Supplementary Figures

Comparative Transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs throughout developmental stages

Patrícia de Cássia Ruy[1¶], Natália Melquie Monteiro Teles[1¶], Rubens Daniel Miserani Magalhães[1], Felipe Freitas de Castro[1], Leandro Dias[1], Tania Paula Aquino Defina[1], Elton José Rosas de Vasconcelos[2], Peter Myler[2], Angela Kaysel Cruz[*1]

[1]Cell and Molecular Biology Department, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil.

[2]Center for Infectious Disease Research, Seattle, United States.

[¶] These authors contributed equally to this work

*Corresponding author: akcruz@fmrp.usp.br

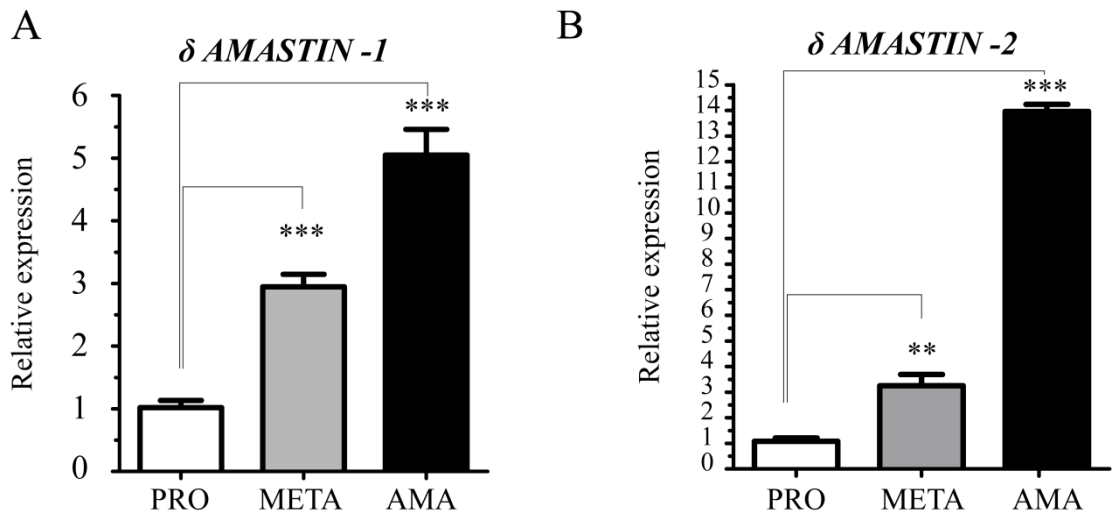
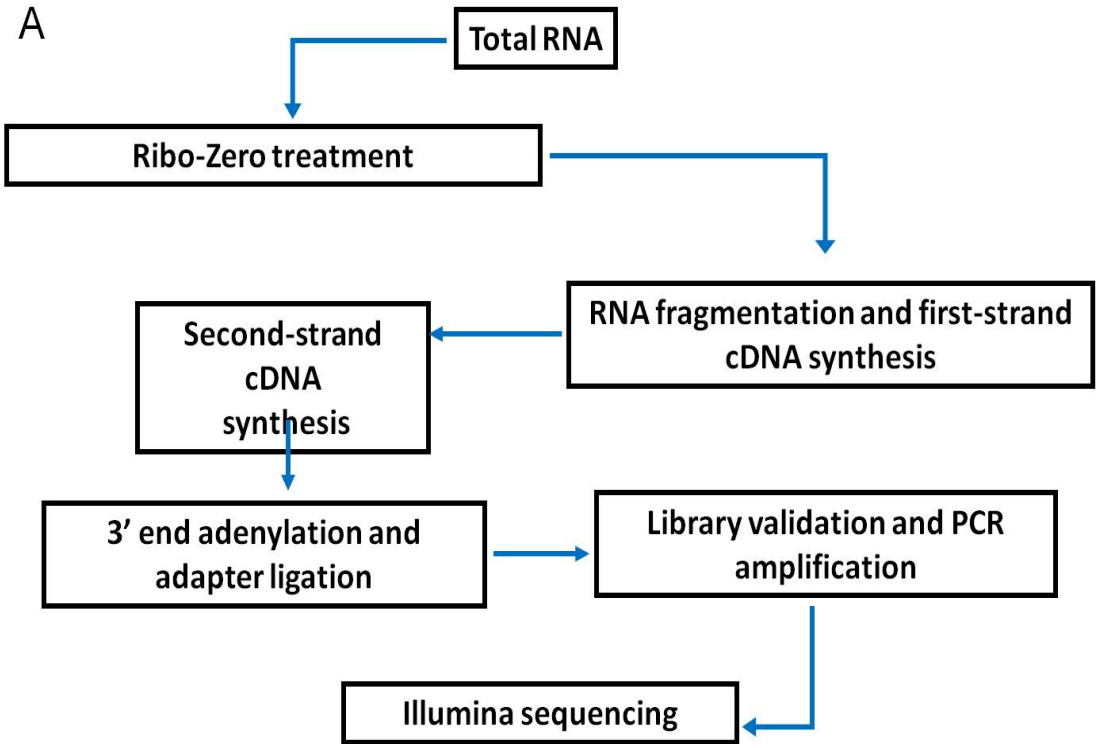
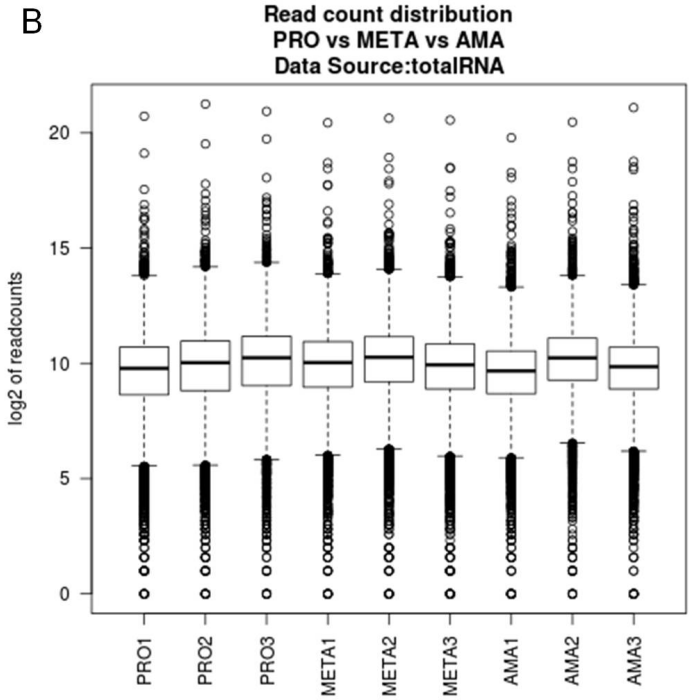


