bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Title:

C-5 DNA methyltransferase 6 does not generate detectable DNA methylation in *Leishmania*

Authors:

Cuypers B[1, 2#], Dumetz F[1*#], Meysman P[2], Laukens K[2], De Muylder G[1], Dujardin J-C[1, 3] and Domagalska MA[1]

Affiliations:

[1] Molecular Parasitology, Institute of Tropical Medicine, Antwerp, Belgium, [2] Adrem Data Lab, Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium, [3] Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium. * Present address: Merrick's Lab, Department of Pathology, University of Cambridge, Cambridge, UK [#] Contributed equally. Corresponding author: mdomagalska@itg.be

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

While Cytosine C5 methylation is an important epigenetic control mechanism in a wide array of Eukaryotic organisms, for many protozoan species its status remains elusive. In this work, we investigated the presence of C5 DNA methylation in *Leishmania donovani*, a Trypanosomatid parasite and the causative agent of the visceral leishmaniasis in humans. We show that the *L. donovani* genome contains a C-5 DNA methyltransferase (*DNMT*) from the *DNMT6* subfamily, of which the function is still unclear, and verified its expression at RNA level. We created a viable overexpressor and knock-out lines of this enzyme and characterised their genome-wide methylation patterns using whole-genome bisulfite sequencing, together with promastigote and amastigote control lines. Interestingly, we found that methylation levels were equal to or lower than 0.0003% at CpG sites, 0.0005% at CHG sites and 0.0126% at CHH sites at genome scale. As none of these methylated sites were retained after manual verification, we concluded that there is no evidence for DNA methylation in this species. Similar results were obtained for the blood form of *T. brucei*, another Trypanosomatid species. We demonstrated that this difference in DNA methylation between the parasite (no detectable DNA methylation) and the vertebrate host (DNA methylation), can be used as the base of an enrichment strategy for parasite DNA. As such, we depleted methylated DNA from 1) mixes of *Leishmania* promastigote and amastigote DNA with human DNA and 2) THP-1 cells infected with *Leishmania* amastigotes. This resulted in 62x to 263x *Leishmania:*human enrichment, depending on the dilution and type of sample studied. These results open a promising avenue for including an unmethylated DNA enrichment step as a pre-enrichment before sequencing clinical samples.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

DNA methylation is an epigenetic mechanism responsible for a diverse set of functions across the three domains of life, Eubacteria, Archeabacteria, and Eukaryota. In Prokaryotes, many DNA methylation enzymes are part of so-called restriction modification systems, which play a crucial role in their defence against phages and viruses. Prokaryotic methylation typically occurs on the C5 position of cytosine (cytosine C5 methylation), the exocyclic amino groups of adenine (adenine-N6 methylation) or cytosine (cytosine-N4 methylation) [1]. In Eukaryotic species, DNA methylation is mostly restricted to 5-methylcytosine (me^5C) and best characterised in mammals, where 70-80% of the CpG motifs are methylated [2]. As such, DNA methylation controls a wide range of important cellular functions, such as genomic imprinting, X-chromosome inactivation (in humans), gene expression and the repression of transposable elements. Consequently, defects in genetic imprinting are associated with a variety of human diseases and changes in DNA methylation patterns are common hallmark of cancer [3,4]. Eukaryotic DNA methylation can also occur at CHG and CHH (where H is A, C or T) sites [5], which was considered to occur primarily in plants. However, studies from the past decade demonstrate that CHG and CHH methylation are also frequent in several mammalian cells types, such as embryonic stem cells, oocytes and brains cells [5-8].

Me^5C methylation is mediated by a group of enzymes called C-5 DNA methyltransferases (DNMTs). This ancient group of enzymes share a common ancestry and their core domains are conserved across Prokaryotes and Eukaryotes [1]. Different DNMT subfamilies have developed distinct roles within epigenetic control mechanisms. For example, in mammals DNMT3a and DNMT3b are responsible for *de novo* methylation, such as during germ cell differentiation and early development, or in specific tissues undergoing dynamic methylation [9]. In contrast, DNMT1 is responsible for maintaining methylation patterns, particularly during the S phase of the cell cycle where it methylates the newly generated hemimethylated sites on the DNA daughter strands [10]. Some DNMTs have also changed substrate over the course of evolution. A large family of DNMTs, called DNMT2, has been shown to methylate the 38th position of different tRNAs to yield ribo-5-methylcytidine (rm5C) in a range of Eukaryotic organisms, including humans [11], mice [12], *Arabidopsis thaliana* [13] and *Drosophila melanogaster* [14]. Therefore, DNMT2s are now often referred to as 'tRNA methyltransferases' or trDNMT and are known to carry out diverse regulatory functions [15]. However, in other Eukaryotic taxa, DNMT2 appears to be a genuine DNMT, as DNMT2 can catalyze DNA methylation in *Plasmodium falciparum* [16], and *Schistosoma mansoni* [17]. In *Entamoeba histolytica* both DNA and RNA can be used as substrates for DNMT2 [18,19]. The increase in available reference genomes of non-model Eukaryotic species has recently also resulted in the discovery of new DNMTs,

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

such as DNMT5, DNMT6 or even SymbioLINE-DNMT, a massive family of DNMTs, so far only found in the dinoflagellate *Symbiodinium* [20].

Indeed, DNMT mediated C5 methylation has been shown to be of major functional importance in a wide array of Eukaryotic species, including also protozoans such as *Toxoplasma gondii* and *Plasmodium* [16,21]. In contrast, studies have failed to detect any C5 DNA methylation in Eukaryotic species such as *Caenorhabditis elegans, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* [22,23]. In other protozoans, the presence and potential role of DNA-methylation remains elusive. This is especially true for *Leishmania*, a Trypanosomatid parasite (Phylum Euglenozoa), despite its medical and veterinary importance. Indeed, *Leishmania* is the causative agent of the leishmaniasis in humans and a wide variety of vertebrate animals, a disease that ranges from self-healing cutaneous lesions to lethal visceral leishmaniasis. Interestingly, the parasite also features a gene expression that is remarkably different from other Eukaryotes and includes a system of constitutive transcription [24]. Despite this constitutive expression system, the parasite extensively regulates its transcript and protein levels between life stages and in response to environmental conditions [25-27]. However, many parts of this responsive regulation system are still poorly understood.

With DNA methylation being an important gene expression regulation mechanism in a wide range of Eukaryotic species, we were particularly interested in the 5-C methylation status of *Leishmania*. A single report shows low levels of Me^5C methylation in *T. brucei* (0.01 %), another Trypanosomatid species [28]. Since this study, new techniques have emerged to study methylation with an unprecedentedly high resolution. Here we present the first deep analysis of genome wide methylation patterns in *Leishmania* across different parasite life stages, making use of whole genome bisulfite sequencing.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Materials and Methods

In silico identification and phylogeny of putative DNMTs

To identify putative C-5 cytosine-specific DNA methylases in *Leishmania donovani*, we obtained the hidden Markov model (hmm) for this protein family from PFAM version 32.0 (Accession number: PF00145) [29]. The hmm search tool contained within hmmer-3.2.1 (hmmer.org) was then used with default settings to screen the LdBPK V2 reference genome for this hmm signature [26]. The initial pairwise alignment between the identified *L. donovani* and *T. brucei* C5 DNA MTase was carried out with T-COFFEE V_11.00.d625267.

