Transcriptomic profiling of gene expression and RNA processing during Leishmania major differentiation

Laura A. L. Dillon¹,², Kwame Okrah³, V. Keith Hughitt¹,², Rahul Suresh¹, Yuan Li¹,², Maria Cecilia Fernandes¹,², A. Trey Belew¹,², Hector Corrada Bravo²,⁴, David M. Mosser¹ and Najib M. El-Sayed¹,²,*

¹Department of Cell Biology and Molecular Genetics, 3128 Bioscience Research Building, University of Maryland, College Park, MD 20742, USA, ²Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA, ³Department of Mathematics, University of Maryland, College Park, MD 20742, USA and ⁴Department of Computer Science, University of Maryland, College Park, MD 20742, USA

Received February 26, 2015; Revised May 21, 2015; Accepted June 15, 2015

ABSTRACT
Protozoan parasites of the genus Leishmania are the etiological agents of leishmaniasis, a group of diseases with a worldwide incidence of 0.9–1.6 million cases per year. We used RNA-seq to conduct a high-resolution transcriptomic analysis of the global changes in gene expression and RNA processing events that occur as L. major transforms from non-infective procyclic promastigotes to infective metacyclic promastigotes. Careful statistical analysis across multiple biological replicates and the removal of batch effects provided a high quality framework for comprehensively analyzing differential gene expression and transcriptome remodeling in this pathogen as it acquires its infectivity. We also identified precise 5’ and 3’ UTR boundaries for a majority of Leishmania genes and detected widespread alternative trans-splicing and polyadenylation. An investigation of possible correlations between stage-specific preferential trans-splicing or polyadenylation sites and differentially expressed genes revealed a lack of systematic association, establishing that differences in expression levels cannot be attributed to stage-regulated alternative RNA processing. Our findings build on and improve existing expression datasets and provide a substantially more detailed view of L. major biology that will inform the field and potentially provide a stronger basis for drug discovery and vaccine development efforts.

INTRODUCTION
Obligate intracellular protozoan parasites of the genus Leishmania are the causative agents of leishmaniasis, a group of diseases with a worldwide incidence of 0.9–1.6 million cases per year. The disease can vary in severity from self-healing skin lesions to disfiguring mucosal manifestations to fatal visceral disease (1). The parasite’s life cycle is divided between its mammalian host, where it resides inside of host macrophages, and its insect vector, the phlebotomine sand fly. When responding to changes in the environment as it moves through its life cycle, such as upon leaving the sand fly vector and infecting host cells, the parasite must adapt to its new surroundings. While some of these adaptations can be seen as changes in morphology (size, shape, position of organelles) and variations in cell surface component (2–4), less is known about the global changes that take place at the transcriptomic level.

Unlike most other eukaryotes, Leishmania and other trypanosomatids, including Trypanosoma brucei and Trypanosoma cruzi, do not regulate the expression levels of individual genes by the differential recruitment of RNA polymerase II influenced by cellular transcription factors. Rather, their genes are arranged as polycistronic clusters of tens to hundreds of functionally unrelated genes which are transcribed at roughly the same rate across the genome (5–7). The trans-splicing of a capped 39-nucleotide (nt) spliced leader (SL) mini-exon sequence to the 5’ end of each nuclear mRNA and the polyadenylation of the 3’ end are used to separate each polycistronic pre-mRNA transcript into its component mature mRNAs (8). A number of other organisms, ranging from dinoflagellates to nematodes to chordates, exhibit evidence of the trans-splicing of an SL sequence to at least a subset of their genes (9–13).

In trypanosomatids, transcription initiation sites occur at divergent ‘strand switch regions’ where polycistronic units originate in opposite directions on opposing DNA strands (7,14,15). Trans-splicing and polyadenylation events are coupled temporally and spatially such that the SL acceptor site of the downstream gene determines the location of the polyadenylation site of the upstream gene.
and both modification events occur simultaneously during post-transcriptional processing (16–18). SL acceptor sites contain a consensus AG dinucleotide that is preceded by polypyrimidine-rich sequence and a G nucleotide excluded from the −3 position (16,17,19). Polyadenylation sites do not appear to contain a specific signal sequence and have been reported to occur about 500–600 nt upstream of the coupled trans-splicing acceptor site (16).

Steady-state mRNA levels for individual genes are largely dependent on gene copy number and the rate of mRNA degradation, with mRNA deadenylation preceding degradation for most mRNAs. Sequence motifs contained in the 3′ UTRs greatly influence mRNA stability and the recruitment of the cellular degradation machinery (20–30). Since kinetoplastids lack introns (with very few exceptions), they do not control gene expression by alternative cis-splicing (6,31). Gene expression is thus predominantly controlled, not at the transcriptional level through the developmental regulation of RNA polymerase II activity, but by gene copy number, post-transcriptional mRNA processing, rates of mRNA degradation and translational efficiency (see (32) for review).

The genome sequences of Leishmania major, T. brucei, and T. cruzi were completed in 2005 (6,33,34), yet much remains unknown about the boundaries of individual genes and the mechanisms directing the expression levels of individual genes. Most previous studies examining Leishmania gene expression have relied on SAGE tags or on microarrays (35–44). While very informative, microarray-based approaches have several inherent limitations such as hybridization and cross-hybridization artifacts, the restriction on genes interrogated to probes included on the array (inhibiting the identification of previously unannotated genes), dye-based detection issues, the need for large amounts of input RNA and the inability to detect 5′ and 3′ UTRs boundaries. Furthermore, comparison of results between studies has been hindered by differences in the developmental stages studied and the probes included on the microarrays. These limitations likely resulted in the identification of an incomplete list of genes that are up- or down-regulated in the various life cycle stages. RNA-seq, which enables a precise and sensitive measurement of mRNA transcript abundance, has begun to be applied to this problem (45), and additional, comprehensive, well-replicated studies examining gene expression across multiple conditions are needed to more fully understand both the gene expression signatures of individual developmental stages and the changes that take place as the parasite progresses through its life cycle.

