 3D structure indicates the merged protein structure with unaligned coil region after the frameshift mutation. The frameshift mutation that leads to insertion of stop codon at amino acid position 284 (Figure 4.7 B), results in reduction of the functional length of the protein Beige/BEACH domain that could be responsible for interrupting various molecular function which is coordinated by factor associated with neutral sphingomyelinase activation such as TNF- α signalling. TNF- α signalling interruption provide as a great escape route for intracellular parasites and could be play a vital role in intracellular parasite survival mechanism or visceral to dermal escaping mechanism.

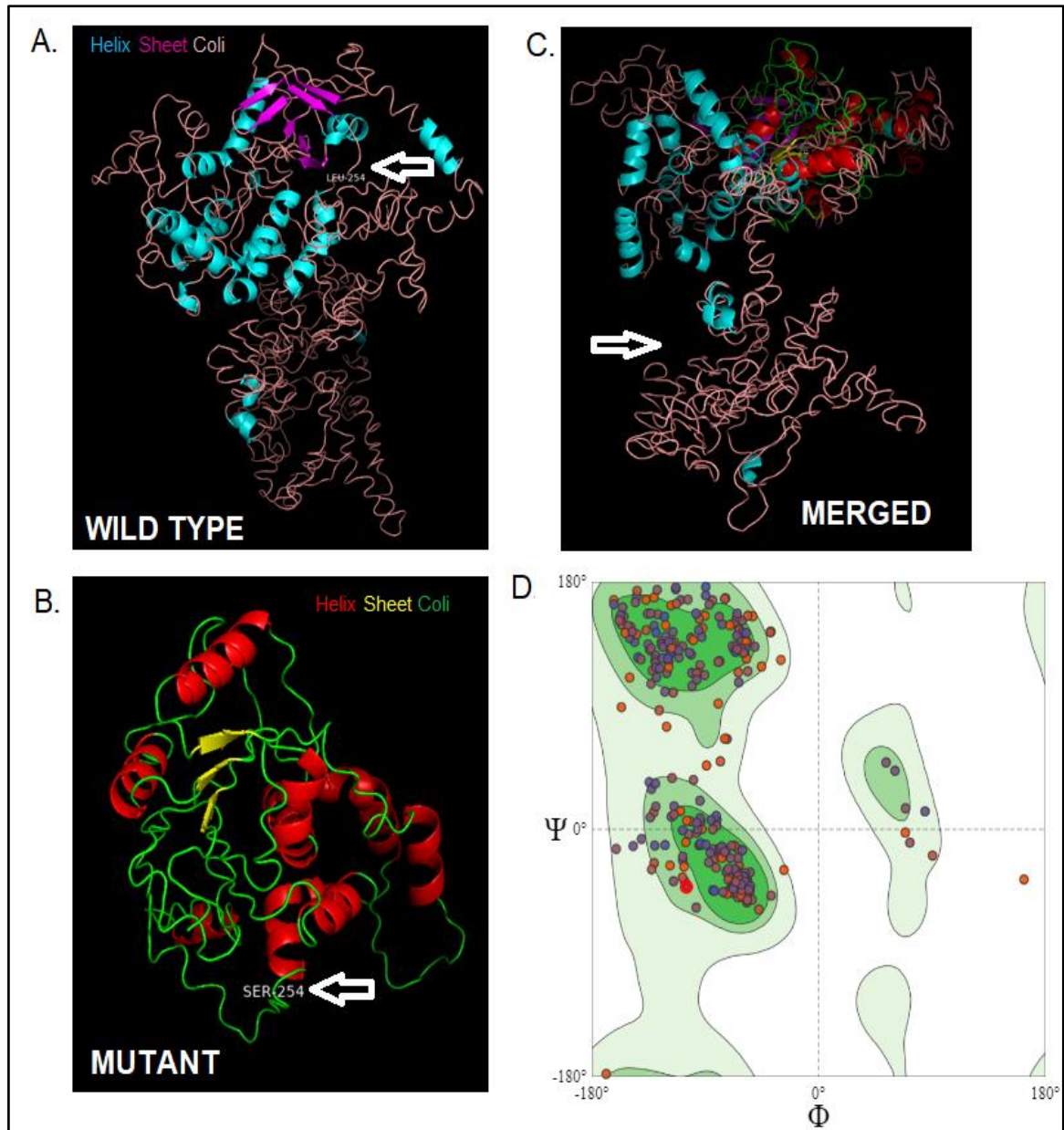
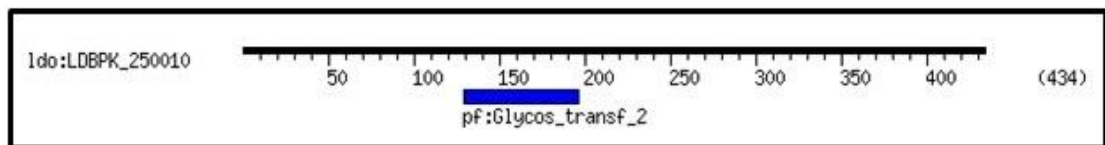


Figure 4.8. I- TASSER homology modelling for the predicted secondary structure of the Neutral sphingomyelinase activation associated factor-like protein. **Predicted tri-dimensional structure** with the position of helix, sheet and coil with different colour code, (A) **Wild type** neutral sphingomyelinase activation associated factor-like protein, (B) **Mutant** neutral sphingomyelinase activation associated factor-like protein. The position of the point mutation in the wild type and mutant are shown by white arrow head, (C) **Predicted tri-dimensional overlay structure** of the wild type and mutant protein. White arrow head are used to locate the mutated positions in the tri-dimensional structure of the protein. (D) **Ramachandran plot**, general case for the predicted neutral sphingomyelinase activation associated factor-like protein structure (wild type) by SWISS-MODEL.

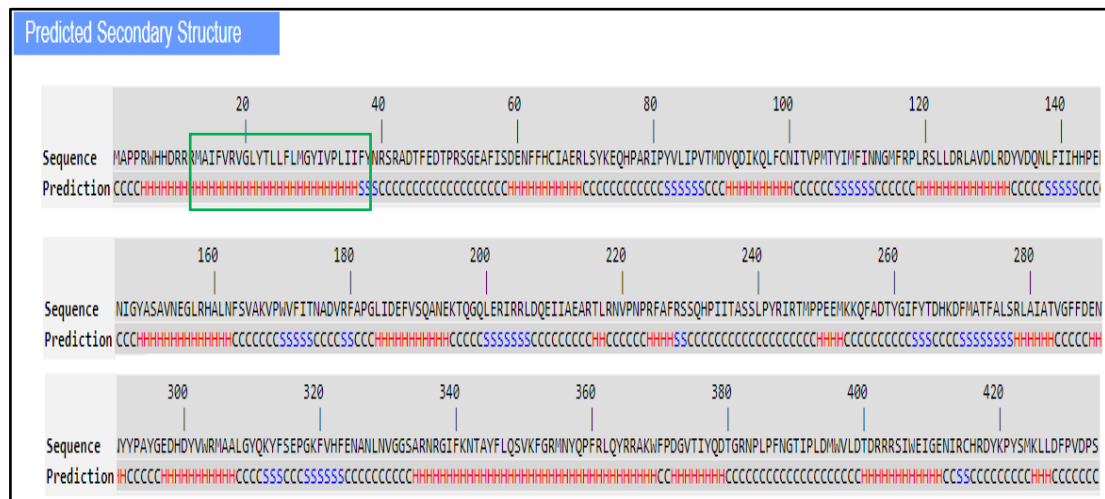
4.3.3 Beta galactofuranosyl transferase

Leishmania donovani gene LDBPK_250010 encoding the protein beta galactofuranosyl transferase or lipophosphoglycan 1 (LPG1). The length of the functional protein is 434 amino acids with 50,957 Da molecular mass. The predicted secondary structure of beta galactofuranosyl transferase protein has one transmembrane domain from amino acid position 15 to 38 (IFVVRVGLYTLLFLMGYIVPLIIFY) and one motif, Glycosyltransferase family 2 from amino acid position 129 to 196 (Fig 4.9A). The observed nucleotide deletion (112delT) leads to frameshift mutation (Tyr38fs) which is present in the transmembrane domain of the protein results truncated protein i.e., 83 amino acids due to insertion of a stop codon at amino acid position 84.

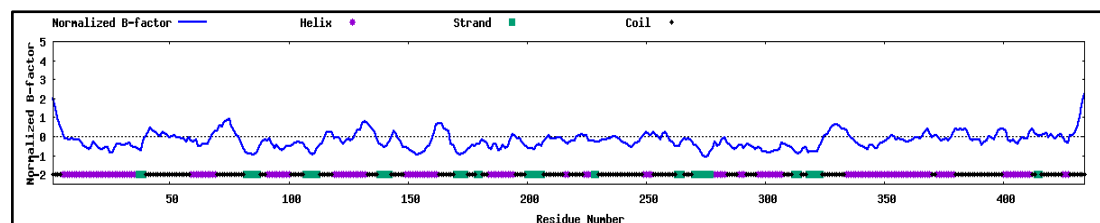
A.



B.



C.



D.

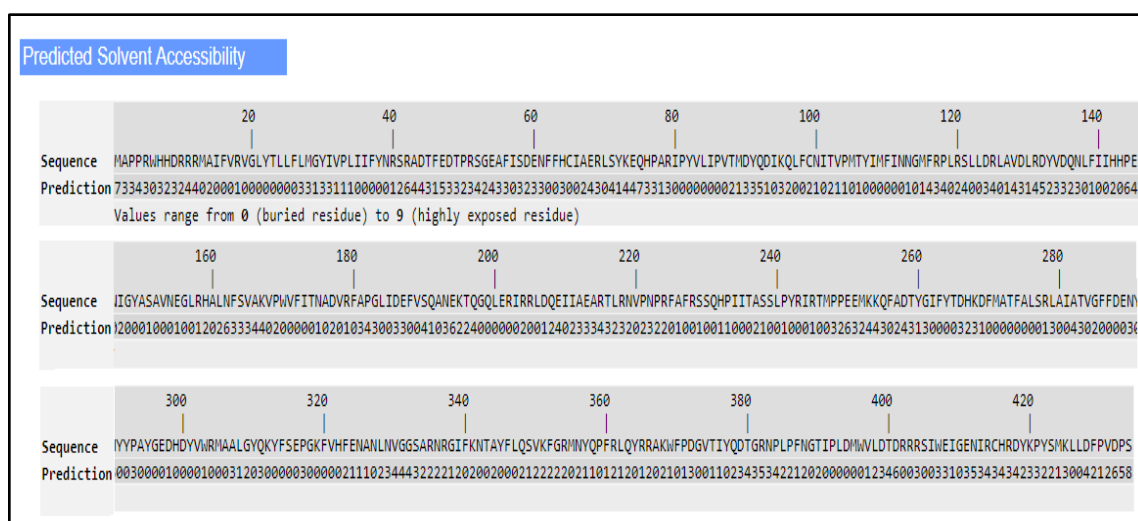


Figure 4.9. A. *L. donovani* specific KEGG database illustrating Beta galactofuranosyl transferase protein single motif, Glycosyltransferase family 2 (129-196 amino acid). **I- TASSER results illustrating, B. Predicted secondary structure** of the 434 amino acids in beta galactofuranosyl transferase protein (Wild type), where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others) and highlighted green box indicates transmembrane domain, **C. Predicted normalized B-factor** of the wild type protein. **D. Predicted solvent accessibility** are encoded in numerical value states where 0 indicates buried and 9 means highly exposed residue in wild type protein.

The predicted secondary structure, solvent accessibility and normalized B-factor of the wild type beta galactofuranosyl transferase protein is explained in the Figure 4.9 B, C and D. Predicted solvent accessibility showing the highly exposed and buried residues (Figure 4.9 D). The predicted secondary structural changes and truncated protein length i.e., 83 amino acids of the mutant beta galactofuranosyl transferase protein is shown in the Figure 4.10 A. The comparative study between the wild type and mutant secondary structure of beta galactofuranosyl transferase shown the frameshift mutation after the mutant position 38 and insertion of stop codon at 84th position. Some secondary structural changes i.e., coil to helix and coil to strand is also observed which indicates this frameshift mutation promote the major secondary structural changes in this protein which truncated the function motif length

C.

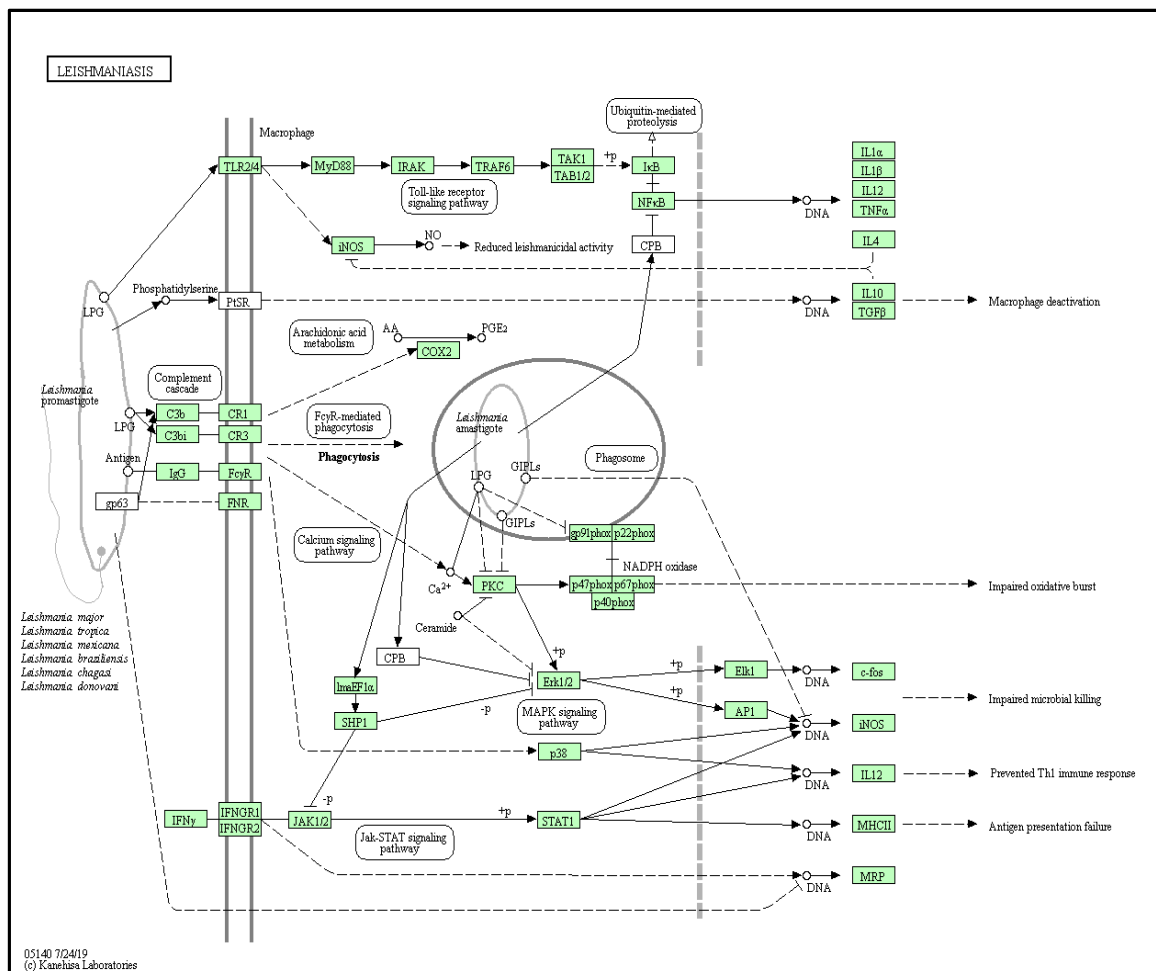


Figure 4.10. I- TASSER results illustrating A. The predicted secondary structure of the mutant Beta galactofuranosyl transferase protein, with 83 amino acids after frame shift mutation where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others), B. Comparison between the predicted secondary structure of wild type and mutant beta galactofuranosyl transferase protein. At position 38, highlighted green box indicates the site of point mutation within the transmembrane domain (IFVRVGLYTLFLMGYIVPLIIFY), and orange box indicates the site of frameshift. Highlighted blue box indicating the changes in secondary structure after the frameshift mutation. C. Metabolic pathways associated with Lipophosphoglycan (LPG) in *Leishmania* during host-parasite interactions.

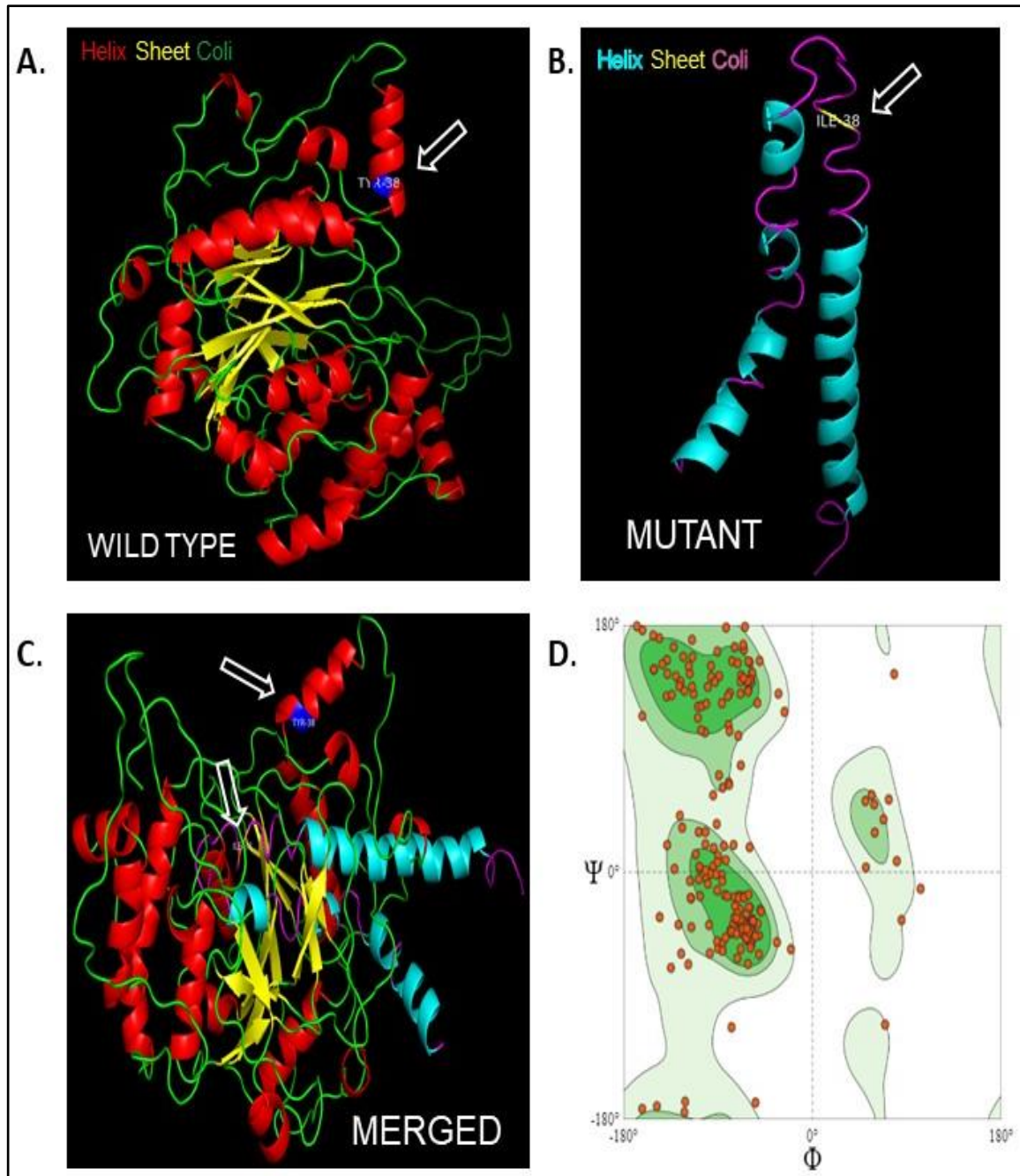


Figure 4.11. I- TASSER homology modelling for the predicted secondary structure of the Beta galactofuranosyl transferase protein. **Predicted tridimensional structure** with the position of helix, sheet and coil with different colour code of the **(A) Wild type** beta galactofuranosyl transferase protein. **(B) Mutant** beta galactofuranosyl transferase protein. The position of the point mutation in the wild type and mutant are shown by white arrow head, **(C) Predicted tri-dimensional overlay structure** of the wild type and mutant protein. The position of the point mutation in the wild type and mutant are shown by white arrow head, **(D) Ramachandran plot**, general case for the predicted beta galactofuranosyl transferase structure (wild type) by SWISS-MODEL.