Figure S1. Comparative expression of amastigote-specific genes during the *L. braziliensis* life cycle using RT-qPCR. Two δ-amastin genes were evaluated: **A)** LbrM.08.0300 (δ AMASTIN-1) and **B)** LbrM.20.1060 (δ AMASTIN-2). The primers used are presented in Table S1. Culture-derived procyclic forms (PROs), metacyclic forms (METAs) and amastigotes (AMAs) were obtained as shown above (see Figure 1). Bars represent the means ± SD from 3 independent experiments. Student's t-test (two-tailed) was used for statistical analysis. Asterisks indicate statistically significant differences between samples, $p \leq 0.002$ (**) and $p \leq 0.0001$ (***).



A



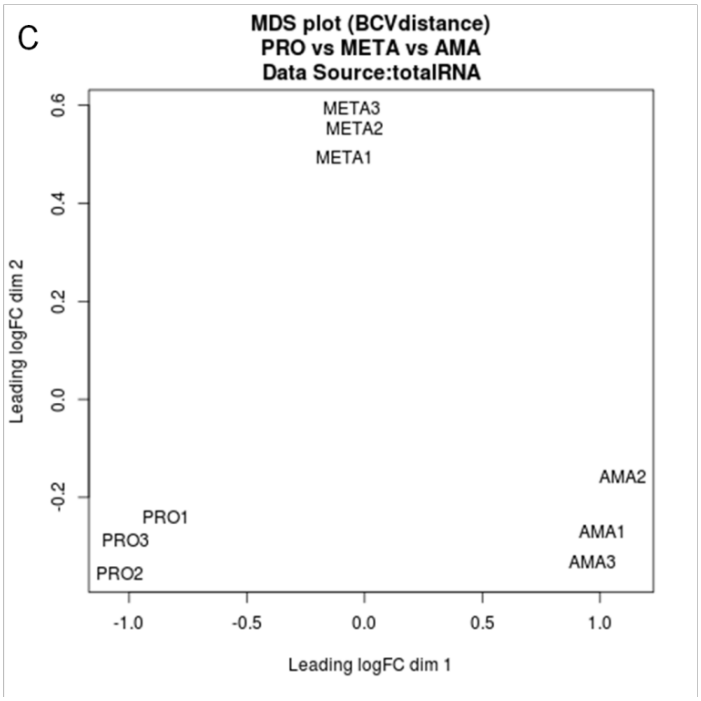


Figure S2. RNA-seq libraries information. A) Pipeline of the protocol to prepare DNA libraries for sequencing. **B)** Reads count distribution within CDSs (annotated protein coding sequences) per library**.** The x-axis depicts each replicate (1, 2 and 3) of procyclics, metacyclics and amastigotes (PRO, META and AMA), as indicated. The y-axis represents the read count (log2) per CDS. **C)** Distance between replicates of sequenced libraries. The MDS plot shows the differential expression (log fold change – logFC) between the transcripts (CDS) of each library. PRO1, PRO2 and PRO3 (procyclic replicates 1, 2 and 3, respectively); META1, META2 and META3 (metacyclic replicates

1, 2 and 3, respectively) and AMA1, AMA2 and AMA3 (amastigote replicates 1, 2 and 3, respectively).