In this study, we performed transcriptome profiling using RNA-seq to identify global changes in gene expression that occur as L. major undergoes metacyclogenesis from the proliferative, non-infective procyclic promastigote form to the non-dividing, infective metacyclic promastigote form, a developmental progression that is well mimicked in vitro using reliable axenic cultivation methods (46). Differential gene expression analysis enabled us to distinguish between the procyclic promastigote and metacyclic promastigote developmental forms and shed light on how the parasite alters gene expression as it achieves infectivity. We precisely identified the 5′ and 3′ UTR boundaries for a majority of Leishmania genes and detected widespread alternative trans-splicing and polyadenylation. A paired-end mRNA sequencing approach was used to allow high confidence read mapping and transcript assembly. Collection of data from multiple biological replicates, careful statistical analysis of variation and removal of batch effects provided us with a unique ability to detect biological differences between the two developmental stages with enhanced confidence and sensitivity. The resources generated by this work will build on and improve existing expression datasets and gene structure annotations and provide a substantially more detailed interpretation of L. major biology that will inform the field and potentially provide additional data for drug discovery and vaccine development efforts.

MATERIALS AND METHODS

Leishmania culture

Leishmania major (clone V1, MHOM/IL/80/Friedlin) was isolated after passage through BALB/c mice. Promastigotes were grown in 50% M199 39% Schneider medium along with 10% Fetal Bovine Serum (FBS) and 1% of Penicillin/streptomycin at 25°C. L. major promastigotes were not split for more than five passages to maintain virulence of the cultures. Enrichment for metacyclic promastigotes from stationary phase cultures was done by Ficoll density gradient centrifugation (47) and by negative selection with peanut agglutinin (PNA) (48). Six biological replicates, each originating from a separate growth and obtained on different days, were collected for procyclic promastigotes and Ficoll-enriched metacyclic promastigotes. For three of the growths, metacyclic promastigotes were also collected following PNA selection.

RNA isolation and cDNA library preparation

Total RNA was isolated using the Trizol® reagent (Invitrogen, CA, USA), treated with DNase and purified using the Qiagen RNeasy mini kit. RNA integrity was assessed using an Agilent 2100 bioanalyzer. Poly(A)⁺-enriched cDNA libraries were generated using the Illumina TruSeq Sample Preparation kit (San Diego, CA, USA) and checked for quality and quantity using the bioanalyzer and qPCR (KAPA Biosystems).

RNA-seq data generation, pre-processing and quality trimming

Paired end reads (100 bp) were obtained from the Illumina HiSeq 1500 platform. Trimmomatic (49) was used to remove any remaining Illumina adapter sequences from reads and to trim bases off the start or the end of a read when the quality score fell below a threshold of 20. Sequence quality metrics were assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Mapping cDNA fragments to the reference genome, abundance estimation and data normalization

 Reads were aligned to the L. major genome (v. 6.0) obtained from the TriTrypDB database (www.tritrypdb.org)
Data quality assessment by statistical sample clustering and visualization

Multiple approaches were used to evaluate replicates and to visualize sample-sample distances. Those included Pearson correlation, median pairwise correlation (MPC) analysis, box plots, principal component analysis (PCA) and Euclidean distances-based hierarchical clustering. Samples that did not pass the following quality assessment procedure were excluded from downstream analyses. For each sample, the MPC to all other samples in the dataset was computed (Supplementary Figure S1C). A standard outlier identification method (52) was applied to remove samples that have low correlation with the other samples. Samples were removed if their median MPC was less than Q1 (MPC)–1.5 IQR (MPC) where Q1 (MPC) and IQR (MPC) represent the first quartile and inter-quartile range of the MPC across all samples, respectively. Two samples from a single sequencing batch (batch A) were removed as a result.

Differential expression analysis

Non-expressed and weakly expressed genes, defined as having \( <1 \) read per million in \( n \) of the samples, where \( n \) is the size of the smallest group of replicates (here \( n = 5 \)), were removed prior to differential expression (DE) analysis (53). A quantile normalization scheme was applied to all samples (54). Following log2 transformation of the data, limma (a Bioconductor package) was used to conduct DE analyses. limma utilizes a standard variance moderated across all genes using a Bayesian model and produces \( P \)-values with greater degrees of freedom (55). The voom module was used to transform the data based on observational level weights derived from the mean-variance relationship prior to statistical modeling (56) (Supplementary Figure S2). Experimental batch effects were adjusted for by including experimental batch as a covariate in the statistical model (57). Differentially expressed genes were defined as genes with a Benjamini–Hochberg multiple-testing adjusted \( P \)-value of \(<0.05\).

Gene Ontology (GO) analysis

Enriched Gene Ontology (GO) categories were identified using the GOseq package in R (58). GOseq was developed specifically to account for transcript length bias in GO analyses using RNA-seq data. Two gene sets were input separately into GOseq: all genes previously identified as up-regulated in metacyclic promastigotes and all genes previously identified as downregulated in metacyclic promastigotes (see Dataset S2). A \( P \)-value cut-off of \(<0.05\) was used.