The role of LPG in parasite intracellular interaction and proliferation within the host cell is been also reported (Forestier et al., 2015). Beta galactofuranosyl transferase is the key enzyme for LPG biosynthesis, hence modulation of the functional activity of such a key enzyme might be adopt by parasites as a part of their survival behaviour inside host cell. The wild type, mutant and the aligned tertiary (3D) structure of the beta galactofuranosyl transferase protein along with their mutational residues are shown in the Figure 4.11 A, B & C. A stereochemical assessment was undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral angles i.e Phi and Psi distribution. The Ramachandran plot statistics of the modelled tri-dimensional wild type beta galactofuranosyl transferase protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (97.0%) and the additional allowed (2.7%) energy regions, meanwhile only the 0.3% were at the disallowed regions (Figure 4.11D). This result suggests that the quality of predicted beta galactofuranosyl transferase protein structure is reliable. In Figure 4.11A white arrow head indicates the position (35th amino acid) of mutation where tyrosine (Tyr) residue is placed but after the frameshift mutation (Figure 4.10, 4.11 B) tyrosine is replaced with isoleucine (Ile) at amino acid position 35 in the protein. In the Figure 4.8C, aligned protein 3D structure indicates the merged protein structure with unaligned coil region after the frameshift mutation. The frameshift mutation that leads to insertion of stop codon at amino acid position 84 (Figure 4.10 B), results reduction in the functional length of the protein motif Glycosyltransferase family 2 that could be responsible for interrupting various molecular functions including the biosynthesis of most abundant surface glycoprotein LPG which is associated with host- parasite interactions could be play a vital role in intracellular parasite survival mechanism.

4.3.4 Amino Acid Transporter aATP11, Putative

Leishmania donovani gene LDBPK_310600 encoding the protein amino acid transporter aATP11, Putative. The length of the functional protein is 509 amino acids with 55,734 Da molecular mass. The predicted secondary structure of amino Acid Transporter aATP11 protein has eleven transmembrane domains and three motifs (Table 4.5, Figure 4.12). The observed nucleotide deletion (810 del T) leads to frameshift mutation (Phe 270 fs) which is present in the fifth transmembrane domain (AVGVFLIVFFVICAIYHSIV) and the first motif i.e., transmembrane amino acid transporter of the protein. The observed mutation results a truncated protein length of 282 amino acids due to insertion of a stop codon at amino acid position 283.

Table.4.5 List of motifs and transmembrane domains in the amino acid transporter aATP11.

Motif id	From	To	Feature	Position(s)	Length
Aa_trans Transmembrane amino acid transporter protein	113	502	Transmembrane	103 – 129	26
			Transmembrane	141 – 162	21
			Transmembrane	188 – 210	22
			Transmembrane	230 – 249	19
			Transmembrane	261 – 280	19
Trp_Tyr_perm Tryptophan/tyrosine permease family	113	249	Transmembrane	300 – 322	25
			Transmembrane	334 – 356	21
			Transmembrane	376 – 399	23
			Transmembrane	420 – 438	18
MaAIMP_sms Putative methionine and alanine importer, small subunit	138	172	Transmembrane	444 – 467	23
			Transmembrane	479 – 507	28

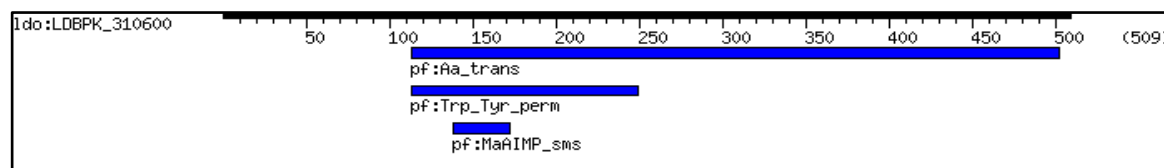
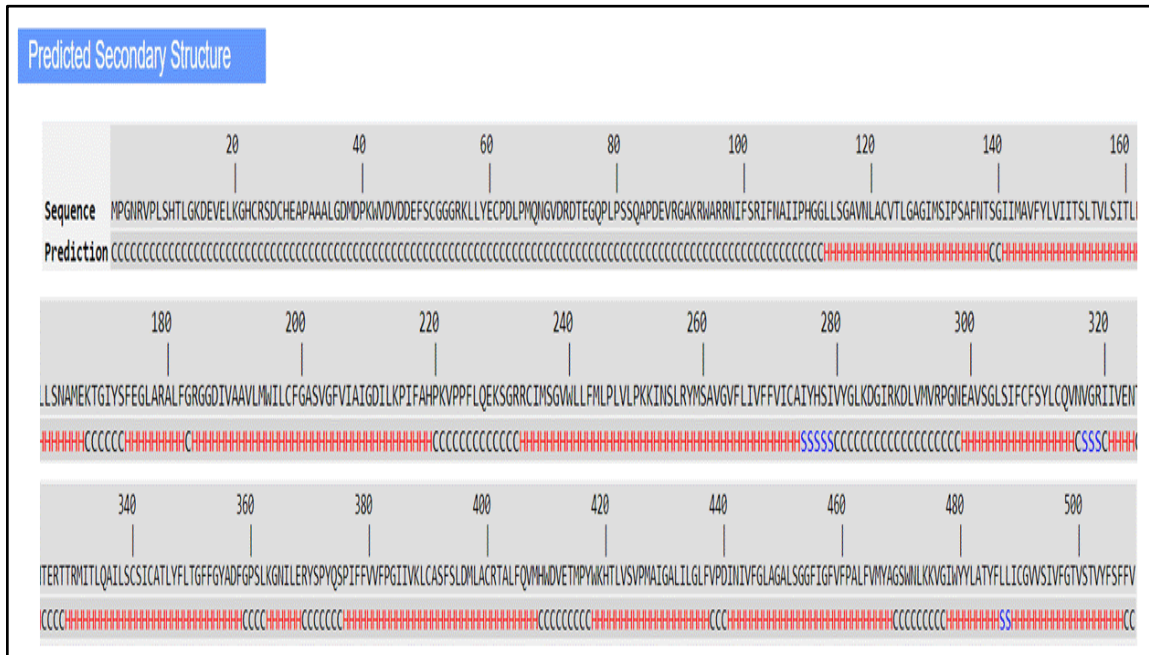


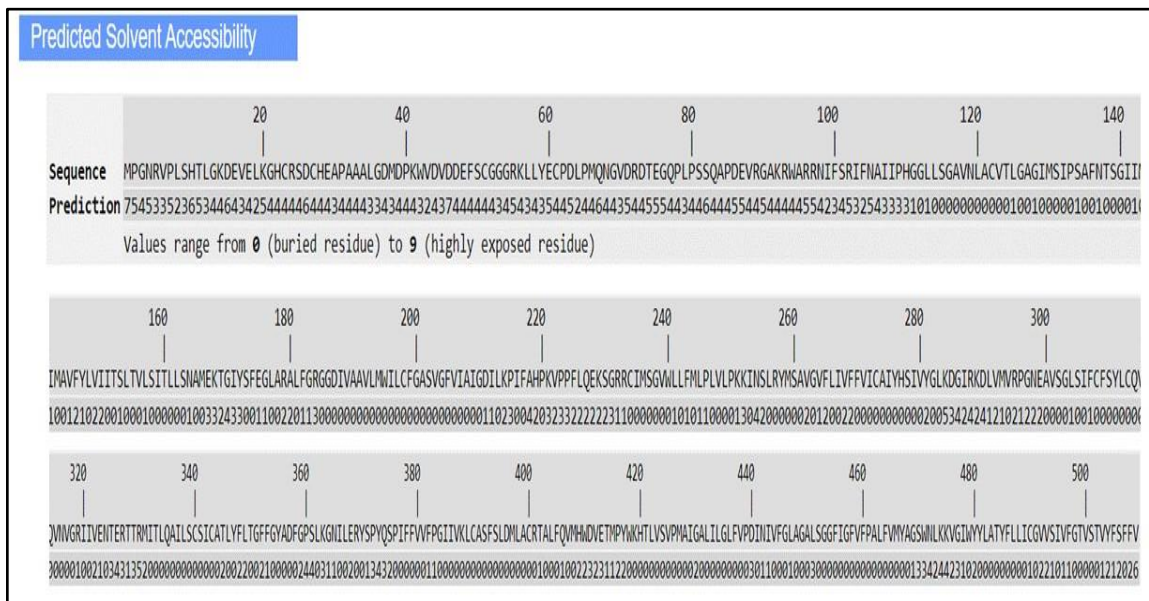
Figure 4.12. *L. donovani* specific KEGG database illustrating three motifs of the amino acid transporter aATP11, putative protein.

The predicted secondary structure, solvent accessibility and normalized B-factor of the wild type amino acid transporter aATP11, putative protein is explained in the Figure 4.13A, B and C. Predicted solvent accessibility showing the highly exposed and buried residues (Figure 4.13 B).

A.



B.



C.

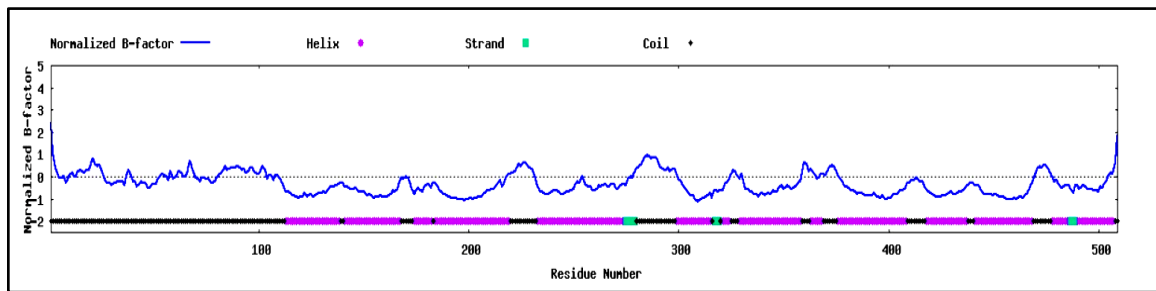
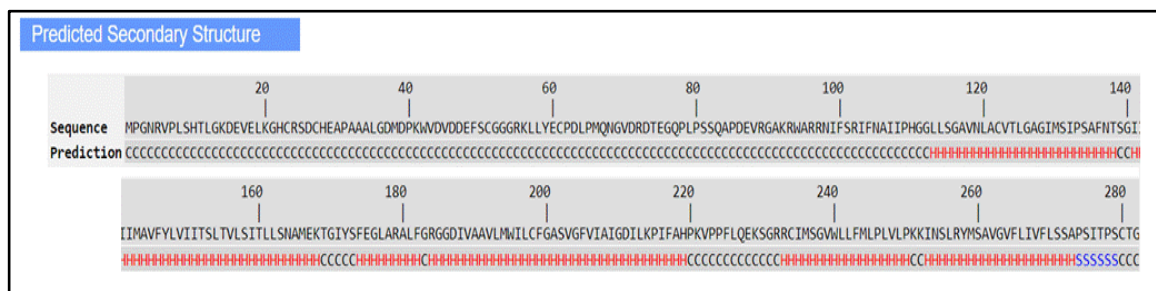


Figure 4.13. I- TASSER results illustrating the 509 amino acids in amino Acid Transporter aATP11, Putative (Wild type), **A. Predicted secondary structure** where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others). **B. Predicted solvent accessibility** are encoded in numerical value states where 0 indicates buried and 9 means highly exposed residue. **C. Predicted normalized B-factor** of the wild type protein,

The predicted secondary structural changes and truncated protein length i.e., 282 amino acids of the mutant amino acid transporter aATP11, putative protein is shown in the Figure 4.14 A. The comparative study between the wild type and mutant secondary structure of amino acid transporter aATP11, putative protein shown the frameshift mutation after the mutant position 270 and leads to insertion of stop codon at 283th position. Some secondary structural changes i.e., helix to coil is also observed which indicates this frameshift mutation promotes the major secondary structural changes in this protein. The molecular function of gene is associated with the transmembrane amino acid transport mostly during amino acid starvation mainly arginine.

A.



B.

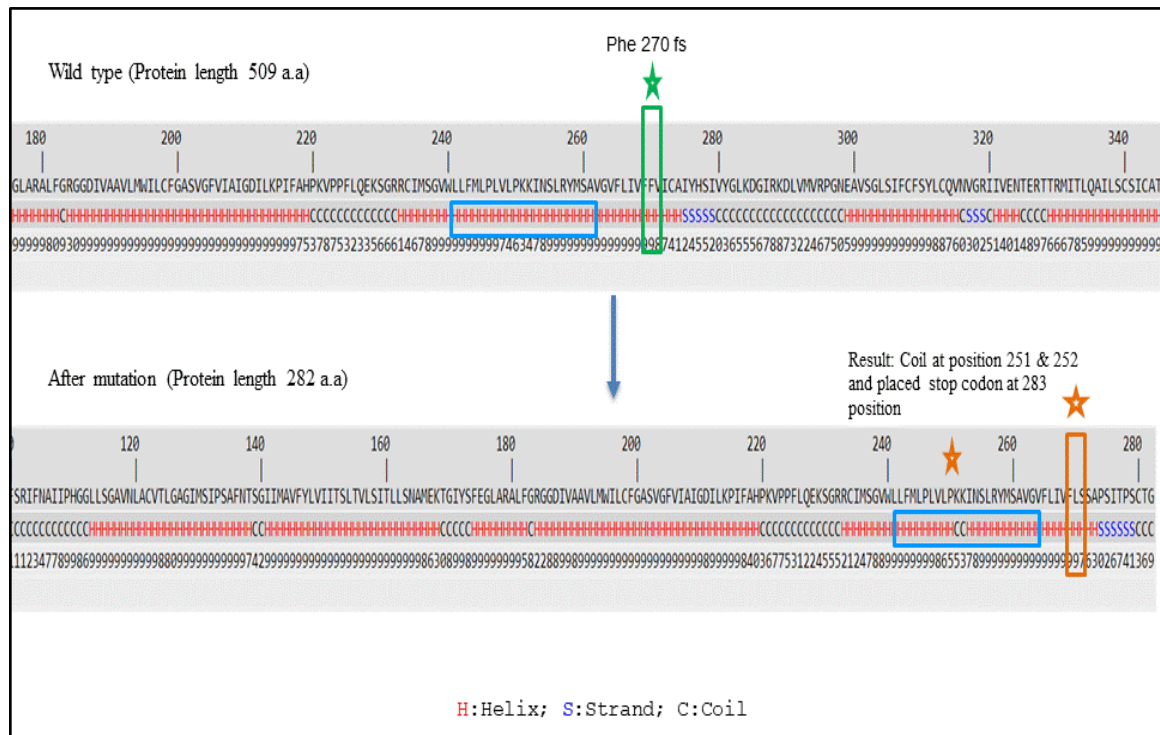


Figure 4.14. I- TASSER results illustrating, A. The predicted secondary structure of the mutant amino acid transporter aATP11, Putative, protein with 282 amino acids after frame shift mutation where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others). B. Comparison between predicted secondary structure of wild type and the mutant amino Acid Transporter aATP11, Putative. Highlighted green box indicates the site of point mutation, and orange box indicates the site of frameshift from amino acid position 270 and stop codon placed at position 283. Highlighted blue box indicating the respective changes in secondary structure after the frameshift mutation.

Previous study suggested that in response to cellular stress inside the host macrophages the *Leishmania* parasites exhibits upregulation and downregulation of the acid transporter aATP11 members that regulates many signalling pathways including mTORC1 signaling which is associated with the macrophage polarisation (Tomiotto et al., 2018). The wild type, mutant and the aligned tertiary (3D) structure of the amino acid transporter aATP11, putative, protein along with their mutational residues are shown in the Figure 4.14 A, B and C.

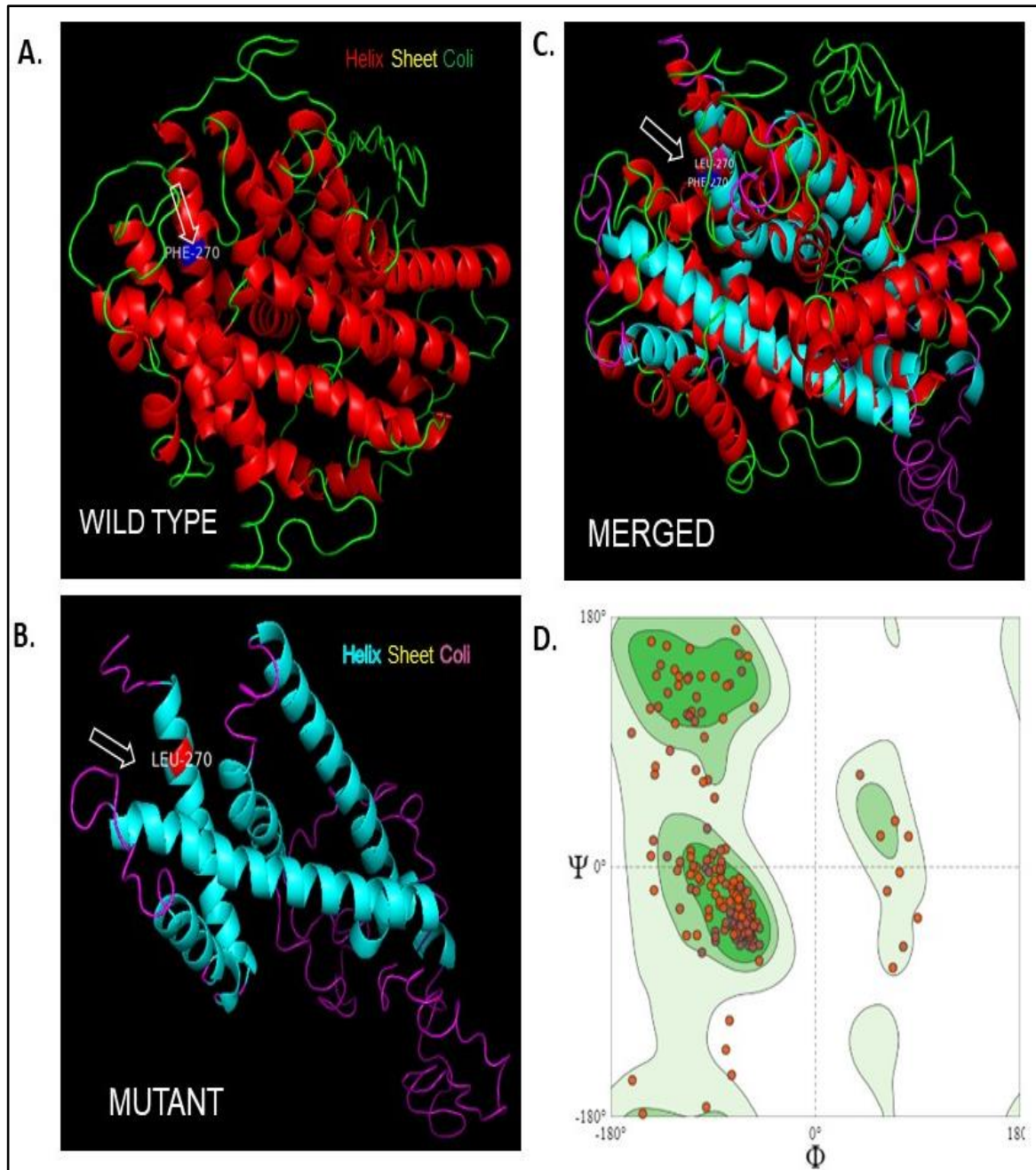


Figure 4.15. I- TASSER homology modelling for the predicted secondary structure of the amino acid transporter aATP11 protein. **Predicted tri-dimensional structure** with the position of helix, sheet and coil with different colour code of the **(A) Wild type** amino acid transporter aATP11. **(B) Mutant** amino acid transporter aATP11. The position of the point mutation in the wild type and mutant are shown by white arrow head, **(C) Predicted tri-dimensional overlay structure** of the wild type and mutant protein. White arrow head are used to locate the mutated positions in the tri-dimensional structure of the protein. **(D) Ramachandran plot**, general case for the predicted secondary structure of the wild type amino acid transporter aATP11 protein by SWISS-MODEL.

A stereochemical assessment was undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral angles i.e Phi and Psi distribution. The Ramachandran plot statistics of the modelled tri-dimensional wild type amino acid transporter aATP11, putative, protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (98.0%) and the additional allowed (2.0%) energy regions, none of them were present in disallowed regions (Figure 4.14D). This result suggests that the quality of amino acid transporter aATP11, putative, protein structure is reliable.

In Figure 4.14 A white arrow head indicates the position (270th amino acid) of mutation where phenylalanine (Phe) residue is placed but after the frameshift mutation (Figure 4.14B, Figure 4.15 B) phenylalanine is replaced with leucine (Leu) at amino acid position 270 in the protein. In the Figure 4.15C, aligned protein 3D structure indicates the merged protein structure with unaligned coil region after the frameshift mutation.

Thus, mutation in acid transporter aATP11 gene that leads to pre-mature release of protein or truncated protein interferes the molecular function of the protein due to early insertion of stop codon at amino acid position 283 (Figure 4.14B), that might be one of the paths taken by parasites to avoid stress responses inside host cell.

4.3.5 p-glycoprotein e, partial

Leishmania donovani gene LDBPK_311290 encodes the protein **p-glycoprotein e, partial** and the length of the functional protein is 1,525 amino acids with 166,805 Da molecular mass. The predicted secondary structure of p-glycoprotein e protein has ten transmembrane domains (Table 4.6 B) and twenty-eight motifs (Table 4.6 A, Figure 4.16). The observed heterozygous transversion (3539G>C) mutation that leads to single nucleotide polymorphism (Ser1180Thr) which is present in the eight transmembrane domains (SYSYLLSTISTFFSTVAMM) and twenty second motif of the protein. The length of protein remains same after the mutation but major changes occur in the secondary structure of the protein. The predicted secondary structure, solvent accessibility and normalized B-factor of the wild type p-glycoprotein e, partial, protein is explained in the Figure 4.17A, B and C. Predicted solvent accessibility showing the highly exposed and buried residues (Figure 4.17 B).