To construct a comprehensive phylogenetic tree of the C5 DNA MTase family, including members found in Trypanosomatid species, we modified the approach from Ponts et al. [16]. Firstly, we download the putative proteomes of a wide range of Prokaryotic and Eukaryotic species. These species were specifically selected to cover the different C5 DNA MTase subfamilies [1]. Specifically, the following proteomes were obtained: *Trypanosoma brucei* TREU92, *Trypanosoma vivax* Y486 and *Leishmania major* Friedlin from TriTrypDB v41 [24,30,31], *Plasmodium falciparum* 3D7, and *Plasmodium vivax* P01 from PlasmodDB v41 [32-34], *Cryptosporidium parvum* Iowa II and *Cryptosporidium hominis* TU502 from CryptoDB v41 [35-37], *Toxoplasma gondii* ARI from ToxoDB v 41 [38,39], *Euglena gracilis* Z1 (PRJNA298469) [40], *Entamoeba histolytica* HM-1:IMSS (GCF_000208925.1) [41], *Schizosaccharomyces pombe* ASM294 (GCF_000002945.1) [42], *Saccharomyces cerevisiae* S288C (GCF_000146045.2), *Neurospora crassa* OR74A (GCF_000182925.2) [43], *Arabidopsis thaliana* (GCF_000001735.4), *Drosophila melanogaster* (GCF_000001215.4), *Homo sapiens* GRCh38.p12 (GCF_000001405.38), *Bacillus subtilis* 168 (GCF_000009045.1), *Clostridium botulinum* ATCC 3502 (GCF_000063585.1) [44], *Streptococcus pneumoniae* R6 (GCF_000007045.1) [45], *Agrobacterium tumefaciens* (GCF_000971565.1) [46], *Salmonella enterica* CT18 (GCF_000195995.1) [47] and *Escherichia coli* K12 (GCF_000005845.2) from NCBI, *Ascobolus immersus* RN42 [48] from the JGI Genome Portal (genome.jgi.doe.gov) and *Danio rerio* (GRCz11) from Ensembl (ensembl.org).

All obtained proteomes were then searched with the hmm signature for C5 DNA MTases, exactly as described above for *L. donovani.* All hits with an E-value < 0.01 (i.e. 1 false positive hit is expected in every 100 searches with different query sequences) were maintained, and all domains matching the query hmm were extracted and merged per protein. This set of sequences was aligned in Mega-X with the MUSCLE multiple sequence alignment algorithm [49,50] and converted to the PHYLIP format with the ALTER tool [51]. Phage sequences and closely related isoforms were removed.

A maximum likelihood tree of this alignment was generated with RAxML version 8.2.10 using the automatic protein model assignment algorithm (option: -m PROTGAMMAAUTO). RAxML was run in

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

three steps: Firstly, 20 trees were generated and only the one with the highest likelihood score was kept. Secondly, 1000 bootstrap replicates were generated. In a final step, the bootstrap bipartions were drawn on the best tree from the first round. The tree was visualised in Figtree v1.4.4 (https://github.com/rambaut/figtree/).

Culturing & DNA extraction for Bisulfite Sequencing

Promastigotes (extracellular life stage) of *Leishmania donovani* MHOM/NP/03/BPK282/0 cl4 (further called BPK282) and its genetically modified daughter lines (see below) were cultured in HOMEM (Gibco) supplemented with 20% (v:v) heat-inactivated foetal bovine serum at 26°C. Amastigotes (intracellular life stage) of the same strain were obtained from three months infected golden Syrian hamster (Charles Rivers) as described in Dumetz *et al.* [26] and respecting BM2013-8 ethical clearance from Institute of Tropical Medicine (ITM) Animal Ethic Committee. Briefly, 5 week old female golden hamsters were infected via intracardiac injection of 5.10^5 stationary phase promastigotes. Three months post infection, hamsters were euthanised and amastigotes were purified from the liver by Percol gradient (GE Healthcare) after homogenisation. *T. brucei gambiense* MBA blood stream forms were obtained from OF-1 mice when the parasitaemia was at its highest, according to ITM Animal Ethic Committee decision BM2013-1. Parasites were separated from the whole blood as described in Tihon *et al.* [52]. Briefly, the parasites were separated from the blood by placing the whole blood on an anion exchanger Diethylaminoethyl (DEAE)-cellulose resin (Whatman) suspended in phosphate saline glucose (PSG) buffer, pH 8. After elution and two washes on PSG, DNA was extracted. DNA of *L. donovani*, both promastigotes and amastigotes, as well as *T. brucei* was extracted using DNeasy Blood & Tissue kit (Qiagen) according manufacturer instructions.

Arabidopsis thaliana Col-0 was grown for 21 days under long day conditions, i.e. 16 hrs light and 8 hrs darkness. DNA was then extracted from the whole rosette leaves using the DNeasy Plant Mini Kit (Qiagen).

Genetic engineering of *L. donovani* BPK282

We generated both an LdDNMT overexpressing (LdDNMT+) and null mutant line (LdDNMT-/-) of *L. donovani* BPK282. The PCR amplicons for LdDNMT+ were generated using Phusion (NEB) and sequenced at the VIB sequencing facility: overexpression construct, pLEXSY-DNMT, was generated by amplifying LdBPK_251230 from BPK282 genomic DNA and cloned inside the expression vector pLEXSY-Hyg2 (JENA bioscience) using NEBuilder (NEB) according to manufacturer's instruction for primer design and cloning instructions. Once generated, 10 µg of pLEXSY-DNMT was electroporated in 5.10^7 BPK282 promastigotes from logarithmic culture using cytomix on a GenePulserX (BioRad) according to LeBowitz (1994) [53] and selected *in vitro* by adding 50 µg/mL hygromycin B (JENA

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bioscience) until parasite growth [54]. Verification of overexpression was carried out by qPCR on a LightCycler480 (Roche) using SensiMix SYBR No-ROX (Bioline) on cDNA. Briefly, 10^8 logarithmic-phase promastigotes were pelleted, RNA extraction was performed using RNAqueous-Micro total RNA isolation kit (Ambion) and quantified by Qubit and the Qubit RNA BR assay (Life Technologies, Inc.). Transcriptor reverse transcriptase (Roche) was used to synthesise cDNA following manufacturer's instructions. qPCRs were run on a LightCycler 480 (Roche) with a SensiMix SYBR No-ROX kit (Bioline); primer sequences available in **Supplementary Table S1**. Normalisation was performed using two transcripts previously described as stable in promastigotes and amastigotes in Dumetz et al [55], LdBPK_340035000 and LdBPK_240021200.

For the generation of LdDNMT-/-, a two-step gene replacement strategy was used: replacing the first allele of LdBPK_250018100.1 by nourseothricin resistance gene (SAT) and the second allele by a puromycin resistance gene (Puro). Briefly, each drug resistance gene was PCR amplified from pCL3S and pCL3P using Phusion (NEB) and cloned between 300 bp of PCR amplified DNA fragments of the LdBPK_250018100.1 5' and 3' UTR using NEBuilder (NEB) inside pUC19 for construct amplification in *E. coli* DH5α (Promega) (cf. primer list in **Supplementary Table S1**). Each replacement construct was excised from pUC19 using SmaI (NEB), dephosphorylated using Antarctic Phosphatase (NEB) and 10 µg of DNA was used for the electroporation in the same conditions as previously described to insert the pLEXSY-DNMT. The knock-out was confirmed by whole genome sequencing.

Bisulfite sequencing and data analysis

For each sample, one microgram of genomic DNA was used for bisulfite conversion with innuCONVERT Bisulfite All-In-One Kit (Analytikjena). Sequencing libraries were prepared with the TruSeq DNA Methylation kit according to the manufacturer's instructions (Illumina). The resulting libraries were paired-end (2 x 100bp) sequenced on the Illumina HiSeq 1500 platform of the University of Antwerp (Centre of Medical Genetics). The sequencing quality was first verified with FastQC v0.11.4. Raw reads generated for each sample were aligned to their respective reference genome with BSseeker 2-2.0.3 [56]: LdBPK282v2 [26] for *L. donovani*, TREU927 [30] for *T. brucei* and Tair10 [57] for the *A. thaliana* positive control. Samtools fixmate (option -m) and samtools markdup (option -r) were then used to remove duplicate reads. CpG, CHG and CHH methylation sites were subsequently called with the BS-Seeker2 'call' tool using default settings and further filtered with our Python3 workflow called 'Bisulfilter' (available at https://github.com/CuypersBart/Bisulfilter). Genome-wide visualisation of methylated regions was then carried out with ggplot2 in R [58]. In *Leishmania*, the positions that passed our detection thresholds (coverage > 25, methylation percentage > 0.8), were then manually inspected in IGV 2.5.0 [59].