Trans-splicing site detection and 5′ UTR analysis

Sequences from each sample were mapped to the \( L.\ major \) genome (v.6.0) using TopHat (v 2.0.10) (50). Only one mismatch per read was allowed and the paired reads were required to be mapped for an alignment to be reported. Reads that did not align to the genome were retained to form a pool of candidate SL-containing reads. These reads were filtered to keep only those containing at least four bases of the end of the \( L.\ major \) SL sequence (AACTAACGCTAT-TATTGATACAGTTTCTGTACTATATTG) or its reverse complement. This target sequence (or its reverse complement) was trimmed from the reads and TopHat was used to align the remaining portions to the \( L.\ major \) genome. Two mismatches per read were allowed and reads were assigned only to a single locus of the gene model annotations provided to TopHat (containing previously annotated genes and novel ORFs). The alignment coordinates of the trimmed reads were used to retrieve the exact locations of the putative trans-splicing sites. The genomic sequence neighboring each putative site was compared against the portion of the read that was removed. Reads for which the trimmed portion (4–39 nt) did not differ by at least two bases from the corresponding genomic sequence were treated as false hits and discarded. Putative sites that were located within a previously annotated coding sequence (CDS) (from TriTrypDB, version 6.0) or within a novel ORF (as defined above) or those with no such feature within 7500 nt downstream of the site were excluded. Trans-splicing sites that remained were assigned to the nearest downstream feature. The length of the 5′ UTR was defined as the distance between the trans-splicing site and the start of the CDS/ORF to which it was assigned. Splice acceptor sites were identified for each gene by extracting the dinucleotide sequence in the genome upstream of each detected trans-splicing site using a custom Python script. Sequence composition was plotted using WebLogo version 3.3 (59). The trans-splicing site detection pipeline was written in Python and made use of the Ruffus pipeline software framework (60) and Biopython library (61). Data visualization was done using ggplot2 (62).

Polypyrimidine tract characterization

A custom Python script was used to scan a window of 250 nt upstream of each primary trans-splicing site to identify the corresponding polypyrimidine (polyPy) tract. A polyPy tract was defined as the longest stretch of sequence consisting of pyrimidines, allowing interruption by no more than a single purine.

Polyadenylation site detection and 3′ UTR analysis

Identification of the polyadenylation sites was done using a process similar to the one used for trans-splicing site detection. The initial filtering step performed on unmapped
reads identified reads containing at least 4 nt of thymine or at least 4 nt of adenine residues. This target sequence was trimmed from the reads and TopHat was used to align the remaining portions of the reads to the \textit{L. major} genome. Two mismatches per read were allowed and reads were assigned only to a single locus in the gene model annotations provided to TopHat (containing previously annotated genes and novel ORFs). The alignment coordinates of the trimmed reads were used to retrieve the exact locations of the putative polyadenylation sites. The sequence neighboring each putative site was compared against the portion of the read that was removed. Any reads for which the trimmed portion (4+ nt) did not differ by at least two bases from the corresponding genomic sequence were treated as false hits and discarded. Putative sites that were located within a previously annotated CDS (from TriTrypDB, version 6.0) or within a novel ORF (as defined above) or those with no such feature within 7500 nt upstream of the site were excluded. Polyadenylation sites that remained were assigned to the nearest downstream feature (CDS or novel ORF). The length of the 3’ UTR was defined as the distance between the stop of the CDS/ORF and the polyadenylation site. Sequence composition was plotted using WebLogo version 3.3 (59). The polyadenylation site detection pipeline was written in Python and made use of the Ruffus pipeline software framework (60) and Biopython library (61). Data visualization was done using ggplot2 (62).

**Alternative RNA processing site analysis**

Counts of \textit{trans}-splicing sites or polyadenylation sites were combined from biological replicates for each developmental stage (procyclic promastigotes and metacyclic promastigotes). The site with the largest number of reads mapped was defined as ‘primary’ for each of the developmental stages. All other sites were considered to be ‘minor’ with the most utilized of the minor sites designated as the ‘secondary’ site. The ratio of reads mapping to the primary site to those mapping to the secondary site (P/S) for a given gene was used to determine the dominance (preference) of the primary site for that gene.

**Data access**

Sequence data are available at the NCBI sequence read archive (SRA) under accession numbers SRR1460763-SRR1460775. All components of the data quality assessment statistical pipeline, named ccbSEQ, were done in R and can be accessed on GitHub (https://github.com/kokrah/ccbSEQ). The code used for the \textit{trans}-splicing and polyadenylation pipelines is freely available at https://github.com/elsayed-lab/utr_analysis. The code used to determine alternative RNA processing sites is available at https://github.com/elsayed-lab/l-major_alternate_acceptor_site_usage.

**RESULTS**

**Experimental design**

Transcriptome profiling by RNA-seq was used to identify global changes in gene expression as \textit{L. major} achieves infectivity. RNA was isolated from cultured \textit{L. major} grown to log phase (procyclic form) or enriched for metacyclic forms using: (i) a Ficoll gradient or (ii) negative selection using PNA. These two methods for metacyclic promastigote enrichment were used to test whether different methods for the procurement of metacyclic parasites could be responsible for different findings in previous studies (35,37,39,43). PolyA enriched cDNA libraries were generated using the Illumina TruSeq protocol and 100-bp paired end sequences were generated. A total of six procyclic promastigote biological replicates and nine metacyclic promastigote biological replicates were collected (Supplementary Table S1). Each procyclic replicate was matched to one or two metacyclic replicates from the same batch/expansion of cells. Phases contrast images, promastigote sample quantification and an infectivity curve for the parasites in murine macrophages are provided in Supplementary Figure S3.