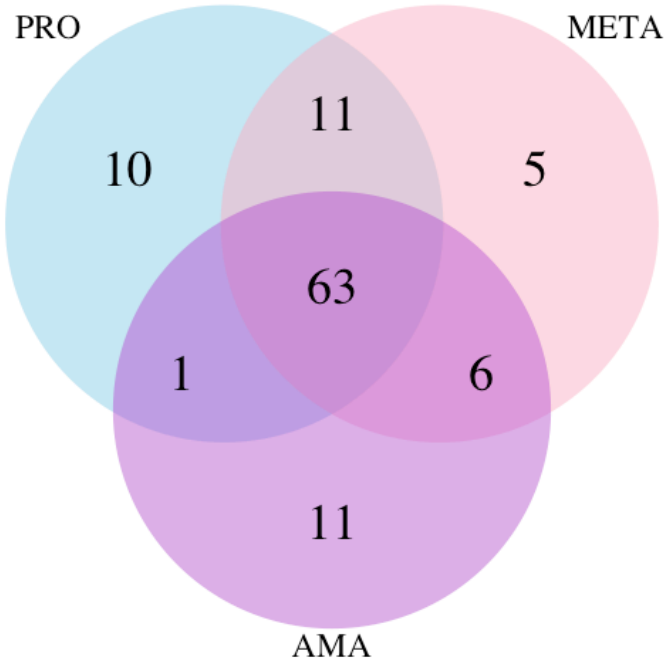
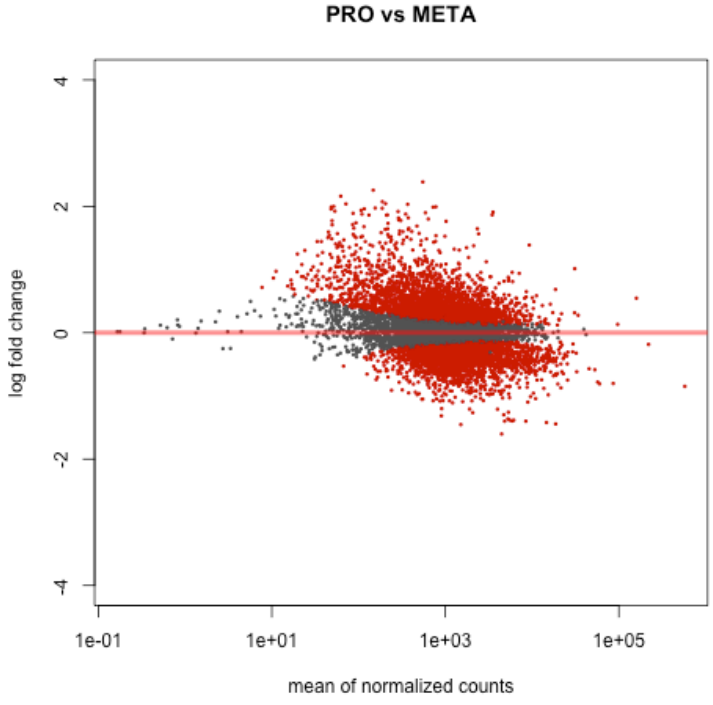
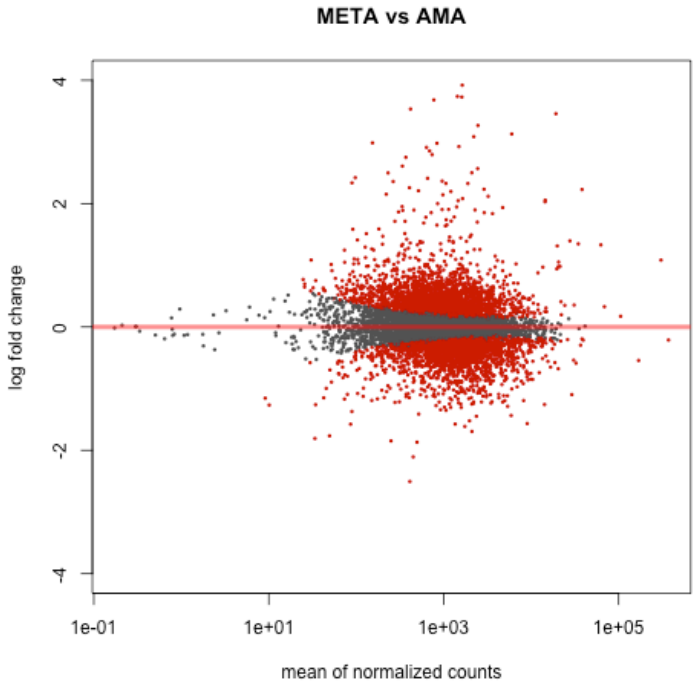


Figure S3. Highly expressed genes from each *L. braziliensis* developmental stage. The top FPKM percentile genes were considered.