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Leishmania DNA enrichment from a mix of human and *Leishmania* DNA

To check whether the lack of detectable DNA methylation in *Leishmania* can be used for the enrichment of *Leishmania* versus (methylated) human DNA, we carried out methylated DNA removal on three types of samples: (1) An artificial mix of *L. donovani* BPK282/0 cl4 promastigote DNA with human DNA (Promega) from 1/15 to 1/150000 (*Leishmania*:human) and (2) *in vitro* infected PMA differentiated THP-1 cell lines with BPK282/0 cl4 stationary phase promastigotes (1:10). For the last experiment (3), linked promastigote and hamster-derived amastigote samples from 3 clinical *Leishmania donovani* strains (BPK275, BPK282 and BPK026), which were generated in previous work [26]. For this experiment, we used a 1/1500 artificial mix of *L. donovani* BPK282/0 cl4 promastigote DNA and human DNA (Promega) to reflect the ratio found in clinical samples. All DNA (except the human DNA from Promega) was extracted with the DNA (DNeasy Blood & Tissue kit, Qiagen). *Leishmania* DNA was then enriched using NEBNext Microbiome DNA Enrichment Kit (NEB) according to manufacturer instructions. Evaluation of the ratio *Leishmania*/human DNA was performed by qPCR on LightCycler480 (Roche) using SensiMix SYBR No-ROX (Bioline) and RPL30 primers provided in the kit to measure human DNA and *Leishmania* CS primers (Cysteine synthase) [60].

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results

The *Leishmania* genome contains a C-5 DNA methyltransferase (DNMT)

Eukaryotic DNA methylation typically requires the presence of a functional C-5 cytosine-specific DNA methylase (C5 DNA MTase). This type of enzymes specifically methylates the C-5 position of cytosines in DNA, using S-Adenosyl methionine as a methyl-donor. To check for the presence of a C5 DNA MTase in *Leishmania donovani*, we carried out a deep search of the parasite's genome. In particular, we used the LdBPK V2 reference genome [26] and searched the predicted protein sequences of this assembly using the hidden-markov-model (hmm) signature of the C5 DNA MTase protein family obtained from PFAM (PF00145) and obtained a single hit: the protein LdBPK_250018100.1, (E-value: 2.7e-40). LdBPK_250018100.1 was already annotated as 'modification methylase-like protein' with a predicted length of 840 amino acids. We will further refer to this protein as LdDNMT. Interestingly, in another Trypanosomatid species, *Trypanosoma brucei*, the homolog of this protein (Tb927.3.1360 or TbDNMT) has been previously been studied in detail by Militello et al [28]. Moreover, these authors showed that TbDNMT has all the ten conserved domains that are present in functional DNMTs. We aligned TbDNMT with LdDNMT using T-Coffee (**Fig. 1**) and found that these 10 domains are present in LdDNMT, including also the putative catalytic cysteine residue in domain IV.

Leishmania and Trypanosomatid C-5 DNA belong to the Eukaryotic DNMT6 family

To learn more about the putative function and evolutionary history of this protein, we wanted to characterise the position of *LdDNMT* and those of related Trypanosomatid species within the DNMT phylogenetic tree. Consequently, we collected the publicly available, putative proteomes of a wide range of Prokaryotic and Eukaryotic species, searched them for the hmm signature of the C5-DNMT family, aligned the identified proteins and generated a RAxML maximum likelihood tree. In total we identified 131 putative family members in the genomes of 24 species (E-value < 0.01), including 4 Prokaryotic (*Agrobacterium tumefaciens, Salmonella enterica*, *Escherichia coli* and *Clostridium botulinum*) and 20 Eukaryotic species. These Eukaryotic species were selected to contain organisms from the Excavata Phylum (of which *Leishmania* is part) and a range of other, often better-characterised Phyla as a reference. The Excavata species included 4 Trypanosomatids (*Leishmania donovani*, *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma vivax*), 1 other, non-Trypanosomatid Euglenozoid species (*Euglena gracilis*) and 1 other non-Euglenozoid species (*Naegleria gruberi*). The other Eukaryotic Phyla included in the analysis were: Apicomplexa (*Plasmodium vivax*, *Plasmodium falciparum*, *Cryptosporidium parvum*, *Cryptosporidium hominis*),

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Amoebozoa (*Entamoeba histolytica*), Angiosperma (*Arabidopsis thaliana*, *Oryza sativa*), Ascomycota (*Ascobolus immerses*, *Neurospora crassa*) and Chordata (*Homo Sapiens, Danio rerio)* (**Figure 2**).

Our phylogenetic tree was able to clearly separate known DNMT subgroups, including DNMT1, DNMT2, DNMT3, DRM (Domain rearranged methyltransferase), DIM and 2 groups of Prokaryotic DNMTs [1,16,61]. Interestingly, the tree also showed that Trypanosomatid DNMTs group together and are part of the much less-characterised DNMT6 group, as has been previously described for *Leishmania major* and *Trypanosoma brucei* [20]. This group of DNMTs has also been found also in diatoms (e.g. *Thalasiosira*) and recently in dinoflaggelates (e.g. *Symbiodinium kawagutii* and *Symbiodinium minutum*), but its function remains elusive [20,62]. The most closely related branch to DNMT6 contains a group of bacterial DNMTs (here represented by *Agrobacterium tumefaciens, Salmonella enterica*, *Escherichia coli*). This highlights that DNMT6 emerged from the pool of Prokaryotic DNMTs independently from the groups previously mentioned. The fact that another Euglenozoid, *Euglena gracilis,* has DNMT1, DNMT2, DNMT4 and DNMT5, while another Excavata species, *Naegleria gruberi* has both a DNMT1 and a DNMT2, suggests that the ancestors of the current Excavata species possessed a wide battery of DNMTs including also DNMT6. In the lineage that eventually led to Trypanosomatids, these were all lost, except DNMT6.

Whole genome bisulfite sequencing reveals no evidence for functional C-5 methylation

As 1) we identified LdBPK_250018100.1 to be from the C5 DNA MTase family, 2) all 10 conserved domains were present, we decided to check also for the presence and functional role of C5 DNA methylation in *L.* donovani. Therefore, we assessed the locations and degree of CpG, CHG and CHH methylation across the entire *Leishmania* genome and within the two parasite life stages: amastigotes (intracellular mammalian life stage) and promastigotes (extracellular, insect life stage). Amastigotes were derived directly from infected hamsters, while promastigotes were obtained from axenic cultures. Promastigotes were divided in two batches, one passaged long-term in axenic culture, the other passaged once through a hamster and then sequenced at axenic passage 3, thus allowing us to study also the effect of long versus- short term *in vitro* passaging. *Arabidopsis thaliana* and *T. brucei* were included as a positive control as the degree of CpG, CHG and CHH methylation in *A. thaliana* is well known [63,64], while *T. brucei* is the only Trypanosomatid in which (low) methylation levels were previously detected by mass spectrometry [28].

An overview of all sequenced samples can be found in **Supplementary Table S2.** All *L. donovani* samples were sequenced with at least 30 million 100bp paired end reads (60 million total) per sample resulting in an average genomic coverage of at least 94X for the *Leishmania* samples. The *T. brucei* was sequenced with 69 million PE reads resulting in 171X average coverage and *A. thaliana* 27

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

million PE reads, resulting in 21X average coverage. Detailed mapping statistics can be found in **Supplementary Table S3**.

We first checked for global methylation patterns across the genome. Interestingly, we could not detect any methylated regions in *Leishmania donovani* promastigotes, both short (P3) and long-term *in vitro* passaged, nor in hamster derived promastigotes or amastigotes (**Fig. 3**). Minor increases in the CHH signal towards the start end of several chromosomes, were manually checked in IGV and attributed to poor mapping in (low complexity) telomeric regions. This was in contrast to our positive control, *Arabidopsis thaliana*, that showed clear highly methylated CpG, CHG and CHH patterns across the genome. This distribution was consistent with prior results with MethGO observed on *Arabidopsis thaliana*, confirming that our methylation detection workflow was working [58].