A total of \(\sim1.1\) billion sequence reads were produced across the 15 samples, 91% of which mapped to the \textit{L. major} reference genome (Supplementary Table S1). For each sample, the number of reads mapping to existing gene annotations was determined. The resulting count table was restricted to the 8486 protein-coding genes in the TriTrypDB \textit{L. major} annotation v. 6.0.

**Statistical evaluation of biological replicates and batch effects**

We used multiple robust statistical methods to evaluate the global characteristics of samples and to identify outlier samples that should be removed prior to DE and gene structure analysis (Supplementary Figure S1). Box plots were used to compare the distribution of per-gene read counts within each sample. All 15 samples showed a similar distribution of these counts with median steady-state expression levels of \(\sim7.2\) log2 counts per million and very few genes (5–10 per sample) expressed at levels of <4 counts per million. This observation is consistent with a lack of gene regulation at the level of transcription and may indicate that very few protein-coding transcripts are completely degraded following polycistronic transcription. A heatmap of Pearson correlations was used to visualize the relationship between each pair of samples. While all samples showed a pairwise correlation (r) of at least 0.85, samples prepared on one experimental date (batch A) were less correlated to samples from other batches, which largely showed r-values of >0.95 when compared to one another. MPC was also computed to assess global correlation between samples and a standard outlier identification method was applied to establish a cutoff for the identification of outliers. Consistent with observations from the Pearson correlation heatmap analysis, this method identified the two samples from batch A as outliers. These two samples were excluded from further analyses.

The dataset used for DE analysis was further restricted to genes expressed at a level of at least 1 read per million in at least 5 of the 13 remaining samples. Of the 8486 protein-coding genes analyzed, 8475 met this threshold, consistent with observations described above that few genes were completely degraded after transcription. No statistical difference was found in protein coding gene expression between metacyclic promastigote samples prepared using the Ficoll or PNA protocols. Consequently, all metacyclic promastig-
ote samples were pooled together for the remainder of the analyses.

The large number of biological replicates used for the analysis necessitated the evaluation of the dataset for batch effects. A batch effect represents experimental variation caused by sub-groups of measurements that are independent of the underlying biology of the system being studied. They have been shown to introduce unwanted variability into biological studies and confound the results, leading to erroneous conclusions. Previous analyses of high-throughput data, like those produced by RNA-seq, have indicated the need to assess and correct batch effects (57). In this study, we used experimental start date as a surrogate for batch when testing for DE between developmental stages of *L. major*.

PCA and Euclidean distance heatmap analysis were used to visualize the relationship between samples both prior to (Supplementary Figure S4) and after (Figure 1) accounting for batch effects. PCA reduces the dimensionality of a dataset while allowing variability to be represented to the greatest extent possible (63). The PCA plots showed the first two principal components, which account for the greatest percent of variability in the data, on the X and Y axes, respectively, with each of the 13 samples represented as a single point. When batch was accounted for, a clear separation between procyclic promastigote and metacyclic promastigote samples was seen along the X axis of the PCA plot (Figure 1A). Separation between the stages was not as pronounced when batch was not considered (Supplementary Figure S4A). Indeed, prior to accounting for batch effects, 25% of the variance represented by PC1 and 77% of the variance represented by PC2 were attributable to the batch of the samples. Likewise, when Euclidean distance between samples was computed and used to create a heatmap color image and dendrogram depicting the closeness between samples, a clear separation between procyclic promastigote and metacyclic promastigote samples was observed after accounting for batch effects (Figure 1B) but not before (Supplementary Figure S4B). As a result of these analyses, batch effects were controlled for in the subsequent DE analysis by including experimental batch in the statistical model used by limma.

**Identification of genes differentially expressed between developmental stages**

DE analysis identified 3138 genes that were expressed at significantly different levels between procyclic and metacyclic promastigotes at an adjusted *P*-value cutoff of <0.05 (Dataset S2). Fold change differences ranged from 3.1-fold downregulated to 3.6-fold upregulated in metacyclic promastigotes. These genes were visualized using an MA plot showing the relationship between mean expression and fold change for each gene (Figure 2). Almost 60% of the DE genes (1829 of 3138) are annotated as hypothetical proteins. The remaining gene products have been characterized to different extents, albeit not always in the context of their possible role(s) in metacyclogenesis.

We extended our DE analysis to a set of 1044 novel ORFs of at least 90 nt in length identified based on evidence of translation in *L. major* by ribosome profiling data (Dataset S1). The addition of these ORFs lead to an increase in the list of differentially expressed genes by ~12% (a total of 368, from 3138 to 3506) with fold changes ranging from 5.7-fold downregulated to 3.6-fold upregulated in the metacyclic stage (Dataset S3). The top 25 down- and upregulated genes are shown in Table 1. Of these, 8 are novel ORFs, including the most downregulated gene.

The list of DE genes was used as input into GO analysis to identify cellular functions and processes that are enriched during *L. major* metacyclogenesis. Genes downregulated in
metacyclic promastigotes were considered separately from upregulated genes. Forty GO categories were identified as being significantly enriched (P-value cutoff of <0.05) for genes downregulated (33 categories) and upregulated (7 categories) in metacyclic promastigotes (Table 2 and Dataset S4).

### Examination of differentially expressed gene lists and gene ontology-based enrichment analyses

Many novel genes were identified among the most downregulated during metacyclogenesis, including multiple genes with unknown function. GO enrichment analysis of these genes reflected a clear reduction in a number of cellular processes including DNA replication and nucleosome assembly, translation-related activities (initiation and elongation), protein metabolism and energy metabolism (i.e. adenosine...
Table 2. Gene ontology (GO) categories enriched during the procyclic to metacyclic transition

| GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID       | GO ID�
in *L. major* metacyclic promastigotes in our study, as were cyclin A and DOT1, which are both involved in cell cycle progression (71,72).

