1 *Short communication*

2 **RNA-seq analysis reveals differences in transcript abundance between**
3 **cultured and sand fly-derived *Leishmania infantum* promastigotes.**

4
5 Pedro J. Alcolea^{a,b*}, Ana Alonso^a, Loren Baugh^b, Carolyn Paisie^{b,c}, Gowthaman Ramasamy^b, Aarthi
6 Sekar^b, Aakash Sur^{b,c}, Maribel Jiménez^d, Ricardo Molina^d, Vicente Larraga^a, Peter J. Myler^{b,c,e*}.

7
8 ^aDepartment of Molecular Microbiology and Biology of Infections, Centro de Investigaciones Biológicas
9 (Consejo Superior de Investigaciones Científicas), Calle Ramiro de Maeztu, 9. Madrid 28040, Spain.

10 ^bCenter for Infectious Disease Research (formerly Seattle Biomedical Research Institute), 307 Westlake
11 Avenue N., Seattle WA 98109-5219, USA

12 ^cDepartment of Biomedical Informatics and Medical Education, University of Washington, Box 358047,
13 Seattle, Washington 98195, USA

14 ^dUnit of Medical Entomology, Service of Parasitology, Centro Nacional de Microbiología, Virología e
15 Inmunología Sanitarias, Instituto de Salud Carlos III. Ctra. Majadahonda-Pozuelo s/n, Majadahonda
16 28220, Spain.

17 ^eDepartment of Global Health, University of Washington, Box 359931, Seattle, Washington, 98195, USA

18
19
20 *Corresponding author at: Center for Infectious Disease Research, 307 Westlake Ave N., Suite 500,
21 Seattle, WA 98109-5219, USA. Tel.: +1 206 256 7332. Fax: +1 206 256 7229. E-mail:
22 peter.myler@cidresearch.org

ABSTRACT

Leishmania infantum is responsible for human and canine leishmaniasis in the Mediterranean basin, where the major vector is *Phlebotomus perniciosus*. Because isolation of sufficient parasites from the sand fly gut is technically challenging, axenic cultivation of promastigotes is routinely used to obtain material for biochemical and genetic analyses. Here, we report the use of Spliced Leader RNA-seq (SL-seq) to compare transcript abundance in cultured promastigotes and those obtained from the whole midgut of the sand fly 5 days after infection. SL-seq allows for amplification of RNA from the parasite avoiding contamination with RNA from the gut of the insect. The study has been performed by means of a single technical replicate comparing pools of samples obtained from sand fly-derived (sfPro) and axenic culture promastigotes (acPro). Although there was a moderate correlation ($R^2=0.83$) in gene expression, 793 genes showed significantly different (≥ 2 -fold, $p < 0.05$) mRNA levels in sand fly-derived promastigotes and in culture, of which 31 were up-regulated ≥ 8 -fold ($p < 10^{-8}$ in most cases). These included several genes that are typically up-regulated during metacyclogenesis, suggesting that sand fly-derived promastigotes contain a substantial number of metacyclics, and/or that their differentiation status as metacyclics is more advanced in these populations. Infection experiments and studies evaluating the proportion of metacyclic promastigotes in culture and within the sand fly gut, previously reported by us, support the last hypothesis.

Keywords: *Leishmania infantum*; promastigotes; *Phlebotomus perniciosus*; axenic culture; RNA-seq; metacyclogenesis.

50 *Leishmania infantum* (Kinetoplastida: Trypanosomatidae) is the etiological agent of
51 human visceral leishmaniasis and canine leishmaniasis in the Mediterranean basin, where the
52 major vector is *Phlebotomus perniciosus* (Psychodidae: Phlebotominae). The study of parasites
53 from the genus *Leishmania* often requires axenic cultivation of both the promastigote (insect)
54 and the amastigote (vertebrate) stages. However, long-term passaging of isolates usually leads to
55 attenuation with reduced infectivity for either sand fly vectors or vertebrate hosts. Consequently,
56 passage through laboratory animals is required to maintain virulence (reviewed in [1]). While
57 promastigote culture is designed to mimic the conditions of the microenvironment of the vector
58 host, previous transcriptome analyses and *in vitro* infection experiments revealed important
59 differences between promastigotes from the stomodeal valve of the sand fly and from cultures in
60 stationary phase [2, 3]. In fact, peanut lectin non-agglutinating (PNA⁻) promastigotes obtained
61 from *L. infantum* promastigote cultures in stationary phase are ~30% less infective than sand fly-
62 derived promastigotes, and the whole populations in stationary phase ~50% less infective. The
63 proportion of *L. infantum* metacyclic promastigotes in culture [3, 4] and within the *P.*
64 *perniciosus* gut [5] is up to 10% depending on the elapsed time of development. Metacyclic
65 promastigotes are found at the stomodeal valve, which is located at the front of the thoracic mid-
66 gut.