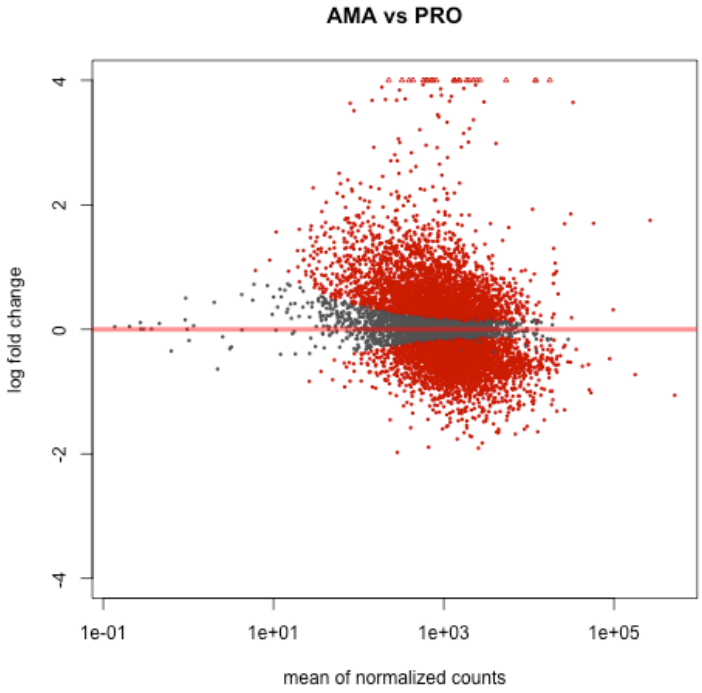
A



B


META vs AMA
log fold change
4
2
0
-2
-4
1e-01
1e+01
1e+03
1e+05
mean of normalized counts

C


AMA vs PRO
log fold change
4
2
0
-2
-4
1e-01
1e+01
1e+03
1e+05
mean of normalized counts

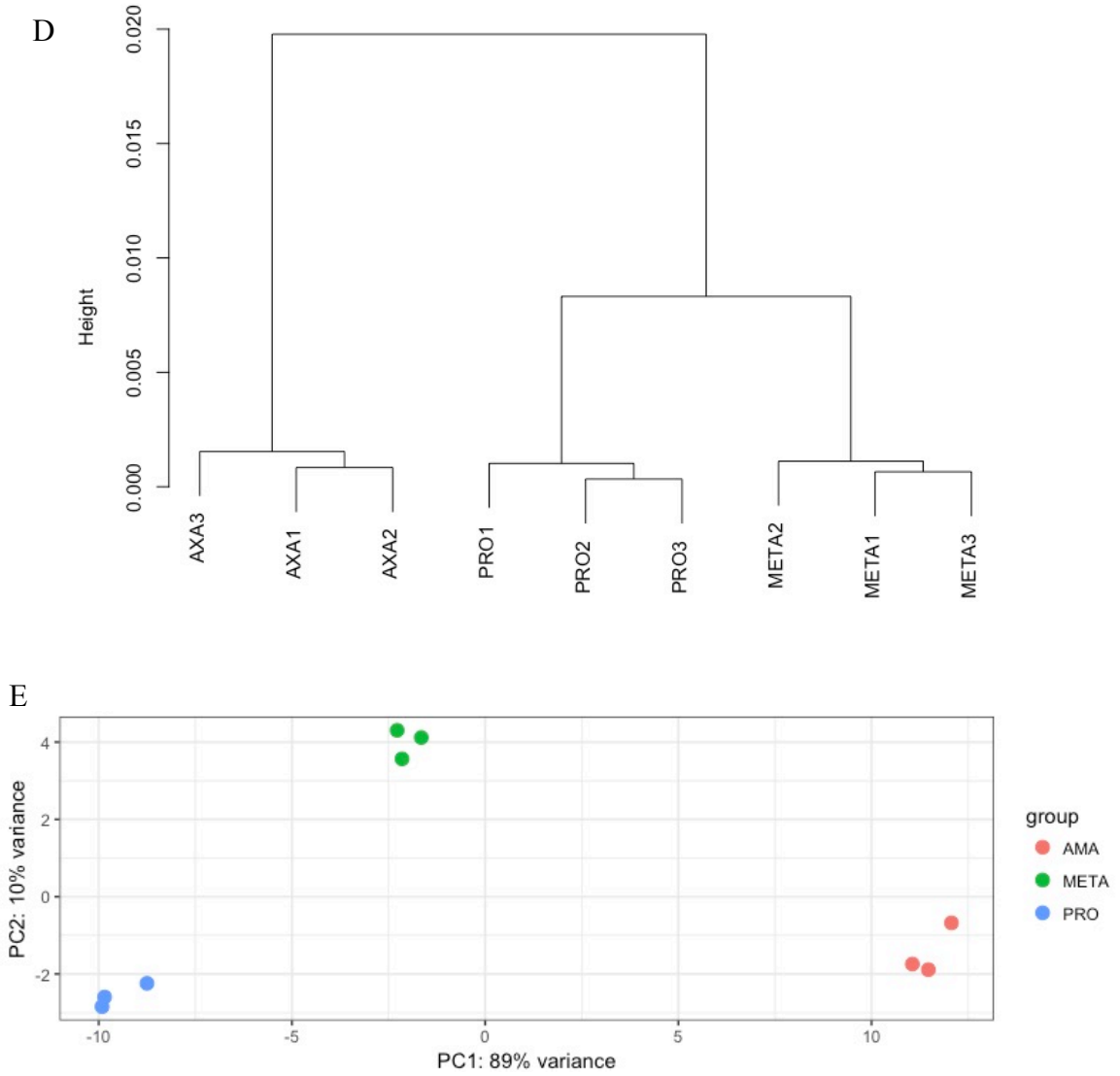


Figure S4. ***L. braziliensis*** **differential expression and similarities/differences throughout developmental stages.** Global expression changes **A)** PROvsMETA; **B)** METAvsAMA and **C)** AMAvsPRO. RNA-seq was performed for *L. braziliensis* procyclic promastigotes, metacyclic promastigotes and amastigotes. DE was analyzed using DESeq. Red points represent 5689 (A), 4856 (B) and 6576 (C) DE genes. Adjusted p value ≤ 0.05 was used. **D)** Hierarchical Clustering Analysis using Pearson correlation distance and log of read counts and **E)** Principle Component Analysis using log of read counts.

