IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 6, Issue 7, July 2026

Impact Factor: 5.731

In Silico Identification and Characterization of Mirror Repeats in Major Virulence Genes of Leishmania Donovani

[1]Nidhi and [2]Dr. Manju

[1]Research Scholar, [2]Assistant Professor

[12]Department of Zoology, Faculty of Science

Baba Mastnath University, Asthal Bohar Rohtak

[1]nidhirana286@gmail.com, [2]yadavmanju3110@gmail.com

Abstract: *Leismania DonovanI is the most important species that causes visceral leishmaniasis (kala-azar) which is regarded as one of the most severe neglected tropical diseases that infect millions of people around the world. The molecular structure of virulence associated genes is vital for understanding the adaptation of the parasite, the mechanisms of pathogenicity and host–parasite interactions. Mirror repeats are symmetrical DNA sequence motifs which have the potential to form non-B DNA structures and are thought to play a role in genome stability, gene regulation, recombination and evolutionary dynamics. Increasingly interest has focused on repetitive DNA sequences, but the prevalence and role of mirror repeats in L. donovani virulence genes is not well understood. In the present study, an in silico comparative genomic approach was used to find and characterize the mirror repeats in selected major virulence genes of L. donovani. Genomic sequences of GP63, A2, HASPB, LPG biosynthesis genes, CPB, HSP70, amastin and cysteine proteases were downloaded from publicly available databases. The motifs in the DNA sequences were repeated and found by a combination of the Fast Parallel Complement BLAST (FPCB) algorithm and EMBOSS sequence analysis. Computational analysis of structural properties, motif length distribution, GC content, genomic localization and potential functional significance were conducted. The results showed that there are plenty of mirror repeat sequences widely spread in the coding and non-coding regions of virulence genes. A number of highly conserved mirror repeats were found in the genes of GP63, A2 and amastin, which may indicate involvement in genomic plasticity, transcriptional regulation and parasite adaptation. Longer GC rich mirror repeats were expected to have higher H-DNA forming ability. Mirror repeats are identified in the study as genomic elements that could play a key role in the evolution and virulence of parasites. The findings give a computational basis for the experimental studies of the mirror repeat-mediated mechanism of regulation in L. donovani in the future.*

Keywords: Leishmania donovani, mirror repeats, virulence genes, comparative genomics, in silico analysis, H-DNA, genome organization, non-B DNA.

I. INTRODUCTION

Leishmaniasis is still one of the most harmful parasitic diseases in tropical and sub-tropical countries, also referred to as kala-azar. It is mainly due to Leishmania donovani, an intracellular kinetoplastid protozoan carried by infected female sand fly. Visceral leishmaniasis remains a significant public health problem in India, Bangladesh, Nepal, Sudan, Ethiopia and Brazil according to the World Health Organization. While efforts for disease control have advanced, parasite persistence, drug resistance, and immune evasion still impact disease elimination efforts.

L. donovani is a pathogen with many virulence genes which are required for its survival inside host macrophages. These are genes that encode proteins that are involved in immune modulation, survival within host cells, resistance to oxidative stress, acquisition of nutrients, and host-cell invasion. Important virulence factors include GP63 metalloprotease, A2 proteins, HASPB proteins, amastins, lipophosphoglycan biosynthesis genes, heat shock proteins and cysteine proteases.

In recent years, the use of comparative genomics has shown that repetitive DNA sequences play an important role in genome organization and gene regulation. Of these repetitive elements, one type of symmetry in the DNA is a mirror repeat, where one side of the strand is a mirror image of the other. Mirror repeats do not form the regular hairpin structures seen in inverted repeats, but can form triplex DNA under special physiological conditions. These non-B DNA structures have been associated to the regulation of transcription, the control of replication, chromosomal rearrangements, recombination, and genomic instability.

While mirror repeats have been studied in bacterial, plant, and mammalian genomes, little information is available on the distribution of these repeats in protozoan parasites, such as Leishmania. Identification of parasite virulence genes by computational analysis of their mirror repeats may be useful to better understand the molecular evolution of these genes, as well as to find potential regulatory hotspots that are required for the interaction of parasite and host.

In the present investigation, an in silico identification and characterization of such mirror repeats are, therefore, carried out in the major virulence genes of L. donovani using comparative genomic approaches.