67 Here, we have adapted our previously described spliced leader RNA sequencing (SL-seq)
68 strategy [6] to gain further insight on the differences in the transcript levels between cultured and
69 sand fly-derived promastigotes. Each sample contained the whole spectrum of promastigote
70 forms occurring during development. For this purpose, sample pools were prepared as detailed
71 below. This experimental design was based on the fact that only about 10% of promastigotes

reach the metacyclic stage when the infection of the sand fly gut is mature, even after two weeks of infection [5].

Promastigotes of the *L. infantum* MCAN/ES/98/10445 isolate were cultured in triplicate at 27°C in RPMI 1640 supplemented with L-glutamine (Life Technologies-Gibco, Karlsbad, CA), 10% heat inactivated fetal bovine serum (Lonza, Karlskoga, Sweden) and 100 µg/ml streptomycin – 100 IU/ml penicillin (Lonza). Culture replicates were started with different inocula. Aliquots containing 2×10^7 promastigotes were centrifuged at 2,000g for 10 min, and washed with PBS. Then, total RNA was isolated immediately as specified below. This procedure was repeated daily during one week and all RNA samples were mixed. This pool is the axenic culture promastigote sample (acPro) used in the RNA-seq experiment.

Sand flies from an established laboratory colony of *P. perniciosus* were maintained in a climatic chamber at 27-28°C, 90-100 % relative humidity, with a 17-hour light/7-hour dark photoperiod and 30% fructose solution. About 200 sand flies were fed over a 3-day chicken skin membrane with rabbit blood containing promastigotes from a late-logarithmic phase culture [7] and maintained in the same conditions, but without the fructose solution. The course of infection was followed every day and finally, sand flies were dissected at day 5 over a PBS droplet in a microscope slide to isolate promastigotes from the whole midgut (sfPro). For this purpose, each whole gut was dissected to isolate the midgut (including the stomodeal valve, the thoracic midgut and the abdominal midgut) and immediately squeezed with a coverslip to release promastigotes. The suspensions were recovered minimizing carryover of gut tissue. Suspensions of promastigotes from 10 different infected sand flies were pooled in 100 µl PBS and immediately harvested by centrifugation at 2,000g for 10 min. Infections were mature because each thoracic midgut was fully populated with promastigotes up to the stomodeal valve, as

shown in our previous experiments [2, 3, 8]. This included daily follow-up of infections at the light microscope after random sampling. The average number of promastigotes obtained per mid-gut was $\sim 2 \times 10^5$.

The samples were immediately re-suspended in TRizol® reagent (Life Technologies, Carlsbad, CA) for total RNA isolation following the manufacturer's instructions. One μg per ml of glycogen (Life Technologies) was added as a carrier prior to isopropanol precipitation. RNA concentration was measured by Qubit RNA broad range assay (Life Technologies) as specified by the manufacturer and sample quality was assessed on an RNA 6000 NanoChip using an Agilent 2100 Bioanalyzer (Life Technologies). A total of $\sim 63\text{ng}$ RNA was obtained from the sand fly pool. Prior to cDNA synthesis, 60 ng purified RNA from each sample was treated at 37°C for 10 min with 2 U DNase I in SuperScript III (SSIII) RT buffer (Life Technologies) in a final volume of 8 μl . The enzyme was inactivated at 75°C for 10 min and 2 μl of 10 μM Random5 primer (ACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN) added to each sample before heating to 65°C for 5 min, then immediately cooled at 4°C . Reagents for first strand cDNA synthesis were then added (SSIII RT buffer, 5 mM DTT, 2 mM each dNTP, and 200 U SSIII reverse transcriptase), brought to a 20- μl final volume and incubated at 40°C for 90 min. Subsequently, the enzyme was inactivated at 70°C for 15 min and RNA was degraded with 2 U RNaseH, which was inactivated at 70°C for 10 min. cDNA was purified with QIAquick PCR Purification Kit (QIAGEN, Valencia, CA). Second strand cDNA synthesis was begun by denaturing 45 μl of purified first strand cDNA together with 10 μl of 10 μM SL-2nd-primer3 (AGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTCAGTTTCTGTACT) at 95°C for 2 min followed by immediate cooling. The reaction was then performed at 37°C for 30 min with 0.5 mM of each dNTP and 15 U 3'-5' exo⁻ Klenow fragment (New England Biolabs,