In a second phase, we checked for individual sites that were fully methylated (>80% of the sequenced DNA at that site) using BS-Seeker2 and filtering the results with our automated Python3 workflow. CpG methylation in all three biological samples for *L. donovani* was lower than 0.0003%, CHG methylation lower than 0.0005% and CHH methylation lower than 0.0126% (**Table 1, Supplementary Table S4**). However, when this low number of detected 'methylated' sites was manually verified in IGV, they could all clearly be attributed to regions where BS-Seeker2 wrongly called methylated bases, either because of poor mapping (often in repetitive, low complexity regions) or of strand biases. In reliably mapped regions, there was clearly no methylation. Similarly, we detected 0.0001% of CpG methylation, 0.0006% of CHG methylation and 0.0040% of CHH methylation for *T. brucei*, which could all be attributed to mapping errors or strand biases. In *A. thaliana*, our positive control, we detected 21.05% of CpG methylation, 4.04% of CHG methylation and 0.31% of CHH methylation, which is similar as reported values in literature [65,66], and demonstrates that our bioinformatic workflow could accurately detect methylated sites. We also checked sites with a lower methylation degree (>40%), which gave higher percentages, but this could be attributed to the increased noise level at this resolution (**Supplementary Table S4**). Indeed, even when applying stringent coverage criteria (>25x) this approach is susceptible for false positive methylation calls, as we are checking millions of positions (in case of *Leishmania*, more than 5.8 million CG sites, 3.9 million CHG sites and 9.3 million CHH sites).

To determine whether *LdDNMT* is essential and/or if it affects the C-5 DNA-methylation pattern, we also sequenced an *L. donovani DNMT* knock-out (LdDNMT-/-) line as well as a *DNMT* overexpressor (LdDNMT+). The successful generation of LdDNMT-/- and LdDNMT+ was verified on the DNA level by calculating their *LdDNMT* copy number based on the sequencing coverage (**Figure 4**). Indeed, the

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

copy number of the *LdDNMT* gene in LdDNMT-/- was reduced to zero, while that of LdDNMT+ was increased to 78 copies. The overexpressor was also verified on the RNA level (**Table 2**) and showed a 2.5-fold higher expression than the corresponding wild type. Although the LdDNMT+ initially seemed to have slightly higher methylation percentages (**Table 1, Supplementary Table S4**), none of these methylation sites passed our manual validation in IGV. Thus, we did not find evidence for methylation in either of these lines. Additionally, the fact that the LdDNMT-/- line was viable shows that *LdDNMT* is not an essential gene in promastigotes.

Absence of C5 DNA Methylation as a *Leishmania* vs host DNA enrichment strategy

The lack or low level of C5 DNA methylation opens the perspective for enriching *Leishmania* DNA in mixed parasite- host DNA samples, based on the difference in methylation status (the vertebrate host does show C5 DNA methylation). This could potentially be an interesting pre-enrichment step before whole genome sequencing analysis of clinical samples containing *Leishmania*.

To test this assumption, we first generated artificially mixed samples using different ratios of *L. donovani* promastigote DNA with human DNA. Ratios were made starting from 1/15 to 1/15000, which reflects the real ratio of *Leishmania* vs human DNA in clinical samples [67]. From these mixes, *Leishmania* DNA was enriched using NEBNext Microbiome DNA Enrichment Kit (NEB) that specifically binds methylated DNA, while the non-methylated remains in the supernatant. We observed an average 263 X enrichment of *Leishmania* versus human DNA (**Figure 5**). This ranged between 378x for the lowest dilution (removing 99.8% of the human DNA) to 164x (removing 99.6% of the human DNA) in the highest diluted condition (1/15000 *Leishmania*:human).

Secondly, we wanted to test if this enrichment technique worked equally well on *L. donovani* amastigotes for (a) fundamental reasons, as an (indirect) second method to detect if there are any methylation differences between promastigotes and amastigotes, and (b) practical reasons, as it the (intracellular) life stage encountered in clinical samples. Therefore, we also carried out this enrichment technique on 3 sets (3 strains) of hamster derived amastigotes and their promastigote controls. Similarly as in the previous experiment, *Leishmania*-human DNA mixes were generated in a 1/1500 (*Leishmania:*human) ratio after which enrichment was carried out with the NEBNext Microbiome DNA Enrichment Kit. The enrichment worked well for both life stages, the promastigote samples were on average 76.22 ± 14.28 times enriched and the amastigote samples 61.68 ± 4.23 times (**Table 3**. We also tested if the enrichment technique worked on non-artificial mixes, i.e. directly on DNA samples of *Leishmania* infected cells without prior purification of amastigotes. For this, we carried out *in vitro* infections of a THP-1 cell line with BPK282 *L. donovani* promastigotes, adding 10 parasites per human cell. qPCR showed that the ratio *Leishmania* amastigote/human DNA

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in these samples was around 1/100 **(Table 4).** The enrichment was successful and increased the amastigote DNA proportion in the mix on average 36.73 times.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Discussion

With this work, we present the first comprehensive study addressing the status of DNA-methylation in *Leishmania*.

We show that the *Leishmania* genome contains a C5-*DNMT* (*LdDNMT*) that contains all 10 conserved *DNMT* domains. We also showed the gene is expressed at the RNA level. As the C5-DNMT family is diverse and several family members are known to have adopted (partially) distinct functions during the course of evolution, we were particularly interested in the position of this DNMT within the evolutionary tree of this family, as it could guide our hypothesis for the function of this protein. We found that LdDNMT is in fact a DNMT6, just as those found in *L. major* and *T. brucei* [20]. Interestingly, all other (non-Trypanosomatid) species studied so far had either multiple *DNMT6* copies and/or other *DNMT* subfamily members in their genomes [20,62]. Therefore, *Trypanosomatids* might be a unique model species to further study the role of this elusive DNMT subfamily, as there can be no interaction with the effects of other DNMTs.

The fact that our LdDNMT knock-out line (verified by sequencing) was viable shows that DNMT6 is not essential for the survival of the parasite, at least in promastigotes in our experimental conditions. However, at the same time one might hypothesize that DNMT6 does offer a selective advantage to the parasite. First of all, the sequence of DNMT core domains is extremely conserved across the tree of life and this is no different from those that we encountered in *Leishmania*. Secondly, *Leishmania* is characterised by a high genome plasticity and features extensive copy number differences [68,69]. Therefore, one might speculate that the parasite would have lost the gene long time ago if it did not provide any selective advantage.

In addition, we aimed to characterise the DNA-methylation patterns of the parasite's genome. Therefore, we carried out the first whole genome bisulfite sequencing experiment on *Leishmania* and Trypanosomatids in general. We checked both the promastigote (both culture and amastigote life stage). Surprisingly, we did not find any evidence for DNA methylation in *L. donovani* even though we checked both for large, regional patterns (sensitive for low levels of methylation over longer distances) and site-specific analyses (sensitive for high levels of methylation at individual sites). This could either mean that there is indeed no DNA-methylation in these species, or that was below our detection threshold. Regarding this detection threshold, two factors should be considered. Firstly, bisulfite sequencing and analysis allows for the detection of specific sites that are consistently methylated across the genomes of a mix of cells. For example, in our case, we looked for sites that are methylated in at least 80% or 40% of the cases. Thus, if *Leishmania* consistently methylates certain genomic positions, our pipeline would have uncovered this. However, if this

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

methylation would be more random, or occurring in only a small subset of cells, we would not be able to distinguish this for random sequencing errors, and as such, we cannot exclude this possibility. Secondly, bisulfite sequencing typically suffers from poor genomic coverage due to the harsh BS treatment of the DNA [70]. In our *L. donovani* samples we covered at least 30.14% of the CpG sites, 29.47% of the CHG sites and 24.23% of the CHH sites (even though having more than 90x average coverage). As there are millions of CpG, CHG and CHH sites in the genome, the chance is very small (0.75^n, with n = number of methylated sites) that we would not have detected methylation, even if present in low numbers. However, this assumes our methylation is distributed randomly, which therefore is a second consideration for the interpretation of the results reported here.