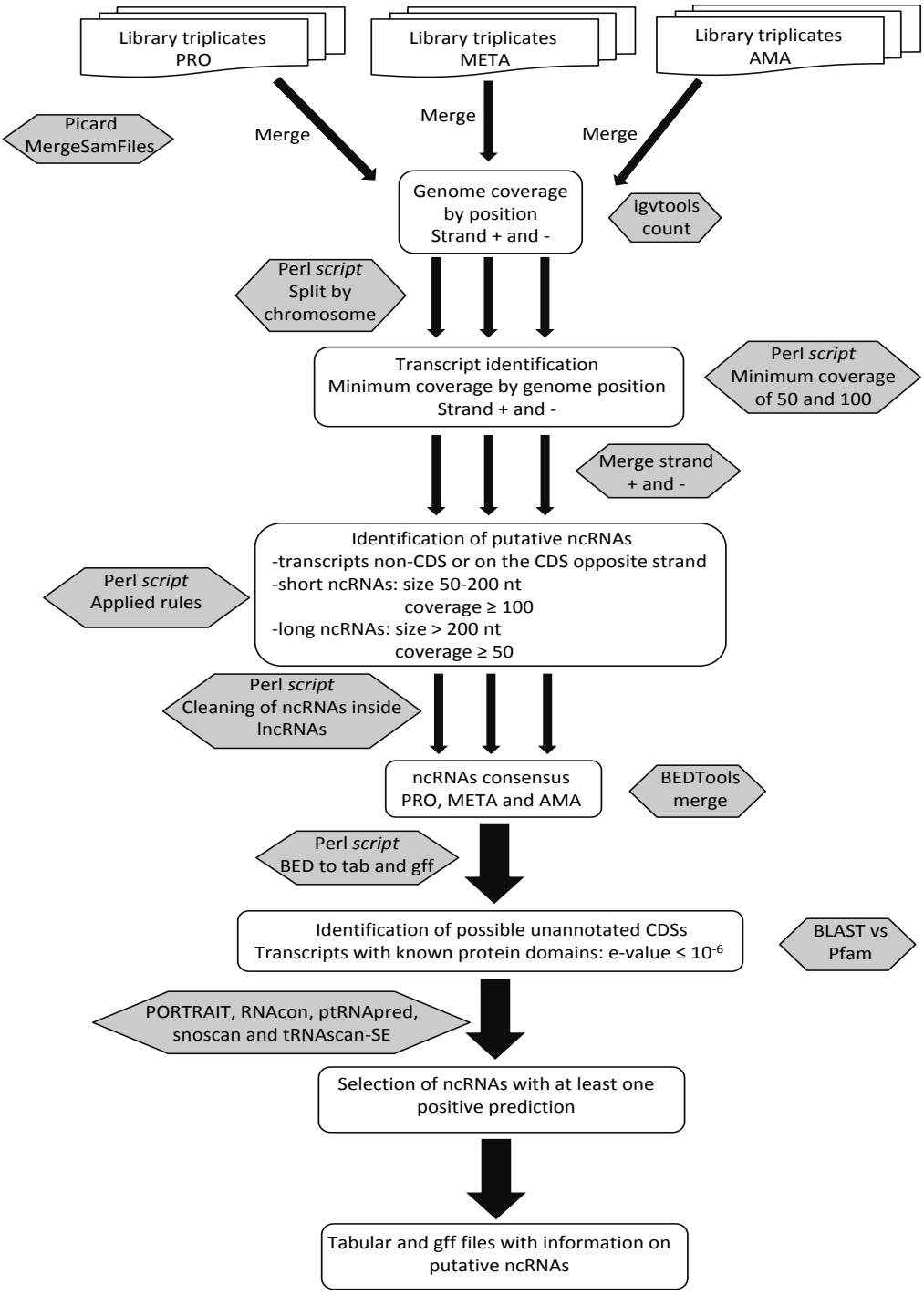


Figure S5. Association of approaches used to identify short and long putative ncRNAs. The designed pipeline started with data generated by sequencing total RNA from *L. braziliensis*. The Picard, IGVTools and BEDTools programs were used together with Perl scripts and programs for the identification of ncRNA characteristics for the definition of putative ncRNAs. PRO, META and AMA stand for procyclic promastigote, metacyclic promastigote and amastigote, respectively.