Ipswich, MA) in NEB buffer 2, with 100 µl final reaction volume. Double-stranded cDNA was then purified using a QIAquick PCR purification Kit (QIAGEN). RNA-seq libraries were generated by PCR amplification of one third of the cDNA, using 5U Expand High Fidelity^{PLUS} enzyme (Roche, Mannheim, Germany) in Reaction Buffer 2 with 1.75 mM MgCl₂ in the final reaction mixture, 0.5 mM of each dNTP, 1 µM Multi-PCR-P1 primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC T) and RP-index primer (CAAGCAGAAGACGGCATACGAGATANNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCG), where the indexing sequence (NNNNNN) in RP-index primer was ATCAGT for sfPro and GCAGCT for acPro. Thermal cycling was as follows: 94°C, 2 min; 2 × (94 °C, 10 s; 40°C, 2 min; 72°C, 1 min); 22 × (94°C, 10 s; 40°C, 30 s; 72°C, 1 min); 72°C, 5 min. Finally, the PCR products were purified using the QIAquick PCR purification Kit (QIAGEN) following the standard protocol specified by the manufacturer, except for an additional wash with 0.75 ml of 35% guanidinium hydrochloride after decantation of the flow-through and before washing with PE buffer.

The SL libraries were sequenced in a HiSeq2500 platform as described [6]. The SL_SEQ primer was CGCTCTTCCGATCTCAGTTTCTGTACTTTATTG. An indication of the high quality of the reads is that the average Q score per base calculated with FastQC was above 30 in every sequencing cycle (Additional file 1) and a uniform ~60% G+C content was observed across cycles, as expected for *Leishmania* spp. After quality check, the reads were aligned against the *L. infantum* JPCM5 genome sequence [9] downloaded from TriTrypDB version 33 using Bowtie2 with previously optimized parameters [6]. SAM-to-BAM conversion was performed with SAMtools v1.5. Since the number of reads in each library varied considerably

(14,042,258 for sfPro *versus* 39,553,738 for acPro), transcript counts and transcripts per million (TPM) were normalized using the Median of Gene Expression Ratios method implemented in Geneious v10.2.3 (<http://www.geneious.com>) [10]. The protein-coding gene boundaries of the *L. infantum* JPCM5 reference sequence were adjusted using a custom script [11] to ensure that the gene models included SL sites within their 5' UTR. Differential gene expression analysis was performed with Geneious using statistical analyses based on a binomial test of the probability of randomly selecting a transcript from a particular gene in each condition, assuming that the gene is not differentially expressed (<https://assets.geneious.com/manual/10.2/index.html>). Raw and normalized data have been published in the NCBI's GEO and SRA databases (GSE70992, SRP115111) within BioProject PRJNA290068.

While the expression profiles were very similar, with a Pearson correlation coefficient of 0.83 (**Figure 1**), 143 genes showed statistically significant ($p < 0.05$) 4-fold or greater differences in relative expression (**Additional file 2**). The great majority (130/143) had higher mRNA levels in sfPro, with only 13 higher in acPro. The 32 genes that are most up-regulated (≥ 8 -fold) in sfPro compared to acPro are shown in **Table 1**, and **Table 2** contains statistics of p-values for genes showing ≥ 2 -fold expression values. Differential expression of 16 genes from Table 1 was analyzed by real-time quantitative RT-PCR (qRT-PCR) using TaqMan probes designed according to the procedure described in [2]. The results of this analysis are provided in **Additional file 3**, along with the sequences of primers and probes used. In general, there was good agreement between the results obtained from RNA-seq and qRT-PCR, although in several cases the fold-changes from the latter were lower than the former. This likely reflects the difficulty in distinguishing mRNA levels for closely related members of the same gene family

(see below), because it was often not possible to design specific primers and probes that do not hybridize with other members of the gene family.