The top upregulated gene, ascorbate peroxidase, is protective against both endogenous and exogenous H2O2 and appears to play a role in differentiation to the metacyclic form as well as in protecting the cell against oxidative stress-induced apoptosis (73). Other genes that were upregulated in metacyclic promastigotes include casein kinase 1, a Ser/Thr protein kinase that exists in multiple isoforms and has been identified as playing a role in *Leishmania* infectivity (74), and *meta1*, which encodes a protein that localizes in the region of the flagellar pocket of stationary phase promastigotes and is thought to play a role in virulence, potentially through altering secretory processes (75,76). The *p27* gene, which encodes a mitochondrial membrane protein that is an important component of the cytochrome oxidase complex, was also more abundantly expressed in metacyclic promastigotes. This result is consistent with previous findings reporting its upregulation in both metacyclics and intracellular amastigotes and its role in promoting parasite survival and virulence in the host (77) and supports the hypothesis that metacyclic promastigotes are pre-adapted to survival within the mammalian host (78). Finally, two known differentiation markers of metacyclic promastigotes, SHERP and HASPB (79–81), were also identified in this analysis, with SHERP upregulated ∼1.9-fold and HASPB upregulated ∼2.3-fold in metacyclic promastigotes.

The results of the DE analysis were compared to the list of differentially expressed genes identified in an earlier study by Saxena *et al.* (43) that used microarrays of PCR-amplified fragments from genomic survey sequence (GSS) clones. Only GSS clones whose 5′ and 3′ sequences could be mapped to the same gene in the *L. major* Friedlin genome sequence (31 in total) were considered in our comparison and 19 of the corresponding genes showed a similar DE trend, albeit to varying degrees and levels of significance. Given the disparate platforms, the level of agreement was reasonable.

**Identification of transcript boundaries**

Deep sequencing of *L. major* procyclic and metacyclic promastigote samples by RNA-seq presented an opportunity to comprehensively annotate transcript boundaries, thereby enhancing the structural annotation of *L. major* genes. We exploited the signal sequences generated by trans-splicing and polyadenylation events to accurately map the 5′ and 3′ UTR boundaries of transcripts by comparing reads containing these signals to the reference genome sequence. Since UTRs are expected to contain motifs that direct the post-transcriptional regulation of individual mRNAs—including degradation, storage and translation rate—determining transcript boundaries is very important for understanding gene regulation in the parasite.

Distinct transcript boundaries were determined for a large majority of previously annotated protein-coding genes and novel genes for which there was evidence of translation by ribosome profiling (Dataset S1). To do this, RNA-seq reads which did not map to the *L. major* genome due to RNA processing events were examined separately for evidence of SL sequence and a polyA tail. Of the ∼960 million reads from the 13 *L. major* samples, ∼3.9% contained evidence of trans-splicing and ∼0.05% contained evidence of polyadenylation (Supplementary Table S1). Once the SL and polyA sequences were removed, the remainders of the reads were mapped to the genome, allowing the identification of coordinates for at least one trans-splicing site for 8981 genes (94.2% of a total of 9530 genes) and at least one polyadenylation site for 8841 genes (92.8%). The coordinates of all identified trans-splicing and polyadenylation sites are provided in Dataset S5.

A sampling of the trans-splicing and polyadenylation sites identified here was compared to existing data in TriTrypDB (Peter Myler’s group, Seattle Biomed) that were generated using an RNA-seq method that specifically enriched for SL-containing sequences (biological sample type unknown). Our trans-splicing site data were highly concordant with these previously reported data. This high degree of agreement is remarkable given the differences in sample type, culture and preparation across different labs and may potentially indicate that the usage of trans-splicing sites in *Leishmania* is fairly consistent across various biological conditions. The observed variability is likely attributable to the differences in coverage, RNA-seq approach and data analysis methodology. Our polyA site data did not generally match the existing data on TriTrypDB down to the specific nucleotide. This could be due to the extreme heterogeneity of these sites (previously reported for *T. brucei* (19)), differ-
Figure 3. Length and position distribution of gene structure components in *Leishmania major*. Data from procyclic and metacyclic promastigote samples were combined to describe the gene structure elements of *L. major*. (A) Distribution of CDS lengths. Start and stop coordinates for coding sequences for previously annotated protein-coding genes (TriTrypDB v. 6.0) and novel ORFs were used to compute CDS lengths. For genes with multiple isoforms, the first isoform listed in TriTrypDB was included in the analysis. (B) Distribution of 5′ UTR lengths. The exact *trans*-splicing sites associated with each CDS were used to determine the coordinates and lengths of 5′ UTRs. A total of 154.046 *trans*-splicing sites were identified, corresponding to 5′ UTRs ranging from 0 to 7252 nt in length. (C) Distribution of 3′ UTR lengths. An analysis of polyadenylation sites associated with each CDS was performed to determine 3′ UTR coordinates and lengths. A total of 64.331 polyadenylation sites were identified, corresponding to 3′ UTRs ranging from 0 to 7133 bases in length. (D) Distribution of polypyrimidine (polyPy) tract lengths. A window of 250 nt upstream of each primary *trans*-splicing site (8981 total) was scanned to identify its corresponding polyPy tract, defined as the longest stretch of pyrimidine residues interrupted by no more than one purine. (E) Distribution of distances between each primary *trans*-splicing site and its corresponding polypyrimidine (polyPy) tract. (F) Distribution of distances between each polyPy tract and the polyadenylation site of the upstream gene. A total of 6174 instances of a neighboring polyadenylation site and polyPy tract were identified. (G) Diagram of a ‘typical’ *L. major* genic region. The median lengths of the gene structure components of *L. major* were used to construct the structure of a ‘typical’ gene region (two genes and the intergenic region). The median values of each component are shown. Colors correspond to the features depicted in panels A–F.

ences in the biological samples studied or differences in the methods used to identify and assign sites.