In any case, it is hard to imagine that any of the typical Eukaryotic DNA methylation systems such as genomic imprinting, chromosome inactivation, gene expression regulation and/or the repression of transposable elements could be of significance with such low methylation levels. On the other hand, given its phylogenetic position, it is perfectly possible that DNMT6 has changed its biological activity and now carries out another function. Indeed, as we described above, a similar phenomenon was observed with DNMT2 that switched it substrate from DNA to tRNA during the course of evolution [16,17].

Correspondingly, we did not observe any detectable DNA methylation for *T. brucei*. These findings are, however, in contrast to what has been reported before by Militello et al., who detected 0.01% of 5MC in the *T. brucei* genome [28]. Also the methylated loci described in this paper could not be confirmed in the current work. However, this is maybe not be surprising as the same authors reported later that TbDNMT might in fact methylate RNA, as they identified methylated sites in several tRNAs [71]. This would indeed explain why we do not observe C5-DNA methylation in *T. brucei* with high resolution, whole genome bisulfite sequencing, and further suggest that a similar substrate switch to tRNA has occurred for DNMT6, just like has occurred for DNMT2. Further functional characterisation of DNMT6 is required to verify this hypothesis.

From an applied perspective, this study opens new avenues for the enrichment of Trypanosmatid DNA from clinical samples, which often have an abundance of host DNA. Indeed, depletion of methylated DNA could be included as pre-enrichment step for existing enrichment approaches. For example, our group has recently obtained excellent sequencing results of clinical samples using SureSelect (97% of the samples for diagnostic SNPs, 83% for genome wide information for sequenced samples), but was not able to sequence samples below 0.006% of *Leishmania* DNA content [71]. Perhaps the removal of methylated DNA could further enhance the sensitivity of this method. In the case of *Leishmania* the technique could even be useful both from enrichments from

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the mammalian hosts and the insect vector, as it was recently shown the phlebotomine vector also carries Me^5C in its genome [72]. The depletion of methylated DNA as a pre-enrichment step before whole genome sequencing has also been successfully used before for the parasite *Plasmodium falciparum* (malaria) and shown to generate unbiased sequencing reads [73].

In conclusion, we demonstrated that the *Leishmania* genome encodes for a DNMT6, but DNA methylation is either absent or present in such low proportion that it is unlikely to have a major functional role. We suggest that more investigation at RNA level is required to address the role of LdDNMT in *Leishmania*.

Data Availability

Raw sequencing data is available in the Sequence Read Archive under project accession numbers PRJNA560731 and PRJNA560871. Individual sample accession numbers are available in **Supplementary Table S2**.

Acknowledgments

We thank Dr. Gaurav Zinta and Prof. Dr. Gerrit Beemster for providing us with the *A. thaliana* DNA, as well as Prof. Dr. Philippe Büscher and Nicolas Bebronne for the blood stream form of *T. brucei gambiense* MBA and Prof. Dr. Joachim Clos for the *Leishmania* expression vectors pCL3S an pCL3P. This work was supported by the Interuniversity Attraction Poles Program of Belgian Science Policy [P7/41 to JC.D.] and by the organisation "Les amis des Instituts Pasteur à Bruxelles, asbl" [F.D.]. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI. This work was also supported by the Department of Economy, Science and Innovation in Flanders ITM-SOFIB (SINGLE project, to J-C D). We thank the Center of Medical Genetics at the University of Antwerp for hosting the NGS facility. BC is a post-doctoral fellow funded from the FWO [12V5319N].

Author contributions

Designed the experiments: B.C., F.D., G.D.M., J-C.D., M.A.D. Performed the experiments: B.C., F.D., M.A.D. Analysed the data: B.C., F.D., P.M., K.L., M.A.D. Wrote the manuscript: B.C., F.D., P.M., K.L., J-C.D, M.A.D. All authors reviewed and approved the final version of the manuscript.

Additional Information

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1 Jurkowski, T. P. & Jeltsch, A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One* **6**, e28104, doi:10.1371/journal.pone.0028104 (2011).

2 Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133-a019133, doi:10.1101/cshperspect.a019133 (2014).

3 Vidal, E. *et al.* A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* **36**, 5648, doi:10.1038/onc.2017.176

https://www.nature.com/articles/onc2017176#supplementary-information (2017).

4 Jin, Z. & Liu, Y. DNA methylation in human diseases. *Genes & diseases* **5**, 1-8, doi:10.1016/j.gendis.2018.01.002 (2018).

5 Guo, W., Chung, W.-Y., Qian, M., Pellegrini, M. & Zhang, M. Q. Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells. *Nucleic acids research* **42**, 3009-3016, doi:10.1093/nar/gkt1306 (2013).

6 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).

7 Tomizawa, S. *et al.* Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development (Cambridge, England)* **138**, 811-820, doi:10.1242/dev.061416 (2011).

8 Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816-831, doi:10.1016/j.cell.2011.12.035 (2012).

9 Chédin, F. in *Progress in Molecular Biology and Translational Science* Vol. 101 (eds Xiaodong Cheng & Robert M. Blumenthal) 255-285 (Academic Press, 2011).

10 Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nature reviews. Genetics* **10**, 805-811, doi:10.1038/nrg2651 (2009).

11 Jurkowski, T. P. *et al.* Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA* **14**, 1663-1670, doi:10.1261/rna.970408 (2008).

12 Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature Structural &Amp; Molecular Biology* **19**, 900, doi:10.1038/nsmb.2357

https://www.nature.com/articles/nsmb.2357#supplementary-information (2012).

13 Burgess, A. L., David, R. & Searle, I. R. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC plant biology* **15**, 199, doi:10.1186/s12870-015-0580-8 (2015).

14 Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic acids research* **37**, e12-e12, doi:10.1093/nar/gkn954 (2009).

15 Jeltsch, A. *et al.* Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA biology* **14**, 1108-1123, doi:10.1080/15476286.2016.1191737 (2016).

16 Ponts, N. *et al.* Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum. *Cell host & microbe* **14**, 696-706, doi:10.1016/j.chom.2013.11.007 (2013).

17 Motorin, Y. & Grosjean, H. Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: identification of the gene and substrate specificity of the enzyme. *RNA* **5**, 1105-1118, doi:10.1017/s1355838299982201 (1999).

18 Jeltsch, A., Nellen, W. & Lyko, F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends in biochemical sciences* **31**, 306-308, doi:10.1016/j.tibs.2006.04.005 (2006).

19 Kaiser, S. *et al.* The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA biology* **14**, 1241-1251, doi:10.1080/15476286.2016.1236170 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

20 Ponger, L. & Li, W. H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular biology and evolution* **22**, 1119-1128, doi:10.1093/molbev/msi098 (2005).

21 Wei, H. *et al.* Characterization of Cytosine Methylation and the DNA Methyltransferases of Toxoplasma gondii. *Int J Biol Sci* **13**, 458-470, doi:10.7150/ijbs.18644 (2017).

22 Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4**, e06205, doi:10.7554/eLife.06205 (2015).

23 Greer, E. L. *et al.* DNA Methylation on N6-Adenine in C. elegans. *Cell* **161**, 868-878, doi:10.1016/j.cell.2015.04.005 (2015).

24 Ivens, A. C. *et al.* The genome of the kinetoplastid parasite, Leishmania major. *Science (New York, N.Y.)* **309**, 436-442, doi:10.1126/science.1112680 (2005).

25 Cuypers, B. *et al.* Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method for RNA-seq in Trypanosomatids. *Scientific Reports* **7**, 3725, doi:10.1038/s41598-017-03987-0 (2017).

26 Dumetz, F. *et al.* Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio* **8**, doi:10.1128/mBio.00599-17 (2017).

27 Biyani, N. & Madhubala, R. Quantitative proteomic profiling of the promastigotes and the intracellular amastigotes of Leishmania donovani isolates identifies novel proteins having a role in Leishmania differentiation and intracellular survival. *Biochimica et biophysica acta* **1824**, 1342-1350, doi:10.1016/j.bbapap.2012.07.010 (2012).

28 Militello, K. T. *et al.* African Trypanosomes Contain 5-Methylcytosine in Nuclear DNA. *Eukaryotic Cell* **7**, 2012-2016, doi:10.1128/ec.00198-08 (2008).