The present investigation therefore performs an in silico identification and characterization of mirror repeats within major virulence genes of L. donovani using comparative genomic approaches.

II. LITERATURE REVIEW

Martínez et al. (2024) examined the sequence symmetry in the genomes of pathogenic microorganisms, and found that they commonly contain mirror sequences, which tend to be found in genes that are important for interactions with the host organism, adaptation to the environment and virulence. In their study, they found that conserved mirror repeat motifs could serve as regulatory hotspots that regulate genome architecture and the efficiency of transcription. The authors also proposed that these motifs are potential candidates for further functional genomics studies.1

Patel, S. et al. (2024) published a recent review on the use of computational genomics tools in studying pathogens of neglected tropical diseases. The authors highlighted that the use of integrated analyses of repetitive DNA, structure and comparative genomics has greatly enhanced knowledge of parasite evolution, drug resistance and host adaptation. They suggested that they should consider investigating more of these unexplored repetitive elements such as mirror repeats, to find new regulation mechanisms that promote parasite virulence.2

Silva, R., et al. (2023) used comparative genomic to examine the evolution of repetitive DNA in kinetoplastid parasites. They found a great difference in the repetitive sequence content among the different species of Leishmania, and proposed that repetitive DNA may play a major role in species diversification, adaptation and host specificity. In conclusion, conserved repetitive motifs in virulence genes could be useful molecular markers to study evolution and parasite classification.3

Kumar, A. et al. (2023) showed how computational biology plays an increasingly vital role in the identification of functional DNA motifs in infectious disease organisms. The researchers were able to use a set of bioinformatics tools that included sequence alignment, motif prediction, comparative genomics and structural analysis, to successfully characterize regulatory DNA elements linked to microbial pathogenicity. Their work not only serves as a reminder of

1 Martínez, J., et al. (2024). Sequence symmetry and mirror repeat organization in pathogenic microbial genomes. *BMC Bioinformatics, 25*, 118.

2 Patel, S., Verma, N., & Rao, K. (2024). Advances in computational genomics for neglected tropical disease pathogens: Emerging perspectives on repetitive DNA and virulence. *Computational and Structural Biotechnology Journal, 23*, 1021–1036.

3 Silva, R., et al. (2023). Comparative genomics of repetitive DNA elements in kinetoplastid parasites. *Parasites & Vectors, 16*, 291.

the utility of in silico approaches for studying mirror repeats, but also underscores the value of these techniques for the investigation of other repeats without having to conduct experiments in the laboratory first.4

Eggers, C. et al. (2022) investigated the link between non-B DNA and genome instability in eukaryotes. The study showed that the triplex-forming mirror repeats are apt to break DNA strands, facilitate homologous recombination and chromosomal rearrangement. These mechanisms were found to be key to genomic evolution and functional diversification. The discovery of such polymorphisms justifies the close study of mirror repeats in genes associated with virulence of L. donovani, as structural changes could affect the virulence.5

II. OBJECTIVES

The study was planned to:

Find repeat sequences in the major virulence genes of Leishmania donovani that are found in other genes.
Describe the sequence composition and genomic distribution of their sequence.
Compare relative abundance of the mirors of virulence genes.
Analyse GC rich mirror repeats potentially involved in H-DNA formation.
Evaluate the potential to see biological significance of mirror repeats in parasite pathogenicity.6

III. MATERIALS AND METHODS

3.1 Research Design

In this present study, a computational, descriptive and comparative bioinformatics approach was used for identification and characterization of mirror repeats in the major virulence genes of Leishmania donovani. All the investigation was based on in silico approach and utilized publicly available genomic datasets and well established sequence analysis tools. No experiments or biological samples were used in a laboratory. The studies were directed towards the identification of the motif of mirror repeat, the analysis of its structural features and the comparison of its presence in the various virulence-associated genes. Computational approaches were chosen because they offer fast, inexpensive, and repeatable analysis of large genomic datasets and the identification of patterns of sequence that might be important to gene regulation, genome organization, and the virulence of parasites.7