**Gene structure features in *L. major***

We sought to determine the length distribution of the elements of each gene—5′ UTR, CDS and 3′ UTR—as well as the intergenic region, including the polypyrimidine (polyPy) tract, for previously annotated protein-coding genes and the novel ORFs. Start and stop coordinates for *L. major* genes were used to determine a median CDS length of 1241 nt with a range from 64 to 52 178 nt (Figure 3A). The boundaries of 5′ UTRs were defined using the coordinates of the SL addition sites and start codon annotations and a similar analysis was done to determine the lengths of 3′ UTRs using stop codon and polyadenylation site coordinates. The median length of all identified 5′ UTRs (not including the 39 nt SL sequence) and 3′ UTRs was 547 and 729 nt, respectively (Figure 3B and C). When only the most-utilized (primary) *trans*-splicing or polyadenylation site for each gene was c...
Figure 4. Characterization of alternative RNA processing. Alternative RNA processing events were detected in both promastigote developmental stages and pooled for this analysis. The distribution of distances between primary and minor trans-splicing sites is shown for minor splice sites that use (A) the canonical AG acceptor sequence or (B) an acceptor sequence other than AG. Panel (C) depicts the distribution of distances between primary and minor polyadenylation sites. Only minor sites within 1000 nt of the primary site are plotted. The full ranges for A, B and C, are $-7234$ to 5973, $-7178$ to 5876 and $-6429$ to 5993, respectively. About 18% (11 201 of 62 098), 29% (24 303 of 82 947) and 10% (7 301 of 75 487) of values fell outside of the plotted range for A, B and C, respectively. (D) Sequence composition of the region spanning from 90 nt upstream to 10 nt downstream of each primary trans-splicing site. (E) Sequence composition of the region spanning from 50 nt upstream to 50 nt downstream of each primary polyadenylation site.

considered, these values were reduced to 233 and 517 nt, respectively (see alternative RNA processing section below). The distribution of both the 5' and 3' UTR lengths was similar in both stages (Supplementary Figure S6) and there did not appear to be a correlation between CDS length and either UTR length or between corresponding UTR lengths.

The length distribution analysis was extended to examine the polyPy tract, which is known to be involved in the regulation of RNA processing events in trypanosomatids (82–84). In this analysis, the polyPy tract was identified as the longest stretch of pyrimidine residues located upstream of each of the (primary) trans-splicing sites and interrupted by no more than one purine. PolyPy tracts ranged from 7 to 123 nt in length, with a median value of 21 nt (Figure 3D) and a clear usage preference for cytosine (54%) over thymine (42%) residues. This observation, which runs counter to what has been found in related species where thymine was preferentially used (>98%) used more than one trans-splicing in at least one developmental stage. We were able to detect alternative splicing in L. major with a greater sensitivity than has been previously reported (45), presumably due to the deeper coverage of this dataset. Indeed, for genes with detectable trans-splicing events, alternative trans-splicing was pervasive with 88, 56 and 18% of genes using more than one trans-splicing in at least one developmental stage. This observation indicates significantly longer 5' and 3' UTRs and longer intergenic distances than what has been reported in either T. cruzi and T. brucei (19,86) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision), and is consistent with previous observations regarding the relative compactness of the species’ genomes (5).

Detection of alternative RNA processing events within and between developmental stages

The sequencing depth of our L. major transcriptome profiling experiments allowed not only the identification of the SL-addition and polyadenylation sites at a single-base resolution, but also the quantification of alternative RNA processing events. Of the 8981 genes with SL-addition sites detected, 8777 (~98%) used more than one trans-splicing site in at least one developmental stage. We were able to detect alternative splicing in L. major with a greater sensitivity than has been previously reported (45), presumably due to the deeper coverage of this dataset. Indeed, for genes with detectable trans-splicing events, alternative trans-splicing was pervasive with 88, 56 and 18% of genes using at least 5, 10 or 20 sites, respectively, in at least one developmental stage. This observation indicates that L. major exhibits a somewhat higher degree of alternative splicing than related species T. cruzi and T. brucei where <90% of genes were identified as alternatively spliced (19) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). This observation persisted even after accounting for differences in sequencing depth. The distribution of the distances between the primary and minor trans-splicing sites revealed that almost half (~48%) of the alternative splice sites are located within 200 bases of the primary site in either direction. Even so, a significant percentage (18%) of minor sites were observed more than 1000 bp from the primary site, with most of these (78%) occurring upstream of the primary site.
Figure 5. Preferential usage of primary site across developmental stages. (A) For each gene, the length of the 5′ UTR determined by the primary trans-splicing site for the metacyclic stage was plotted against the length of the 5′ UTR determined by the primary trans-splicing site for the procyclic stage. Each point represents a single gene. Points that do not fall along the diagonal represent a change in the primary site between the stages. The color of each point represents the dominance of the primary site (preference for usage of the primary site over other sites within a stage), as determined by the ratio of primary site read counts to secondary site read counts (P/S), averaged for both stages. The size of each point depicts average read count for the primary sites from both stages, thereby showing expression levels and providing a measure of confidence in the data. Three genes highlighted in Supplementary Figure S7 are circled with the gene identifier labeled in blue. (B) A similar plot was done for 3′ UTR length as determined by the primary polyadenylation site for each stage.

