1 **Title: Genome instability drives epistatic adaptation in the human pathogen *Leishmania***

2 **Authors:** Giovanni Bussotti[1,2], Laura Piel[2,a], Pascale Pescher[2], Malgorzata A. Domagalska[3], K.

3 Shanmugha Rajan[4], Tirza Doniger[4], Disha-Gajanan Hiregange[5], Peter J. Myler[6-10], Ron Unger[4],

4 Shulamit Michaeli[4], and Gerald F. Späth[2,*]

5 **Affiliations:**

6 [1]Institut Pasteur – Bioinformatics and Biostatistics Hub – C3BI, USR 3756 IP CNRS – Paris,

7 France.

8 [2]Institut Pasteur, INSERM U1201, Unité de Parasitologie moléculaire et Signalisation, Paris,

9 France.

10 [3]Institute of Tropical Medicine, Molecular Parasitology Unit, B-2000, Antwerp, Belgium.

11 [4]The Mina and Everard Goodman Faculty of Life Sciences and Advanced Materials and

12 Nanotechnology Institute, Bar-Ilan University, Ramat-Gan 52900, Israel.

13 [5] Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot,

14 7610001 Israel.

15 [6] Seattle Structural Genomics Center for Infectious Disease, Seattle, Washington, USA.

16 [7] Department of Biomedical Informatics and Medical Education, University of Washington,

17 Seattle, Washington, USA.

18 [8] Department of Pediatrics, University of Washington, Seattle, Washington, USA.

19 [9] Department of Global Health, University of Washington, Seattle, Washington, USA.

20 [10] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle,

21 Washington, USA peter.myler@seattlechildrens.org.

[a] Current address: Current address: Institut de Biologie Intégrative de la Cellule (I2BC) UMR 9198, Gif-sur-Yvette, France.

*corresponding author: Gerald Spaeth, Gerald.spaeth@pasteur.fr

Abstract. How genome instability is harnessed for fitness gain despite its potential deleterious effects is largely elusive. An ideal system to address this important open question is provided by the protozoan pathogen *Leishmania* that exploits frequent variations in chromosome and gene copy number to regulate expression levels. Using ecological genomics and experimental evolution approaches we provide first evidence that *Leishmania* adaptation relies on epistatic interactions between functionally associated gene copy number variations that can inform on pathways driving fitness gain in a given environment. We further uncover post-transcriptional regulation as a key mechanism that compensates for deleterious gene dosage effects and provides phenotypic robustness to genetically heterogenous parasite populations. Finally, we correlate dynamic variations in snoRNA gene dosage to changes in rRNA 2'-*O*-methylation and pseudouridylation, proposing translational control as an additional layer of parasite adaptation. *Leishmania* genome instability is thus harnessed for fitness gain by genome-dependent variations in gene expression, and genome-independent, compensatory mechanisms. This allows for polyclonal adaptation and maintenance of genetic heterogeneity despite strong selection. Epistatic adaptation described here needs to be considered in *Leishmania* epidemiology and biomarker discovery, and may be relevant to other fast evolving, eukaryotic cells that exploit genome instability for adaptation, such as fungal pathogens or cancer.

One Sentence Summary:

Epistatic interactions harness genome instability for *Leishmania* fitness gain.

46

47 **Main Text:**

48

49 Darwinian evolution plays a central yet poorly understood role in human disease. Iterations
50 between genetic mutation and environmental selection drive tumor development, microbial fitness
51 and therapeutic failure. Genome instability is a key source for genetic and phenotypic diversity,
52 often defining disease outcome (*1-4*). However, how genome instability can be harnessed for
53 fitness gain despite its potential deleterious effects is largely elusive. Here we investigate this
54 question in the protozoan parasite *Leishmania* that causes devastating human infections. These
55 parasite cycle between insect-stage promastigotes that infected *Phlebotome* sand flies, and
56 intracellular amastigotes that proliferate inside phagolysosomes of mammalian macrophages.
57 Genome instability is hallmark of *Leishmania* biology. These parasites lack promoter-dependent
58 gene regulation (*5, 6*), but exploit chromosome and gene copy number variations to regulate
59 mRNA abundance by gene dosage (*7-11*). In the absence of confounding transcriptional control,
60 *Leishmania* thus represents an ideal system to investigate the role of genome instability in fast-
61 evolving eukaryotic cells.
62 Here we uncover complex epistatic interactions between gene copy number variations and
63 compensatory, transcriptomic responses as key processes that harness genome instability for
64 fitness gain in *Leishmania*. Our data may be broadly applicable to pathogenic fungi or cancer cells,
65 known to exploit genome instability for adaptation.

66

67 **Results**

68 ***Leishmania* genomic adaptation is governed by epistatic interactions between gene copy**
69 **number variations.** We first assessed the level of copy number variation across 204 *Leishmania*
70 *donovani* clinical isolates from the Indian sub-continent (ISC) (*12*). This collection includes a core

71 group of 191 strains that are genetically highly homogenous as judged by the small number of
72 single nucleotide variants (SNVs) (<2,500 total), which provided us with a useful benchmark to
73 study the dynamics of copy number variations (CNVs) across a large number of quasi-clonal
74 populations. DNA read depth analysis of these isolates revealed important CNVs in both coding
75 and non-coding regions, with amplifications and deletions affecting respectively 14% and 4% of
76 the genome (**Fig. 1A-B, Tables S1-S2**). Analysing the statistical association of observed CNVs
77 with repetitive sequence elements uncovered 10 novel DNA repeats that could drive *Leishmania*
78 genome instability through microhomology-mediated, break-induced replication as observed for
79 human CNVs (*13*) (**Fig. 1C**, **Fig. S1, Table S3**). Gene dosage changes were not random but clearly
80 under selection as judged by the reproducibility of genetic interactions across independent isolates
81 and the enrichment of amplified genes in biological functions associated with fitness gain.
82 Statistically significant interactions were observed between positive (correlating) and negative
83 (anti-correlating) read depth variations (**Fig. 1D**, **Fig. S2-S3**, **Table S4-S5**), including a highly
84 connected Network Cluster (NC) containing 60 co-amplified tRNA genes encoded on 16 different
85 chromosomes (NC1, **Fig. 1D**, **Fig. S4**, **Tables S6-S7**). Natural selection of gene CNVs is further
86 supported by (i) their independent emergence across phylogenetically distinct strains providing
87 evidence for evolutionary convergence, (ii) the very high copy number observed for certain genes
88 (up to 30-fold) suggesting strong, positive selection, and (iii) the global enrichment of read depth
89 variations in phenotypically silent, intergenic regions, suggesting purifying selection against
90 deleterious effects caused by gene CNVs (**Fig. 1E-G**, **Fig. S5, Table S8-S10**). Together our data
91 provide first evidence that *Leishmania* genomic adaptation is governed by gene CNVs through
92 highly dynamic, functional interactions that are under natural selection. These interactions define