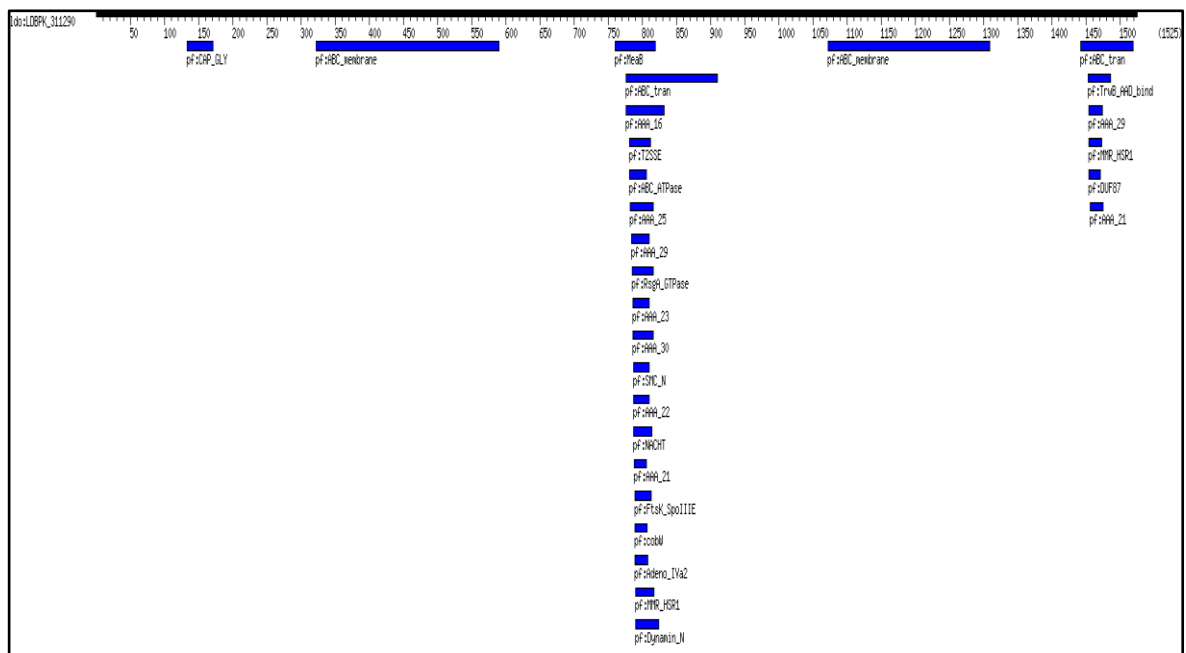


Figure 4.16. *L. donovani* specific KEGG database illustrating twenty-eight motifs of the p-glycoprotein e protein.

Table.4.6. (A) List of motifs in the protein p-glycoprotein e.

Motif id	From	To	Definition
<u>pf:CAP_GLY</u>	133	171	CAP-Gly domain
<u>pf:ABC_membrane</u>	322	590	ABC transporter transmembrane region
<u>pf:MeaB</u>	760	819	
<u>pf:ABC_tran</u>	776	910	ABC transporter
<u>pf:AAA_16</u>	776	832	AAA ATPase domain
<u>pf:T2SSE</u>	781	812	Type II/IV secretion system protein
<u>pf:ABC_ATPase</u>	781	806	Predicted ATPase of the ABC class
<u>pf:AAA_25</u>	782	816	AAA domain
<u>pf:AAA_29</u>	784	810	P-loop containing region of AAA domain
<u>pf:RsgA_GTPase</u>	785	816	
<u>pf:AAA_23</u>	786	810	AAA domain
<u>pf:AAA_30</u>	786	816	AAA domain
<u>pf:SMC_N</u>	787	810	RecF/RecN/SMC N terminal domain
<u>pf:AAA_22</u>	787	810	AAA domain
<u>pf:NACHT</u>	787	814	NACHT domain
<u>pf:AAA_21</u>	788	806	AAA domain, putative AbiEii toxin, Type IV TA system
<u>pf:FtsK_SpoIIIE</u>	789	813	FtsK/SpoIIIE family
<u>pf:cobW</u>	789	807	CobW/HypB/UreG, nucleotide-binding domain
<u>pf:Adeno_IVa2</u>	789	808	Adenovirus IVa2 protein
<u>pf:MMR_HSR1</u>	790	817	50S ribosome-binding GTPase
<u>pf:Dynamin_N</u>	790	824	Dynamain family
<u>pf:ABC_membrane</u>	1072	1309	ABC transporter transmembrane region
<u>pf:ABC_tran</u>	1442	1519	ABC transporter
<u>pf:TrwB_AAD_bind</u>	1453	1486	Type IV secretion-system coupling protein DNA-binding domain
<u>pf:AAA_29</u>	1454	1474	P-loop containing region of AAA domain
<u>pf:MMR_HSR1</u>	1454	1473	50S ribosome-binding GTPase
<u>pf:DUF87</u>	1454	1471	Domain of unknown function DUF87
<u>pf:AAA_21</u>	1456	1475	AAA domain, putative AbiEii toxin, Type IV TA system

Table.4.6. (B) List of transmembrane domains in the protein p-glycoprotein e.

Feature	Position(s)	Length
Transmembrane	358 – 375	18
Transmembrane	434 – 453	20
Transmembrane	459 – 481	23
Transmembrane	543 – 565	23
Transmembrane	577 – 600	24
Transmembrane	1066 – 1093	28
Transmembrane	1105 – 1126	22
Transmembrane	1180 – 1198	19
Transmembrane	1204 – 1224	21
Transmembrane	1296 – 1317	22

C.

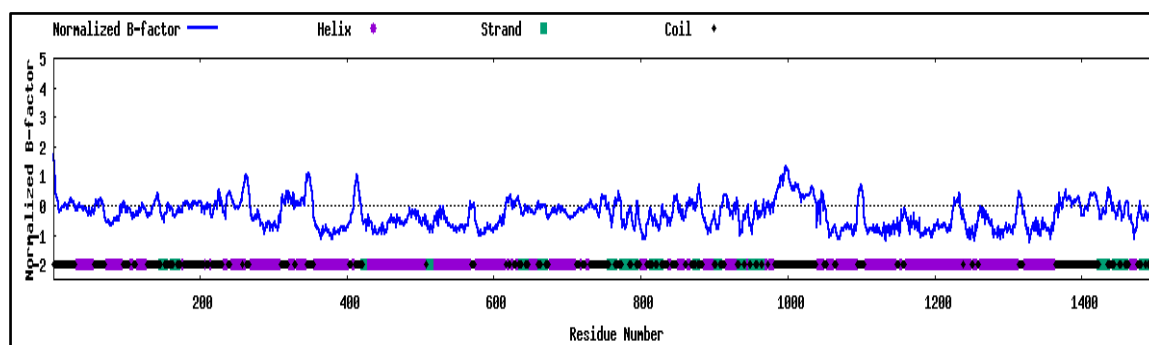
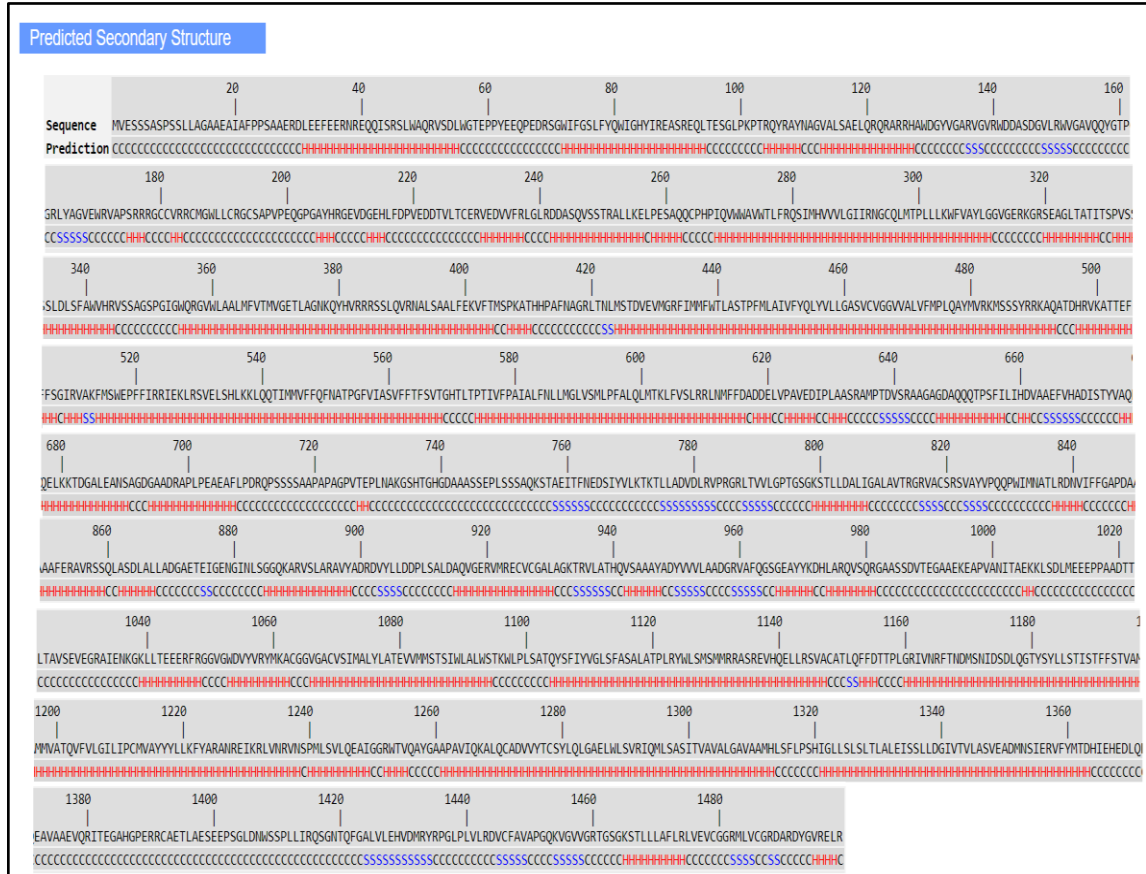


Figure 4.17. I- TASSER results illustrating the secondary structure of the surface and transporter protein p-glycoprotein e, partial (Wild type), A. Predicted secondary structure where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others), **B. Predicted solvent accessibility** are encoded in numerical value states where 0 indicates buried and 9 means highly exposed residue, **C. Predicted normalized B-factor** of the wild type protein,

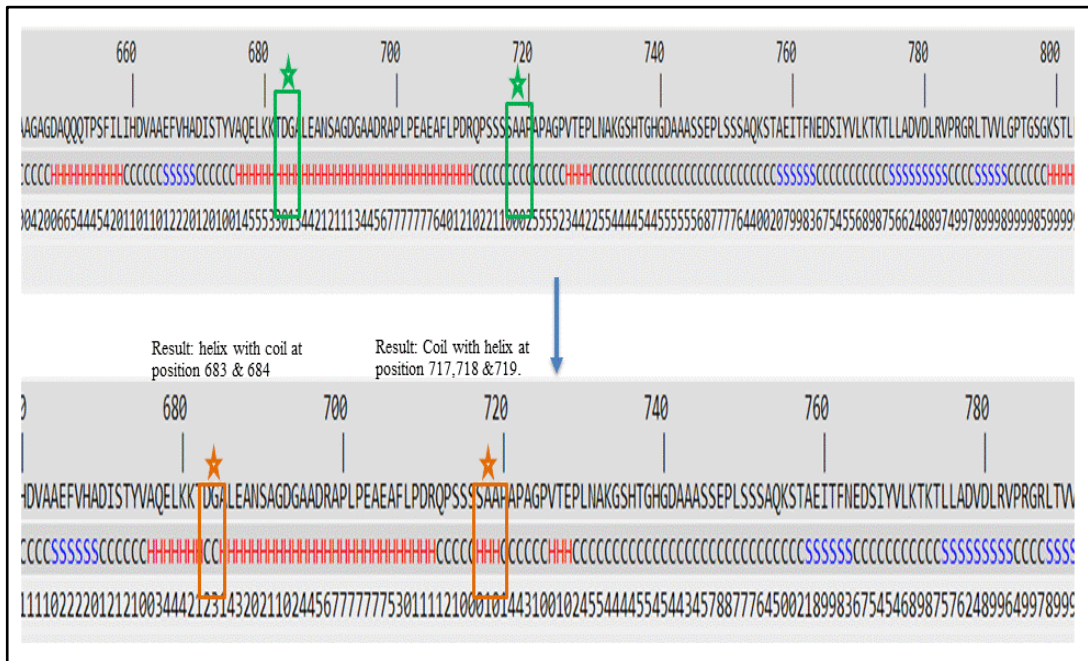
The predicted secondary structural changes with same protein length i.e., 1525 amino acids of the mutant p-glycoprotein e, partial protein is shown in the Figure 4.18 A. The comparative study between the wild type and mutant secondary structure of p-glycoprotein e, partial protein is explained in the Figure 4.18 B. The observed homozygous transversion (3539G>C) mutation that leads to single nucleotide polymorphism i.e., serine (Ser) residue is replaced with threonine (Thr) residue, at amino acid position 1180 indicates coil to helix changes in the secondary structure of mutant p-glycoprotein e, partial protein at amino acid position 1238 & 1239. In some position secondary structure changes are also observed which is shown in figure 4.18 C. The wild type, mutant and the aligned tertiary (3D) structure of the p-glycoprotein e, partial protein along with their mutational residues are shown in the Figure 4.19 A, B & C. A stereochemical assessment was undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral angles i.e. Phi and Psi distribution. The Ramachandran plot statistics of the modelled tri-dimensional wild p-glycoprotein e, partial protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (97.0%)

and the additional allowed (2.2%) energy regions, meanwhile only the 0.8% were at the disallowed regions (Figure 4.19 D). This result suggests that the quality of p-glycoprotein e, partial protein structure is reliable. In Figure 4.19 A white arrow head indicates the position (1180 amino acid) of mutation where serine (Ser) residue is placed, in Figure 4.19B white arrow head indicates that the serine is replaced with threonine (Thr) at amino acid position 1180 in the protein after transverse mutation and position of 1238 and 1239 amino acid i.e., asparagine (N) and serine (S). In the Figure 4.15C, aligned protein 3D structure indicated the merged protein structure along with unaligned coil region after the homozygous transversion mutation. Thus, the SNP in p-glycoprotein e, partial gene leads to major secondary structural changes in the 28th motif i.e., ABC transporter transmembrane region of the protein which is an important surface membrane transporter that play role during multi drug transport.

A.



B.



C.

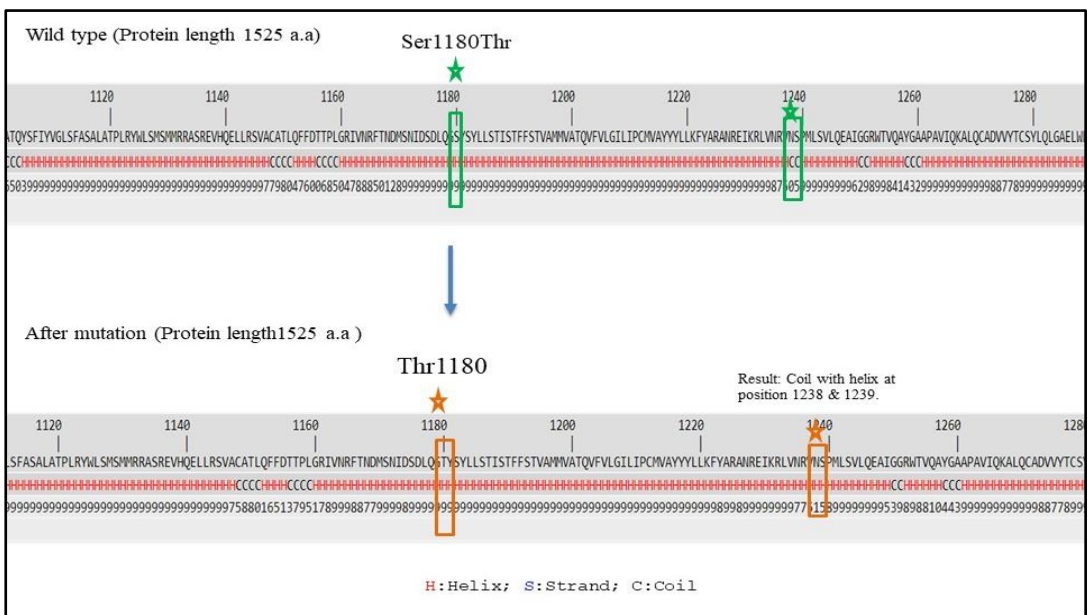


Figure 4.18. I- TASSER results illustrating, A. The predicted secondary structure of the mutant p-glycoprotein e, partial protein after homozygous transversion (3539G>C) mutation that leads amino acid change at position 1180 from serine (Ser) to threonine (Thr), where H is an (alpha) helix, S is a (beta) strand, and C is a coil (all others). B. & C. Comparison between the predicted secondary structure, that leads to SNPs (Ser1180Thr) which is present in the 8th transmembrane domain (SYSYLLSTISTFFSTVAMM) of the protein. Highlighted green and orange box indicates the site of secondary structural changes in the wild type as well as mutant p-glycoprotein e, partial protein respectively.

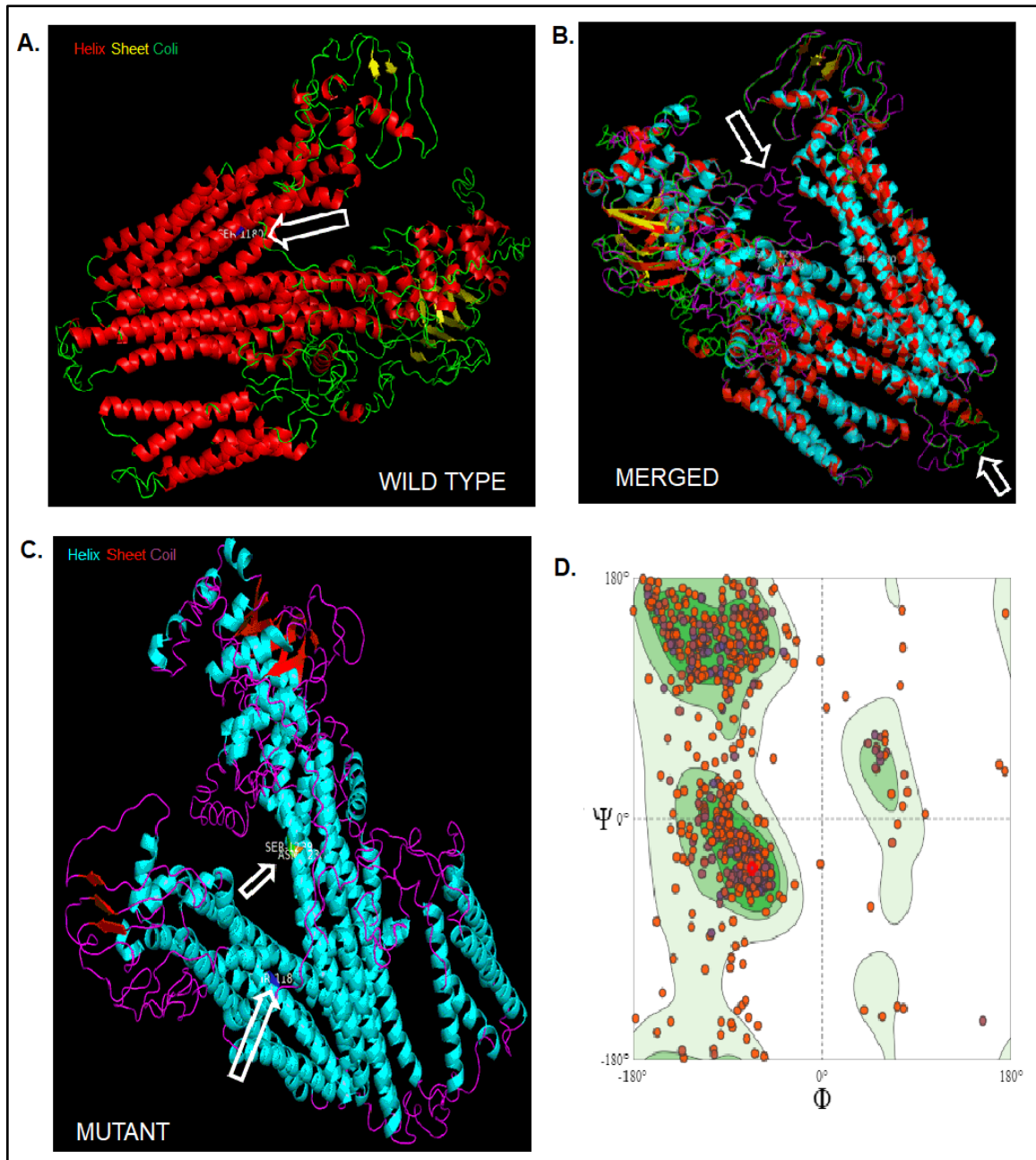


Figure 4.19. I- TASSER homology modelling for the predicted secondary structure of the p-glycoprotein e, partial protein. **Predicted tri-dimensional structure** with the position of helix, sheet and coil with different colour code of the **(A) Wild type** p-glycoprotein e, partial protein. **(B) Mutant** p-glycoprotein e, partial protein. **(C) Predicted tri-dimensional overlay structure** of the wild type and mutant protein. The position of the point mutation and change in the **tri-dimensional** structure of protein due to that mutation are shown by white arrow head. **(D) Ramachandran plot**, general case for the predicted secondary structure of wild type p-glycoprotein e, partial protein by SWISS-MODEL.