29 Bateman, A. *et al.* The Pfam protein families database in 2019. *Nucleic acids research* **47**, D427-D432, doi:10.1093/nar/gky995 (2018).

30 Berriman, M. *et al.* The genome of the African trypanosome Trypanosoma brucei. *Science (New York, N.Y.)* **309**, 416-422, doi:10.1126/science.1112642 (2005).

31 Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic acids research* **38**, D457-462, doi:10.1093/nar/gkp851 (2010).

32 Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic acids research* **37**, D539-543, doi:10.1093/nar/gkn814 (2009).

33 Auburn, S. *et al.* A new Plasmodium vivax reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. *Wellcome open research* **1**, 4-4, doi:10.12688/wellcomeopenres.9876.1 (2016).

34 Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498-511, doi:10.1038/nature01097 (2002).

35 Puiu, D., Enomoto, S., Buck, G. A., Abrahamsen, M. S. & Kissinger, J. C. CryptoDB: the Cryptosporidium genome resource. *Nucleic acids research* **32**, D329-331, doi:10.1093/nar/gkh050 (2004).

36 Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, Cryptosporidium parvum. *Science (New York, N.Y.)* **304**, 441-445, doi:10.1126/science.1094786 (2004).

37 Xu, P. *et al.* The genome of Cryptosporidium hominis. *Nature* **431**, 1107-1112, doi:10.1038/nature02977 (2004).

38 Gajria, B. *et al.* ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic acids research* **36**, D553-556, doi:10.1093/nar/gkm981 (2008).

39 Lorenzi, H. *et al.* Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. *Nature communications* **7**, 10147, doi:10.1038/ncomms10147 (2016).

40 Ebenezer, T. E. *et al.* Transcriptome, proteome and draft genome of Euglena gracilis. *BMC Biology* **17**, 11, doi:10.1186/s12915-019-0626-8 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

41 Lorenzi, H. A. *et al.* New assembly, reannotation and analysis of the Entamoeba histolytica genome reveal new genomic features and protein content information. *PLoS Negl Trop Dis* **4**, e716, doi:10.1371/journal.pntd.0000716 (2010).

42 Wood, V. *et al.* The genome sequence of Schizosaccharomyces pombe. *Nature* **415**, 871-880, doi:10.1038/nature724 (2002).

43 Galagan, J. E. *et al.* The genome sequence of the filamentous fungus Neurospora crassa. *Nature* **422**, 859-868, doi:10.1038/nature01554 (2003).

44 Sebaihia, M. *et al.* Genome sequence of a proteolytic (Group I) Clostridium botulinum strain Hall A and comparative analysis of the clostridial genomes. *Genome research* **17**, 1082-1092, doi:10.1101/gr.6282807 (2007).

45 Hoskins, J. *et al.* Genome of the bacterium Streptococcus pneumoniae strain R6. *Journal of bacteriology* **183**, 5709-5717, doi:10.1128/jb.183.19.5709-5717.2001 (2001).

46 Huang, Y. Y. *et al.* Complete Genome Sequence of Agrobacterium tumefaciens Ach5. *Genome announcements* **3**, doi:10.1128/genomeA.00570-15 (2015).

47 Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-852, doi:10.1038/35101607 (2001).

48 Murat, C. *et al.* Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle. *Nature ecology & evolution* **2**, 1956-1965, doi:10.1038/s41559-018-0710-4 (2018).

49 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).

50 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

51 Glez-Peï¿½a, D., Gï¿½mez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F. & Posada, D. ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic acids research* **38**, W14-W18, doi:10.1093/nar/gkq321 (2010).

52 Tihon, E., Imamura, H., Dujardin, J. C., Van Den Abbeele, J. & Van den Broeck, F. Discovery and genomic analyses of hybridization between divergent lineages of Trypanosoma congolense, causative agent of Animal African Trypanosomiasis. *Molecular ecology* **26**, 6524-6538, doi:10.1111/mec.14271 (2017).

53 LeBowitz, J. H. Transfection experiments with Leishmania. *Methods in cell biology* **45**, 65-78 (1994).

54 Morales, M. A. *et al.* Phosphoproteome dynamics reveal heat-shock protein complexes specific to the Leishmania donovani infectious stage. *Proc Natl Acad Sci U S A* **107**, 8381-8386, doi:10.1073/pnas.0914768107 (2010).

55 Dumetz, F. *et al.* Molecular Preadaptation to Antimony Resistance in Leishmania donovani on the Indian Subcontinent. *mSphere* **3**, e00548-00517, doi:10.1128/mSphere.00548-17 (2018).

56 Guo, W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC genomics* **14**, 774, doi:10.1186/1471-2164-14-774 (2013).

57 Berardini, T. Z. *et al.* The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *genesis* **53**, 474-485, doi:doi:10.1002/dvg.22877 (2015).

58 Liao, W. W. *et al.* MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data. *BMC genomics* **16 Suppl 12**, S11, doi:10.1186/1471-2164-16-s12-s11 (2015).

59 Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer research* **77**, e31-e34, doi:10.1158/0008-5472.CAN-17-0337 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

60 Decuypere, S. *et al.* Gene expression profiling of Leishmania (Leishmania) donovani: overcoming technical variation and exploiting biological variation. *Parasitology* **135**, 183-194, doi:10.1017/s0031182007003782 (2008).

61 Cao, X. *et al.* Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. *Current Biology* **13**, 2212-2217, doi:https://doi.org/10.1016/j.cub.2003.11.052 (2003).

62 de Mendoza, A. *et al.* Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nature communications* **9**, 1341, doi:10.1038/s41467-018-03724-9 (2018).

63 Dubin, M. J. *et al.* DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife* **4**, e05255, doi:10.7554/eLife.05255 (2015).

64 Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215-219, doi:10.1038/nature06745 (2008).

65 Hsieh, T. F. *et al.* Genome-wide demethylation of Arabidopsis endosperm. *Science (New York, N.Y.)* **324**, 1451-1454, doi:10.1126/science.1172417 (2009).

66 Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science (New York, N.Y.)* **328**, 916-919, doi:10.1126/science.1186366 (2010).

67 Verma, S. *et al.* Quantification of Parasite Load in Clinical Samples of Leishmaniasis Patients: IL-10 Level Correlates with Parasite Load in Visceral Leishmaniasis. *PLOS ONE* **5**, e10107, doi:10.1371/journal.pone.0010107 (2010).

68 Cuypers, B. *et al.* Integrated genomic and metabolomic profiling of ISC1, an emerging Leishmania donovani population in the Indian subcontinent. *Infection, Genetics and Evolution* **62**, 170-178 (2018).

69 Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife* **5**, doi:10.7554/eLife.12613 (2016).

70 Olova, N. *et al.* Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol* **19**, 33-33, doi:10.1186/s13059-018-1408-2 (2018).

71 Domagalska, M. A. *et al.* Genomes of intracellular Leishmania parasites directly sequenced from patients. *bioRxiv*, 676163, doi:10.1101/676163 (2019).

72 Bewick, A. J., Vogel, K. J., Moore, A. J. & Schmitz, R. J. The evolution of DNA methylation and its relationship to sociality in insects. *bioRxiv*, 062455, doi:10.1101/062455 (2016).

73 Oyola, S. O. *et al.* Efficient depletion of host DNA contamination in malaria clinical sequencing. *Journal of clinical microbiology* **51**, 745-751, doi:10.1128/jcm.02507-12 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Tables and Figures

Tables

Table 1: CpG, CHG and CHH methylation percentages in different *Leishmania donovani* lines (Ld), *Trypanosoma brucei* and *Arabidopsis thaliana* (positive control).

	CpG (%)	CHG (%)	CHH (%)
LdPro	0.0003	0.0005	0.0126
LdAmas	0.0001	0.0003	0.0073
LdHamPro	0.0002	0.0005	0.0113
LdDNMT+	0.0013	0.0026	0.0627
LdDNMT-/-	0.0002	0.0006	0.0079
Tbrucei	0.0001	0.0006	0.0040
Athaliana	21.0473	4.0401	0.3141

Table 2: qPCR estimation of LdBPK_251230 expression level (copy number) of Ldo-Pro and Ldo-DNMToverex.