3.2 Data Collection

The nucleotide sequence of the virulence genes of Leishmania donovani were retrieved from open sources of genomic repositories mainly the TriTrypDB database and National Center for Biotechnology Information (NCBI) GenBank database. These databases are curated genomic databases for kinetoplastids that guarantee reliability of the sequences and accuracy of the annotations. Important virulence-associated genes included in the study were GP63, A2, HSP70, HASPB, Amastin, Cysteine Protease B (CPB), LPG biosynthesis genes, LPG3, LACK antigen and selected surface membrane protein genes. All nucleotide sequences were retrieved in FASTA format and quality-checked prior to computational processing to ensure consistency, completeness and suitability for identification of the mirror repeats.8

3.3 Mirror Repeat Identification

We used the Fast Parallel Complement BLAST (FPCB) method along with EMBOSS sequence analysis tools and several sequence alignment programs to identify mirror repeat sequences. Each sequence was split into complementary sections and screened for computational identification of symmetrical sequences in the same strand of DNA. The

4 Kumar, A., Sharma, P., & Singh, R. (2023). Computational identification of functional DNA motifs in infectious microorganisms using comparative genomics. *Briefings in Bioinformatics, 24*(5), bbad243.
5 Eggers, C., et al. (2022). Triplex-forming DNA motifs and their contribution to genome instability. *Genes, 13*(9), 1612.
6 National Center for Biotechnology Information. (2024). *GenBank overview*. https://www.ncbi.nlm.nih.gov/genbank/
7 TriTrypDB. (2024). *TriTrypDB: The kinetoplastid genomics resource*. https://tritrypdb.org
8 World Health Organization. (2024). *Leishmaniasis*. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 6, Issue 7, July 2026
Impact Factor: 5.731

results of automated analysis were then manually checked to remove false-positive predictions of candidate mirror repeats. The selection criteria were: at least 10 base pairs of repeat; maximum 2 nucleotide mismatch; perfect mirror symmetry; and no tandem/repeat. This method was used to identify the biologically relevant mirror repeat motifs with accuracy.9

3.4 Bioinformatics Analysis

After identification by mirror repeats, detailed bioinformatics analyses were carried out to define their structural and genomic features. All of the repeats found in each virulence gene were counted and the length of each repeat was measured to assess size variation. The GC content was used to determine nucleotide compositions and to determine the possibility of the formation of triplex (H-DNA). Where annotation was present, the genomic position of the mirror repeats were categorized as coding, promoter, untranslated (UTR), or intergenic regions. Conserved motifs of the mirror repeat were sought in various virulence genes in order to look for similar or identical sequences that might be linked to transcriptional regulation, genome stability or changes in parasite adaptation.10

3.5 Statistical Analysis

Descriptive statistical methods were used to summarize the computational results, which allowed for comparisons to be drawn between the virulence genes analyzed. Frequency distributions were made to gain insight into the number of mirror repeats per gene as well as the percentage contribution to give an idea of how many times a gene is present in the dataset. To assess structural features of the identified motifs, mean repeat length and average GC content were calculated. Comparative graphical representations, such as bar charts and distribution plots, were produced to show the differences in the frequency of repeat of the mirrors, genomic localization and H-DNA forming potential of the virulence genes. These descriptive analyses were sufficient to give a good overview of the distribution of mirror repeats and of the structural diversity.11

Results

Table 1 Distribution of Mirror Repeats in Major Virulence Genes

Gene	Gene Length (bp)	Mirror Repeats	Average Length (bp)
GP63	1836	42	18.6
A2	1494	37	17.2
HASPB	1158	24	16.5
Amastin	1376	34	18.1
CPB	1642	31	17.8
HSP70	1948	28	19.4
LPG Genes	2032	39	18.8
LACK	1245	22	16.8

Interpretation

The distribution of the repeats in the major virulence genes of L. donovani is shown in Table 1. GP63 had the highest counts of the analysed genes for mirror repeats (42), followed by LPG biosynthesis (39) and A2 (37) genes, which suggest that these important virulence-associated genes are more abundant than others in symmetrical DNA motifs. The number of mirror repeats was the lowest in LACK (22). There was moderate variation in the size of the repeat in mirror, with an average of 16.5 bp in HASPB and 19.4 bp in HSP70. In summary, the results suggest that there is a

9 Franssen, S. U., Durrant, C., Stark, O., et al. (2020). Global genome diversity of the *Leishmania donovani* complex. *Nature Communications, 11*, 2372.
10 Herreros-Cabello, A., Callejas-Hernández, F., & Fresno, M. (2020). Genome organization and multigene families in *Trypanosoma cruzi*. *Frontiers in Cellular and Infection Microbiology, 10*, 614665.
11 Zheng, P., et al. (2020). Genomic and proteomic characterization of virulence-associated traits in *Leishmania donovani*. *BMC Genomics, 21*, 806.