93 a novel form of epistasis at gene rather than nucleotide levels, with the phenotypic effect of a given
94 gene amplification being dependent on co-amplification of functionally related genes.
95
96 **Epistatic adaptation drives polyclonal fitness gain *in vitro*.** We next used an experimental
97 evolution approach to directly assess the link between epistatic interactions and fitness gain in
98 hamster-derived *L. donovani* parasites during adaptation to *in vitro* culture (*10*). Following
99 normalization for karyotypic variations (**Fig. S6A, Table S11**), changes in read depth were
100 monitored between passages 2 (P2, two weeks in culture) and 135 (P135, 36 weeks in culture),
101 corresponding to approximately 20 and 3,800 generations. Our analysis revealed co-amplification
102 of coding and non-coding (nc) genes and gene clusters that are functionally linked to fitness gain
103 in culture (i.e. accelerated cell proliferation), i.e. genes encoding for rRNAs, tRNAs, snRNAs,
104 snoRNAs, SLRNAs, and ribosomal proteins (**Fig. 2A**, **Fig. S6B** and **Tables S12-S14**). Epistatic
105 adaptation thus increases translation efficiency, thus overcoming the major rate limiting steps for
106 fast growing populations (*14*). This functional link between a given fitness phenotype and its
107 underlying epistatic network opens unexplored venues for the discovery of *Leishmania*
108 biomarkers, which may be represented by complex genetic interactions rather than individual loci.
109
110 Unexpectedly, aside gene amplification, we also revealed gene depletion as a major driver
111 for environmental adaptation and fitness gain. We identified a genomic deletion of 11kb containing
112 a single gene encoding for a NIMA-related kinase gene (Ld1S_360735700), which is gradually
113 selected in the adapting promastigote population from a pre-existing mutant that was detected in
114 splenic amastigotes (**Fig. 2B**, **Fig. S7A** and **B**). Clonal analysis of the P20 population revealed the
115 presence of the spontaneous knock out (spo-KO) in six out of eight individual clones (**Fig. 2C**).

116 These spo-KO clones were clearly not the descendants of a single founder cell but were of
117 independent, polyclonal origin as judged by their distinct gene CNV profiles (**Fig. 2C**, **Table S15**),
118 and their polyphyletic clustering based on SNVs compared to wild-type (WT) clones (**Fig. 2D,**
119 **Table S16**). This evolutionary convergence strongly supports natural selection of the deletion
120 during culture adaptation, and suggests a potential role of the deleted NIMA kinase in growth
121 restriction, which we assessed in the following by gene editing.
122
123 **Toxic gene dosage effects are compensated at post-transcriptional levels.** Unlike expected
124 from the spo-KO phenotype, CRISPR/Cas9-generated, NIMA knock out mutants (cri-KO) (**Fig.**
125 **S7C** and **D**) did not show increased proliferation but on the contrary were not viable, and
126 heterozygous mutants showed a strong growth defect that was partially rescued by episomal over-
127 expression of the NIMA kinase gene (**Fig. 3A**). This paradoxical result reveals that spo-KO cells
128 must have evolved mechanisms that can compensate for the loss of this essential gene. Read depth
129 analyses of the spo-KO and WT clones ruled out genetic compensation (**Table S15**). RNAseq
130 analysis in contrast revealed highly reproducible, compensatory transcript profiles in the six
131 independent spo-KO clones (**Fig. 3B**, **Table S17**). Our analysis revealed reduced stability in spo-
132 KO clones of 23 transcripts implicated in flagellar biogenesis (**Fig. S8**), which correlated with
133 reduced motility (data not shown). Loss of motility may represent a fitness tradeoff, providing
134 energy required for accelerated *in vitro* growth. On the other hand, we identified 350 transcripts
135 with increased abundance (Table S17, second sheet), which either result from increased gene
136 dosage or mRNA post-transcriptional stabilization. Direct comparison of DNA and RNA reads
137 depth variations allowed us to dissociate between these two possibilities and identified a set of
138 transcripts whose expression changes between WT and spo-KO did not correlate with gene dosage.

139 In absence of transcriptional regulation in *Leishmania* (*5*), the abundance of these transcripts is
140 likely regulated by differential mRNA turn-over. Increased abundance was observed in spo-KO
141 clones for the mRNA of another NIMA-related kinase (Ld1S_360735800) encoded adjacent to the
142 deleted region, suggesting a direct, post-transcriptional compensation of kinase functions (**Fig. 3C,**
143 **Table S18**). Likewise, we observed increased stability for functionally related small, non-coding
144 RNAs, including 43 snoRNAs, and various rRNAs and tRNAs involved in (e.g. rRNAs, snoRNAs)
145 (**Fig. 3D, Table S18**), suggesting ribosomal biogenesis and translational regulation as yet another
146 level of non-genomic adaptation. Together our data identify gene deletion and compensatory, post-
147 transcriptional responses as novel drivers of *Leishmania* fitness gain.
148
149 **The non-coding RNome as a novel driver of *Leishmania* fitness gain.** The selective stabilization
150 of snoRNAs during early culture adaption (P2-P20) suggests these non-coding RNAs as novel key
151 drivers in *Leishmania* fitness gain. We confirmed this possibility in long-term adapted parasites
152 that were continuously cultured to passage 135 (P135). Read depth analysis revealed selective
153 amplification of snoRNAs, which was confirmed in an independent, long-term evolutionary
154 experiment conducted until passage 125 (P125, **Fig. 4A**, **Fig. S9** and **Table S19**). Surprisingly,
155 rather than amplification of individual snoRNA genes, increased read depth was caused by the
156 recovery of a single locus on chromosome 33 containing a cluster of 15 snoRNA genes, which
157 was depleted in the original amastigotes population (**Fig. 4B**). The restoration of this locus between
158 P20 and P135 is further proof that snoRNAs are under positive selection during culture adaptation.
159 Rather than gene amplification, the recovery of this locus is more likely due to the selection of a
160 small sub-population that penetrated the culture between P20 and P135, similar to what we
161 observed for the NIMA kinase spo-KO. The mosaic structure of this locus is further supported by