Five genes, which belong to a single ATG8 family, encode ubiquitin-like proteins that bind to phosphatidylethanolamine during formation of autophagosomes. Up-regulation of this gene family is in agreement with the requirement and predominance of autophagy in the promastigote differentiation process of metacyclogenesis [12]. In fact, increased levels of autophagy lead to higher infectivity [13].

One gene encoding paraflagellar rod protein 1D (*PFR1D*) is up-regulated >150-fold and several other genes encoding flagellar rod proteins are also expressed at higher levels (>2-fold) in sfPro (see **Additional file 2**), indicating that formation of the flagellum and motility are enhanced in this promastigote population. In addition, three members of a gene family encoding a microtubule-associated protein are also highly up-regulated in sfPro, suggesting other processes, including signaling, (reviewed in [14]) may also be involved.

Two genes encoding a 3-hydroxyisobutyryl-coenzyme A hydrolase-like protein are up-regulated in sfPro. This enzyme, which is involved in valine degradation, as well as β -alanine (pantothenate and CoA biosynthesis) and propionate metabolism, has been previously shown to be up-regulated in metacyclic PNA⁻ promastigotes in culture [4].

Six genes encoding proteases are up-regulated in sfPro, including four from a tandem array encoding metallopeptidase Clan MA(M) Family M8 (also known as GP63 or leishmanolysin), and three other GP63 genes are also up-regulated 2- to 4-fold (see Additional file 2). GP63 family proteases are among the most abundant proteins in the surface of promastigotes and are critical virulence factors [15], consistent with up-regulation in sfPro. Their

mechanisms of action include cleavage of components of the complement system and host macrophage transcription factors and signaling proteins (reviewed in [16]).

Three genes encoding two different surface antigen-like proteins and an amastin-like protein are also substantially over-expressed in sfPro (**Table 1**), while several other surface and membrane-associated proteins are also up-regulated in sfPro, albeit only 2- to 8-fold relative to acPro (**Additional file 2**). These include a membrane bound acid phosphatase (MBAP), the acidocalcisomal pyrophosphatase, the hydrophilic acylated surface proteins (HASPA1, HASPA2, HASPB), the apical membrane antigen 1 (AMA1), and the META2 calpain (LinJ.17.0970). Like the ATG8 and the GP63 proteins, these have been previously associated with metacyclic promastigotes [4, 17, 18]. HASPs are involved in resistance to the complement system and survival within the macrophage [19], AMA1 is involved in interaction with and invasion of the host cell [20], the MBAP is associated to metacyclogenesis and increased virulence in mice [21] and META2 is a potential drug target [22]. In a recent study, the transcriptomes of procyclic, nectomonad and metacyclic *L. major* promastigotes obtained from the gut of *P. duboscqi* were compared [23]. This *Leishmania*-sand fly model is different to *L. infantum*-*P. perniciosus*. In fact, promastigote-gut interaction mechanisms are unknown in all *Leishmania*-sand fly pairs, except for LPG-galectins in *L. major*-*P. duboscqi* and *L. major*-*P. papatasi* [24]. In addition, the comparisons performed in both studies are not equivalent. Up-regulation of HASP, GP63 and autophagy genes in *L. major* nectomonads and metacyclics [23], is coincident with the results found in the pooled *L. infantum* samples from sand-fly and culture (Table 1), whereas amastins and SHERP are only increased in metacyclic *L. major* promastigotes but not in the *L. infantum* pools. In a previous comparison of *L. infantum* metacyclics from culture (peanut lectin non-agglutinating) and the stomodeal valve of *P. perniciosus* [3], some