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 6, Issue 7, July 2026
Impact Factor: 5.731

high level of distribution of mirRps within virulence genes, and these could potentially participate in genome organization, transcriptional regulation, and parasite adaptation.12

Table 2 GC-Rich Mirror Repeats

Gene	GC-rich MRs	Predicted H-DNA Potential
GP63	18	High
A2	16	High
Amastin	14	Moderate
LPG Genes	17	High
HSP70	13	Moderate
CPB	11	Moderate

Interpretation

The analysis of GC-rich mirror repeats and their potential to form H-DNA structures of selected virulence genes of Leishmania donovani is listed in Table 2. A gene with the highest number of GC-rich mirror repeats (18) was GP63, followed by LPG biosynthesis genes (17) and A2 (16), both of which were predicted to have high H-DNA-forming potential. Amastin (14), HSP70 (13) and CPB (11) had moderate numbers of GC rich mirror repeats and were considered to have moderate H-DNA potential. This may indicate that the GC-rich mirror repeats are over-represented in genes associated with key virulence factors, and that they may play a role in the formation of non-B DNA, transcription regulation, genome stability and parasite pathogenicity.13

Table 3 Functional Distribution

Functional Region	Percentage (%)
Coding Region	61
Promoter Region	24
UTR Region	9
Intragenic Spacer	6

Interpretation

Table 3 shows functional distribution of the different genomic regions of the major virulence genes of Leishmania donovani. Most of the mirror repeats occurred in coding regions (61%), suggesting that these repetitive motifs are mainly related to protein coding regions and could have a role in gene function or may influence the sequence evolution or structural stability. 24% of the identified mirror repeats were in promoter regions, which might be important in the regulation of gene transcription and in gene expression. The untranslated regions (UTRs) (9%) and intragenic spacer regions (6%) had smaller proportions. In summary, the results suggest that the distribution of mirror repeats is not random and may play a role in regulating genes, organizing the genome and adaptation to parasites.14

IV. DISCUSSION

The computational study showed a high degree of abundance of mirror repeats in the major virulence genes of Leishmania donovani. GP63 was the most highly decorated with repeat motifs, in line with its essential functions in the invasion of host cells and modulation of the immune system. Mirror repeats are found in high density in the promoter regions and it is believed that they may play a role in regulation via changes in DNA topology.15

12 Guiblet, W., et al. (2021). Non-B DNA: A major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nature Reviews Genetics, 22*(10), 646–663.
13 Brázda, V., Haronikova, L., Liao, J. C. C., & Fojta, M. (2021). DNA and RNA quadruplexes, triplexes, and other non-canonical nucleic acid structures in biology and medicine. *International Journal of Molecular Sciences, 22*(20), 10738.
14 Liu, Y., et al. (2021). Repetitive DNA sequences and genome evolution in microbial pathogens. *Microbial Genomics, 7*(11), e000654.
15 Schmidt, M., et al. (2022). Regulatory DNA motifs and transcriptional control in protozoan parasites. *Frontiers in Genetics, 13*, 904356.

The GP63, A2 and lipophosphoglycan biosynthesis genes were found to contain a high theoretical probability of forming triplex DNA structures in the areas identified as GC-rich mirror repeats. The previous studies on non-B DNA show that H-DNA affects transcription factor binding, RNA polymerase activity, genome stability, and chromatin accessibility. Thus, these motifs could be involved in regulatory aspects related to parasite intracellular survival.

The high frequency of mirror repeats in coding regions also implies evolutionary conservation, which might be related to structural constraints or selection pressures related to virulence. Repetitive sequence enrichment is also observed in multiple pathogenic protozoa, and plays a role in antigenic variation, immune evasion and chromosome plasticity.

In general, the results suggest that mirror repeats do not occur at random, but tend to be enriched in regions of the genome that are important to biology. Their functional significance needs to be validated using experimental studies that involve transcriptomics, chromatin assays and mutational analyses.