162 the CNVs we detected in independent amastigotes isolates (**Fig. 4C**, **Fig. S10**). Thus, while short-
163 term adaptation in *Leishmania* is governed by control of RNA stability, long-term adaptation
164 occurs through more cost-efficient gene dosage regulation. Increased snoRNA abundance likely
165 satisfies a quantitative need for ribosomal biogenesis in our fast-growing cultures. However,
166 changes in snoRNA expression can also affect the nature and quality of ribosomes (*15, 16*). In the
167 following we assessed the possibility of such fitness-adapted ribosomes in *Leishmania* especially
168 since snoRNA genes guiding 2'-*O*-methylation (Nm) and pseudouridine (Ψ) modifications were
169 extensively amplified. Amplification of snoRNAs should lead to increased modifications on sites
170 that are accessible for modification. The mapping of Nm via RiboMeth-seq (*17*) revealed an
171 increased level of modification for 18 sites by at least two fold (**Table S20**), while the mapping of
172 pseudouridylation by Ψ-seq (*18, 19*) showed increase for 5 Ψ sites (**Table S21**) during adaptation
173 from P2 to P135. Interestingly, the hyper-modified Nm sites are localized around the peptidyl-
174 transferase centre (PTC) and mRNA entrance tunnel, whereas the hyper-modified Ψ sites are
175 located in the PTC itself (**Fig. 4D-E**, **Fig. S11-S14**).
176 Together, our data reveal a complex model of *Leishmania* fitness gain, where epistatic interactions
177 between gene amplifications and compensatory responses at post-transcriptional levels harnesses
178 *Leishmania* genome instability for polyclonal adaptation (**Figure 4F**).
179
180
181
182
183

184 **Discussion**

185 A common strategy in microbial evolutionary adaptation is known as 'bet-hedging', where the
186 fitness of a population in changing environments is increased by stochastic fluctuations in gene
187 expression that are regulated at epigenetic levels (*20-24*). The protozoan pathogen *Leishmania*
188 largely lacks transcriptional regulation, raising the question on how these parasites generate
189 variability in transcript abundance and phenotype required for adaptation (*5, 6*). Our data uncover
190 an alternative mechanism of bet-hedging evolved in these parasites that relies on the unique
191 biology of *Leishmania*.

192 First, we provide evidence that *Leishmania* genome instability may be driven by 10 new
193 repetitive DNA elements that were associated with genome-wide amplifications and deletions,
194 both of which are under positive and purifying selection. *Leishmania* thus compensates for the
195 absence of stochastic gene regulation by the generation of stochastic gene CNVs, which are known
196 to cause dosage-dependent changes in transcript abundance (*7-11*). Second, we reveal an important
197 role for post-transcriptional regulation in *Leishmania* fitness gain, which (i) compensates for the
198 deleterious deletion of a NIMA kinase by selectively stabilizing the transcripts of an orthologous
199 kinase, (ii) allows for gene dosage-independent increase in the abundance of non-coding RNAs
200 (SLRNAs, snoRNAs) required for proliferation, and (iii) provides phenotypic robustness to
201 genetically heterogenous populations as documented by the converging transcript profiles of
202 independent NIMA spo-KO mutants. This adaptation process guards against toxic gene dosage
203 effects and simultaneously increases the phenotypic landscape available to *Leishmania* for
204 adaptation via gene deletions and compensatory transcriptional responses. Significantly, both the
205 NIMA ortholog as well as the depleted SLRNA and snoRNA loci are amplified during long-term
206 culture (**Fig. 4B**, **Fig. S6B**, **Fig. S9**), revealing a two-step adaption process reminiscent to yeast

207 (*25*), implicating a first, post-transcriptional mechanism via transient changes in RNA stability,
208 followed by a second, genomic mechanism via the selection for stable gene CNVs.

209 The dynamic changes in snoRNA stability and gene copy number observed in our
210 experimental evolution system identifies this class of ncRNAs as an unexpected driver of
211 *Leishmania* fitness gain. snoRNAs guide rRNA modification and processing, as well as
212 modification of snRNAs and additional non-coding ncRNAs (*26, 27*). Changes in snoRNA gene
213 copy number and transcript abundance during parasite adaptation correlated with
214 hypermodification on rRNA, known to affect the quality and translation specificity of the
215 ribosome. It is interesting to speculate that these specialized ribosomes represent an additional
216 layer of regulation at translational level that can further counteract toxic gene dosage effects,
217 provide phenotypic robustness and adapt the proteome profile to a given environment, much like
218 it was observed in cancer cells or during differentiation of stem cells (*28, 29*).

219 Our findings have important implications for *Leishmania* clinical infection. *Leishmania*
220 adapts to various environmental cues, notably the presence of anti-leishmanial drugs. In contrast
221 to high frequency amplifications observed during experimental drug treatment in culture (*30-32*),
222 treatment failure and drug resistance observed during natural infection may evolve through multi-
223 locus, epistatic interactions such as described in our systems, that can balance the fitness trade-off
224 between drug resistance and infectivity (*33*). Our data thus define biological networks rather than
225 individual loci as novel biomarkers with potential diagnostic or prognostic value. Conceivably, the
226 epistatic mechanisms we uncover in *Leishmania* can be of broader relevance to other human
227 pathologies caused by fast evolving eukaryotic cells exploiting genome instability for polyclonal
228 adaptation, such as cancer cells. While single nucleotide, epistatic interactions are recognized as
229 important drivers of tumor development (*34, 35*), the role of epistatic interactions between

230 structural mutations and between the genome and transcriptome in drug resistant cancer cells
231 remains to be elucidated (*36*).

232 In conclusion, our results propose a first model of *Leishmania* fitness gain (**Fig. 4F**), where
233 polyclonal adaptation of mosaic populations is driven by epistatic interactions that (i) buffer the
234 detrimental effects of genome instability, (ii) coordinate expression of functionally related genes,
235 and (iii) generate beneficial phenotypes for adaptation. This mechanism of fitness gain avoids
236 genetic death by maintaining heterogeneity in competing parasite populations under environmental
237 selection, and may be generally applicable to other eukaryotic systems that adapt through genome
238 instability.

239

240

Acknowledgments:

Funding: This study was supported by a seeding grant from the Institut Pasteur International Department to the LeiSHield Consortium, the EU H2020 project LeiSHield-MATI - REP-778298-1, the Fondation pour la Recherche Médicale (grant number FDT201805005619), the Flemish Ministry of Science and Innovation (MADLEI, SOFI Grant 754204) and a grant from CAMPUS France and the Israeli Ministry of Science and Technology PHC MAIMONIDE 2018-2019-Projet N° 41131ZD. We thank Cedric Notredame and Jean-Claude Dujardin for critical reading of the manuscript.