genes of the amastin superfamily were also found as up-regulated in metacyclics from *P. perniciosus*, which were shown to be more infective than metacyclics from culture. In fact, the proportion of *L. infantum* metacyclics within the gut of *P. perniciosus* is not higher than 10% even after 2 weeks of sand fly infection by means of bloodfeeding from dogs [5], which is comparable to the percentage of metacyclics obtained from culture [3, 4]. Caution advised for interpretation of all these studies, we postulate that: i) up-regulation of HASP, GP63 and autophagy genes in nectomonads and metacyclics of *L. major* as well as in pooled *L. infantum* promastigote samples confirm their contribution to the metacyclogenesis process; ii) only certain amastin genes are up-regulated in metacyclics for preparation in advance for life in the parasitophorous vacuole of the mammalian host phagocyte as described [25, 26]; iii) the SHERP gene cluster is up-regulated only in *L. major* metacyclics, which is in agreement with the fact that SHERP is the only well validated marker of metacyclics exclusively in *L. major* [27]; iv) up-regulation of genes involved in metacyclogenesis in *L. infantum* promastigote pools derived from the gut of *P. perniciosus* with respect to cultured ones (Table 1) is due to more efficacious metacyclogenesis, yielding metacyclic promastigotes showing exacerbated infectivity [2, 3] presumably due to the influence of the sand fly gut microenvironment. Three genes in tandem array that encode a nucleoside transporter are among the mostly highly up-regulated genes in sfPro. Since *Leishmania* is unable to synthesize purine *de novo*, up-regulation of these genes is in agreement with the increasing nucleoside starvation [28] in the microenvironment of the sand fly gut [29]. Kinetoplastids are also auxotrophic for pteridines [30], which likely accounts for the up-regulation of a pteridine transporter gene in sfPro.

In conclusion, the extracellular milieu exerts important influence in the transcriptome of promastigotes, according to this SL-seq study performed with a single technical replicate

repeated, or how general our conclusions are more widely, except for those genes validated in terms of expression profile by means of qRT-PCR (Table 1). Namely, these genes are up-regulated at least 8-fold in promastigotes isolated from the whole sand fly gut, and most Promastigotes isolated from the whole sand fly gut up-regulate genes encode surface components or molecules directly involved in metacyclogenesis (*ATG8*, *GP63*, *HASPA1*, *HASPB*, *MBAP*, *AMA1*, *META2*), amastins, surface antigens and calpains. While this does not invalidate the culture model, since most of genes have similar expression levels in the two populations, we suggest some caution should be exercised when using cultured promastigotes as surrogates for insect-derived parasites.

ACKNOWLEDGEMENTS

This work was funded by the Ramón Areces Foundation (contract 050204100014, OTT code 20100338).