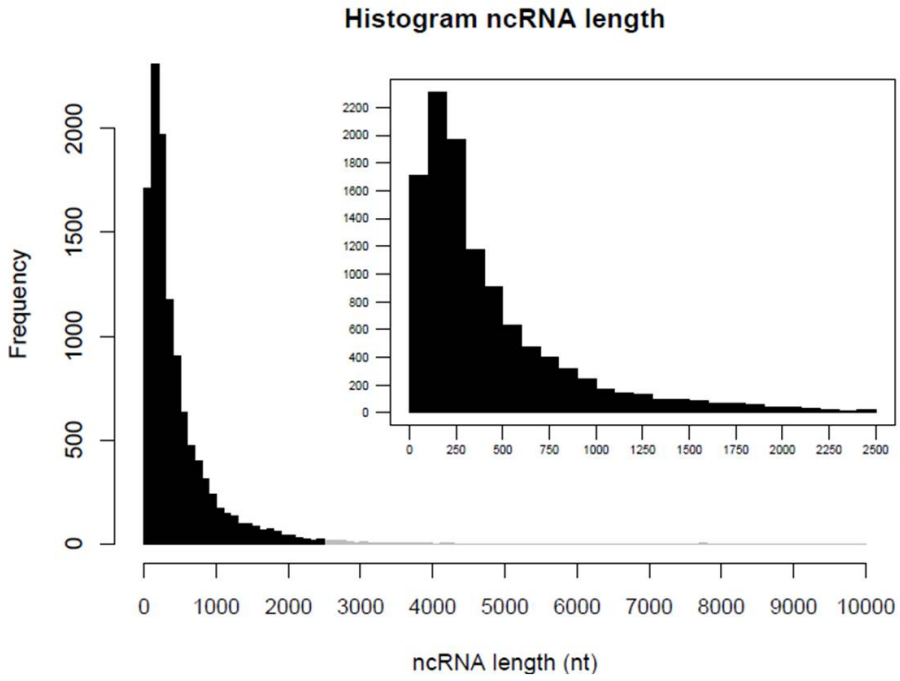
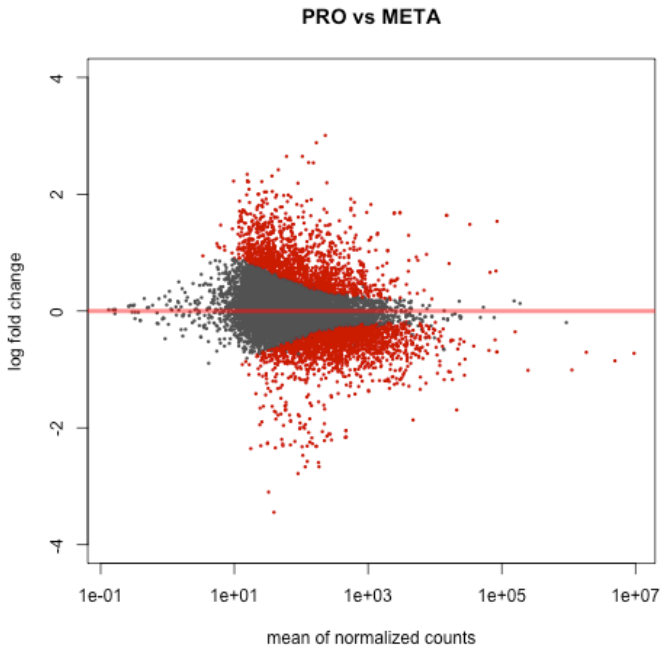
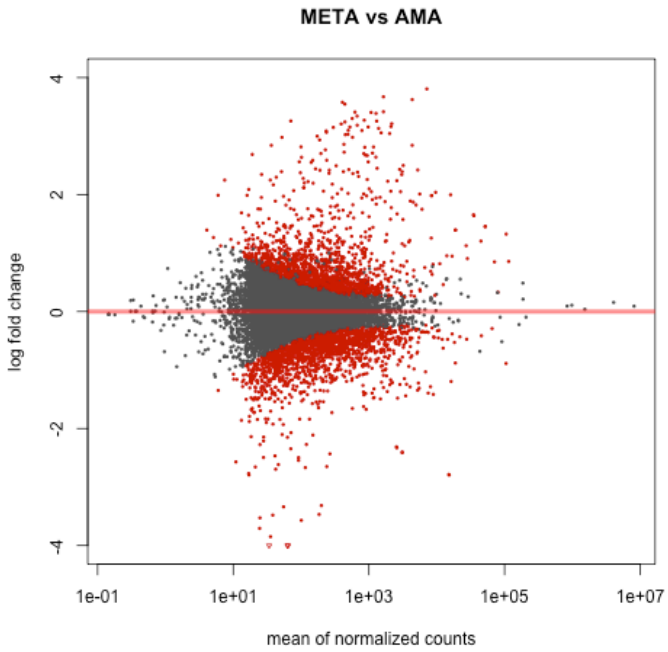


Figure S6. Putative ncRNA length distribution. The upper panel represents the zoom in the region of the most frequent size. The size of each bin is 100 nt.