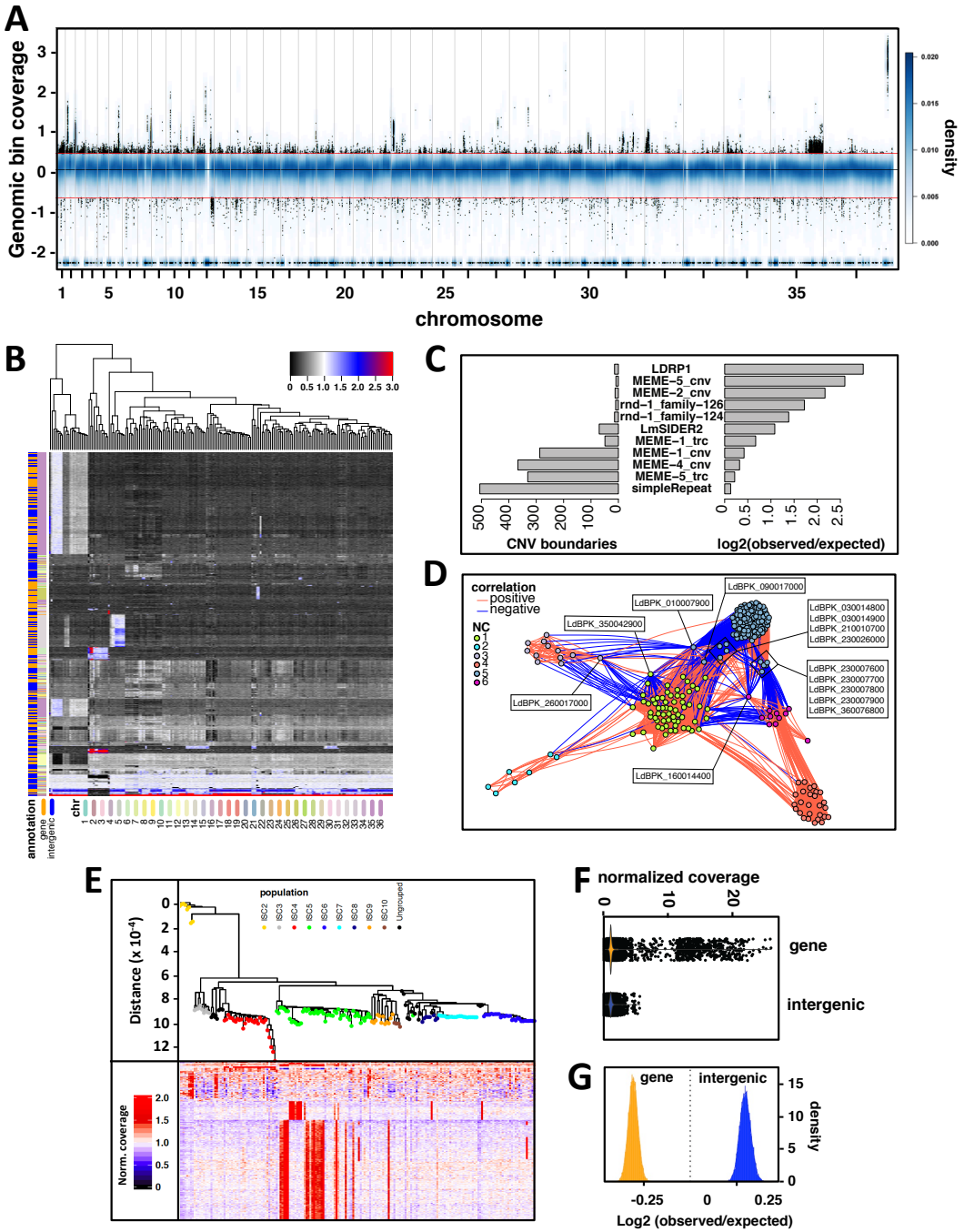
Author contributions: Conceptualization, G.B., L.P., P.P., M.A.D., T.D., R.U. S.M. and G.F.S.; Methodology, G.B., L.P., P.P., T.D., R.U. and G.F.S.; Software, G.B., T.D. and R.U.; Formal Analysis, G.B., L.P., P.P. and G.F.S.; Investigation, G.B., L.P., P.P. and G.F.S.; Writing – Original Draft, G.B., L.P., P.P. and G.F.S.; Writing – Review & Editing, G.B., L.P., P.P., M.A.D., T.D., R.U. S.M. and G.F.S.; Supervision, G.F.S., R.U. and S.M.; Funding Acquisition, G.F.S., S.M.

Competing interests: Authors declare no competing interests.

Data and materials availability:

Reads were deposited in the Sequence Read Archive (SRA) database (*37*) and are publicly available under accession no PRJNA605972. All data is available in the main text or the supplementary materials.

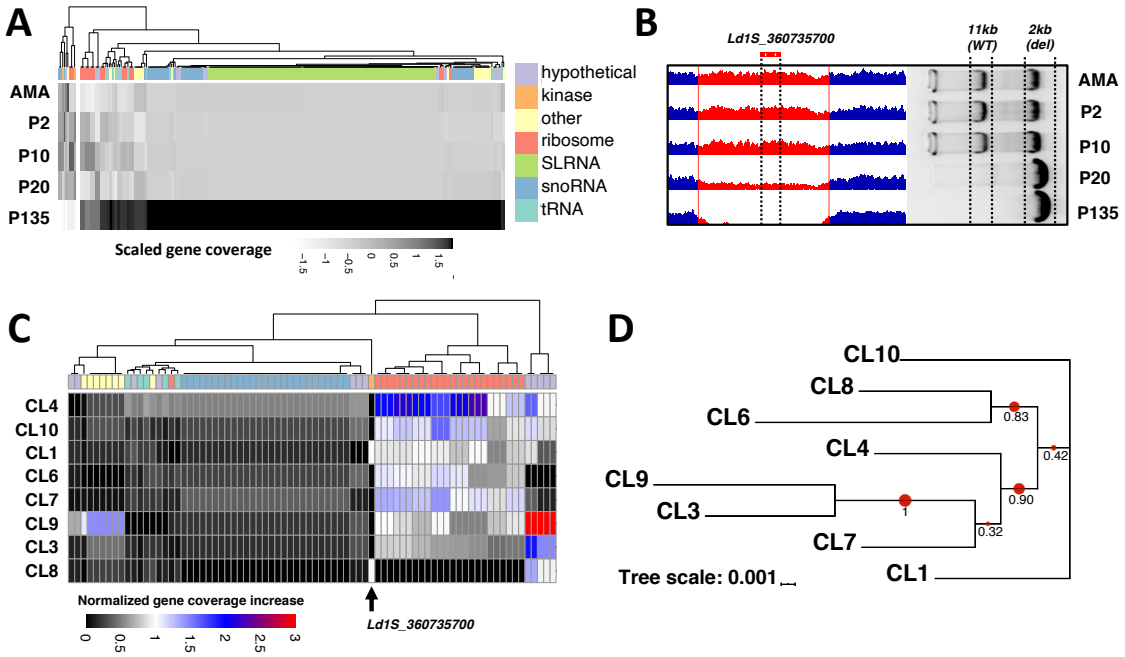
Fig. 1



Genome-wide mapping of copy number variations (CNVs), their environmental selection, and epistatic interactions. (**A**) Genome-wide normalized coverage values in natural logarithm