253 **REFERENCES**

- 254 1. Zilberstein, D., *Physiological and biochemical aspects of Leishmania development*, in *Leishmania*
255 *after the genome*, P. Myler and N. Fassel, Editors. 2008, Caister Academic Press: Norfolk. p. 107-
256 122.
- 257 2. Alcolea, P.J., et al., *Influence of the Microenvironment in the Transcriptome of Leishmania*
258 *infantum Promastigotes: Sand Fly versus Culture*. PLoS Negl Trop Dis, 2016. **10**(5): p. e0004693.
- 259 3. Alcolea, P.J., et al., *In vitro infectivity and differential gene expression of Leishmania infantum*
260 *metacyclic promastigotes: negative selection with peanut agglutinin in culture versus isolation*
261 *from the stomodeal valve of Phlebotomus perniciosus*. BMC Genomics, 2016. **17**: p. 375.
- 262 4. Alcolea, P.J., et al., *Genome-wide analysis reveals increased levels of transcripts related with*
263 *infectivity in peanut lectin non-agglutinated promastigotes of Leishmania infantum*. Genomics,
264 2009. **93**(6): p. 551-64.
- 265 5. Killick-Kendrick, R., et al., *A laboratory model of canine leishmaniasis: the inoculation of dogs*
266 *with Leishmania infantum promastigotes from midguts of experimentally infected phlebotomine*
267 *sandflies*. Parasite, 1994. **1**(4): p. 311-8.
- 268 6. Haydock, A., et al., *RNA-seq approaches for determining mRNA abundance in Leishmania*.
269 Methods Mol Biol, 2015. **1201**: p. 207-19.
- 270 7. Molina, R., *Laboratory adaptation of an autochthonous colony of Phlebotomus perniciosus*
271 *Newstead, 1911 (Diptera: Psychodidae)*. Res Rev Parasitol, 1991. **51**: p. 87-89.
- 272 8. Alcolea, P.J., et al., *Stage-specific differential gene expression in Leishmania infantum: from the*
273 *foregut of Phlebotomus perniciosus to the human phagocyte*. BMC Genomics, 2014. **15**: p. 849.
- 274 9. Peacock, C.S., et al., *Comparative genomic analysis of three Leishmania species that cause*
275 *diverse human disease*. Nat Genet, 2007. **39**(7): p. 839-47.
- 276 10. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for*
277 *the organization and analysis of sequence data*. Bioinformatics, 2012. **28**(12): p. 1647-9.
- 278 11. Cuypers, B., et al., *Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method*
279 *for RNA-seq in Trypanosomatids*. Sci Rep, 2017. **7**(1): p. 3725.
- 280 12. Besteiro, S., et al., *Endosome sorting and autophagy are essential for differentiation and*
281 *virulence of Leishmania major*. J Biol Chem, 2006. **281**(16): p. 11384-96.
- 282 13. Pinheiro, R.O., et al., *Induction of autophagy correlates with increased parasite load of*
283 *Leishmania amazonensis in BALB/c but not C57BL/6 macrophages*. Microbes Infect, 2009. **11**(2):
284 p. 181-90.
- 285 14. Portman, N. and K. Gull, *The paraflagellar rod of kinetoplastid parasites: from structure to*
286 *components and function*. Int J Parasitol, 2010. **40**(2): p. 135-48.
- 287 15. Yao, C., J.E. Donelson, and M.E. Wilson, *The major surface protease (MSP or GP63) of Leishmania*
288 *sp. Biosynthesis, regulation of expression, and function*. Mol Biochem Parasitol, 2003. **132**(1): p.
289 1-16.
- 290 16. Isnard, A., M.T. Shio, and M. Olivier, *Impact of Leishmania metalloprotease GP63 on*
291 *macrophage signaling*. Front Cell Infect Microbiol, 2012. **2**: p. 72.
- 292 17. Jackson, A.P., *The evolution of amastin surface glycoproteins in trypanosomatid parasites*. Mol
293 Biol Evol, 2010. **27**(1): p. 33-45.
- 294 18. Aguirre-Garcia, M.M., et al., *Leishmania major : detection of membrane-bound protein tyrosine*
295 *phosphatase*. Parasitology, 2006. **132**(Pt 5): p. 641-9.
- 296 19. McKean, P.G., et al., *Phenotypic changes associated with deletion and overexpression of a stage-*
297 *regulated gene family in Leishmania*. Cell Microbiol, 2001. **3**(8): p. 511-23.
- 298 20. S, S.K., K.G. R, and M. Ghosh, *Comparative in-silico genome analysis of Leishmania (Leishmania)*
299 *donovani: A step towards its species specificity*. Meta Gene, 2014. **2**: p. 782-98.