4.4 Discussion

In this chapter, we have focussed on the *in silico* structural analysis of some mutated surface and transporter genes in the *Leishmania donovani* parasites which was collected (2013-2014) from the para-KDL patients. These patients are clinically diagnosed with PKDL manifestation in which parasites are also present in their visceral organs, hence this disease taken an intermediate stage where PKDL is associated with the active VL cases (Zijlstra et al., 2003). As far the elimination programme of VL is concerned, we have to give attention towards the evolution of PKDL manifestations because of the fact that the PKDL patients act as a parasite reservoir which is thought to play an important role during the parasite's transmission cycle. Now, this arising intermediate stage of disease i.e., para-KDL cases make the situation more complicated. Although, we know that the etiopathogenesis of the PKDL is still an unsolved question (Dürbaum et al., 2016) which needs advanced molecular biological approach to explore the disease pathophysiology. Thus, to illustrate the parasites journey from visceral organ to dermal area, it is important to understand the genomic modifications of the parasites during this journey. WGS analysis of the para-KDL isolates is one of the modern and advanced tools to investigate the parasite genome wide modifications. Interestingly, after performing WGS and comparing the genome of the ten (three VL-SSG-S, three VL-SSG-R, three para-KDL-SSG-S and one VL-MIL-R) *Leishmania donovani* isolates, we found 24 novel genes mutations, which were only present in the para-KDL strains. Among them, 13 mutated genes are associated with the previously known cellular functions including transporter and surface proteins. Previously, it has been reported that the surface protein and transporter play a significant role in the host-parasite interface and also act as vital factor during parasite intracellular survival, drug metabolism and virulence (Maria Valente et al., 2019).

Therefore, among the known cellular function of 13 mutated genes, we have only taken five genes (three surface protein, two transporter) for the *in silico* structural analysis. The secondary and tertiary structural analyses were performed and a stereochemical assessment was also undertaken to further validate the model by constructing a Ramachandran plot to analyse the dihedral Φ (ϕ) and Ψ (ψ) angles. The mean Ramachandran plot statistics of the all-modelled tri-dimensional protein structure displayed that the amino acids of the predicted wild type protein structure were found mainly within the most favoured (96.0%) and the additional allowed (3.05%) energy regions, meanwhile only the 0.8% were at the disallowed regions. Hence, the quality of protein structures was reliable. In our study, the unique sets of observed mutations in the para-KDL strains are based on the non-synonymous mutation profiles in the coding sequences, which corroborate the functional role of the novel mutations that could be significantly related with the manifestation of the phenotype. Non-synonymous SNPs present in the coding sequences are very crucial, because they might be responsible for changing the structural and functional properties of the encoded proteins with possible modification in their expression level. Hence, the identified surface and transporter genes *in silico* secondary structural analysis provides an idea about their expression level. All the studied genes namely calcium-translocating P-type ATPase (LDBPK_040010), neutral sphingomyelinase activation associated factor-like protein (LDBPK_231970), beta galactofuranosyl transferase (LDBPK_250010), amino acid transporter aATP11, putative (LDBPK_310600) except one i.e., p-glycoprotein e, partial (LDBPK_311290), have shown the frameshift mutation that leads to the truncated length of functional protein which may affect their expression level. In the wider range, molecular function of these surface proteins and transporter are associated with different metabolic pathways including the host macrophages interaction, drug metabolism, biosynthesis of glycosylphosphatidylinositol (GPI)-anchor glycoprotein,

Lipophosphoglycan (LPG) and regulation of some signalling cascade that links with macrophage polarization. Interestingly, a heterozygous transversion mutation was observed in the surface protein gene LDBPK_311290 which encodes p-glycoprotein e, partial. The observed single nucleotide polymorphism in the gene LDBPK_311290 leads to polar-to-polar change in amino acid i.e., serine to threonine at amino acid position 1180 (Figure 4.15 C). The comparative study between the predicted secondary and tertiary structure of the wild type and mutant protein have been carried out to elucidate their possible modifications in the residues due to observed mutation. Here, we observed coil to helix changes in the secondary structure of the mutant protein at amino acid position 1238 & 1239. In some positions coil to helix and helix to coil structural changes are also observed (Figure 4.18 C). Thus, the single nucleotide polymorphism in the protein p-glycoprotein e, leads to a major secondary structural change in the 28th motif i.e., ABC transporter transmembrane region of the surface membrane protein that plays vital role during drug transport. The other function of this protein is associated with the ATPase activity and the multi drug resistance (Forestier et al., 2015). The comparative study between the predicted secondary and tertiary structure of the wild type and mutant protein of two surface proteins neutral sphingomyelinase activation associated factor-like protein & beta galactofuranosyl transferase and two transporter proteins calcium-translocating P-type ATPase & amino acid transporter aATP11, putative illustrate the early insertion of stop codon due to frame shift mutation at amino acid position 284, 84 and 986, 283 respectively. Hence, the observed non-synonymous coding sequences SNPs could change the structure-function properties of the encoded proteins with possible modification in expression level due to insertion of stop codon which promotes many secondary structure changes in the truncated proteins. In the protein calcium-translocating P-type ATPase helix (wild type) to strand (mutant) changes are observed in the transmembrane domain and the 8th motif of the protein, while in the

beta galactofuranosyl transferase and neutral sphingomyelinase activation associated factor-like protein, coil (wild type) to helix (mutant) changes are observed in their single motif respectively. Previously, the major functions of these surface proteins are reported as their association with the host macrophages interaction during the parasites infection, the biosynthesis of lipophosphoglycan and in the tumour necrosis factor (TNF)- α signalling pathway (John et al., 2002, Forestier et al., 2015, Liu et al., 2012) respectively. While in the amino acid transporter aATP11 protein helix (wild type) to coil (mutant) changes are observed in the fifth transmembrane domain and the first motif of the predicted secondary and tertiary structure of the protein. The molecular function of protein is associated with the transmembrane amino acid transport mostly during amino acid starvation mainly arginine and the mTORC1 signaling (Tomiotto et al., 2018). The influence of unique genomic alterations on the structural and functional features of encoded proteins would be intriguing to investigate, since this would reveal the molecular mechanism of parasite persistence that could be responsible for the PKDL manifestation in the Indian subcontinent. In our study the *L. donovani* parasites of para-KDL isolates are sensitive towards the drug SSG. Hence, the observed mutations in the surface and transporter proteins could be adopted by parasites either as an escaping phenomenon due to drug pressure or to favour macrophages M1 to M2 polarization for intracellular survival mechanism inside the host cell. Our study highlights the changes in the secondary and tertiary structure of the protein due to appearance of the unique set of mutations in the gene that encodes surface protein and transporters of the para-KDL strains and could be important for viscerotropic to dermatotropic transition of the parasites. This study provides a small contribution towards the future clinical management of the PKDL.



DISCUSSION

Leishmaniasis are complex disease forms of global concern associated with varied clinical manifestations ranging from self-healing Cutaneous Leishmaniasis (CL) to disfiguring Mucocutaneous Leishmaniasis (MCL) to fatal Visceral Leishmaniasis (VL) in humans. The causal agent of the disease is a unicellular Kinetoplastid protozoan parasite belonging to the Genus *Leishmania*. The bite of infected female sandflies transmits the disease. They are endemic in 98 countries, with 350 million people at risk and around 2 million new cases reported annually worldwide (Alvar et al., 2012). Both cutaneous and visceral forms of the diseases are present in India though Visceral Leishmaniasis (VL) or Kala-azar (KA) is a much more prominent disease than CL in India and still, is a serious public health problem in the country. The VL is endemic in the eastern part of India and 90% of VL cases are reported from the state of Bihar alone. At the same time, adjoining districts of West Bengal, Jharkhand and Uttar Pradesh are also endemic to the disease. Historically, VL is caused in India by *Leishmania donovani*, that belongs to the genus *Leishmania*. Recently, the association of other species (*L. tropica*) and other genus (*Leptomonas*) with the disease made the control strategy complicated and inadequate (Singh et al., 2013; Khanra et al., 2011, 2012; Srivastava et al., 2010). Moreover, HIV/VL co-infection has worsened the situation more (Lindoso et al., 2018). The control strategy is further besieged by the problem of about 70% of drug unresponsiveness cases towards Sodium stibogluconate (SSG), the first-line drug. The second-line drugs, Amphotericin B, Pentamidine etc., are costly & extremely toxic (Croft et al., 2006). Miltefosine (MIL), an oral cancer drug was primarily introduced to treat SSG unresponsive cases (Jha, 2006). However, within a few years, MIL-resistant cases have also been reported (Arif et al., 2008). Hence, the emergence of drug-resistant parasites and the association of other species with the disease have made elimination strategies much more challenging.

Thus, the identification of species is an essential step in the initial phase of the diagnosis and drug treatment, which would further help to combat the disease. The introduction of PCR technique-based strategy has brought a paradigm shift in the field of biological sciences, especially in the area of species identification, characterization, molecular evolutionary relationship analysis and epidemiological surveillance of the disease. The present study attempted to employ sophisticated molecular biological methods and tools for characterizing both drug-sensitive and drug-resistant clinical isolates of Indian KA along with para-KDL isolates and to establish differentiation at the genome level, if any, among them so that they may be exploited as genetic markers for identification and diagnosis. The molecular biological typing methods, namely PCR-RFLP method, Multilocus Sequence Typing (MLST) and Whole Genome Sequencing (WGS) were used and then compared the genome-wide genetic variations with their possible roles in the observed drug-sensitive and drug-resistant isolates that could be responsible for the different kinds of disease manifestation in the Indian subcontinent. Moreover, attention has been given to the role of para-KDL isolates in relation to the disease condition, mainly when PKDL patients serve as reservoirs for the parasites before initiating the subsequent cycle of infection.

In the initial part of the study, we utilized the PCR- RFLP method by taking the ITS1 and HSP70 regions as targets for diagnosis and directly identifying *Leishmania* clinical isolates collected from KA and para-KDL patients. PCR-RFLP data of ITS1 and HSP70 regions demonstrated that the causative agent of all recently collected clinical isolates from Indian KA and clinically diagnosed PKDL (para-KDL) patients were *Leishmania donovani*, possessed similar PCR-RFLP profiles to the WHO reference strain of *L. donovani* (DD8). This supports the previous report (Burza et al., 2018, Zijlstra et al., 2003) that suggested the involvement of *L. donovani* as a causative agent of VL and PKDL and also in the

complex association of VL and PKDL infection that leads to the manifestation of para Kala-azar Dermal Leishmaniasis (para-KDL) in the Indian subcontinent (Sarraf et al., 2021). Further attempts were made by the PCR-RFLP analysis of ITS regions to characterise ten restriction enzymes. Fascinatingly, for the ITS amplicon, two restriction enzymes MseI & FspI unambiguously differentiated *L. donovani* from *L. tropica* in the Indian clinical isolates collected from KA and para-KDL patients. We also observed that four restriction enzymes, Hinf I, BstDE I, BstSC I and Bst2U I have the same RFLP pattern for *L. donovani* and *L. tropica* making them inappropriate as species marker. It may be mentioned here that four restriction enzymes, Aco I, FauND I, Hpa I and MnaI did not have any restriction sites in the ITS amplicon of the studied clinical isolates. This may be for the fact that RFLP analysis only examines a part of the total variable sites so that sequence variation may go undetected (Stothard et al., 1997). So, out of ten restriction enzymes examined, two were found out to be useful for species differentiation between *L. donovani* and *L. tropica*. With these enzymes, fourteen clinical isolates were found out *L. donovani*. The rest one isolate, T5 (MHOM/IN/2010/T5) was previously found out *L. tropica* with other such restriction enzymes in our laboratory (Khanra et al., 2012) and here also by these two new restriction enzymes used, the isolate was found to share same profile as that of *L. tropica*. So, accurate marker selection for the species identification is a must before implementation of the advanced tools to detect the sequence variations in the clinical samples, particularly when dealing with VL cases in the Indian subcontinent where VL caused by *L. donovani* and *L. tropica* has already been reported (Sacks et al., 1995; Khanra et al., 2011). Various DNA sequence analysis methods have been proven beneficial for studying minor genetic variations and their application has resulted in valuable molecular biological data for clinical research.

Currently, to examine the minor genetic variations in the housekeeping genes of microorganisms, including the genus *Leishmania*, an advanced molecular biological approach of sequence typing, i.e., Multilocus Sequence Typing (MLST), is available (Miles et al., 2009, Akhouni et al., 2016) which is a highly discriminatory alternative to MLEE (Marco et al., 2015). The MLST approach mainly focuses on the housekeeping genes because they are sensitive to purifying selection (selective eradication of deleterious alleles) and evolutionary variation within them are practically neutral, i.e., they are conserved in nature (Marco et al., 2015; Margos, Gabriele, et al., 2008; Mauricio et al., 2006; Zemanova et al., 2007).

To examine the variation in the DNA sequences of the studied Indian *Leishmania spp.* population, we performed MLST analysis of the fifteen housekeeping gene coding regions, namely aconitase (*aco*), alanine aminotransferase (*alat*), enolase (*enol*), phosphoglucomutase (*pgm*), spermidine synthase (*spdsyn*), aspartate aminotransferase (*asat*), glucose-6-phosphate isomerase (*gpi*), nucleoside hydrolase 1 (*nhl*), isocitrate dehydrogenase (*icd*), mannose phosphate isomerase (*mpi*), fumarate hydratase (*fh*), 6-phosphogluconate dehydrogenase (*pgd*), glucose-6-phosphate dehydrogenase (*g6pdh*), hypoxanthine-guanine phosphoribosyl transferase (*hgprt*) and phosphomannomutase (*pmm*) for the characterization and identification of the DNA polymorphisms which would help us to explore the various genetic diversity parameters and phylodynamics.

We assessed intraspecific and interspecific genetic diversity parameters, Sequence Type (ST) and phylogenetic analysis of the fifteen clinical isolates (n=15) from Indian KA and para-KDL patients, along with the retrieved sequences of other *Leishmania* species from NCBI GenBank, USA. Total fifty reference strains were used that belong to the two subgenera *Leishmania* and *Viannia* and include 12 species.

From our DNA polymorphism data of the all studied fifteen housekeeping genes (*aco*, *alat*, *enol*, *hgprt*, *pgm*, *spdsyn*, *icd*, *fh*, *mpi*, *g6pdh*, *pgd*, *asat*, *nh1*, *gpi*, *pmm*), we concluded that in the intraspecies analysis phosphomannomutase (*pmm*) locus showed statistically significant ($P < 0.05$) Tajima's D value (Table 3.2) while in the interspecies analysis at level $P < 0.001$, significant Tajima's D Value was observed for variable regions of *pgd*, *asat*, and at level $P < 0.05$ significant Tajima's D Value was observed for variable regions of *gpi*, *mpi* and *g6pdh* gene loci. This indicates the presence of a greater number of rare alleles (Table 2.3 A) in the above mentioned loci. To decipher the intraspecies polymorphism of Indian *L. donovani*, we analysed intraspecies allelic profiles of the seven housekeeping genes *aco*, *alat*, *asat*, *enol*, *gpi*, *nh1* and *pgm*, which denotes 10, 2, 11, 6, 9, 7 and 9 number of alleles respectively. Each Indian *L. donovani* isolates (H1, A2, KK, BK, P2, P6, T2, T8, T9, T10) encodes a specific Sequence Type (ST) with a discriminatory power value of 1.0, indicating that the typing method could distinguish each member of a strain population from all other members. Furthermore, intraspecies phylogenetic analysis based on the concatenated sequence of these seven housekeeping genes represents the four distinct clusters in the Indian clinical isolates studied, which further supported the existence of intraspecies polymorphism in the Indian *Leishmania donovani* isolates. On the other hand, the intraspecific genetic diversity analysis of the spermidine synthase (*spdsyn*) locus was found out to be conserved. In the interspecific genetic diversity parameters analysis, we observed one neutrally evolved gene, 6-phosphogluconate dehydrogenase (*pgd*), which has a significantly high value (200) of the gamma parameter, impaling a very low variation of substitution rates in the coding sequence. In contrast, the shape parameter of gamma distribution was remarkably very low (0.0854) for nucleoside hydrolase 1 (*nh1*), indicating a very high variation of substitution rates among sites (Table 2.3 A).

In the interspecific non-synonymous (dN) to synonymous (dS) substitution rate analysis per nucleotide site, we observed non-synonymous substitutions in the four loci *asat*, *gpi*, *mpi* and *g6pdh* with a significant number of rare alleles (Figure 2.4). The analysed sequence type (ST) of each fifteen housekeeping gene indicated that the *aco* gene has the highest ST, i.e., 23 and the *hgprt* gene has the lowest ST, i.e., 9 in the studied interspecific *Leishmania* strains. Interestingly, we observed that only two gene loci, *alat* and *enol*, were amplified in the Indian *L. tropica* (T5) isolate that has unique sequence type ST3 for *alat* locus (Figure 2.7A) and for *enol* locus, the same sequence type ST6 was shared by two *L. donovani* (T3 & T7) isolates (Figure 2.7 B). While for *icd* loci, *L. donovani* isolate, T9 showed unique sequence type ST2. The sequence type, ST3 for the Indian T5 isolate is being reported for the first time, which was isolated from a confirmed Kala-azar patient and found to be *L. tropica*. MLST based phylogenetic analysis is a powerful method to explain the relatedness between the *Leishmania sp.* from the large geographical area and to differentiate species complex within the genus *Leishmania* (Banu et al., 2019). Here we performed interspecies phylogenetic analysis to explain the relatedness between the *Leishmania sp.* of Indian isolates with the worldwide isolates/strains and to differentiate species complex within the genus *Leishmania*. For the interspecies phylogenetic analysis, we selected all the fifteen individual gene loci and also a concatenated sequence of three groups of gene loci (Scheme-A; *aco*, *alat*, *enol*, *hgprt*, *pgm*, *spdsyn*), (Scheme-B; *pgd*, *asat*, *nh1*, *gpi*) and (Scheme-C; *icd*, *fh*, *mpi*, *g6pdh*) respectively. Among these three groups of schemes, we observed that the phylogenetic analysis of Scheme-A, a group of six genes loci *aco*, *alat*, *enol*, *hgprt*, *pgm* and *spdsyn* concatenated sequences, could clearly differentiate two species within *Leishmania* species complexes and *L. (Viannia)* subgenus level organization with high bootstrap support (Figure 2.8).

Interestingly, one common cluster was formed by one Nepal (BPK282A1) and one Canada (LdCL) *L. donovani* strain (Figure 2.8). No statistically significant incongruence was detected among loci (ILDp = 0.009). Distance matrices had high and statistically significant Kendall's $W = 0.8970$ (p value= 0.01), supporting the concatenating loci dendrogram of MLST Scheme A. Among all the studied fifteen gene loci, we found that the phylogenetic tree of individual locus isocitrate dehydrogenase (*icd*) sequence could differentiate *Leishmania* strains according to their geographical origin with maximum bootstrap support (Figure 2.9 A).

Globally, *Leishmania donovani* complex is the foremost etiologic agent of the systemic form of the disease Visceral Leishmaniasis (Burza et al., 2018), which can be potentially fatal if not treated. Worldwide around 50,000 to 90,000 new cases are reported annually, and more than half of the world's disease burden comes from the Indian subcontinent (Burza et al., 2018; Croft et al., 2006). Visceral leishmaniasis affects the poorest communities and causes epidemics in several regions (Boelaert et al., 2009). In 10%-20% of cases, asymptomatic visceral infection represents Post Kala-azar Dermal Leishmaniasis (PKDL), a dermal sequel of VL. Recently, para Kala-azar Dermal Leishmaniasis (para-KDL) cases were reported (Zijlstra et al., 2003) which is a co-association of active VL with PKDL. This further increases the complexity of disease management. The presence of active VL in an increasing number of PKDL cases requires a greater focus on the pathophysiology, diagnosis and treatment regimen of these new instances of para-KDL manifestation. Though the PKDL symptom may not appear dangerous, soon the patient's skin harbour parasites that can be transferred by sand fly bites (Zijlstra et al., 2017). As a result, PKDL patients are thought to play a key role in the parasite transmission cycle as reservoirs and are a serious issue in KA eradication programmes. Hence, divulging the mechanism that helps parasites invade visceral and dermal areas simultaneously is crucial.