An examination of the trans-splicing sites revealed a propensity for usage of the canonical acceptor sequence (AG) at both the primary (~97%) and minor (~43%) splicing sites (Supplementary Table S2), consistent with previous findings in T. cruzi and T. brucei (19) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). A sequence composition analysis of the region upstream of the SL-addition site allowed the visualization of the tail end of the polypyrrimidine tract through the trans-splicing acceptor site (Figure 4D). As reported previously (45,87), a C nucleotide was preferred prior to the AG acceptor sequence. When minor sites that do not use the canonical AG acceptor sequence were considered, this phenomenon was largely absent and the percentage of minor sites that are downstream of the primary site drops to 36.3% (Figure 4B). This observation was maintained when procyclic and metacyclic promastigotes were considered separately.

Alternative polyadenylation sites were detected for 8391 (~95%) of the 8841 genes for which polyadenylation events were observed with 61, 21 and 5% of genes using at least 5, 10 or 20 polyadenylation sites. As observed for trans-splicing events, this observation indicates a higher degree of alternative polyadenylation in L. major than what has been reported in either T. cruzi or T. brucei where 63 and 92% of genes had detectable alternative polyadenylation, respectively (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). A sequence composition analysis was done to visualize the region surrounding the polyA-addition site. Even though no consensus motif was observed upstream of the polyadenylation site, such as the AAUAAA required for polyadenylation in higher eukaryotes, we did note an (A/G)(A/G) motif preceded by 1–2 thymines abutting the polyA addition site for both primary and minor polyadenylation sites (Figure 4E). Similar to what was found for SL addition sites above, the analysis of the distribution of the distances between primary and alternative sites revealed that ~49% of the minor polyadenylation sites were located within a 200 nt window of the primary site (Figure 4C).

Alternative trans-splicing or polyadenylation are suspected to play a role in the regulation of gene expression in L. major, but instances of regulation through alternative RNA processing between developmental stages have not been systematically identified. We sought to identify the subset of genes that change the use of their primary trans-splicing or polyadenylation sites between the procyclic and metacyclic stages and to investigate possible correlations between these changes and DE. We were specifically interested in genes that showed a strong preferential usage for the primary site over other sites within a given stage (dominance), as determined using the ratio of reads that map to the primary site to those that map to the secondary site (P/S). Of the 8797 genes that had at least one trans-
splicing site identified in both stages. 523 showed preferen-
tial usage of different primary trans-splicing sites between
the stages. We plotted the lengths of the UTRs for each
gene, as determined by the primary trans-splicing site in
each stage (Figure 5A). Each gene was represented by a sin-
gle point with the color indicating the average P/S ratio for
the two stages (thereby providing a measure of a primary
site’s dominance) and the size indicating the average num-
ber of reads mapping to that gene’s primary sites (thereby
indicating expression level and an indirectly providing con-
fidence in the data). Data points along the diagonal repre-
sent genes that did not exhibit a change in the primary trans-
splicing site between the stages. Largely, genes that had high
dominance did not exhibit a change in primary site location
between the stages, but instead used the same primary site
in both procyclic and metacyclic promastigotes. Genes that
did change primary site tended to have only a slight prefer-
eence for each stage-specific primary site. A few interesting
genes did not follow this trend and showed both a change in
primary site (location away from the diagonal), high domi-
nance (red) and high confidence/expression (large). Exam-
plses of the alternative usage of trans-splicing sites for a sub-
set of these interesting genes—LmjF.31.0710, LmjF.33.0310
and LmjF.36.3810—are depicted in Supplementary Figure
S7.

We compared the expression profiles for genes that dis-
played a change in trans-splicing site preference versus those
that did not. The proportion of differentially expressed
genes in both sets was nearly identical (39.4% for genes that
changed primary site and 37.0% for genes that did not), in-
dicating that there was no association between changes in
primary site used by individual genes and their expression
levels ($\chi^2 = 0.28$).

A similar analysis was done to assess alternative
polyadenylation between procyclic and metacyclic promas-
tigotes (Figure 5B). Strikingly, this analysis revealed that
most of the genes that showed high dominance (red) and high confidence (large) of their primary polya sites did
not exhibit a change in primary site location between the
stages. For the large number of genes that showed a change
in the primary polyA site between developmental stages
(4377 of 8337), very few reads were mapped to the primary
sites, resulting in low confidence in these data points. The
low numbers of mapped polyA-containing reads was likely
due to the extensive heterogeneity of polyadenylation sites
or the relative low coverage of polyA-containing reads map-
ing at unique sites. Of the 4377 genes that changed primary
polyadenylation sites between stages, the portion that was
differentially expressed (37.2%) was remarkably similar to
the portion that was not (38.7%, $\chi^2 = 0.15$). This lack of
systematic association with DE also supports that conclu-
sion that differences in expression levels in the samples ex-
amined by this study cannot be attributed to stage-regulated
alternative RNA processing.

DISCUSSION

The work reported here represents a comprehensive charac-
terization of the global transcriptional changes that define
the transition of the human pathogen _L. major_ from its non-
infective to human-infective forms. Through the exploita-
tion of massively parallel sequencing to detect the most sub-
tle changes in steady-state levels of mRNA, the use mul-
tiple biological replicates to derive robust statistical anal-
yses, the careful consideration of batch effects that often
confound and mask true biological effects, and the exploita-
tion of signal sequences added during RNA processing of
polycistronic pre-mRNA, we observed changes in expres-
sion profiles and identified RNA processing sites with un-
precedented depth and reliability.