- 300 21. Espiau, B., et al., *A soluble pyrophosphatase, a key enzyme for polyphosphate metabolism in*
301 *Leishmania*. J Biol Chem, 2006. **281**(3): p. 1516-23.
- 302 22. Branquinho, M.H., et al., *Calpains: potential targets for alternative chemotherapeutic*
303 *intervention against human pathogenic trypanosomatids*. Curr Med Chem, 2013. **20**(25): p.
304 3174-85.
- 305 23. Inbar, E., et al., *The Transcriptome of Leishmania major Developmental Stages in Their Natural*
306 *Sand Fly Vector*. MBio, 2017. **8**(2).
- 307 24. Dostalova, A. and P. Volf, *Leishmania development in sand flies: parasite-vector interactions*
308 *overview*. Parasit Vectors, 2012. **5**: p. 276.
- 309 25. Alcolea, P.J., et al., *Transcriptomics throughout the life cycle of Leishmania infantum: high down-*
310 *regulation rate in the amastigote stage*. Int J Parasitol, 2010. **40**(13): p. 1497-516.
- 311 26. Alcolea, P.J., et al., *Temperature increase prevails over acidification in gene expression*
312 *modulation of amastigote differentiation in Leishmania infantum*. BMC Genomics, 2010. **11**: p.
313 31.
- 314 27. Sadlova, J., et al., *The stage-regulated HASPB and SHERP proteins are essential for differentiation*
315 *of the protozoan parasite Leishmania major in its sand fly vector, Phlebotomus papatasi*. Cell
316 Microbiol, 2010. **12**(12): p. 1765-79.
- 317 28. Carter, N.S., et al., *Adaptive responses to purine starvation in Leishmania donovani*. Mol
318 Microbiol, 2010. **78**(1): p. 92-107.
- 319 29. Bates, P.A., *Leishmania sand fly interaction: progress and challenges*. Curr Opin Microbiol, 2008.
320 **11**(4): p. 340-4.
- 321 30. Trager, W., *Pteridine requirement of the hemoflagellate Leishmania tarentolae*. J Protozool,
322 1969. **16**(2): p. 372-5.

FIGURE LEGEND.

Figure 1. Differential gene expression between insect-derived and cultured promastigotes.

The normalized TPM counts for sfPro and acPro are plotted on a binary log scale with the Pearson Correlation indicated. Significantly up-regulated genes ($\alpha=0.01$) in sfPro are represented in different shades of red according to their fold-change value (>2-fold, >4-fold and >8-fold). Significantly down-regulated genes are represented using green color similarly (<-2-fold and <-4-fold).

ADDITIONAL FILES.

Additional file 1. Quality check of sfPro and acPro raw data.

Additional file 2. Differential gene expression analysis of sfPro and acPro by SL-seq.

Additional file 3. qRT-PCR analysis. Primers, TaqMan FAM-NFQ probes, and calculations by using the efficiency-corrected ΔC_t method. GOI, gene of interest. GOR, gene of reference; NTC, negative control. Calculation of SD ($\log_2 F$): since $\log_a m = \ln m / \ln a$, for $y = f(m) = \log_a m$, $dy = (1/\ln a) * (dx/x)$; the outcome for each sample is shown in column SD($\log_2 \text{MeanQnorm}$); the final SD($\log_2 F$) values are calculated considering that SD propagation for the quotient is the square root of the sum of the SD squares.

343 **Table 1. Genes up-regulated more than 8-fold in sfPro and in acPro.** Geneious v10.2.3 was used to calculate normalized
344 transcripts per million counts (Norm. TPM), binary logarithm of the fold-change between sfPro and acPro ($\log_2F$), and the statistical
345 significance of relative gene expression values (p-value). Differential expression data obtained by qRT-PCR are also detailed herein,
346 including $\log_2F$ and standard deviation (SD) of these values. p-values corresponding to the binomial test were $p < 10^{-8}$ for all genes
347 contained in this table, except for LinJ.27.2510 ($p = 0.025$) and LinJ.10.0590 ($p = 3.7 \times 10^{-5}$). In fact, most p-values are null.
348