266 scale (*y*-axis) across the 36 chromosomes (*x*-axis) for 204 *Leishmania donovani* field isolates from
267 the Indian Sub-continent (*12*). The *x*-axis reports the position of genomic windows along the
268 chromosomes. The smoothed blue color represents the 2D kernel density estimate of genomic bins.
269 A sample of 50,000 genomic bins with normalized coverage ≥ 1.5 or ≤ 0.5 are materialized as
270 black dots. The black horizontal line and the two red lines indicate normalized coverage values of
271 1.5, 1 and 0.5. (**B**) Heatmap generated for the 204 clinical isolates (columns) showing their
272 phylogenetic relationship as a function of CNV bins (rows). The color scale reflects the deviation
273 from the minimum bin coverage observed across all genomes. The presence of annotated genes
274 and the chromosomal location of the CNVs are indicated by the two colored columns on the left.
275 The color code is defined in the legend below the plot. (**C**) Association between CNVs and
276 repetitive elements. The bar plots show the number of observed overlap instances between the
277 boundaries of the CNV regions and repetitive elements (left panel), and the log2 ratio between the
278 observed and expected overlap events over 10,000 randomizations (right panel). (**D**) Gene CNV
279 network analysis. The nodes represent gene CNVs while the edges indicate statistically significant
280 positive (red) and negative (blue) correlations observed in the 204 field isolates. The nodes are
281 colored according to the predicted network clusters (NC). (**E**) Phylogenetic tree based on SNVs
282 (>90% frequency) (upper panel) for the ISC core population comprising 191 isolates (*12*) (not
283 including the distant ISC1 strains (*38*) – for full phylogeny see **Fig. S5**). The heatmap (lower panel)
284 shows the normalized gene sequencing coverage across genomes (rows). To ease the visualization,
285 gene amplifications with normalized coverage > 2 are indicated as 2. (**F**) Violin plot showing the
286 distributions of the normalized genomic coverage values of the collapsed CNV positions (dots)
287 matching genic and intergenic regions. (**G**) Log2 ratio distributions of observed and expected
288 nucleotide overlap between collapsed CNV regions and gene/intergenic annotations.

289 **Fig. 2**

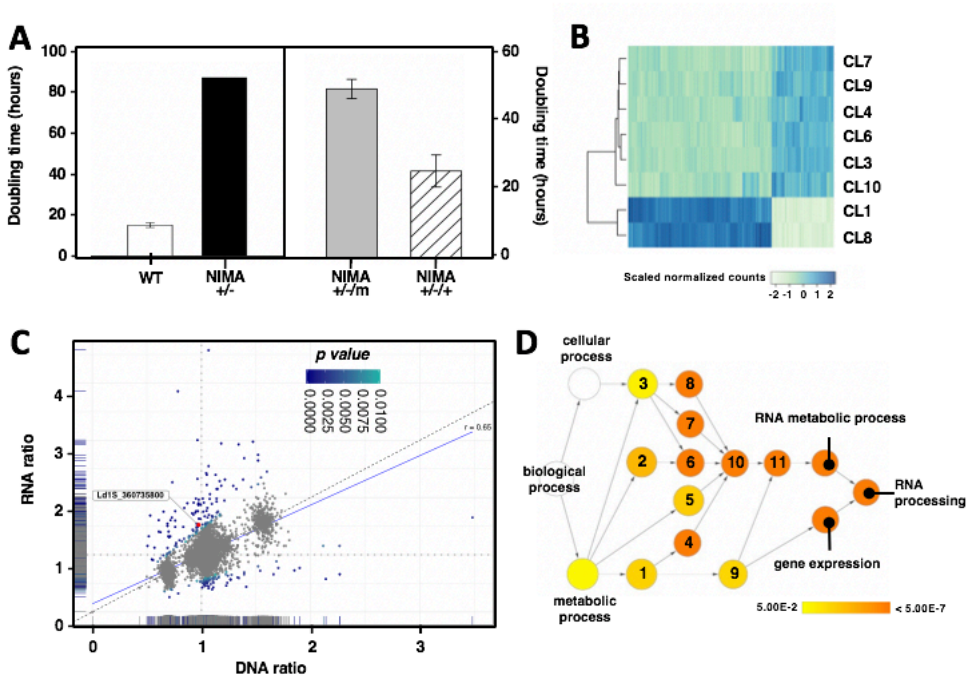


290
291

292 **Longitudinal analysis of gene CNVs during fitness gain in culture.** (**A**) Heatmap generated by
293 plotting gene read depth values (columns) across *L. donovani* amastigotes isolated from infected
294 hamster spleen (AMA) and derived promastigotes evolving in continuous culture for 2, 10, 20 and
295 135 passages (P2, P10, P20, P135) (rows). The gray level reflects the scaled normalized gene
296 coverage as indicated in the figure. (**B**) Screenshot of the IGV genome browser showing gradual
297 loss of the NIMA-like kinase gene Ld1S_360735700 during culture adaptation between splenic
298 amastigotes (AMA), and derived promastigotes at passages P2 and P135. The right panel shows
299 the gel-electrophoretic analysis of PCR fragments obtained from lesional amastigotes (AMA) and
300 derived promastigotes at the indicated culture passages that are diagnostic for the WT (11kb) and
301 the deleted NIMA-like kinase locus (2kb). (**C**) Heatmap generated by plotting gene read depth
302 variation (columns) across eight clones isolated from the P20 population (rows). The color code is
303 defined in the legend and corresponds to the deviation from the minimum sequencing coverage

304 measured for that gene in all clones. The colored ribbon indicates the simplified gene annotation
305 as shown in the legend of panel A. The deletion of the NIMA-like kinase Ld1S_360735700 is
306 indicated by the arrowhead. (**D**) Analysis of the phylogenetic relation of the P20 clones. The red
307 dots and numbers indicate the node bootstrap support (1 = 100%). The genetic distance is indicated
308 by the branch length and scaled as indicated in the figure.
309