Although VL was nearly eradicated from the Indian sub-continent in the 1960s (Thakur, 2007), it reappeared in 1977 and caused several successive major epidemics (Dye and Wolpert, 1988). One major reason for such incidences could be the increasing number of cases unresponsive to sodium stibogluconate (SSG), the first line of drugs for Indian KA (Jackson et al., 1990; Sundar, 2001; Sundar et al., 2006) . Unfortunately, most alternative drugs, such as Amphotericin B and Pentamidine have similar limitations and have proven to be very toxic (Amato et al., 1998; den Boer et al., 2011). Miltefosine (MIL), though introduced to treat SSG unresponsive cases, reports were coming up with unresponsive cases towards the drug (Arif et al., 2008). *L. donovani* is responsible for a significant portion of the leishmaniasis burden in terms of morbidity and mortality, particularly in East Africa and the Indian subcontinent (Downing et al., 2011; Sacks et al., 1995).

Despite the fact that many research have been conducted by utilising advanced molecular biological tools to explore the genome of *Leishmania*, but little is known about the diversity of parasite populations in the clinical aspect and how they change their genome during drug treatment (Alam et al., 2009; Schonian et al., 2008). Therefore, the emergence of drug unresponsive cases and the arrival of para-KDL cases are the major concern in the management of VL eradications. A magnificent contribution of computational science and development of database system in the biomedical fields with the emergence of advanced molecular biological techniques and multiple bioinformatics tools. The applications of these modern tools and associated methods provide an effective approach to *Leishmania* epidemiological research. Nowadays, the fourth-generation molecular biological technique, Whole Genome Sequencing (WGS) has been introduced which provides a multidisciplinary field to understand parasite biology.

To investigate the genetic variation that could be responsible for the manifestation of para-KDL and the development of drug resistance in Indian VL patients, we performed whole genome sequencing of *L. donovani* VL isolates (n=7) and para-KDL isolates (n=3) collected from the bone marrow or splenic aspirates from the Indian KA and para-KDL patients of states Bihar and West Bengal respectively during 2010 to 2014 (Table 4, Materials and Methods section). The isolates were genotyped by PCR-RFLP or PCR-ITS1 sequencing methods and their drug sensitivities were confirmed. For WGS, we selected three cloned strains of SSG sensitive VL (VL-SSG-S), three cloned strains of SSG resistant VL (VL-SSG-R) and three cloned strains of SSG sensitive para-KDL (para-KDL-S) and one cloned strain of Miltefosine resistant VL (VL-MIL-R). The WGS of the cloned clinical isolates were carried out on the Ion Proton plus platform, followed by mapping to the genomic reference sequence of LdBPK282/Oc14 (Downing et al., 2011). 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.

To analyse the association of aneuploidy with the development of PKDL and drug resistance in Indian VL patients, we estimated the chromosome copy numbers using the whole chromosome normalized median read depths (Domagalska et al., 2019; Downing et al., 2011; Dumetz et al., 2017). The result indicates the varied somy values across all studied *L. donovani* strains (Table 3.1). On an average, thirty chromosomes were disomic in all *L. donovani* strains except in the VL-MIL-R strain, where the number was only twenty-three. In para-KDL-S strains, the number of trisomy and tetrasomy chromosomes was estimated to be 4 and 1, respectively, whereas the numbers were 6 and 1, respectively in VL-SSG-S strains. Notably, we observed that the proportion of tetrasomy was more in VL-SSG-R strains.

A higher proportion of aneuploidy was observed in the VL-MIL-R strain with monosomic and pentasomic chromosomes (Table 3.1), which could be significant in developing resistance toward Miltefosine. Analysis of the some values revealed the presence of significant variability in aneuploidy across all strains. However, the drug sensitive (VL and para-KDL) and resistant (SSG and MIL) strains were broadly clustered together (Figure 1). We also observed a distinct change between the VL-SSG-S and VL-SSG-R strains in an average chromosome aneuploidy (Figure 1 and Table 2), with a more significant number of chromosomes showing varying ploidy. We mainly performed the analyses of non-synonymous SNVs and InDels to estimate the contribution of the genome-wide mutation profile of the protein coding genes in the development of PKDL and drug resistance. We have identified an average of 6700 mutations in the non-coding regions of each studied strain (Table 3.5A), ranging from 2859 mutations in VL-SSG-R strain R2 to 11600 mutations in VL-MIL-R strain T9 (Table 3.5B). On the other hand, 3837 unique non-synonymous mutations (InDels, MNVs and SNVs) were identified in the protein coding genes across all strains (Table 3.5B).

The extent of similarities among the different strains based on whole genome non-synonymous mutation profiles in the coding sequences and the profiles of the affected genes by a change in some values were shown by heat map in Figure 2. As shown in the heat maps, the para-KDL-S strains K1, K2 and K3 formed a distinct cluster based on total mutation profiles in coding sequences and profiles of mutated genes (3×3 cluster in the top left corners of Figures 2A and B). The para-KDL-S strains also showed fewer similarities with the other strains (Figure 2). Similarly, VL-SSG-S strain S1 and VL-SSG-R strain R3 also indicate less resemblance with most of the different strains. While the single VL-MIL-R strain T9 indicates the highest mutational burden, which was displayed by a distinct profile with a mixed similarity pattern (Figure 2).

We observed extensive aneuploidy across the strains with specific changes in the VL-SSG-R and the para-KDL-S strains compared to that of the VL-SSG-S strains. The variation of gene dosage through aneuploidy and mRNA turnover is essential to maintain a varied pool of messages for responding quickly to stressful and changing conditions in various hosts in the absence of transcriptional regulation of gene expression in kinetoplastid parasites (Dumetz et al., 2017). Formerly, it has been reported that the higher copy number of a chromosome increases the mean gene expression from it, and expectedly, the higher gene expression from the trisomic chromosomes rather than that from disomic chromosomes (Iantorno et al., 2017; Ubeda et al., 2008). In the para Kala-azar Dermal Leishmaniasis strains, trisomy was exclusively noticed for chromosomes 5 and 6. Several essential protein-coding genes are located on chromosomes 5 and 6, whose products are involved in various metabolic activities (Manzano et al., 2017; Tovar et al., 1998; Vickers and Beverley, 2011), including three ABCG subfamily of transporters - LABCG1 (LdBPK_060080), LABCG2 (LdBPK_060090) and LABCG3 (LdBPK_060100) on chromosome 6. Recent studies show that LABCG2 (ATP-binding cassette G subfamily) transporter is required for the externalization of phosphatidylserine (PS), leading to induction of PPAR γ in parasitized macrophages and activation of M2 phenotype (Campos-Salinas et al., 2013). Such polarization of the M2 phenotype has been shown to promote *Leishmania* survival and PKDL manifestation (Mukhopadhyay et al., 2015). But the three ABCG subfamily of transporters on chromosome 6 is 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. Interestingly, chromosome 15 is trisomic in the para kala-azar dermal leishmaniasis strains, implicating higher expression of LABCG4 and possibly leading to the survival of the parasite inside the host.

Interestingly, we also reported (Sarraf et al., 2021) the occurrence of 24 genes mutation which were only present in the para-KDL-S strains, among them 13 being homologous to genes with previously known cellular functions (Table 3.7). To examine the significance of genetic differences across all strains studied here, the Principal Component Analysis (PCA) was performed based on the frequency of occurrence of the type of mutation among the genes in the individual strains. Further, clustering analysis was performed with the sets of the mutated genes which were likely to get mutated in the respective *Leishmania* strains. As shown in Figure 3A, we observed and reported the two most significant principal components that resulted in a distinct clustering of para-KDL strains with a slight deviation of K2 from K1 and K3, indicating an overall covariance of the mutation profiles. Interestingly, different locations in the PCA plots have also been observed for VL-SSG-S strain S1, VL-SSG-R strain R3 and VL-MIL-R strain T9, implying a unique mutation burden for each of these strains.

Also, the aneuploidy profiles of the VL-SSG-S strain S1 and VL-MIL-R strain T9 strains have been distinct, indicating a unique genetic nature of the respective strains (Sarraf et al., 2021). The association of *L. donovani* with viscerotropic 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 were also observed for some of these gene loci in the Indian strains studied here (Tables 3.9, Table 3.10). However, no differential CNVs were observed among the drug sensitive and resistant VL strains and para-KDL strains except for one major facilitator superfamily protein in chromosome 29 (LdBPK_291620), which is found to be absent in SSG-R-VL strains. The absence of ABC transporter-like protein (LdBPK_111220) and phosphoglycerate kinase B (LdBPK_200120) were the other notable features found among the Indian strains.

Moreover, a similar average copy number of A2 and A2rel repeat cluster (LdBPK_220670) was found in all the Indian strains (Table 3.9). However, its lower copy number in *L. donovani* strains of Sri Lanka, where it causes Cutaneous Leishmaniasis (CL), was shown to be associated with the impaired survival of the parasite in visceral organs (Zhang et al., 2014). Among the genes that were mutated in *L. donovani* strains of Sri Lanka causing CL (Lypaczewski et al., 2018), two were found to be mutated in all the Indian strains. However, the mutation profiles were different (Table 3.9). On the other hand, in the Indian para-KDL strains, specifically, 24 genes were found to be mutated. Therefore, we conclude that the genomic variations exist in *Leishmania donovani* strains between two neighbouring countries on the same continent, and the findings strongly imply the presence of different underlying mechanisms for the development of PKDL in India and CL in Sri Lanka, even though both are skin manifestations caused by the same species of the parasite. In the para-KDL-S strains, among the 24 genes, 13 novel gene mutations are homologous to genes with previously known functions which are associated with various cellular processes, including metabolism, ion transport, repair and translation. Out of 13 genes, the mutation type in seven genes is associated with single nucleotide variation. The nature of mutation is transversion in the three genes and transition in the case of four genes. The other six genes' mutations are associated with the deletion. Hence, the functional gene attributes could be affected due to the observed mutations because these mutations may change proteins secondary and tertiary structure, which possibly affects their expression level and might play a significant role in the development of para-KDL manifestations. It has been reported that the surface protein and transporter play a significant role in the host-parasite interface and act as a vital factor during intracellular parasite survival, drug metabolism and virulence (Maria Valente et al., 2019).

Therefore, among the known cellular function of 13 mutated genes in the Indian para-KDL strains, we selected five genes (three surface proteins, two transporters) to assess the respective changes in proteins secondary and tertiary structure by *in silico* structural analysis. Through rigorous bioinformatics analysis, we have successfully analysed the secondary and tertiary structure of the five proteins, namely calcium-translocating P-type ATPase (LDBPK_040010), amino acid transporter aATP11, putative (LDBPK_310600), beta galactofuranosyl transferase (LDBPK_250010), neutral sphingomyelinase activation associated factor-like protein (LDBPK_231970) and p-glycoprotein e, partial (LDBPK_311290).

A stereochemical assessment was also performed to further validate the model by constructing a Ramachandran plot to analyse the dihedral Phi (ϕ) and Psi (ψ) angles. The mean Ramachandran plot statistics of the all-modelled tri-dimensional protein structure displayed that the amino acids of the predicted wild-type protein structure were found mainly within the most favoured (96.0%) and the additional allowed (3.05%) energy regions, meanwhile only the 0.8% energy regions were disallowed. That suggests the quality of the protein structures was reliable. The observed non-synonymous SNPs present in the coding sequences could be particularly crucial because they could change the structural and functional properties of the encoded proteins with possible modification in expression level due to the insertion of a stop codon, that promotes many secondary structure changes in the truncated proteins. In the studied two genes of surface protein (neutral sphingomyelinase activation associated factor-like protein & beta galactofuranosyl transferase) and transporter protein (calcium-translocating P-type ATPase & amino acid transporter aATP11, putative), we observed frameshift mutation that leads to the truncated length of functional protein due to the insertion of a stop codon at amino acid position 284, 84, 986 and 283 respectively.

Hence, the observed non-synonymous coding sequences SNPs could change the structural and functional properties of the encoded proteins with possible modification in expression level due to changes in their secondary and tertiary protein structures (Figure 4.4, 4.8, 4.11 & 4.15). In the broader range, the molecular function of these surface proteins and transports are associated with different metabolic pathways, including the host macrophages interaction, drug metabolism, biosynthesis of glycosylphosphatidylinositol (GPI)-anchor glycoprotein, Lipophosphoglycan (LPG) and regulation of some signalling cascade that linked with macrophage polarization.

Interestingly, a heterozygous transversion mutation was observed in the one gene LDBPK_311290, encoding p-glycoprotein e, a putative protein that acts as a surface and transporter protein. The observed single nucleotide polymorphism in the gene LDBPK_311290 leads to a polar-to-polar change in amino acid i.e., serine to threonine at amino acid position 1180 (Figure 4.15 C). The comparative study between the predicted secondary and tertiary structure of the wild-type and mutant protein has been carried out to elucidate their possible modifications in the residues due to observed mutation.

Here, we observed coil to helix changes in the secondary structure of the mutant protein at amino acid positions 1238 & 1239 (Figure 4.18 C). In some positions, coil to helix and helix to coil structural changes were also observed (Figure 4.18 C). Thus, the single nucleotide polymorphism in the protein p-glycoprotein e (partial) leads to a significant secondary structural change in the 28th motif, i.e., the ABC transporter transmembrane region (Figure 4.19) of the surface membrane protein that plays a vital role during drug transport. The other function of this protein is associated with ATPase activity and multidrug resistance (Forestier et al., 2015).

The comparative study between the predicted secondary and tertiary structure of the wild-type and mutant proteins indicates many secondary structure changes in the truncated proteins. In the calcium-translocating, P-type ATPase protein, helix (wild type) to strand (mutant) changes were observed in the transmembrane domain and the 8th motif of the protein, while in the beta galactofuranosyl transferase and neutral sphingomyelinase activation associated factor-like protein, coil (wild type) to helix (mutant) changes were observed in their single motif respectively. Previously, the major function of these surface proteins is reported as their association with the host macrophages interaction during the infection of the parasite, the biosynthesis of lipophosphoglycan and in the tumour necrosis factor (TNF)- α signalling pathway (John et al., 2002, Forestier et al., 2015, Liu et al., 2012). While in the amino acid transporter aATP11 protein, helix (wild type) to coil (mutant) changes were observed in the protein's fifth transmembrane domain and the first motif of the predicted secondary and tertiary structure of the protein. The molecular function of protein is associated with the transmembrane amino acid transport mostly during amino acid starvation, mainly arginine and the mTORC1 signaling (Tomiotto et al., 2018). The influence of unique genomic alterations on the structural and functional features of encoded proteins would be intriguing to investigate since this would reveal the molecular mechanism of parasite persistence that could be responsible for the PKDL manifestation in the Indian subcontinent. In our study, the *L. donovani* parasites of para-KDL isolates are sensitive to the drug SSG. Hence, the observed mutations in the surface and transporter proteins could be adopted by parasites either as an escaping phenomenon due to drug pressure or to favour macrophages M1 to M2 polarization for intracellular survival mechanism inside the host cell.

The study revealed a strong correlation between the development of drug resistance and the clinical manifestation of para-KDL, with special emphasis on aneuploidy and existing non-synonymous genetic mutations in the coding sequences.

The major findings of the present study are:

- The clinical status of recently collected six clinical isolates of KA and para-KDL has been confirmed as *Leishmania donovani*, including nine from our previous studies where one isolate (T5) was confirmed as *L. tropica*.
- Out of the ten restriction enzymes analyzed, only six markers have restriction sites in the ITS amplicon. Among Six markers, four markers, Hinf I, BstDE I, BstSC I, and Bst2U I, have the same RFLP patterns in both *L. donovani* and *L.tropica* and two markers, Fsp I and Mse I produced different patterns with ITS amplicon that unambiguously differentiated *L. donovani* from *L. tropica*.
- The four restriction enzymes, Aco I, FauND I, Hpa I and Mna1 did not show any restriction cut in the ITS amplicon, after restriction digestion.
- Some housekeeping gene loci indicated the significant intraspecies and interspecies genetic diversity with their specific haplotype number.
- The genus *Leishmania* displayed high nucleotide diversity among the coding regions of the studied housekeeping genes throughout the genomic sequence.
- Individual locus spermidine synthase (*spdsyn*) was found to be conserved in the Indian clinical isolates.

- In the intraspecies allelic profile analysis, the sequence typing method was able to differentiate each isolate of the *Leishmania* population from all other members of that population taking the seven housekeeping gene loci (*aco*, *alat*, *asat*, *enol*, *gpi*, *nhl*, *pgm*) in consideration, in the Indian *Leishmania donovani* clinical isolates, irrespective of their drug susceptibility.
- The intraspecies phylogenetic analysis based on the concatenated sequence of the seven housekeeping genes (*aco*, *alat*, *asat*, *enol*, *gpi*, *nhl*, *pgm*), represented the four distinct clusters of the Indian *Leishmania donovani* isolates.
- Neighbour Joining (NJ) phylogeny analysis of the concatenated sequences of six genes (*aco*, *alat*, *enol*, *pgm*, *spdsyn* and *hgprt*) could clearly differentiate two species within *Leishmania* species complexes and species level organisation with high bootstrap support.
- The interspecies phylogenetic analysis based on the individual locus Isocitrate dehydrogenase (*icd*) sequences could clearly distinguish the geographical origin of the *Leishmania sp.* complexes.
- Interspecies phylogenetic analysis of the Indian *L. donovani* and *L. tropica* indicated that the *L. tropica* (T5) isolate had a unique sequence type ST3 for *alat* locus. This is reported for the first time.
- For *enol* locus, the same ST6 was shared by two *L. donovani* (T3 & T7) isolates.

- Somy estimation analysis by whole genome sequencing of three SSG sensitive VL and para-KDL strains along with three SSG resistant and one MIL resistant VL strains from the Indian clinical isolates showed that each strain contained an average of 30 disomic chromosomes.
- In Indian VL-SSG-S and VL-SSG-R strains, somy variation was found in chromosomes 8,23,26 and 36 respectively. The proportion of tetrasomy was estimated to be more in VL-SSG-R strains.
- Notably, the absence of one important ABC transporter-like protein (LdBPK_111220) was reported in the Indian VL-SSG-R strains.
- In the Indian *L. donovani* VL-MIL-R strain (T9), a higher proportion of aneuploidy (including one tetrasomy at chromosome 23 & one pentasomy at chromosome 3) was observed with 170 unique mutations. Significant CNVs were also observed in the multi-drug resistant protein A (MRPA, LDBPK_230290) loci.
- In the Indian para-KDL-S strains, 24 unique gene mutations are reported for the first time, with 13 being homologous to genes with previously known functions related to the various cellular processes, including metabolism, ion transport, repair and translation.
- In the Indian para-KDL strains, out of 13 mutations, the observed novel mutations in the two surface proteins, namely neutral sphingomyelinase activation associated factor-like protein (Leu254fs), beta galactofuranosyl transferase (Tyr38fs) and two transporter proteins, namely amino acid transporter aATP11, putative (Phe270fs), calcium-translocating P-type ATPase (Met982fs) changes the functional length of respective proteins which affects their secondary and tertiary structure.

- A single nucleotide polymorphism (Ser1180Thr) in the gene p-glycoprotein e, partial leads to a polar-to-polar change in amino acid i.e., serine to threonine at amino acid position 1180 that leads to a significant secondary structural change in the 28th motif, i.e., the ABC transporter transmembrane region of the protein.

Overall, the study emphasized that the MLST approach of the studied fifteen housekeeping gene loci is the most effective method to explain the phylogenetic relationship of *Leishmania* species. Whole genome sequencing analysis elucidated about the specific genomic modifications that leads to structural alterations of the encoded proteins, potentially revealing molecular information about the parasite persistence that results in the coexistence of the VL and PKDL in the para-KDL patients of the Indian subcontinent. In future, the findings of our work may aid research on *Leishmania* genomics, especially in the domains of drug resistance and PKDL manifestation.