The genome of _L. major_, which defined the boundaries
of CDS for the large majority of _L. major_ genes, was com-
pleted in 2005 (6). While it has been an invaluable resource
for researchers in the field, the lack of defined UTR bound-
aries has hampered the ability to look for sequence motifs
contained in the UTRs that may be involved in the regu-
lation of individual _L. major_ genes. The RNA-seq datasets
generated in this work enabled us to precisely delineate the
5’ and 3’ UTR boundaries of _L. major_ transcripts, provid-
ing a substantial additional resource for the _Leishmania_ re-
search community. In addition, we were able to evaluate
how transcript structure compares to related species, to ex-
amine trans-splicing and polyadenylation events within and
between _L. major_ developmental stages, and to assess the
possible relationship between alternative RNA processing
and gene regulation in the context of the same biological
samples.

This analysis resulted in the generation of reliable and
substantially deep lists of differentially expressed genes that
include RNAs of low abundance, which may have fallen
outside of the limits of detection in past studies. Indeed,
the individual fold changes observed were relatively mod-
est (3.1-fold downregulated to 3.6-fold upregulated in meta-
cyclics) and may have been missed except for the sensitiv-
ity afforded by the RNA-seq technology and the statisti-
cal power provided by the use of five biological replicates in
the DE analyses. The extension of the DE analysis to novel
ORFs identified from ribosome profiling data points to
some genes that were not included in the initial annotation
of _L. major_ which may have functionally important roles
in the differentiation of the parasite. These genes should be
added to the list of candidates to be included in future anal-
eses.

GO analysis using the lists of differentially expressed
genes revealed how they are collectively involved in a num-
cber of cellular processes as the parasite transforms into
its infective form. Specifically, as the parasites become
infective, processes including translation, replication and
metabolism decrease while cell signaling and stress re-
ponses increase. This observation supports previous work
on _Leishmania_ virulence (73,75–77) and specific processes
involved in metacyclogenesis, such as iron metabolism (64),
cell cycle regulation (65,66,70–72) and cell signaling (74),
while at the same time implicating large numbers of previ-
ously unannotated genes in these processes, thereby provid-
ing evidence as to their possible functions.

Since _Leishmania_ and related trypanosomatids employ
polycistronic transcription across their entire genomes,
post-transcriptional RNA processing is thought to be a
likely mechanism for regulating the mRNA expression
levels of individual genes. While we were able to detect
widespread alternative trans-splicing and polyadenylation
for the large majority of genes, the observed heterogeneity of RNA processing sites was not systematically associated with the DE of the genes that showed the alternative RNA processing. Thus, in this analysis, RNA processing of premRNA did not appear to be a driving force for determining the expression levels of individual genes.

Trans-splicing of a specific SL sequence is not itself unique to trypanosomatids, but has evolved in parallel in a range of diverse organisms including Euglenozoa and dinoflagellates, Caenorhabditis elegans and related nematodes, Platyhelminthes and primitive chordates (90). Trypanosomatids, however, are distinct because they rely on trans-splicing to express all genes transcribed by RNA polymerase II and because they do so without the use of any sequence-specific transcription factors. Indeed, it was this exclusive use of trans-splicing that enabled the precise and quantitative approach used here to define the UTR boundaries for almost all protein-coding genes. This model of transcriptional regulation differs from organisms like C. elegans in which ~70% of the genes, sometimes contained in operons of functionally-related genes, are trans-spliced (91).

While the precise role of trans-splicing is also unknown in C. elegans and other species, some evidence exists that it may play a role in responding to changes in nutrient levels (92) and that trans-spliced genes are enriched for Ca2+ homeostasis, cytoskeleton and plasma/endomembrane system function (93). Insights into RNA processing in trypanosomatids may also shed light on gene regulation in other species that rely on similar mechanisms of post-transcriptional control, even if only for a subset of genes. In summary, transcriptome profiling of two major developmental stages of L. major provided a robust set of markers for the L. major procyclic and metacyclic developmental stages, revealed genes and processes involved in the transition between stages as the parasite becomes capable of infecting mammalian host cells, provided evidence for the function of hundreds of genes of unknown function, defined precise UTR boundaries and detailed how post-transcriptional RNA processing differs between the stages. Additionally, the transcriptome profiles reported here set the stage for the construction of co-expression networks, which are useful for identifying driver mechanisms underlying co-regulation and for tentatively annotating unknown genes through guilt-by-association inferences. Our ongoing work is focused on exploring these inferences and extending our current dataset with a simultaneous interrogation of the expression profiles of the intracellular stages of L. major as well as its host (macrophage) cells in both murine and human systems. We have also initiated comparative transcriptome analyses of different Leishmania species that cause different disease outcomes. Finally, the precise definition of UTR boundaries opens up opportunities for regulatory motif analyses and comparative analyses of UTR usage across Leishmania species.

ACKNOWLEDGEMENT

All the sequencing was performed at the University of Maryland Institute for Bioscience and Biotechnology Research (IBBR) sequencing core.

Author contributions: L.A.L.D., D.M.M. and N.E.S. designed research; L.A.L.D., R.S. and M.C.F. performed experiments; L.A.L.D., V.K.H., Y.L. and N.E.S. analyzed data; K.O., V.K.H., Y.L. and H.C.B. contributed statistical and analytic tools; A.T.B. contributed data; L.A.L.D. and N.E.S. wrote the paper; N.E.S and D.M.M. conceived the project.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institutes of Health [AI094773 to N.E.S., D.M.M.]. Funding for open access charge: National Institutes of Health [AI094773 to N.E.S., D.M.M.]. Conflict of interest statement. None declared.

REFERENCES

Comparison of the effects of Leishmania major or Leishmania donovani infection on macrophage gene expression. Infect. Immun., 76, 1186–1192.