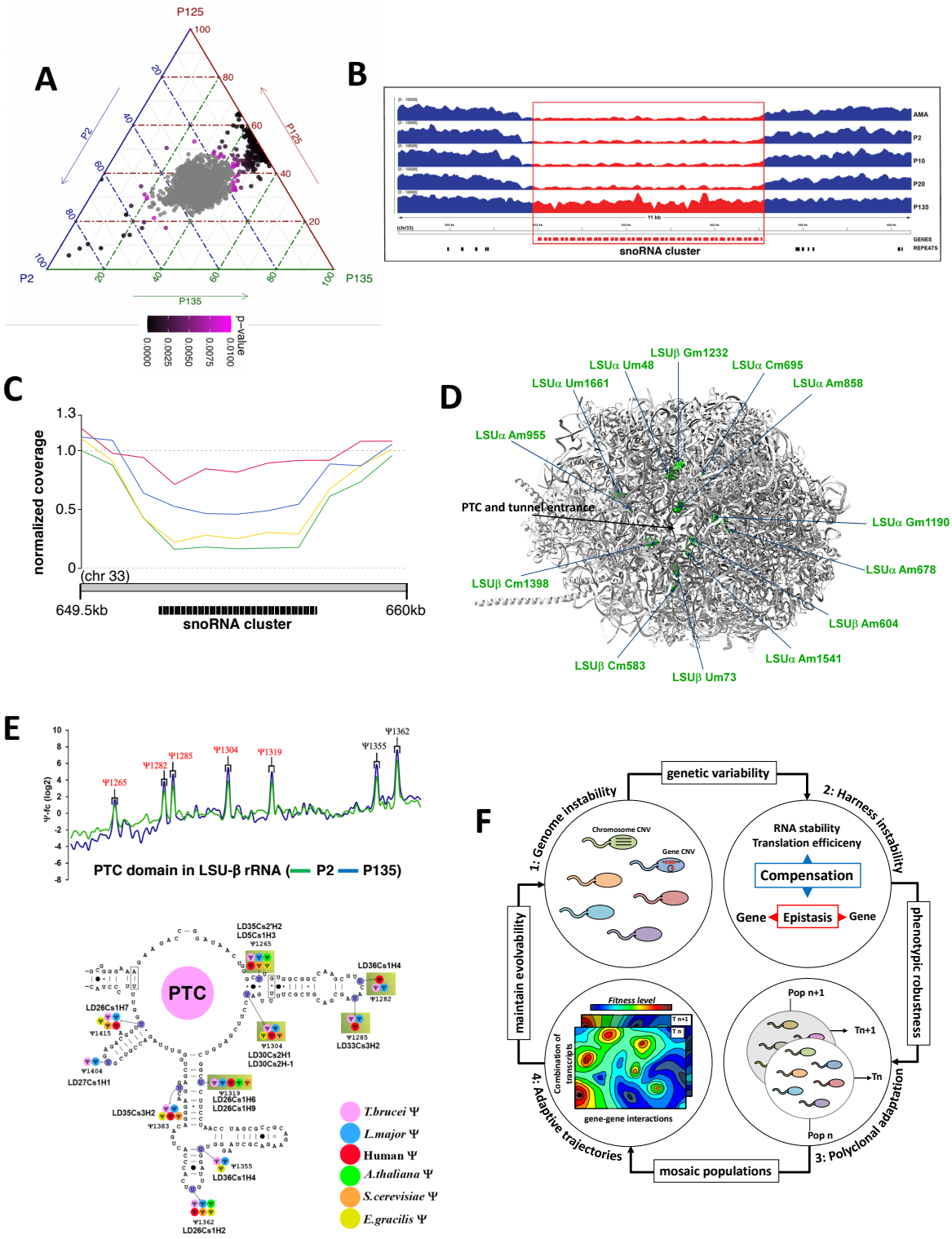
310 **Fig. 3**



311
312 **Genetic and transcriptomic analyses of the NIMA-like kinase null mutant.** (**A**) Growth
313 analysis. WT and heterozygous NIMA +/- mutants generated by CRISPR/Cas9 gene editing (left
314 panel). Transgenic NIMA+/- parasites transfected with empty vector (NIMA+/-/m) or vector
315 encoding for the NIMA-like kinase gene (NIMA+/-/+) (right panel). The doubling time of parasites
316 in logarithmic culture phase is shown. (**B**) Heatmap of the scaled normalized RNAseq counts of
317 the genes differentially expressed in the P20 clones 1 and 8 (WT) with respect to the spontaneous
318 NIMA null mutant (spo-KO) clones 3, 6, 4, 7, 9, 10. A log2 fold change > 0.5 with adjusted p-
319 value < 0.01 was considered significant. Darker levels of blue reflect higher expression levels as
320 indicated by the legend. (**C**) Double-ratio scatter plot. The plot represents the ratio of the mean
321 DNA (*x*-axis) and RNA (*y*-axis) sequencing read counts between the clones that lost the NIMA-
322 kinase gene (CL3-4-6-7-9-10) and the NIMA-kinase wild-type clones (CL1 and CL8). Each dot
323 represents an individual gene. The marginal distributions for DNA and RNA ratio values are
324 displayed along the *x*- and *y*-axes. The color indicates the statistical significance level of the genes'
325 double-ratio scores (i.e. RNA ratio divided by DNA ratio) as indicated in the legend. The NIMA-

326 like kinase homolog Ld1S_360735800 is labeled in red. The vertical and horizontal dotted lines
327 indicate DNA and RNA ratio values of 1, while the diagonal dashed line specifies the bisector.
328 The blue line represents a linear regression model built on the DNA and RNA ratio values and
329 measuring a Pearson correlation value of 0.65. (**D**) Functional enrichment analysis of the biological
330 process Gene Ontology (GO) terms for all genes in panel C showing a statistically significant
331 double-ratio score. The node color mapping is ranging from yellow to dark orange to represent
332 increasing significance levels, or lower adjusted p-values. White nodes are not significant. 1:
333 organic substance metabolic process, 2: nitrogen compound metabolic process, 3: cellular
334 metabolic process, 4: organic cyclic compound metabolic process, 5: primary metabolic process,
335 6: cellular nitrogen compound metabolic process, 7: heterocycle metabolic process, 8: cellular
336 aromatic compound metabolic process, 9: macromolecule metabolic process, 10: nucleobase-
337 containing compound metabolic process, 11: nucleic acid metabolic process.
338

Fig. 4:


A
P125
P2
P135
p-value
0.0000
0.0025
0.0050
0.0075
0.0100
B
AMA
P2
P10
P20
P135
11 kb
snoRNA cluster
GENES
REPEATS
C
normalized coverage
1.3
1.0
0.5
0
(chr 33)
649.5kb
660kb
snoRNA cluster
D
LSUβ Gm1232
LSUα Um48
LSUα Cm695
LSUα Um1661
LSUα Am858
LSUα Am955
PTC and tunnel entrance
LSUα Gm1190
LSUα Am678
LSUβ Cm1398
LSUβ Am604
LSUα Am1541
LSUβ Cm583
LSUβ Um73
E
Ψ-fc (log2)
Ψ1265
Ψ1282
Ψ1285
Ψ1304
Ψ1319
Ψ1355
Ψ1362
PTC domain in LSU-β rRNA (P2 P135)
PTC
LD35Cs2'H2
LD5Cs1H3
Ψ1265
LD36Cs1H4
Ψ1282
LD26Cs1H7
Ψ1415
Ψ1304
LD30Cs2H1
LD30Cs2H-1
Ψ1285
LD33Cs3H2
Ψ1404
LD27Cs1H1
Ψ1319
LD26Cs1H6
LD26Cs1H9
LD35Cs3H2
Ψ1383
Ψ1355
LD36Cs1H4
Ψ1362
LD26Cs1H2
T.brucei Ψ
L.major Ψ
Human Ψ
A.thaliana Ψ
S.cerevisiae Ψ
E.gracilis Ψ
F
genetic variability
1: Genome instability
2: Harness instability
Chromosome CNV
Gene CNV
RNA stability
Translation efficiceny
Compensation
Gene
Epistasis
Gene
maintain evolvability
phenotypic robustness
Fitness level
Combination of transcripts
gene-gene interactions
Pop n+1
Tn+1
Tn
Pop n
4: Adaptive trajectories
3: Polyclonal adaptation
mosaic populations