A



B



C

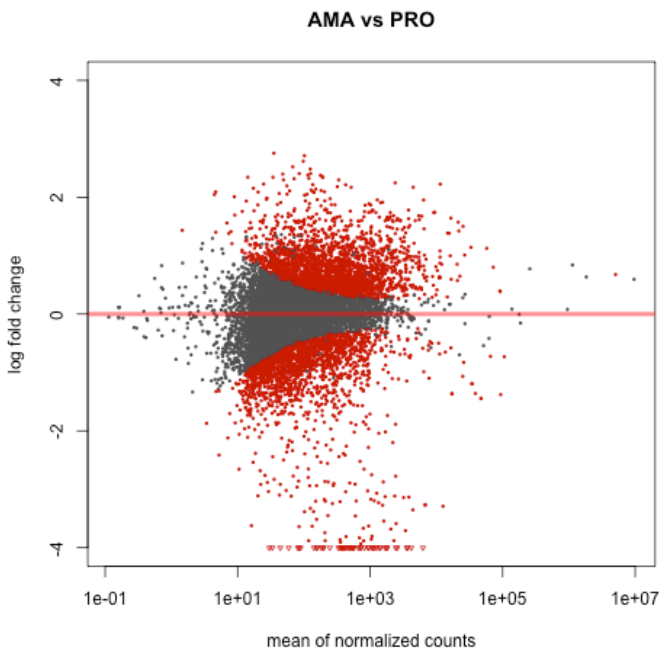


Figure S7. ncRNA differential expression analysis. Global expression changes **A)** PROvsMETA; **B)** METAvsAMA and **C)** AMAvsPRO. Differential expression was analyzed using DESeq. Red points represent 3266 (A), 3058 (B) and 4380 (C) DE ncRNAs. Adjusted p value < 0.05 was used.

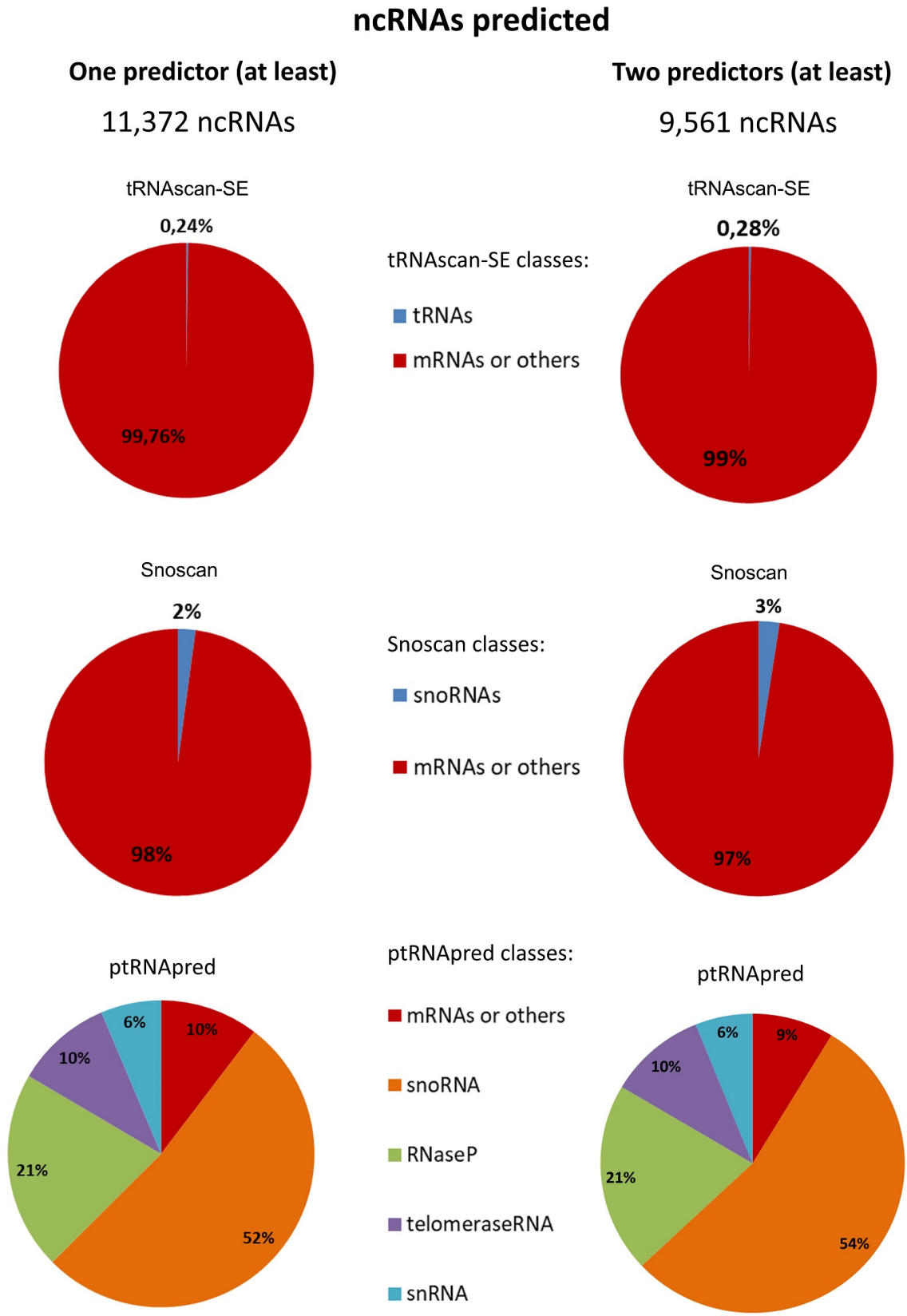


Figure S8. Graphic representation of ncRNA predictions by tRNAscan-SE, Snoscan and ptRNApred. Percentages and classes, according to each program (as indicated). On the left the 11,372 putative

ncRNAs that had been predicted by one of the 5 predictors (at least) and on the right, in a similar organization, the 9,561 ncRNAs predicted by at least two of the 5 programs are depicted and distributed among the different classes. mRNAs or others = messenger RNA or other ncRNA classes.