REFERENCES

- Adhya, S., Chatterjee, M., Hassan, M. Q., Mukherjee, S., & Sen, S. (1995). Detection of Leishmania in the blood of early kala-azar patients with the aid of the polymerase chain reaction. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 89(6), 622-624. [https://doi.org/10.1016/0035-9203\(95\)90416-6](https://doi.org/10.1016/0035-9203(95)90416-6)
- Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., & Sereno, D. (2016). A historical overview of the classification, evolution, and dispersion of Leishmania parasites and sandflies. *PLoS neglected tropical diseases*, 10(3), e0004349. <https://doi.org/10.1371/journal.pntd.0004349>
- Alam, M. Z., Kuhls, K., Schweynoch, C., Sundar, S., Rijal, S., Shamsuzzaman, A. K. M., ... & Schönian, G. (2009). Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of Leishmania donovani strains in the Indian subcontinent. *Infection, Genetics and Evolution*, 9(1), 24-31. <https://doi.org/10.1016/j.meegid.2008.09.005>
- Alcolea, P. J., Alonso, A., Gómez, M. J., Moreno, I., Domínguez, M., Parro, V., & Larraga, V. (2010). Transcriptomics throughout the life cycle of Leishmania infantum: high down-regulation rate in the amastigote stage. *International journal for parasitology*, 40(13), 1497-1516. <https://doi.org/10.1016/j.ijpara.2010.05.013>
- Alvar, J., Canavate, C., Molina, R., Moreno, J., & Nieto, J. (2004). Canine leishmaniasis. *Advances in parasitology*, 57(3), 1-88. [https://doi.org/10.1016/s0065-308x\(04\)57001-x](https://doi.org/10.1016/s0065-308x(04)57001-x)
- Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., ... & WHO Leishmaniasis Control Team. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS one*, 7(5), e35671. <https://doi.org/10.1371/journal.pone.0035671>
- Amato, V., Amato, J., Nicodemo, A., Uip, D., Amato-Neto, V., & Duarte, M. (1998, August). Treatment of mucocutaneous leishmaniasis with pentamidine isothionate. In *Annales de Dermatologie et de Venereologie* (Vol. 125, No. 8, pp. 492-495). <https://doi.org/10.1590/s0036-46651998000100006>
- Ansari, N. A., Kumar, R., Gautam, S., Nylén, S., Singh, O. P., Sundar, S., & Sacks, D. (2011). IL-27 and IL-21 are associated with T cell IL-10 responses in human visceral leishmaniasis. *The Journal of Immunology*, 186(7), 3977-3985. <https://doi.org/10.4049/jimmunol.1003588>

-
- Arango Duque, G., & Descoteaux, A. (2014). Macrophage cytokines: involvement in immunity and infectious diseases. *Frontiers in immunology*, 5, 491. <https://doi.org/10.3389/fimmu.2014.00491>
- Arif, S. M., Haque, M. A., Choudhury, A. M., & Ahmed, H. (2008). Unresponsiveness to miltefosine in visceral leishmaniasis (VL)-an experience of seven cases. *Journal of Medicine*, 9(2), 105-107. <https://doi.org/10.3329/jom.v9i2.1440>
- Ashford, R. W., Desjeux, P., & Deraadt, P. (1992). Estimation of population at risk of infection and number of cases of leishmaniasis. *Parasitology today*, 8(3), 104-105. [https://doi.org/10.1016/0169-4758\(92\)90249-2](https://doi.org/10.1016/0169-4758(92)90249-2)
- Ato, M., Stäger, S., Engwerda, C. R., & Kaye, P. M. (2002). Defective CCR7 expression on dendritic cells contributes to the development of visceral leishmaniasis. *Nature immunology*, 3(12), 1185-1191. <https://doi.org/10.1038/ni861>
- Bankoti, R., Gupta, K., Levchenko, A., & Stäger, S. (2012). Marginal zone B cells regulate antigen-specific T cell responses during infection. *The Journal of Immunology*, 188(8), 3961-3971. <https://doi.org/10.4049/jimmunol.1102880>
- Banu, S. S., Meyer, W., Ferreira-Paim, K., Wang, Q., Kuhls, K., Cupolillo, E., ... & Lee, R. (2019). A novel multilocus sequence typing scheme identifying genetic diversity amongst *Leishmania donovani* isolates from a genetically homogeneous population in the Indian subcontinent. *International journal for parasitology*, 49(7), 555-567. <https://doi.org/10.1016/j.ijpara.2019.02.010>
- Banuls, A. L., Hide, M., & Prugnolle, F. (2007). *Leishmania* and the leishmaniasis: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in parasitology*, 64, 1-458. [https://doi.org/10.1016/s0065-308x\(06\)64001-3](https://doi.org/10.1016/s0065-308x(06)64001-3)
- Bhattacharyya, A., Mukherjee, M., & Duttagupta, S. (2002). Studies on Stibionate unresponsive isolates of *Leishmania donovani*. *Journal of biosciences*, 27(5), 503-508. <https://doi.org/10.1007/bf02705047>
- Boelaert, M., Meheus, F., Sanchez, A., Singh, S. P., Vanlerberghe, V., Picado, A., ... & Sundar, S. (2009). The poorest of the poor: a poverty appraisal of households affected by visceral leishmaniasis in Bihar, India. *Tropical medicine & international health*, 14(6), 639-644. <https://doi.org/10.1111/j.1365-3156.2009.02279.x>

- Bossolasco, S., Nozza, S., Gaiera, G., Bestetti, A., Lazzarin, A., & Cinque, P. (2007). Lack of immune recovery in HIV/Leishmania co-infection treated with human recombinant IL-2. *AIDS*, 21(9), 1223-1225. <https://doi.org/10.1097/qad.0b013e32810c8d27>
- Brittingham, A., Morrison, C. J., McMaster, W. R., McGwire, B. S., Chang, K. P., & Mosser, D. M. (1995). Role of the Leishmania surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *The Journal of Immunology*, 155(6), 3102-3111. [https://doi.org/10.1016/0169-4758\(95\)80054-9](https://doi.org/10.1016/0169-4758(95)80054-9)
- Burza, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis *Lancet* 392: 951–970. [https://doi.org/10.1016/s0140-6736\(18\)31204-2](https://doi.org/10.1016/s0140-6736(18)31204-2)
- Burza, S., Croft, S. L., & Boelaert, M. (2019). Leishmaniasis—Authors' reply. *The Lancet*, 393(10174), 872-873. [https://doi.org/10.1016/s0140-6736\(18\)33057-5](https://doi.org/10.1016/s0140-6736(18)33057-5)
- Campos-Salinas, J., León-Guerrero, D., González-Rey, E., Delgado, M., Castanys, S., Pérez-Victoria, J. M., & Gamarro, F. (2013). LABCG2, a new ABC transporter implicated in phosphatidylserine exposure, is involved in the infectivity and pathogenicity of Leishmania. *PLoS neglected tropical diseases*, 7(4), e2179. <https://doi.org/10.1371/journal.pntd.0002179>
- Chappuis, F., Rijal, S., Soto, A., Menten, J., & Boelaert, M. (2006). A meta-analysis of the diagnostic performance of the direct agglutination test and rK39 dipstick for visceral leishmaniasis. *Bmj*, 333(7571), 723. <https://doi.org/10.1136/bmj.38917.503056.7c>
- Crandall, M. C. D. P. K., Clement, M., & Posada, D. (2000). TCS: a computer program to estimate gene genealogies. *Molecular ecology*, 9, 1657-1660. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- Croft, S. L., Sundar, S., & Fairlamb, A. H. (2006). Drug resistance in leishmaniasis. *Clinical microbiology reviews*, 19(1), 111-126. <https://doi.org/10.1128/cmr.19.1.111-126.2006>
- Cruz, I., Canavate, C., Rubio, J. M., Morales, M. A., Chicharro, C., Laguna, F., ... & Spanish HIV-Leishmania Study Group. (2002). A nested polymerase chain reaction (Ln-PCR) for diagnosing and monitoring Leishmania infantum infection in patients co-infected with human immunodeficiency virus. *Transactions of the royal society*

- of tropical medicine and hygiene, 96, S185-S189. [https://doi.org/10.1016/s0035-9203\(02\)90074-x](https://doi.org/10.1016/s0035-9203(02)90074-x)
- Cruz, I., Chicharro, C., Nieto, J., Bailo, B., Canavate, C., Figueras, M. C., & Alvar, J. (2006). Comparison of new diagnostic tools for management of pediatric Mediterranean visceral leishmaniasis. *Journal of Clinical Microbiology*, 44(7), 2343-2347. <https://doi.org/10.1128/jcm.02297-05>
- Daniel, C., Sartory, N. A., Zahn, N., Radeke, H. H., & Stein, J. M. (2008). Immune modulatory treatment of trinitrobenzene sulfonic acid colitis with calcitriol is associated with a change of a T helper (Th) 1/Th17 to a Th2 and regulatory T cell profile. *Journal of Pharmacology and Experimental Therapeutics*, 324(1), 23-33. <https://doi.org/10.1124/jpet.107.127209>
- De Doncker, S., Hutse, V., Abdellati, S., Rijal, S., Singh Karki, B. M., Decuypere, S., ... & Dujardin, J. C. (2005). A new PCR—ELISA for diagnosis of visceral leishmaniasis in blood of HIV-negative subjects. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(1), 25-31. <https://doi.org/10.1016/j.trstmh.2004.01.015>
- De Godoy Natalia, S. D., Demarchi, A. V., de Souza Regina, M. A. I. A., Thelma, O., & Almeida, B. L. M. (2018). Unusual Clinical Manifestations of *Leishmania (L.) infantum chagasi* in an HIV-coinfected Patient and the Relevance of ITS1-PCR-RFLP: A Case Report. *Iranian Journal of Parasitology*, 13(4), 655. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6348215/>
- de Godoy, N. S., Andrino, M. L. A., de Souza, R. M., Gakiya, E., Amato, V. S., Lindoso, J. Â. L., & Almeida Braz, L. M. (2016). Could kDNA-PCR in peripheral blood replace the examination of bone marrow for the diagnosis of visceral leishmaniasis?. *Journal of Parasitology Research*, 2016. <https://doi.org/10.1155/2016/1084353>
- Deborggraeve, S., Laurent, T., Espinosa, D., Van der Auwera, G., Mbuchi, M., Wasunna, M., ... & Büscher, P. (2008). A simplified and standardized polymerase chain reaction format for the diagnosis of leishmaniasis. *The Journal of Infectious Diseases*, 198(10), 1565-1572. <https://doi.org/10.1086/592509>

-
- Den Boer, M., Argaw, D., Jannin, J., & Alvar, J. (2011). Leishmaniasis impact and treatment access. *Clinical Microbiology and Infection*, 17(10), 1471-1477. <https://doi.org/10.1111/j.1469-0691.2011.03635.x>
- Desjeux, P., Piot, B., O'Neill, K., & Meert, J. P. (2001). Co-infections à Leishmania/VIH dans le sud de l'Europe [Co-infections of leishmania/HIV in south Europe]. *Medecine tropicale : revue du Corps de sante colonial*, 61(2), 187–193. <https://pubmed.ncbi.nlm.nih.gov/11582878/>
- Despommier, Dickson D., Robert W. Gwadz & Peter, H. J. (1994). *Parasitic Diseases: Third Edition*. Springer-Verlag, New York. 203-219.
- Docampo, R., Vickers, T. J., & Beverley, S. M. (2011). Folate metabolic pathways in Leishmania. *Essays in biochemistry*, 51, 63-80. <https://doi.org/10.1042/bse0510063>
- Domagalska, M. A., Imamura, H., Sanders, M., Van den Broeck, F., Bhattarai, N. R., Vanaerschot, M., ... & Dujardin, J. C. (2019). Genomes of Leishmania parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. *PLoS neglected tropical diseases*, 13(12), e0007900. <https://doi.org/10.1371/journal.pntd.0007900>
- Downing, T., Imamura, H., Decuypere, S., Clark, T. G., Coombs, G. H., Cotton, J. A., ... & Berriman, M. (2011). Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome research*, 21(12), 2143-2156. <https://doi.org/10.1101/gr.123430.111>
- Downing, T., Stark, O., Vanaerschot, M., Imamura, H., Sanders, M., Decuypere, S., ... & Schönian, G. (2012). Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the Leishmania donovani species complex. *Infection, Genetics and Evolution*, 12(1), 149-159. <https://doi.org/10.1016/j.meegid.2011.11.005>
- 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. *Molecular BioSystems*, 10(9), 2441-2447. <https://doi.org/10.1039/c4mb00287c>

- Dumetz, F., Imamura, H., Sanders, M., Seblova, V., Myskova, J., Pescher, P., ... & 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(3), e00599-17. <https://doi.org/10.1128/mbio.00599-17>
- Dürbaum, M., & Storchová, Z. (2016). Effects of aneuploidy on gene expression: implications for cancer. *The FEBS Journal*, 283(5), 791-802. <https://doi.org/10.1111/febs.13591>
- Dye, C., & Wolpert, D. M. (1988). Earthquakes, influenza and cycles of Indian kala-azar. *Transactions of the royal society of tropical medicine and hygiene*, 82(6), 843-850. [https://doi.org/10.1016/0035-9203\(88\)90013-2](https://doi.org/10.1016/0035-9203(88)90013-2)
- El Kebir, D., & Filep, J. G. (2010). Role of neutrophil apoptosis in the resolution of inflammation. *TheScientificWorldJournal*, 10, 1731-1748. <https://doi.org/10.1100/tsw.2010.169>
- El Tai, N.O., Osman, O.F., El Fari, M., Presber, W., & Schonian, G. (2000). *Trans. R Soc. Trop. Med. Hyg.* 94, 575–57. [https://doi.org/10.1016/s0035-9203\(00\)90093-2](https://doi.org/10.1016/s0035-9203(00)90093-2)
- Fisa, R., Riera, C., Ribera, E., Gallego, M., & Portus, M. (2002). A nested polymerase chain reaction for diagnosis and follow-up of human visceral leishmaniasis patients using blood samples. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96, S191-S194. [https://doi.org/10.1016/s0035-9203\(02\)90075-1](https://doi.org/10.1016/s0035-9203(02)90075-1)
- Forestier, C. L., Gao, Q., & Boons, G. J. (2015). *Leishmania* lipophosphoglycan: how to establish structure-activity relationships for this highly complex and multifunctional glycoconjugate?. *Frontiers in cellular and infection microbiology*, 4, 193. <https://doi.org/10.3389/fcimb.2014.00193>
- Fraga, J., Montalvo, A. M., De Doncker, S., Dujardin, J. C., & Van der Auwera, G. (2010). Phylogeny of *Leishmania* species based on the heat-shock protein 70 gene. *Infection, Genetics and Evolution*, 10(2), 238-245. <https://doi.org/10.1016/j.meegid.2009.11.007>
- Ganguly, S., Das, N. K., Barbhuiya, J. N., & Chatterjee, M. (2010). Post-kala-azar dermal leishmaniasis—an overview. *International journal of dermatology*, 49(8), 921-931. <https://doi.org/10.1111/j.1365-4632.2010.04558.x>

- Gelanew, T., Hailu, A., Schónian, G., Lewis, M. D., Miles, M. A., & Yeo, M. (2014). Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of *Leishmania donovani*. *International journal for parasitology*, 44(10), 751-757. <https://doi.org/10.1016/j.ijpara.2014.05.008>
- Gibson, M. E. (1983). The identification of kala-azar and the discovery of *Leishmania donovani*. *Medical history*, 27(2), 203-213. <https://doi.org/10.1017/s0025727300042691>
- Gorak, P. M., Engwerda, C. R., & Kaye, P. M. (1998). Dendritic cells, but not macrophages, produce IL-12 immediately following *Leishmania donovani* infection. *European journal of immunology*, 28(2), 687-695. [https://doi.org/10.1002/\(sici\)1521-4141\(199802\)28:02<687::aid-immu687>3.0.co;2-n](https://doi.org/10.1002/(sici)1521-4141(199802)28:02<687::aid-immu687>3.0.co;2-n)
- Grünebast, J., & Clos, J. (2020). *Leishmania*: responding to environmental signals and challenges without regulated transcription. *Computational and Structural Biotechnology Journal*, 18, 4016-4023. <https://doi.org/10.1016/j.csbj.2020.11.058>
- Gupta, A. K., & Gupta, U. D. (2020). Next generation sequencing and its applications. In *Animal Biotechnology* (pp. 395-421). Academic Press. <https://doi.org/10.1016/b978-0-12-8111710-1.00018-5>
- Gupta, G., Oghumu, S., & Satoskar, A. R. (2013). Mechanisms of immune evasion in leishmaniasis. In *Advances in applied microbiology* (Vol. 82, pp. 155-184). Academic Press. <https://doi.org/10.1016/b978-0-12-407679-2.00005-3>
- Hasegawa, M., Kishino, H., & Yano, T. A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of molecular evolution*, 22(2), 160-174. <https://doi.org/10.1007/bf02101694>
- 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 in parasitology*, 33(3), 162-174. <https://doi.org/10.1016/j.pt.2016.11.003>
- Houghton, R. L., Petrescu, M., Benson, D. R., Skeiky, Y. A., Scalone, A., Badaró, R., ... & Gradoni, L. (1998). A cloned antigen (recombinant K39) of *Leishmania chagasi* diagnostic for visceral leishmaniasis in human immunodeficiency virus type 1 patients and a prognostic indicator for monitoring patients undergoing drug therapy.

- The Journal of infectious diseases, 177(5), 1339-1344.
<https://doi.org/10.1086/515289>
- Iantorno, S. A., Durrant, C., Khan, A., Sanders, M. J., Beverley, S. M., Warren, W. C., ... & Grigg, M. E. (2017). Gene expression in *Leishmania* is regulated predominantly by gene dosage. *MBio*, 8(5), e01393-17. <https://doi.org/10.1128/mbio.01393-17>
- Imamura, H., Downing, T., Van den Broeck, F., Sanders, M. J., Rijal, S., Sundar, S., ... & Cotton, J. A. (2016). Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *elife*, 5, e12613. <https://doi.org/10.7554/elife.12613>
- Jackson, J. E., Tally, J. D., Ellis, W. Y., Mebrahtu, Y. B., Lawyer, P. G., Were, J. B., ... & Limmer, B. L. (1990). Quantitative in vitro drug potency and drug susceptibility evaluation of *Leishmania* ssp. from patients unresponsive to pentavalent antimony therapy. *The American journal of tropical medicine and hygiene*, 43(5), 464-480. <https://doi.org/10.4269/ajtmh.1990.43.464>
- Jeffares, D. C., Tomiczek, B., Sojo, V., & Reis, M. D. (2015). A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. In *Parasite genomics protocols* (pp. 65-90). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-1438-8_4
- Jha, T. K. (2006). Drug unresponsiveness & combination therapy for kala-azar. *Indian Journal of Medical Research*, 123(3), 389. https://doi.org/10.1007/978-94-007-0277-6_4
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27-30. <https://doi.org/10.1093/nar/28.1.27>
- Katakura, K., Kawazu, S. I., Sanjyoba, C., Naya, T., Matsumoto, Y., Ito, M., ... & Hashiguchi, Y. (1998). *Leishmania* mini-exon genes for molecular epidemiology of leishmaniasis in China and Ecuador. *The Tokai Journal of Experimental and Clinical Medicine*, 23(6), 393-399. <https://doi.org/10.1128/jcm.36.8.2173-2177.1998>
- Kaufer, A., Ellis, J., & Stark, D. (2019). Identification of Clinical Infections of *Leishmania* Imported into Australia: Revising Speciation with Polymerase Chain Reaction-RFLP of the Kinetoplast Maxicircle. *The American journal of tropical medicine and hygiene*, 101(3), 590. <https://doi.org/10.4269/ajtmh.19-0095>

- Kaur, G., & Rajput, B. (2014). Comparative analysis of the omics technologies used to study antimonial, amphotericin B, and pentamidine resistance in leishmania. *Journal of parasitology research*, 2014. <https://doi.org/10.1155/2014/726328>
- Khanra, S., Bandopadhyay, S. K., Chakraborty, P., Datta, S., Mondal, D., Chatterjee, M., ... & Manna, M. (2011). Characterization of the recent clinical isolates of Indian Kala-azar patients by RAPD-PCR method. *Journal of Parasitic Diseases*, 35(2), 116-122. <https://doi.org/10.1007/s12639-011-0048-1>
- 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 tropica*, 124(3), 229-234. <https://doi.org/10.1016/j.actatropica.2012.08.017>
- 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. *Scientific reports*, 7(1), 1-12. <https://doi.org/10.1038/s41598-017-09720-1>
- 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 Tropica*, 164, 177-184. <https://doi.org/10.1016/j.actatropica.2016.09.012>
- Koltas, I. S., Eroglu, F., Uzun, S., & Alabaz, D. (2016). A comparative analysis of different molecular targets using PCR for diagnosis of old world leishmaniasis. *Experimental parasitology*, 164, 43-48. <https://doi.org/10.1016/j.exppara.2016.02.007>
- Koonin, E. V., & Galperin, M. Y. (1997). Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Current opinion in genetics & development*, 7(6), 757-763. [https://doi.org/10.1016/s0959-437x\(97\)80037-8](https://doi.org/10.1016/s0959-437x(97)80037-8)
- Korber, B. (2000). HIV signature and sequence variation analysis. *Computational analysis of HIV molecular sequences*, 4, 55-72. https://doi.org/10.1007/0-306-46900-6_4
- Kumar, P., Chatterjee, M., & Das, N. K. (2021). Post kala-azar dermal leishmaniasis: Clinical features and differential diagnosis. *Indian Journal of Dermatology*, 66(1), 24. https://doi.org/10.4103/ijd.ijd_602_20

- Kumar, R., Chauhan, S. B., Ng, S. S., Sundar, S., & Engwerda, C. R. (2017). Immune checkpoint targets for host-directed therapy to prevent and treat leishmaniasis. *Frontiers in immunology*, 8, 1492. <https://doi.org/10.3389/fimmu.2017.01492>
- Kumar, R., Das, V. N. R., Topno, R. K., Pal, B., Imam, A., Agrawal, K., ... & Pandey, K. (2016). Para-kala-azar dermal Leishmaniasis cases in Indian subcontinent—A case series. *Pathogens and Global Health*, 110(7-8), 326-329. <https://doi.org/10.1080/20477724.2016.1258163>
- Lahav, T., Sivam, D., Volpin, H., Ronen, M., Tsigankov, P., Green, A., ... & Myler, P. J. (2011). Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *The FASEB Journal*, 25(2), 515-525. <https://doi.org/10.1096/fj.10-157529>
- Lanave, C., Preparata, G., Sacone, C., & Serio, G. (1984). A new method for calculating evolutionary substitution rates. *Journal of molecular evolution*, 20(1), 86-93. <https://doi.org/10.1007/bf02101990>
- Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110-1116. <https://doi.org/10.1111/2041-210x.12410>
- Leprohon, P., Légaré, D., & Ouellette, M. (2009). Intracellular localization of the ABCC proteins of *Leishmania* and their role in resistance to antimonials. *Antimicrobial agents and chemotherapy*, 53(6), 2646-2649. <https://doi.org/10.1128/aac.01474-08>
- Leprohon, P., Legare, D., Raymond, F., Madore, E., Hardiman, G., Corbeil, J., & Ouellette, M. (2009). Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic acids research*, 37(5), 1387-1399. <https://doi.org/10.1093/nar/gkn1069>
- Lindoso, J. A. L., Cunha, M. A., Queiroz, I. T., & Moreira, C. H. V. (2016). Leishmaniasis–HIV coinfection: current challenges. *Hiv/aids (Auckland, NZ)*, 8, 147. <https://doi.org/10.2147/hiv.s93789>
- Lindoso, J. A. L., Moreira, C. H. V., Cunha, M. A., & Queiroz, I. T. (2018). Visceral leishmaniasis and HIV coinfection: current perspectives. *Hiv/aids (Auckland, NZ)*, 10, 193. <https://doi.org/10.2147/hiv.s143929>