	Ldo-Pro	LdDNMT+
RNA	1.53 ±0.2	3.78 ±0.3

Table 3: Enrichment (X) of *Leishmania* DNA in artificial mixtures of *Leishmania* and human DNA for promastigotes and amastigotes of 3 clinical isolates (BPK026, BPK275 and BPK282). Enrichments were carried out with the NEBNext Microbiome DNA Enrichment Kit (NEB).

	BPK026	BPK275	BPK282	Average Enrichment (X)	St.Dev
Promastigotes	79.85	88.32	60.47	76.22	14.28
Amastigotes	64.83	56.87	63.33	61.68	4.23

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 4: Enrichment of unmethylated *Leishmania* DNA from human THP-1 cells infected with *Leishmania donovani* BPK282/0 cl4. Two independent infections were carried out and two enrichments per infection to obtain an idea of the variation at both levels (standard deviations are shown). Enrichments were carried out with the NEBNext Microbiome DNA Enrichment Kit (NEB).

	Leishmania/Human DNA ratio		Removed Human DNA (%)	Enrichment (x)
	Before Enrichment	After Enrichment		
THP-1 + BPK282/0 cl4 inf.1	0.0089	0.3976 ± 0.0048	98.15 ± 0.07	44.70 ± 0.57
THP-1 + BPK282/0 cl4 inf.2	0.0097	0.2787 ± 0.0103	97.05 ± 0.21	28.75 ± 1.06
Average				**36.73 ± 11.28**

Figures

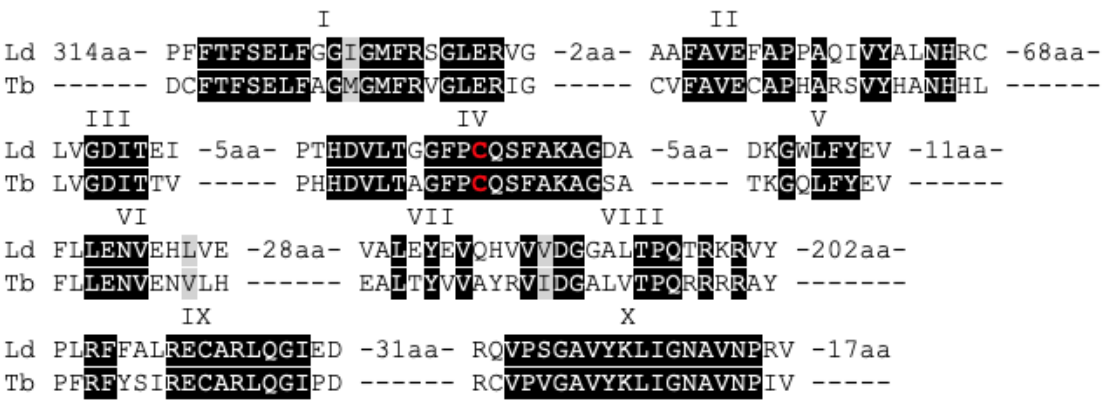

Figure 1: Protein alignment of LdDNMT (LdBPK_250018100) and TbDNMT generated with T-coffee picturing the similarities between the 10 homologous domains of C5 DNA methyltransferases. Black highlights homology and the red character displays the position of the catalytic cysteine residue.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

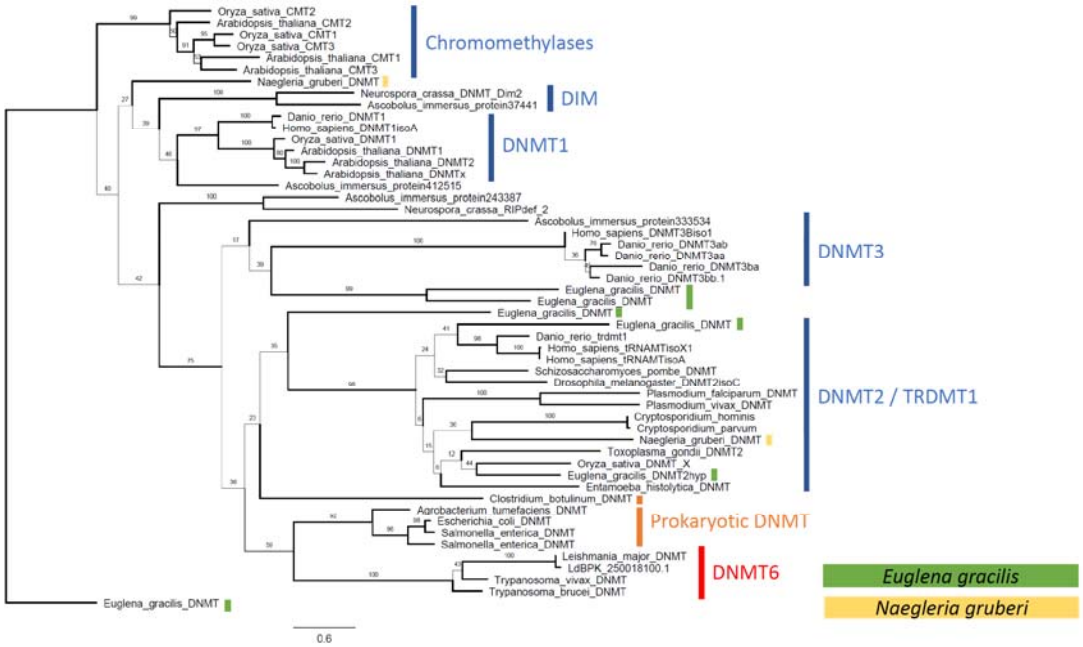

Figure 2: RAxML Maximum Likelihood tree showing the position of Trypanosomatid DNMT (DNMT 6) within the DNMT family. Displayed branch bootstrap values are based on 1000 bootstraps.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