Gene information			SL-Seq		qRT-PCR		
Gene ID	Product name	Function	Norm. TPM		$\log_2F$	p-value	$\log_2F \pm SD(\log_2F)$
			sfPro	acPro			
LinJ.19.0850	ATG8/AUT7/APG8/PAZ2, putative (ATG8B.2)	Autophagy	588	23	4.7	0	N.D.
LinJ.19.0820	ATG8/AUT7/APG8/PAZ2, putative (ATG8B.1)	Autophagy	1118	67	4.1	0	N.D.
LinJ.19.0860	ATG8/AUT7/APG8/PAZ2, putative (ATG8B.3)	Autophagy	4785	375	3.7	0	1.6±0.3
LinJ.19.0840	ATG8/AUT7/APG8/PAZ2, putative (ATG8A.1)	Autophagy	4181	336	3.6	0	1.9±0.2
LinJ.09.0180	ATG8/AUT7/APG8/PAZ2, putative (ATG8C.1)	Autophagy	4422	448	3.3	0	1.5±0.2
LinJ.29.1880	paraflagellar rod protein 1D, putative	Cytoskeleton	792	5	7.3	0	1.9±0.2
LinJ.09.0170	microtubule associated protein-like protein	Cytoskeleton	7211	380	4.2	0	1.2±0.1
LinJ.09.0190	microtubule associated protein-like protein	Cytoskeleton	3840	203	4.2	0	N.D.
LinJ.09.0200	microtubule associated protein-like protein	Cytoskeleton	615	39	4.0	0	N.D.
LinJ.32.3820	3hydroxyisobutyrylcoenzyme a hydrolase-like protein	Metabolism	754	48	4.0	0	N.D.
LinJ.32.3810	3hydroxyisobutyrylcoenzyme a hydrolase-like protein	Metabolism	1563	103	3.9	0	1.3±0.3
LinJ.11.0640	metallopeptidase, Clan MF, Family M17	Proteolysis	146	2	6.4	0.34x10 ⁻¹⁴	1.9±0.1
LinJ.10.0500	metallopeptidase, Clan MA(M), Family M8	Proteolysis	3167	370	3.1	0	N.D.
LinJ.10.0510	metallopeptidase, Clan MA(M), Family M8	Proteolysis	2954	352	3.1	0	2.1±0.4
LinJ.10.0530	metallopeptidase, Clan MA(M), Family M8	Proteolysis	2903	346	3.1	0	3.9±0.7
LinJ.20.1220	cysteine peptidase, Clan CA, family C2, putative	Proteolysis	361	42	3.1	6.2x10 ⁻²⁰	1.5±0.5
LinJ.10.0520	metallopeptidase, Clan MA(M), Family M8	Proteolysis	2844	344	3.0	0	1.6±0.1
LinJ.27.2510	DEAD-box helicase-like protein (fragment)	RNA binding	36	2	4.4	0.0025	N.D.
LinJ.05.0900	surface antigen-like protein	Surface	1552	151	3.4	0	3.6±0.2
LinJ.04.0190	surface antigen-like protein	Surface	1527	164	3.2	0	4.3±0.5
LinJ.08.1320	amastin-like protein	Surface	743	82	3.2	0	N.D.
LinJ.15.1250	nucleoside transporter 1, putative	Transport	732	33	4.5	0	N.D.

LinJ.15.1230	nucleoside transporter 1, putative	Transport	771	43	4.2	0	1.3±0.5
LinJ.15.1240	nucleoside transporter 1, putative	Transport	656	37	4.1	0	N.D.
LinJ.10.1450	pteridine transporter, putative	Transport	1231	85	3.9	0	4.4±0.3
LinJ.31.3190	iron/zinc transporter protein-like protein	Transport	768	69	3.5	0	4.5±0.3
LinJ.04.0160	hypothetical protein, conserved in <i>Leishmania</i>	Unknown	1144	50	4.5	0	N.D.
LinJ.20.1270	Domain of unknown function (DUF1935), putative	Unknown	736	71	3.4	0	N.D.
LinJ.04.1090	hypothetical protein, conserved	Unknown	497	54	3.2	5.9x10 ⁻²⁷	N.D.
LinJ.31.1470	hypothetical protein, unknown function	Unknown	164	19	3.1	1.3x10 ⁻⁹	N.D.
LinJ.36.2920	hypothetical protein, conserved	Unknown	164	20	3.0	3.4x10 ⁻⁹	N.D.
LinJ.10.0590	hypothetical protein, unknown function	Unknown	92	12	3.0	3.7x10 ⁻⁵	N.D.

Table 2. Differential gene expression statistics.

p-value (Fold change ≥ 2 or ≤ -2)	Number of differentially expressed genes
<0.05	793
<0.01	621
<0.001	464
<0.0001	375
<01	319

Additional file 1. Quality check of sfPro and acPro raw data.

sfPro (PA001)

acPro (PA002)

sfPro

Adjacent Base Repeat Distribution:
Lane: 1 File: PA001_FCC161D_lane1_indexATCAGT.fastq
Total Reads: 14,042,258

acPro

Adjacent Base Repeat Distribution:
Lane: 1 File: PA002_FCC161D_lane1_indexGCAGCT_ALL.fastq
Total Reads: 39,553,738