- Liu, D., & Uzonna, J. E. (2012). The early interaction of *Leishmania* with macrophages and dendritic cells and its influence on the host immune response. *Frontiers in cellular and infection microbiology*, 2, 83. <https://doi.org/10.3389/fcimb.2012.00083>
- Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., ... & Remuzzi, G. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, 380(9859), 2095-2128.
- Lypaczewski, P., Hoshizaki, J., Zhang, W. W., McCall, L. I., Torcivia-Rodriguez, J., Simonyan, V., ... & Matlashewski, G. (2018). A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. *Scientific reports*, 8(1), 1-14. <https://doi.org/10.1038/s41598-018-34812-x>
- Lytsy, B., Engstrand, L., Gustafsson, Å., & Kaden, R. (2017). Time to review the gold standard for genotyping vancomycin-resistant enterococci in epidemiology: comparing whole-genome sequencing with PFGE and MLST in three suspected outbreaks in Sweden during 2013–2015. *Infection, Genetics and Evolution*, 54, 74–80. <https://doi.org/10.1016/j.meegid.2017.06.010>
- Mahdy, M. A., Al-Mekhlafi, A. M., Abdul-Ghani, R., Saif-Ali, R., Al-Mekhlafi, H. M., Al-Eryani, S. M., ... & Mahmud, R. (2016). First molecular characterization of *Leishmania* species causing visceral leishmaniasis among children in Yemen. *PLoS One*, 11(3), e0151265. <https://doi.org/10.1371/journal.pone.0151265>
- Mann, K. (2021). The structure of homeomorphism and diffeomorphism groups. *Notices Amer. Math. Soc*, 68, 482-492. <https://doi.org/10.1090/noti2252>
- Mannaert, A., Downing, T., Imamura, H., & Dujardin, J. C. (2012). Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. *Trends in parasitology*, 28(9), 370-376. <https://doi.org/10.1016/j.pt.2012.06.003>
- Manzano, J. I., Perea, A., León-Guerrero, D., Campos-Salinas, J., Piacenza, L., Castanys, S., & Gamarro, F. (2017). *Leishmania* LABC1 and LABC2 transporters are involved in virulence and oxidative stress: functional linkage with autophagy. *Parasites & vectors*, 10(1), 1-12. <https://doi.org/10.1186/s13071-017-2198-1>

- Marco, J. D., Barroso, P. A., Locatelli, F. M., Cajal, S. P., Hoyos, C. L., Nevot, M. C., ... & Ruybal, P. (2015). Multilocus sequence typing approach for a broader range of species of *Leishmania* genus: describing parasite diversity in Argentina. *Infection, Genetics and Evolution*, 30, 308-317. <https://doi.org/10.1016/j.meegid.2014.12.031>
- Marlow, M. A., Boite, M. C., Ferreira, G. E. M., Steindel, M., & Cupolillo, E. (2014). Multilocus sequence analysis for *Leishmania braziliensis* outbreak investigation. *PLoS Neglected Tropical Diseases*, 8(2), e2695. <https://doi.org/10.1371/journal.pntd.0002695>
- Mauricio, I. L., Yeo, M., Baghaei, M., Doto, D., Pralong, F., Zemanova, E., ... & Miles, M. A. (2006). Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). *International journal for parasitology*, 36(7), 757-769. <https://doi.org/10.1016/j.ijpara.2006.03.006>
- Maurya, R., Singh, R. K., Kumar, B., Salotra, P., Rai, M., & Sundar, S. (2005). Evaluation of PCR for diagnosis of Indian kala-azar and assessment of cure. *Journal of clinical microbiology*, 43(7), 3038-3041. <https://doi.org/10.1128/jcm.43.7.3038-3041.2005>
- McConville, M. J., & Naderer, T. (2011). Metabolic pathways required for the intracellular survival of *Leishmania*. *Annual review of microbiology*, 65(1), 543-561. <https://doi.org/10.1146/annurev-micro-090110-102913>
- Meade, J. C., Coombs, G. H., Mottram, J. C., Steele, P. E., & Stringer, J. R. (1991). Conservation of cation-transporting ATPase genes in *Leishmania*. *Molecular and biochemical parasitology*, 45(1), 29-38. [https://doi.org/10.1016/0166-6851\(91\)90024-z](https://doi.org/10.1016/0166-6851(91)90024-z)
- Metsalu, T., & Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research*, 43(W1), W566-W570. <https://doi.org/10.1093/nar/gkv468>
- Meinecke, C. K., Schottelius, J., Oskam, L., & Fleischer, B. (1999). Congenital transmission of visceral leishmaniasis (Kala Azar) from an asymptomatic mother to her child. *Pediatrics*, 104(5), e65-e65. <https://doi.org/10.1542/peds.104.5.e65>

-
- Miles, M. A., Llewellyn, M. S., Lewis, M. D., Yeo, M., Baleela, R., Fitzpatrick, S., ... & Mauricio, I. L. (2009). The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on Leishmania: looking back and to the future. *Parasitology*, *136*(12),1509-1528.
<https://doi.org/10.1017/s0031182009990977>
- Montalvo, A. M., Fraga, J., Monzote, L., Montano, I., De Doncker, S., Dujardin, J. C., & Van der Auwera, G. (2010). Heat-shock protein 70 PCR-RFLP: a universal simple tool for *Leishmania* species discrimination in the New and Old World. *Parasitology*, *137*(8),1159-1168.
<https://doi.org/10.1017/s0031182010000089>
- Montfort, A., Martin, P. G., Levade, T., Benoist, H., & Ségui, B. (2010). FAN (factor associated with neutral sphingomyelinase activation), a moonlighting protein in TNF-R1 signaling. *Journal of leukocyte biology*, *88*(5), 897-903.
<https://doi.org/10.1189/jlb.0410188>
- Mukhopadhyay, D., Mukherjee, S., Roy, S., Dalton, J. E., Kundu, S., Sarkar, A., ... & Chatterjee, M. (2015). M2 polarization of monocytes-macrophages is a hallmark of Indian post kala-azar dermal leishmaniasis. *PLoS neglected tropical diseases*, *9*(10), e0004145. <https://doi.org/10.1371/journal.pntd.0004145>
- Murray, H. W. (2001). Clinical and experimental advances in treatment of visceral leishmaniasis. *Antimicrobial agents and chemotherapy*, *45*(8), 2185-2197.
<https://doi.org/10.1128/aac.45.8.2185-2197.2001>
- Murray, H. W., Hariprasad, J., & Coffman, R. L. (1997). Behavior of visceral *Leishmania donovani* in an experimentally induced T helper cell 2 (Th2)-associated response model. *The Journal of experimental medicine*, *185*(5), 867-874.
<https://doi.org/10.1084/jem.185.5.867>
- Murray, H. W., Oca, M. J., Granger, A. M., & Schreiber, R. D. (1989). Requirement for T cells and effect of lymphokines in successful chemotherapy for an intracellular infection. Experimental visceral leishmaniasis. *The Journal of clinical investigation*, *83*(4), 1253-1257. <https://doi.org/10.1172/jci114009>
- Nuzum, E., White Iii, F., Thakur, C., Dietze, R., Wages, J., Grogl, M., & Berman, J. (1995). Diagnosis of symptomatic visceral leishmaniasis by use of the polymerase

- chain reaction on patient blood. *Journal of infectious diseases*, 171(3), 751-754.
<https://doi.org/10.1093/infdis/171.3.751>
- Ochsenreither, S., Kuhls, K., Schaar, M., Presber, W., & Schönian, G. (2006). Multilocus microsatellite typing as a new tool for discrimination of *Leishmania infantum* MON-1 strains. *Journal of clinical microbiology*, 44(2), 495-503.
<https://doi.org/10.1128/jcm.44.2.495-503.2006>
- Osman, O. F., Oskam, L., Zijlstra, E. E., Kroon, N. C., Schoone, G. J., Khalil, E. T., ... & Kager, P. A. (1997). Evaluation of PCR for diagnosis of visceral leishmaniasis. *Journal of clinical microbiology*, 35(10), 2454-2457.
<https://doi.org/10.1128/jcm.35.10.2454-2457.1997>
- Osorio, Y., Rodriguez, L. D., Bonilla, D. L., Peniche, A. G., Henao, H., Saldarriaga, O., & Travi, B. L. (2012). Congenital transmission of experimental leishmaniasis in a hamster model. *The American journal of tropical medicine and hygiene*, 86(5), 812–820. <https://doi.org/10.4269/ajtmh.2012.11-0458>
- Pal, S., Aggarwal, G., Haldar, A., Majumdar, A., Majumdar, H. K., & Duttgupta, S. (2004). Diagnosis of symptomatic kala-azar by polymerase chain reaction using patient's blood. *Medical science monitor: international medical journal of experimental and clinical research*, 10(1), MT1-5.
https://doi.org/10.5005/jp/books/12385_4
- Pal, S., Mandal, A., & Duttgupta, S. (2001). Studies on stibionate resistant *Leishmania donovani* isolates of Indian origin.
- Palić, S., Bhairoosing, P., Beijnen, J. H., & Dorlo, T. P. (2019). Systematic review of host-mediated activity of miltefosine in leishmaniasis through immunomodulation. *Antimicrobial agents and chemotherapy*, 63(7), e02507-18.
<https://doi.org/10.1128/aac.02507-18>
- Pérez-Losada, M., Arenas, M., & Castro-Nallar, E. (2017). Multilocus Sequence Typing of Pathogens: Methods, Analyses, and Applications. *Methods, Analyses, and Applications*. In *Genetics and Evolution of Infectious Diseases: Second Edition* (pp. 383-404). Elsevier Inc.. <https://doi.org/10.1016/b978-0-12-799942-5.00016-0>
- Peters, W., & Killick-Kendrick, R. (1987). *Leishmaniasis in biology and medicine*. Academic press.

-
- Pitta, M. G., Romano, A., Cabantous, S., Henri, S., Hammad, A., Kouriba, B., ... & Dessein, A. (2009). IL-17 and IL-22 are associated with protection against human kala azar caused by *Leishmania donovani*. *The Journal of clinical investigation*, *119*(8), 2379-2387. <https://doi.org/10.1172/jci38813>
- Pradhan, S., Schwartz, R. A., Patil, A., Grabbe, S., & Goldust, M. (2022). Treatment options for leishmaniasis. *Clinical and experimental dermatology*, *47*(3), 516-521. <https://doi.org/10.1111/ced.14919>
- Ranasinghe, S., Zhang, W. W., Wickremasinghe, R., Abeygunasekera, P., Chandrasekharan, V., Athauda, S., ... & Pratlong, F. (2012). *Leishmania donovani* zymodeme MON-37 isolated from an autochthonous visceral leishmaniasis patient in Sri Lanka. *Pathogens and global health*, *106*(7), 421-424. <https://doi.org/10.1179/2047773212y.0000000054>
- Reimão, J. Q., Coser, E. M., Lee, M. R., & Coelho, A. C. (2020). Laboratory diagnosis of cutaneous and visceral leishmaniasis: current and future methods. *Microorganisms*, *8*(11), 1632. <https://doi.org/10.3390/microorganisms8111632>
- Rijal, S., Uranw, S., Chappuis, F., Picado, A., Khanal, B., Paudel, I. S., ... & Boelaert, M. (2010). Epidemiology of *Leishmania donovani* infection in high-transmission foci in Nepal. *Tropical Medicine & International Health*, *15*, 21-28. <https://doi.org/10.1111/j.1365-3156.2010.02518.x>
- Rioux, J. A., Lanotte, G., Serres, E., Pratlong, F., Bastien, P., & Perieres, J. (1990). Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Annales de parasitologie humaine et comparee*, *65*(3), 111-125. <https://doi.org/10.1051/parasite/1990653111>
- Rotureau, B., Ravel, C., Couppié, P., Pratlong, F., Nacher, M., Dedet, J. P., & Carme, B. (2006). Use of PCR-restriction fragment length polymorphism analysis to identify the main new world *Leishmania* species and analyze their taxonomic properties and polymorphism by application of the assay to clinical samples. *Journal of clinical microbiology*, *44*(2), 459-467. <https://doi.org/10.1128/JCM.44.2.459-467.2006>

-
- Rozas, J. (2009). DNA sequence polymorphism analysis using DnaSP. In *Bioinformatics for DNA sequence analysis* (pp. 337-350). Humana Press. https://doi.org/10.1007/978-1-59745-251-9_17
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, *34*(12), 3299-3302. <https://doi.org/10.1093/molbev/msx248>
- Sacks, D. L., Kenney, R. T., Neva, F. A., Kreutzer, R. D., Jaffe, C. L., Gupta, A. K., ... & Saran, R. (1995). Indian kala-azar caused by *Leishmania tropica*. *The Lancet*, *345*(8955), 959-961. [https://doi.org/10.1016/s0140-6736\(95\)90703-3](https://doi.org/10.1016/s0140-6736(95)90703-3)
- Saha, S., Mondal, S., Ravindran, R., Bhowmick, S., Modak, D., Mallick, S., ... & Ali, N. (2007). IL-10-and TGF- β -mediated susceptibility in kala-azar and post-kala-azar dermal leishmaniasis: the significance of amphotericin B in the control of *Leishmania donovani* infection in India. *The Journal of Immunology*, *179*(8), 5592-5603. <https://doi.org/10.4049/jimmunol.179.8.5592>
- Salotra, P., Sreenivas, G., Pogue, G. P., Lee, N., Nakhasi, H. L., Ramesh, V., & Negi, N. S. (2001). Development of a species-specific PCR assay for detection of *Leishmania donovani* in clinical samples from patients with kala-azar and post-kala-azar dermal leishmaniasis. *Journal of Clinical Microbiology*, *39*(3), 849-854. <https://doi.org/10.1128/jcm.39.3.849-854.2001>
- Sarraf, N. R., Mukhopadhyay, S., Banerjee, A., Das, A. K., Roy, S., Chakrabarti, S., ... & Saha, P. (2021). Genome wide comparison of *Leishmania donovani* strains from Indian visceral leishmaniasis and para-kala-azar dermal leishmaniasis patients. *Acta Tropica*, *223*, 106086. <https://doi.org/10.1016/j.actatropica.2021.106086>
- Sassi, A., Louzir, H., Ben Salah, A., Mokni, M., Ben Osman, A., & Dellagi, K. (1999). Leishmanin skin test lymphoproliferative responses and cytokine production after symptomatic or asymptomatic *Leishmania major* infection in Tunisia. *Clinical & Experimental Immunology*, *116*(1), 127-132. <https://doi.org/10.1046/j.1365-2249.1999.00844.x>

- Schönian, G., Kuhls, K., & Mauricio, I. L. (2011). Molecular approaches for a better understanding of the epidemiology and population genetics of *Leishmania*. *Parasitology*, *138*(4),405-425.
<https://doi.org/10.1017/s0031182010001538>
- Schönian, G., Mauricio, I., Gramiccia, M., Cañavate, C., Boelaert, M., & Dujardin, J. C. (2008). Leishmaniasis in the Mediterranean in the era of molecular epidemiology. *Trends in parasitology*, *24*(3), 135-142.
<https://doi.org/10.1016/j.pt.2007.12.006>
- Schönian, G., Nasereddin, A., Dinse, N., Schweynoch, C., Schallig, H. D., Presber, W., & Jaffe, C. L. (2003). PCR diagnosis and characterization of *Leishmania* in local and imported clinical samples. *Diagnostic microbiology and infectious disease*, *47*(1), 349-358. [https://doi.org/10.1016/s0732-8893\(03\)00093-2](https://doi.org/10.1016/s0732-8893(03)00093-2)
- Schwartz, E., Hatz, C., & Blum, J. (2006). New world cutaneous leishmaniasis in travellers. *The Lancet infectious diseases*, *6*(6), 342-349.
[https://doi.org/10.1016/s1473-3099\(06\)70492-3](https://doi.org/10.1016/s1473-3099(06)70492-3)
- Scorza, B. M., Carvalho, E. M., & Wilson, M. E. (2017). Cutaneous manifestations of human and murine leishmaniasis. *International journal of molecular sciences*, *18*(6), 1296. <https://doi.org/10.3390/ijms18061296>
- Siddig, M., Ghalib, H., Shillington, D. C., & Petersen, E. A. (1988). Visceral leishmaniasis in the Sudan: comparative parasitological methods of diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *82*(1), 66-68.
[https://doi.org/10.1016/0035-9203\(88\)90265-9](https://doi.org/10.1016/0035-9203(88)90265-9)
- Singh, N., & Thomas, R. (1989). Matching grants versus block grants with imperfect information. *National Tax Journal*, *42*(2), 191-203.
<https://doi.org/10.1086/ntj41788788>
- Singh, N., Chikara, S., & Sundar, S. (2013). SOLiD™ sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm *Leptomonas* co-infection and raise some key questions. *Plos one*, *8*(2), e55738.
<https://doi.org/10.1371/journal.pone.0055738>
- Singh, N., Curran, M. D., Rastogil, A. K., Middleton, D., & Sundar, S. (1999). Diagnostic PCR with *Leishmania donovani* specificity using sequences from the variable

- region of kinetoplast minicircle DNA. *Tropical medicine & international health*, 4(6), 448-453. <https://doi.org/10.1046/j.1365-3156.1999.00416.x>
- Srivastava, P., Dayama, A., Mehrotra, S., & Sundar, S. (2011). Diagnosis of visceral leishmaniasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(1), 1-6. <https://doi.org/10.1016/j.trstmh.2010.09.006>
- Srivastava, P., Prajapati, V. K., Vanaerschot, M., Van der Auwera, G., Dujardin, J. C., & Sundar, S. (2010). Detection of *Leptomonas* sp. parasites in clinical isolates of Kala-azar patients from India. *Infection, Genetics and Evolution*, 10(7), 1145-1150. <https://doi.org/10.1016/j.meegid.2010.07.009>
- Sundar, S. (2001). Drug resistance in Indian visceral leishmaniasis. *Tropical Medicine & International Health*, 6(11), 849-854. <https://doi.org/10.1046/j.1365-3156.2001.00778.x>
- Sundar, S., & Singh, A. (2018). Chemotherapeutics of visceral leishmaniasis: present and future developments. *Parasitology*, 145(4), 481-489. <https://doi.org/10.1017/s0031182017002116>
- Sundar, S., Kumar, K., Chakravarty, J., Agrawal, D., Agrawal, S., Chhabra, A., & Singh, V. (2006). Cure of antimony-unresponsive Indian post-kala-azar dermal leishmaniasis with oral miltefosine. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(7), 698-700. <https://doi.org/10.1016/j.trstmh.2005.09.015>
- Sundar, S., Singh, A., Rai, M., Prajapati, V. K., Singh, A. K., Ostry, B., ... & Chakravarty, J. (2012). Efficacy of miltefosine in the treatment of visceral leishmaniasis in India after a decade of use. *Clinical infectious diseases*, 55(4), 543-550. <https://doi.org/10.1093/cid/cis474>
- Sundar, S., Singh, R. K., Maurya, R., Kumar, B., Chhabra, A., Singh, V., & Rai, M. (2006). Serological diagnosis of Indian visceral leishmaniasis: direct agglutination test versus rK39 strip test. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(6), 533-537. <https://doi.org/10.1016/j.trstmh.2005.08.018>
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, 10(3), 512-526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>

- Thakur, C. P. (2007). A new strategy for elimination of kala-azar from rural Bihar. *Indian Journal of Medical Research*, 126(5), 447-452.
- Tomasini, N., Lauthier, J. J., Llewellyn, M. S., & Diosque, P. (2013). MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. *Infection, Genetics and Evolution*, 20, 188-196. <https://doi.org/10.1016/j.meegid.2013.08.029>
- Tomiotto-Pellissier, F., Bortoleti, B. T. D. S., Assolini, J. P., Gonçalves, M. D., Carloto, A. C. M., Miranda-Sapla, M. M., ... & Pavanelli, W. R. (2018). Macrophage polarization in leishmaniasis: broadening horizons. *Frontiers in immunology*, 2529. <https://doi.org/10.3389/fimmu.2018.02529>
- Topno, R. K., Das, V. N., Ranjan, A., Pandey, K., Singh, D., Kumar, N., ... & Das, P. (2010). Asymptomatic infection with visceral leishmaniasis in a disease-endemic area in Bihar, India. *The American journal of tropical medicine and hygiene*, 83(3), 502. <https://doi.org/10.4269/ajtmh.2010.09-0345>
- 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. *Molecular microbiology*, 29(2), 653-660. <https://doi.org/10.1046/j.1365-2958.1998.00968.x>
- Ubeda, J. M., Légaré, D., Raymond, F., Ouameur, A. A., Boisvert, S., Rigault, P., ... & Ouellette, M. (2008). Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome biology*, 9(7), 1-16. <https://doi.org/10.1186/gb-2008-9-7-r115>
- Valente, M., Castillo-Acosta, V. M., Vidal, A. E., & González-Pacanowska, D. (2019). Overview of the role of kinetoplastid surface carbohydrates in infection and host cell invasion: Prospects for therapeutic intervention. *Parasitology*, 146(14), 1743-1754. <https://doi.org/10.1017/s0031182019001355>
- Wanderley, J. L., Moreira, M. E., Benjamin, A., Bonomo, A. C., & Barcinski, M. A. (2006). Mimicry of apoptotic cells by exposing phosphatidylserine participates in the establishment of amastigotes of *Leishmania (L) amazonensis* in mammalian hosts. *The Journal of Immunology*, 176(3), 1834-1839. <https://doi.org/10.4049/jimmunol.176.3.1834>
- Welsh, J., Pretzman, C., Postic, D., Saint Girons, I., Baranton, G., & McClelland, M. (1992). Genomic fingerprinting by arbitrarily primed polymerase chain reaction