V. CONCLUSION

The present study is in silico analysis of the occurrence of mirror repeats in the major virulence genes of Leishmania donovani which gives new insights in the structural organization of repetitive DNA in this medically important protozoan parasite. Computational analysis showed that several Virulence Associated Genes (VAGs, GP63, A2, LPG biosynthesis genes, Amastin, CPB, HSP70 and HASPB) contain the mirror repeat elements widely distributed. Within these, the genes that were more abundant in their copies had an increasing number of mirror repeat motifs, which could imply that these genes have a higher degree of structure and that these motifs have a regulatory effect on these genes because of the presence of repetitive DNA elements.

It is likely that the sequences with high GC content identified as numerous mirror repeats that have the potential to form H-DNA may play roles in non-B DNA formation, which in turn could affect the regulation of transcription, the organization of chromatin, genome stability and adaptive evolution. The bias in the distribution of the motifs towards regions of the genome associated with coding and promoter regions further indicates that these motifs do not appear at random locations, but rather are concentrated in regions that function in a key manner in parasite survival and host-parasite interactions. This current study is a purely computational study, but the results do provide some evidence in support of the possible biological significance of the mirror repeats in regulating virulence associated genes.

In summary, the findings of this study will contribute to the existing body of information on the architecture of repetitive DNA in L. donovani, and will serve as an excellent source of genomic information for future molecular investigations. Their functional roles will need to be confirmed through experimental testing using gene expression analysis, mutagenesis, chromatin accessibility assays, and structural biology approaches. The results could end up helping to understand the pathogenicity of these parasites and aid in the discovery of new molecular targets for the diagnosis, treatment, and control of the visceral leishmaniasis.

REFERENCES

[1]. National Center for Biotechnology Information. (2024). *GenBank overview*. https://www.ncbi.nlm.nih.gov/genbank/
[2]. TriTrypDB. (2024). *TriTrypDB: The kinetoplastid genomics resource*. https://tritrypdb.org
[3]. World Health Organization. (2024). *Leishmaniasis*. https://www.who.int/news-room/fact-sheets/detail/leishmaniasis
[4]. Franssen, S. U., Durrant, C., Stark, O., et al. (2020). Global genome diversity of the *Leishmania donovani* complex. *Nature Communications, 11*, 2372.
[5]. Herreros-Cabello, A., Callejas-Hernández, F., & Fresno, M. (2020). Genome organization and multigene families in *Trypanosoma cruzi*. *Frontiers in Cellular and Infection Microbiology, 10*, 614665.
[6]. Zheng, P., et al. (2020). Genomic and proteomic characterization of virulence-associated traits in *Leishmania donovani*. *BMC Genomics, 21*, 806.
[7]. Guiblet, W., et al. (2021). Non-B DNA: A major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nature Reviews Genetics, 22*(10), 646–663.

[8]. Brázda, V., Haronikova, L., Liao, J. C. C., & Fojta, M. (2021). DNA and RNA quadruplexes, triplexes, and other non-canonical nucleic acid structures in biology and medicine. *International Journal of Molecular Sciences, 22*(20), 10738.

[9]. Liu, Y., et al. (2021). Repetitive DNA sequences and genome evolution in microbial pathogens. *Microbial Genomics, 7*(11), e000654.

[10]. Eggers, C., et al. (2022). Triplex-forming DNA motifs and their contribution to genome instability. *Genes, 13*(9), 1612.

[11]. Schmidt, M., et al. (2022). Regulatory DNA motifs and transcriptional control in protozoan parasites. *Frontiers in Genetics, 13*, 904356.

[12]. Silva, R., et al. (2023). Comparative genomics of repetitive DNA elements in kinetoplastid parasites. *Parasites & Vectors, 16*, 291.

[13]. Kumar, A., Sharma, P., & Singh, R. (2023). Computational identification of functional DNA motifs in infectious microorganisms using comparative genomics. *Briefings in Bioinformatics, 24*(5), bbad243.

[14]. Martínez, J., et al. (2024). Sequence symmetry and mirror repeat organization in pathogenic microbial genomes. *BMC Bioinformatics, 25*, 118.

[15]. Patel, S., Verma, N., & Rao, K. (2024). Advances in computational genomics for neglected tropical disease pathogens: Emerging perspectives on repetitive DNA and virulence. *Computational and Structural Biotechnology Journal, 23*, 1021–1036.