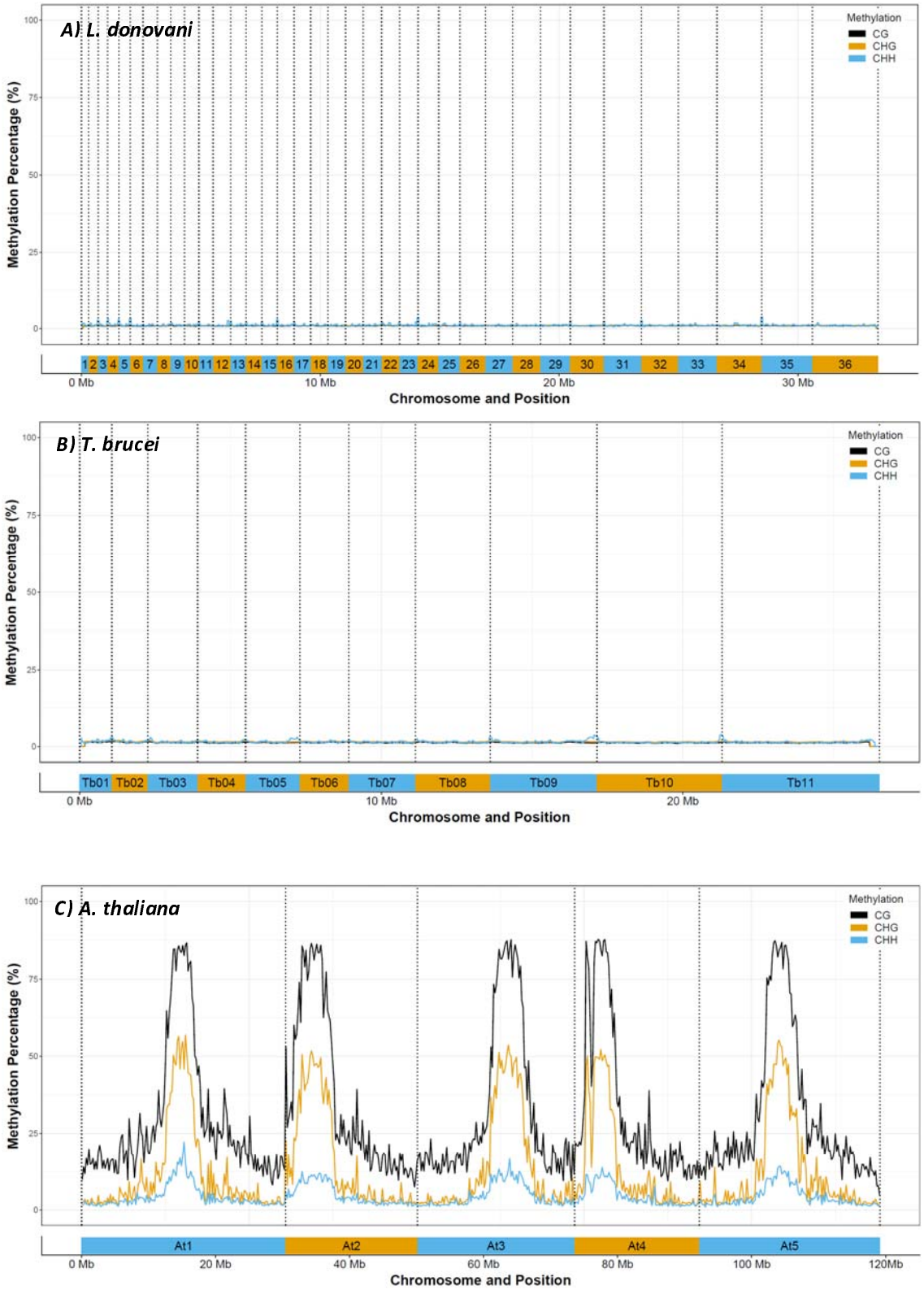

Figure 3: CpG, CHG and CHH genome-wide methylation patterns in A) *Leishmania donovani* BPK282 (36 chromosomes), B) *Trypanosoma brucei brucei* TREU927 (11 chromosomes) and C) *Arabidopsis thaliana* Col-0 (5 chromosomes). Data was binned over 10 000 positions to remove local noise and variation.

bioRxiv preprint doi: https://doi.org/10.1101/747063; this version posted August 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

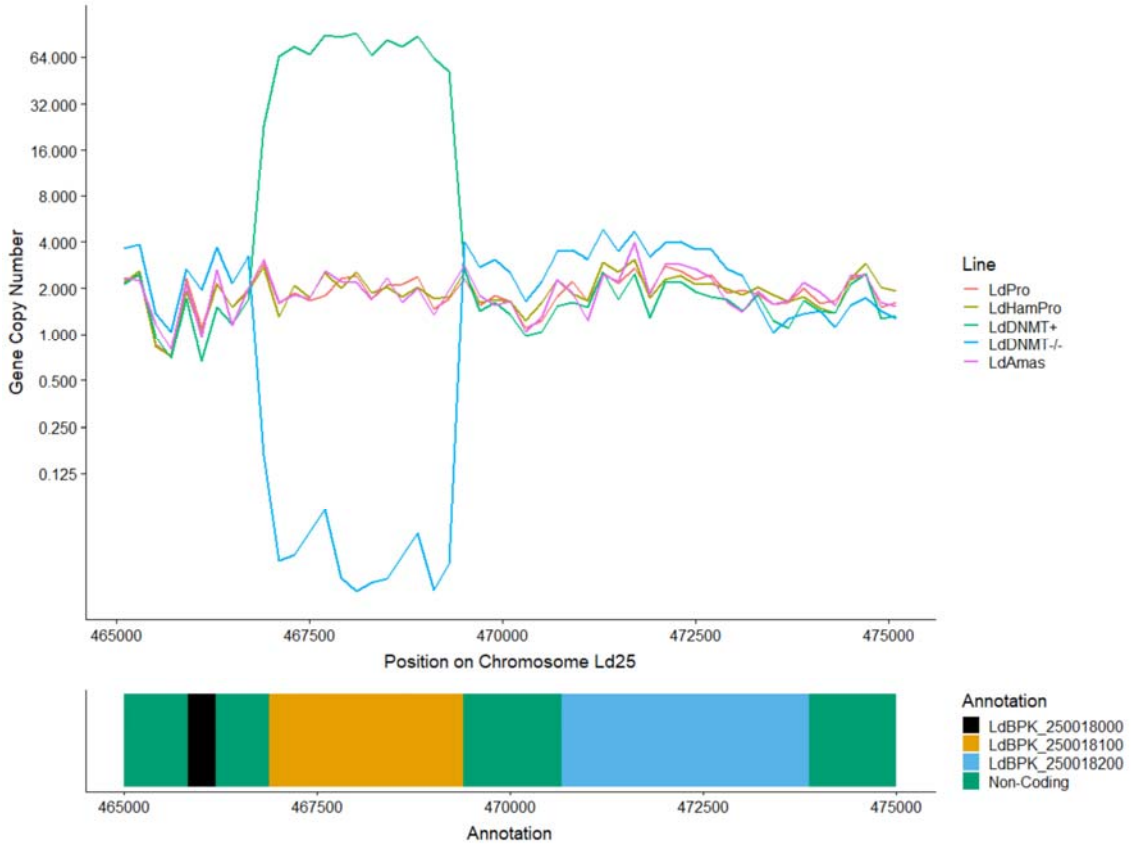

Figure 4: DNA/Gene copy number based on genomic sequencing depth on chromosome 25 position 465000-475000. Both the LdDNMT knock-out (LdDNMT-/-) and LdDNMT overexpressor line (LdDNMT+) were successful with respectively 0 and 64 copies of the gene. The plot shows also that the neighbouring genes LdBPK_250018000 and LdBPK_250018200 are unaffected and have the standard disomic pattern.

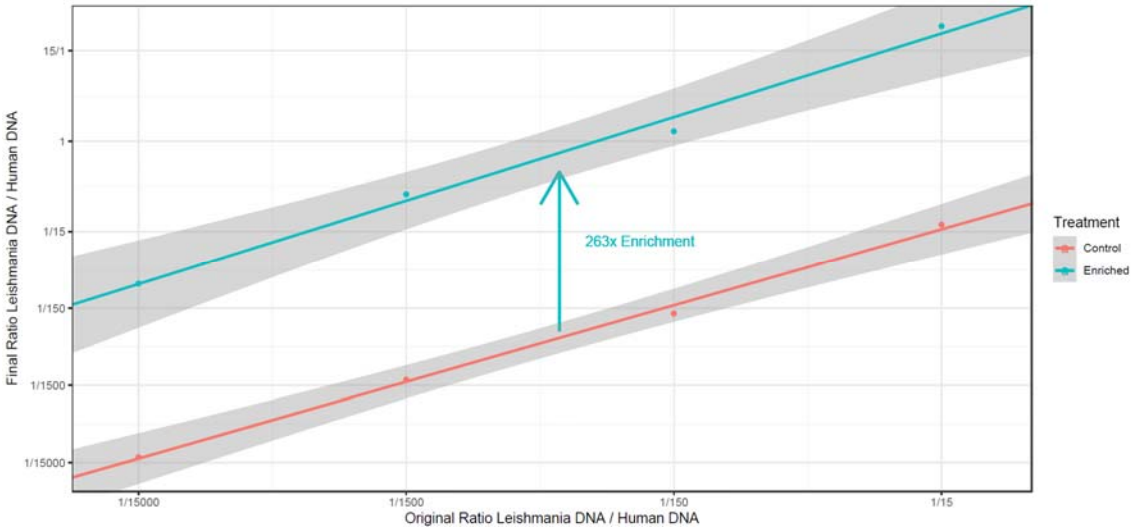

Figure 5: Enrichment (X) of *Leishmania* DNA in artificial mixtures of *Leishmania* promastigote DNA and human DNA, with the mixtures ranging from 1:15 to 1:15000 *Leishmania:*human DNA. Enrichments were carried out with the NEBNext Microbiome DNA Enrichment Kit (NEB) and the unmethylated *Leishmania* DNA was enriched on average 263 times.