- resolves *Borrelia burgdorferi* into three distinct phyletic groups. *International Journal of Systematic and Evolutionary Microbiology*, 42(3), 370-377. <https://doi.org/10.1099/00207713-42-3-370>
- Welsh, J., & McClelland, M. (1991). Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. *Nucleic acids research*, 19(19), 5275–5279. <https://doi.org/10.1093/nar/19.19.5275>
- Welsh, J., Pretzman, C., Postic, D., Saint Girons, I., Baranton, G., & McClelland, M. (1992). Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. *International Journal of Systematic and Evolutionary Microbiology*, 42(3), 370-377. <https://doi.org/10.1099/00207713-42-3-370>
- WHO . Control of neglected tropical diseases (2015).
- WHO . Control of neglected tropical diseases (2017).
- World Health Organization. (1995). *Report on the Consultative Meeting on Leishmania* (No. WHO/LEISH/95.35. Unpublished). World Health Organization.
- Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research*, 18(22), 6531-6535. <https://doi.org/10.1093/nar/18.22.6531>
- Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research*, 43(W1), W174-W181. <https://doi.org/10.1093/nar/gkv342>
- 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 genetics*, 14(1), e1007133. <https://doi.org/10.1371/journal.pgen.1007133>
- Zanluqui, N. G., Wowk, P. F., & Pinge-Filho, P. (2015). Macrophage polarization in Chagas disease. *J Clin Cell Immunology*, 6, 1–6. <https://doi.org/10.4172/2155-9899.1000317>
- Zemanová, E., Jirků, M., Mauricio, I. L., Horák, A., Miles, M. A., & Lukeš, J. (2007). The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence

- typing. *International journal for parasitology*, 37(2), 149-160. <https://doi.org/10.1016/j.ijpara.2006.08.008>
- Zhang, C. Y., Lu, X. J., Du, X. Q., Jian, J., Shu, L., & Ma, Y. (2013). Phylogenetic and evolutionary analysis of Chinese Leishmania isolates based on multilocus sequence typing. *Plos one*, 8(4), e63124. <https://doi.org/10.1371/journal.pone.0063124>
- Zhang, C., Freddolino, P. L., & Zhang, Y. (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic acids research*, 45(W1), W291-W299. <https://doi.org/10.1093/nar/gkx366>
- Zhang, W. W., Ramasamy, G., McCall, L. I., Haydock, A., Ranasinghe, S., Abeygunasekara, P., ... & Matlashewski, G. (2014). Genetic analysis of Leishmania donovani tropism using a naturally attenuated cutaneous strain. *PLoS pathogens*, 10(7), e1004244. <https://doi.org/10.1371/journal.ppat.1004244>
- Zijlstra, E. E. (2016). The immunology of post-kala-azar dermal leishmaniasis (PKDL). *Parasites & vectors*, 9(1), 1-9. <https://doi.org/10.1186/s13071-016-1721-0>
- Zijlstra, E. E. (2019). Biomarkers in post-kala-azar dermal leishmaniasis. *Frontiers in Cellular and Infection Microbiology*, 228. <https://doi.org/10.3389/fcimb.2019.00228>
- Zijlstra, E. E., Ali, M. S., El-Hassan, A. M., El-Toum, I. A., Satti, M., Ghalib, H. W., & Kager, P. A. (1992). Kala-azar: a comparative study of parasitological methods and the direct agglutination test in diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86(5), 505-507. [https://doi.org/10.1016/0035-9203\(92\)90086-r](https://doi.org/10.1016/0035-9203(92)90086-r)
- 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 neglected tropical diseases*, 11(11), e0005877. <https://doi.org/10.1371/journal.pntd.0005877>
- Zijlstra, E. E., Musa, A. M., Khalil, E. A. G., El Hassan, I. M., & El-Hassan, A. M. (2003). Post-kala-azar dermal leishmaniasis. *The Lancet infectious diseases*, 3(2), 87-98. [https://doi.org/10.1016/s1473-3099\(03\)00517-6](https://doi.org/10.1016/s1473-3099(03)00517-6)



PUBLICATION AND PROCEEDINGS

LIST OF PUBLICATIONS

1. **Nibedeeta Rani Sarraf**, Saikat Mukhopadhyay, Anindyajit Banerjee, Anjan Kumar Das, Syamal Roy, Saikat Chakrabarti, Madhumita Manna, Partha Saha, *Genome wide comparison of Leishmania donovani strains from Indian visceral leishmaniasis and para-kala-azar dermal leishmaniasis patients*, Acta Tropica, Volume 223,2021,106086, ISSN 0001706X, <https://doi.org/10.1016/j.actatropica.2021.106086>.
2. Khanra S, **Nibedeeta Rani Sarraf**, Das AK, Roy S, Manna M. *Miltefosine Resistant Field Isolate from Indian Kala-Azar Patient Shows Similar Phenotype in Experimental Infection*. Sci Rep. 2017 Sep 4;7(1):10330. doi: 10.1038/s41598-017-09720-1. PMID: 28871097; PMCID: PMC5583325.
3. Khanra S, **Nibedeeta Rani Sarraf**, Das S, Das AK, Roy S, Manna M. *Genetic markers for antimony resistant clinical isolates differentiation from Indian Kala-azar*. Acta Trop. 2016 Dec; 164:177-184. doi: 10.1016/j.actatropica.2016.09.012. Epub 2016 Sep 11. PMID: 27629023.
4. *Leishmania genomics: a brief account*. Khanra S, **Nibedeeta Rani Sarraf**, Lahiry S, Roy S, Manna M. Nucleus. (Springer) 2017; 60(2): 227-235. doi10.1007/s13237-017-0210-y.

Submitted Manuscript:

Antimony Resistance Mechanism in Genetically Diverse Clinical Isolates of Indian Kala-azar Patients. Supriya Khanra, Shantanabha Das, **Nibedeeta Rani Sarraf**, Sanchita Datta, Anjan Kumar Das, Madhumita Manna, Syamal Roy. Frontiers in Cellular and Infection Microbiology, section Parasite and Host; Manuscript ID: 1021464.

Manuscript Under Preparation:

Multilocus Sequence Typing of The Clinical Isolates of Indian Kala-azar: Genetic diversity and Phylogeny analysis. **Nibedeeta Rani Sarraf**, Anjan Kumar Das, Parimal Karmakar, Syamal Roy, Madhumita Manna.

ABSTRACT IN THE PROCEEDING OF INTERNATIONAL
&
NATIONAL LEVEL CONFERENCES/SYMPOSIA

International Level:

- **Poster Presentation** in Indo-Brazil Symposium on Biochemistry of Kinetoplastid Parasites held in Indian Institute of Chemical Biology, Kolkata on 19-20th September **2016**.

National Level:

- **Oral Presentation** in the 26th West Bengal State Science and Technology Congress held at Science City, Kolkata, on 28th February-1st March **2019**.
- **Oral Presentation** in the 3rd Regional Science and Technology Congress (Southern Region) held at Bidhannagar College, Kolkata, on 18-19th December **2018**.
- **Poster Presentation** in the Conference on “Advances of Life Sciences & Radiation Biology” held at Burdwan University on 17-18th February **2017**.

International Workshop

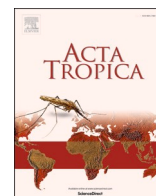
Attended “**Bioinformatics and Computational Biology**” organized by AIIST, Amplicon Biosciences & The Biome from 23rd -25th September **2016**, held at The Biome Research Facility, Salt Lake, Kolkata.

PRIZES AND AWARDS

Got **Outstanding Paper Awards** in the Oral Presentation entitled “*Application of Whole Genome Sequencing to Explore Mechanistic Origins of Sodium Stibogluconate Resistivity in Clinical Isolates from Indian Kala-Azar Patients*” at the 26th WB State Science and Technology Congress held at Science City, Kolkata, on 28th February-1st March **2019**.

National and International Level Conferences / Seminars Attended

- International Conference on “**Exploring Genomes: The New Frontier**” held in Indian Institute of Chemical Biology, Kolkata on 22nd -24th December **2015**.
- International Conference on “**100 years of Antimonials**” held in Indian Institute of Chemical Biology, Kolkata on 23-25th December **2014**.
- National Symposium on “**Therapeutic Targets for Neurodegenerative Disorders**” held at Bidhanagar College on 21st-22nd January **2020**.
- National one day seminar on “**Interdisciplinary admittance of chemical science**” held at Bidhanagar College on 6th September **2017**.
- National seminar on biodiversity on “**Exploration, Exploitation, Conservation and Management -Vision & mission**” held at Barasat Government College, Barasat on 19-20th November **2016**.



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 copy numbers 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

SCIENTIFIC REPORTS

OPEN

Miltefosine Resistant Field Isolate From Indian Kala-Azar Patient Shows Similar Phenotype in Experimental Infection

Supriya Khanra^{1,4}, Nibedeeta R. Sarraf¹, Anjan K. Das², Syamal Roy^{3,5} & Madhumita Manna^{1,6}

Emergence of resistance to drugs used to treat the Indian Kala-azar patients makes control strategy shattered. In this bleak situation, Miltefosine (MIL) was introduced to treat mainly antimonial unresponsive cases. Within years, resistance to MIL has been reported. While checking the MIL sensitivity of the recent KA clinical isolates ($n = 26$), we came across one isolate which showed four times more EC_{50} for MIL than that of MIL-Sensitive (MIL-S) isolates and considered as putative MIL-Resistant (MIL-R). The expressions of LdMT and LdRos3 genes of this isolate were found down regulated. Th1/Th2 cytokines, ROS and NO, FACS dot plots and mitochondrial trans membrane potential measurement were performed. *In vivo* hamster model with this MIL-R isolate showed much lesser reduction in liver weight (17.5%) compared to average reduction in liver weight (40.2%) of the animals infected with MIL-S isolates. The splenic and hepatic stamps smears of MIL-R infected hamsters revealed the retention of parasite load of about 51.45%. The splenocytes of these animals failed to proliferate anti leishmanial T-cells and lack of cell mediated immunity hampered recovery. Thus, these phenotypic expressions of experimental model may be considered similar to that of the MIL unresponsive patients. This is first such kind of report.

The disease, Visceral Leishmaniasis (VL) or Kala-azar (KA) is endemic in the Indian subcontinent and broadening its base on the Gangetic plains of Bangladesh, India and Nepal. KA is the most fatal disease if left untreated¹⁻⁵. Occurrence of immune suppression in the host (eg, human immunodeficiency virus [HIV] co infection) and emergence of resistance to first line antimonial drugs^{6,7} are two serious health problems associated with the disease. Earlier reports suggested that resistance of *Leishmania* parasites to antimonials is related to Sodium Stibo Gluconate (SSG) treatment failure in the Indian state, Bihar^{8,9}. Second line treatment with Amphotericin B (AmB) was highly efficacious¹⁰ but unresponsiveness towards AmB have started been reported¹¹. Other substitute drugs such as Pentamidine, Paromomycin remain largely inadequate due to high cost, high toxicity or side effects^{2,12}. The existing oral drug Miltefosine (MIL) (hexadecylphosphocholine), a lysophospholipid analog, was mainly developed as an anticancer drug but it is effective *in vitro* against a variety of species of *Leishmania*^{13,14} and other protozoan parasites including *Trypanosoma cruzi*, *T. brucei*¹⁵, *Entamoeba histolytica*¹⁶ and *Acanthamoeba sp.*¹⁷. MIL, had been approved in India, in 2002, for the treatment of Indian VL patients¹⁸, including cases unresponsive to antimonial and has achieved more than 97% cure rate¹⁹. Within few years, unresponsiveness to MIL in VL patients^{20,21} and failure of MIL in the treatment of KA in Nepal have been reported²². The above observations revealed the crude reality that any drug, how efficacious it may be, after some years of its use, resistance would appear. Thus, the parasite resistance to the commonly used drugs must be monitored carefully to combat

¹Department of Zoology, Barasat Govt. College, 10, K.N.C Road, Kolkata, 700124, India. ²Department of Pathology, Calcutta National Medical College, 32, Gorachand Road, Kolkata, 700014, India. ³Department of Infectious Diseases & Immunology, Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata, 700032, India. ⁴Present address: Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata, 700064, India. ⁵Present address: Cooch Behar Panchanan Barma University, Vivekananda Road, Cooch Behar, West Bengal, 736101, India. ⁶Present address: Bidhannagar College, EB 2, Salt Lake, Sector I, Kolkata, 700064, India. Correspondence and requests for materials should be addressed to S.R. (email: drsyamalroy@yahoo.com) or M.M. (email: madhumita.manna09@gmail.com)



Genetic markers for antimony resistant clinical isolates differentiation from Indian Kala-azar



Supriya Khanra^a, Nibedeeta Rani Sarraf^a, Shantanabha Das^{b,1}, Anjan Kumar Das^c,
Symal Roy^{b,2}, Madhumita Manna^{a,*}

^a Department of Zoology, Barasat Govt. College, 10, K.N.C Road, Kolkata 700124, India

^b Department of Infectious Diseases & Immunology, Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata 700032, India

^c Department of Medicine, Calcutta National Medical College, 32, Gorachand Road, Kolkata 700014, India

ARTICLE INFO

Article history:

Received 15 May 2016

Received in revised form 9 August 2016

Accepted 9 September 2016

Available online 11 September 2016

Keywords:

Indian kala-azar

Leishmania donovani

L. tropica

Drug resistance

Single stranded conformation

polymorphism

Genetic markers

ABSTRACT

Visceral Leishmaniasis or Kala-azar is caused by the protozoan parasites belonging to the Genus *Leishmania*. Once thought eradicated from the Indian subcontinent, the disease came back with drug resistance to almost all prevalent drugs. Molecular epidemiological studies revealed the polymorphic nature of the population of the main player of the disease, *Leishmania donovani* and involvement of other species (*L. tropica*) and other genus (*Leptomonas*) with the disease. This makes control measures almost futile. It also strongly demands the characterization of each and every isolate mandatory which is not done. In this background, the present study has been carried out to assess the genetic attributes of each clinical isolates (n = 26) of KA and PKDL patients from India and Bangladesh. All the isolates were characterized through Restriction Fragment Length Polymorphism (RFLP) analysis to ascertain their species identity. 46.2% of the isolates were found to be Sodium Stibogluconate (SSG) resistant by amastigote-macrophage model. When the clinical isolates were subjected to Single Stranded Conformation Polymorphism (SSCP) of Internal Transcribed Spacer 1 (ITS1), Internal Transcribed Spacer 2 (ITS2) and some anonymous markers, the drug resistant *Leishmania* isolates of SSG can be distinguished from the sensitive isolates distinctly. This study showed for the first time, the genetic markers for SSG drug resistance of Indian Kala-azar clinical isolates.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Leishmaniasis are a spectrum of protozoan diseases spanning a variety of disease forms that differ in their epidemiology, pathogenesis and clinical manifestations. The visceral form of the disease known as Visceral Leishmaniasis (VL) or Kala-azar (KA) occurs in India, Bangladesh, Brazil, Iran, Nepal and Sudan (Sundar, 2001a) and is most fatal, if left untreated. A proportion of apparently cured KA patients (6–10%) may expand to Post Kala-azar Dermal Leishmaniasis (PKDL) (Thakur and Kumar, 1992). Along with immune suppression and co-infection with HIV, another serious health problem related to it is the emergence and progression of resistance to the commonly used antimonial drugs (Desjeux and Alvar,

2003; Dujardin, 2006). Approximately eight decades ago, the pentavalent antimonial drug, urea stibamine was brought to the Indian subcontinent (Brahmachari, 1922) for the treatment of KA but the effectiveness of this drug had reduced with time, even with increasing the dose and period of treatment (Lira et al., 1999; Croft et al., 2006). It was reported that *Leishmania* parasites resistant to antimonials is related to Sodium Stibogluconate (SSG) treatment breakdown in the Indian province of Bihar (Lira et al., 1999; Dube et al., 2005) and accounts for 60%–70% treatment failure in this region (Sundar et al., 2000). On the other hand, in the neighboring country Nepal, VL patients from endemic regions, infected with SSG-resistant parasites were found to have only a 25% unsuccessful SSG treatment (Rijal et al., 2007). Interestingly, recent report suggested the development of *Leishmania* antimonial resistance in the population of Bihar occurred in some cases was through the distribution of arsenic contamination of ground water (Perry et al., 2015). Antimony and arsenic are the two important elements as they have a long therapeutic history and in the periodic table, they are related to each other through distribution of numerous similar chemical properties (Yan et al., 2005). The naturally occurring triva-

* Corresponding author.

E-mail address: madhumita.manna09@gmail.com (M. Manna).

¹ Present address: Dept. of Zoology, Govt. General Degree College, Kharagpur II, Paschim Medinipur, West Bengal 721149, India.

² Present address: Cooch Behar Panchanan Barma University, Vivekananda Road, Cooch Behar, West Bengal 736101, India.

Leishmania genomics: a brief account

Supriya Khanra^{1,3} · Nibedeeta Rani Sarraf¹ · Sangita Lahiry¹ · Syamal Roy^{2,4} ·
Madhumita Manna^{1,5}

Received: 1 August 2016 / Accepted: 5 June 2017
© Archana Sharma Foundation of Calcutta 2017

Abstract Leishmaniasis, one of the neglected tropical diseases is serious health concern globally. The disease is caused by protozoan parasites belonging to genus *Leishmania*. The main forms of disease are Cutaneous Leishmaniasis (CL), Mucocutaneous Leishmaniasis and Visceral Leishmaniasis (VL). VL or Kala-azar is the most severe form and 90% of global VL cases occur in India, Bangladesh, Nepal, Sudan, Ethiopia and Brazil, while most cases (70–75%) of CL occur in Afghanistan, Brazil, Iran, Ethiopia, Costa Rica and Peru etc. They are spread by the bites of female sand flies of the genus *Phlebotomus* in the Old World and of the genus *Lutzomyia* in the New World. It is essential to determine whether genetic variability of

the parasites is associated with the different clinical manifestations and drug resistance of *Leishmania* sp. Various molecular biological methods have been standardized to study the genomes of the parasites in order to understand the parasites better. Most updated high-throughput approaches are whole genome sequencing, comparative genomics, transcriptomics and proteomics. The present review gives an overview of the advancement in the field of the *Leishmania* genome analysis which would help workers in the field to understand the problem of emergence of drug resistance, current epidemiological status, host parasite interaction and designing the drugs.

Keywords Leishmaniasis · Polymerase chain reaction · Amplified fragment length polymorphism · Whole genome sequencing · Transcriptomics

This article is based on the presentation made during the 17th All India Congress of Cytology and Genetics and International Symposium on “Exploring Genomes: The New Frontier” held at CSIR-Indian Institute of Chemical biology, Kolkata in collaboration with Archana Sharma Foundation of Calcutta during December 22–24, 2015.

✉ Madhumita Manna
madhumita.manna09@gmail.com

¹ Department of Zoology, Barasat Govt. College, 10, K.N.C Road, Kolkata 700124, India

² Department of Infectious Diseases and Immunology, Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata 700032, India

³ Present Address: Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, Sector - 1, Block - AF, Bidhannagar, Kolkata 700064, India

⁴ Present Address: Cooch Behar Panchanan Barma University, Vivekananda Road, Cooch Behar, West Bengal 736101, India

⁵ Present Address: Bidhannagar College, EB 2, Sector 1, Salt Lake, Kolkata 700064, India

Introduction

Leishmania is a genus of trypanosomatids that is responsible for the disease leishmaniasis. They are spread by sandflies of the genus *Phlebotomus* in the Old World and of the genus *Lutzomyia* in the New World [2]. Their primary hosts are vertebrates: *Leishmania* sp. commonly infects hyraxes, canids, rodents and humans. Leishmaniasis is a diversity of diseases and the main forms of disease are Cutaneous Leishmaniasis (CL), Mucocutaneous Leishmaniasis (MCL) and Visceral Leishmaniasis (VL). VL or Kala-azar (KA) is the most severe form of illness characterized by hepatosplenomegaly, anemia, edema, abdominal swellings and liver cirrhosis and it is fatal if left untreated. On the other hand, CL which is the most common occurring form results in skin lesions which could occur in all parts of the body. MCL could be caused by the