341 **snoRNA genes are amplified in long-term adapted parasites and promote rRNA**
342 **modification.** (**A**) Ternary plot showing for each gene the relative abundance in culture passage
343 P2 – P125 and P135. The axes report the fraction of the normalized gene coverage in each sample,
344 with each given point adding up to 100. Dots with color ranging from pink to black indicate
345 significant gene CNVs (p-value < 0.001). (**B**) Recovery of the snoRNA gene cluster CNVs during
346 culture adaptation. The panel illustrates a genome browser representation of the sequencing depth
347 measured in the samples at the indicated passage. Gene annotations and the predicted repetitive
348 elements are indicated. (**C**) Line plot showing the normalized sequencing coverage (y-axis) over
349 the snoRNA cluster region on chromosome 33 (x-axis) for the amastigote sample (AMA, green)
350 and three independent amastigote isolates (AMAH154,red; AMA07142, yellow; AMA1992, blue)
351 obtained from different hamster infections. (**D**) Hypermodified Nm sites are located around the
352 functional domains of the ribosome. The complete stoichiometry of each Nm site was measured
353 by RiboMeth-seq in P2 and P135 *L. donovani* strains as presented in **Table S21**. The
354 hypermodified Nm sites are highlighted in green and their identity is indicated on the 3D structure
355 of the *L. donovani* large subunit ribosome based on the previously deposited cryo-EM coordinates
356 (Protein Data Bank (*39*) accession 6AZ3) (*40*). The location of the PTC and mRNA entrance tunnel
357 are shown. (**E**) Hypermodified Ψs are present in the functional domains of the ribosome. The
358 relative changes in Ψ level was measured using Ψ-seq and is presented in **Table S22**.
359 Representative line graph of the fold change in rRNA pseudouridinylation level (Ψ-fc, log2) is
360 presented for P2 (green line) and P135 (blue line) for the PTC domain in LSU- Ψ rRNA. The
361 positions where the Ψ level is increased in four replicates are indicated in red. The location of Ψ
362 sites in the rRNA is depicted on the secondary structure (*19*). Hypermodified sites are highlighted
363 in boxes. The snoRNA guiding on each Ψ is indicated. The color code for each Ψ site is indicative

364 of the organism where it was already reported. (**F**) Model of *Leishmania* polyclonal adaptation.
365 (1) *Leishmania* intrinsic genome instability generates constant genetic variability. (2) Epistatic
366 interactions between gene CNVs and compensatory responses at the level of RNA stability and
367 translation efficiency eliminate toxic gene dosage effects and harness genome instability. (3) This
368 mechanism generates phenotypic robustness while at the same time maintaining genetic
369 variability, thus allowing for polyclonal adaptation. (4) The genetic mosaic structure allows for
370 distinct adaptive trajectories (T) inside and in-between adapting populations, a process that
371 conserves genetic heterogeneity and thus evolvability of the population despite constant selection.
372

Supplementary Materials for

Genome instability drives epistatic adaptation in the human pathogen *Leishmania*

Giovanni Bussotti[1,2], Laura Piel[2], Pascale Pescher[2], Malgorzata A. Domagalska[3], K. Shanmugha Rajan[4], Tirza Doniger[4], Disha-Gajanan Hiregange[5], Peter J Myler[6-10], Ron Unger[4], Shulamit Michaeli[4], and Gerald F. Späth[2,*]

Correspondence to: Gerald.spaeth@pasteur.fr

This PDF file includes:

Materials and Methods
Figs. S1 to S14
Captions for Data S1 to S21

Other Supplementary Materials for this manuscript include the following:

Data S1 to S21

[**Table S1: Mapping statistics of the dataset for the 204 clinical isolates.**

Table S2: Position and normalized coverage values of the collapsed CNV regions.

Table S3: Repetitive elements.

Table S4: Most positive and most negative LdBPK gene coverage correlations.

Table S5: Correlation map data.

Table S6: Correlation network data.

Table S7: Network hub genes.

Table S8: Collapsed CNVs showing highest read depth in the 204 L. donovani field isolates.

Table S9: Genes amplified in phylogenetically distinct clades.

Table S10: Normalized gene coverage values across ISC samples.

Table S11: Polysomy level analyses of two sets of Ld1S samples: (i) AMA to P135 isolates; (ii) four AMA isolates.

405 **Table S12: Ld1S annotation.**

406 **Table S13: Definitions of gene clusters in two sets of Ld1S samples: (i) AMA-P135 isolates;**

407 **(ii) P20 clones isolates.**

408 **Table S14: Gene CNVs and cluster CNVs detected during Ld1S culture adaptation.**

409 **Table S15: Gene CNVs and cluster CNVs detected in the eight Ld1S P20 clones.**

410 **Table S16: Single nucleotide variants (SNVs) in Ld1S P20 clones.**

411 **Table S17: Data of the differential expression analysis.**

412 **Table S18: DNA and RNA double-ratio data.**

413 **Table S19: P125/P135 gene coverage ratio values.**

414 **Table S20: The complete stoichiometry of Nm sites in *L. donovani* rRNA.**

415 **Table S21: The relative fold-change of Ψs in *L. donovani* rRNA between P135 and P2**

416 **parasites.**]

417
418
419
420

421 **Materials and Methods**

422 ***L. donovani* strains, culture conditions and cell cloning**

423 Culture-adapted *L. donovani* field isolates maintained for more than 20 *in vitro* passages (P) from
424 the Indian Sub-Continent (ISC) and previously subjected to whole genome sequencing analysis
425 were used in our study (*12*). *Leishmania (L.) donovani* strain 1S2D (MHOM/SD/62/1S-CL2D)
426 (Ld1S) was obtained from Henry Murray, Weill Cornell Medical College, New York, USA and
427 maintained by serial passages in hamsters. Amastigotes were recovered from infected hamster
428 spleen and differentiated into promastigotes in M199 medium supplemented with 10% FCS,
429 25mM HEPES pH 6.9, 4mM $NaHCO_3$, 1 mM glutamine, 1x RPMI 1640 vitamin mix, 0.2 µM
430 folic acid, 100 µM adenine, 7.6 mM hemin, 8 µM biopterin, 50 U ml^{-1} of penicillin and 50 µg ml^{-}
431 ^{1} of streptomycin. Promastigotes were serially passaged once stationary phase was reached for
432 either 2, 10, 20, or 135 passages corresponding to respectively 20, 60, 190, and 3,800 generations.
433 An independent *L. donovani* isolate was grown in culture for over 33 months until P125
434 (corresponding to approximately 3,500 generations). Three independent splenic amastigotes
435 populations were recovered from as many infected hamsters using the same culture conditions.
436 Serial dilutions of passage 20 (P20) promastigotes were plated on M199 Agar plates and 8 clones
437 were recovered and expanded for 2 passages in liquid culture (*10*).