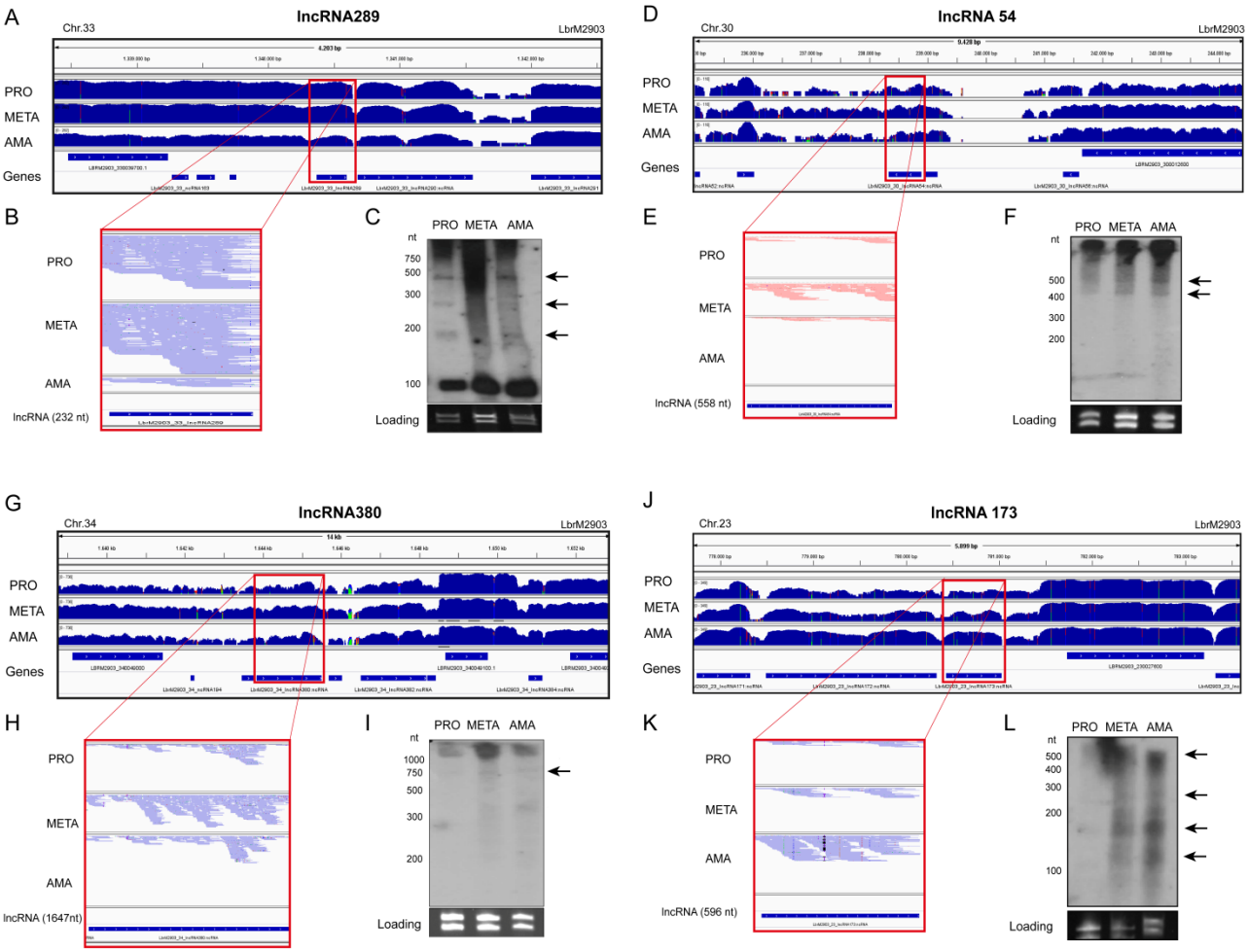


Figure S9: Identification and validation of four ncRNAs in *L. braziliensis*. The ncRNAs were submitted to northern blotting analysis using total RNA from procyclic, metacyclic and amastigote forms of *L. braziliensis*. Upper panel **(A, D, G and J):** regions of chromosomes 33, 30, 34 and 23 respectively encompassing predicted ncRNAs LbrM2903_33_lncRNA289, LbrM2903_30_lncRNA54, LbrM2903_34_lncRNA380 and LbrM2903_23_lncRNA173. Thin multicolored regions are divergences between reads and genome sequences. **(B, E,** H and L) extracted and zoomed areas for the corresponding lncRNAs depicting the number of reads in *L. braziliensis* procyclic, metacyclic and amastigote stages. (C, F, I and M) Northern blot of *L. braziliensis* total RNA using antisense oligonucleotides specific to each putative ncRNA. Arrows show multiple bands with approximate sizes to those predicted for the putative ncRNA. Ama: amastigote; Meta: metacyclic; Pro: procyclic; Gene: annotated genes, localized next to the putative ncRNAs; ncRNA: noncoding RNAs.