438

439 **Nucleic acid extraction and deep sequencing**

440 All sequencing analyses were performed using Illumina short-read technology. DNA extraction
441 and sequencing protocols of the *L. donovani* field isolates from the Indian Sub-Continent (ISC)
442 are described in (*12*) DNA extraction and sequencing analysis of *L. donovani* LD1S amastigotes
443 (AMA) and derived promastigotes at passages P2, 10, and 20 are described in (*10*). For new
444 amastigote isolates (AMAH154, AMA1992, AMA07142) and promastigote strains (P125, P135),

445 DNA extractions were performed using DNeasy blood and tissue kits from Qiagen according to
446 the manufacturer's instructions. Nucleic acid concentrations were measured and the DNA quality
447 was evaluated by measuring the OD ratios 260/280 and 260/230 and by agarose gel-
448 electrophoresis. Between 2 to 5µg of DNA were used for sequencing.

449 For samples AMAH154, AMA1992, AMA07142, P125 and P135, short-insert, paired-end
450 libraries were prepared with the KAPA Hyper Prep Kit (Kapa Biosystems). The libraries were then
451 sequenced using TruSeq SBS kit v3-HS (Illumina Inc., CA, USA). Multiplex sequencing was
452 performed using HiSeq 2000 flowcell v3, generating 2×101bp paired-end reads, according to
453 standard Illumina procedures. Reads were deposited in the Sequence Read Archive (SRA)
454 database (*37*) and are publicly available under accession no PRJNA605972**.**

455

456 **Whole genome sequencing data analysis**

457 Genomic reads were aligned and the nucleotide sequencing depth was measured and normalized
458 as previously described (*41*). For the genome-wide detection of CNVs across the 204 ISC isolates,
459 the LdBPKv2 reference genome was partitioned into adjoining intervals of a fixed length of 300bp
460 (bins), and the depth of coverage of each bin was measured and compared across samples. The
461 mean sequencing coverage normalized by the median chromosome coverage was computed for
462 each bin as already described (*41*). In **Fig. 1A** and **1B** and to compute the genome fraction of
463 amplified or depleted regions, all maxi-circle DNA bins as well as the bins with a median
464 normalized coverage > 0.01 but mean read mapping quality score (MAPQ) < 50 were discarded.
465 We referred to genomic bins as being deleted, depleted or amplified respectively for normalized
466 sequencing coverage values ≤ 0.01, < 0.5 and > 1.5. **Fig. 1B** shows the copy number variant bins
467 across the samples. To account for differences in ploidy levels, the bin copy number was evaluated

468 as the bin sequencing coverage normalized by the chromosome median sequencing coverage (i.e.
469 copy number per haploid genome). Bins showing a copy number variation of at least 1 across the
470 204 ISC isolates were considered in the analysis. The heatmap color scale reflects the difference
471 in bin normalized read depth with respect to the minimum value measured across all genomes.
472 Black indicates no variation. White, blue and red correspond to increases in normalized coverage
473 levels of respectively 1, 2 and 3. Coverage variations greater than 3 were down sized to 3. The
474 analyses shown in **Fig. 1F** and **1G** rely on CNV regions potentially spanning multiple genomic
475 bins aggregated into individual larger units. We refer to such grouped bins as "collapsed bins".
476 Groups of genomic CNV bins separated by less than 1,200 bases were merged together, and their
477 normalized sequencing coverage averaged. To account for potential unannotated UTR regions in
478 **Fig. 1F** and **1G**, CNV regions were considered intergenic if located at a distance of at least 100bp
479 from annotated genes. The line plot in **Fig. 4C** illustrates the sequencing coverage in 1,000bp long
480 bins normalized by chromosome 33 median sequencing coverage.

481 Likewise, sequencing coverage was used to estimate and compare gene copy numbers. The
482 mean sequencing coverage of each gene was estimated and normalized as previously published
483 (*41*). Reads with a mean MAPQ score < 50 were filtered because not directly quantifiable with the
484 sequencing technology we used. Genes with a variation of normalized coverage > 1 in the 204
485 field isolates were considered to be copy number variant. The heatmap in **Fig. 1E** includes the
486 genes (rows) showing convergent amplification (normalized coverage > 1.5) across 191 ISC core
487 group (*12*) isolates (columns). Genes amplified only in isolates sharing a common ancestor were
488 excluded from the analysis. **Fig. 2A**, **2C, 4A** and **S9** show gene copy number variations across
489 different data sets. Genes supported by reads with and average MAPQ score < 50 were grouped
490 into clusters based on sequence similarity, quantified as cluster entities and evaluated for possible

491 variation across samples (gene cluster CNVs). Briefly, the stranded sequences of low MAPQ genes
492 were grouped with cd-hit-est (version 4.6.8) (*42*) with options "-s 0.9 -c 0.9 -r 0 -d 0 -g 1" selecting
493 clusters of highly similar genes. Samtools view (version 1.3) and BEDTools coverage (version
494 2.25.0) were used to measure the mean sequencing depth of the individual low MAPQ genes and
495 were run, respectively, with options "-F 1028" and "-d -split". For the mean coverage estimate
496 possible intragenic gap regions were not considered. The mean coverage of each gene was
497 normalized by the median coverage of its chromosome. Eventually, the normalized mean coverage
498 of all genes in a cluster was averaged to compute the cluster support. **Fig. 2A** shows the gene and
499 cluster increasing in copy number between amastigote (AMA) and P135, using as a cutoff an
500 increment in normalized coverage of at least 0.5. **Fig. 2C** represents the gene and cluster CNV
501 detected in the clones using as a cutoff a minimum variation in normalized coverage of at least
502 0.5. Black indicates no variation. White, blue and red reflect increasing values in normalized
503 coverage levels of respectively 1, 2 and 3. Coverage variations greater than 3 were down sized to
504 3. For the analysis illustrated in **Fig. 4A** the coverage of each gene was evaluated as the number
505 of mapping reads excluding duplicates, increased by a pseudo-count of 1, and normalized by the
506 median number of reads per gene in the respective samples. In **Fig. S9** the mean gene sequencing
507 coverage values were normalized by the median of the mean gene coverage values in the respective
508 P125 and P135 samples. The genes with mean MAPQ < 50 in either sample were discarded.

509 To represent the sequencing coverage of selected genes and genomic areas we used the
510 Integrative Genomics Viewer (IGV) (*43*). In **Fig. 2B**, **4B**, **S6B** and **S7A** the sequencing coverage
511 tracks were produced with bamCoverage from the deepTools suite (*44*) (version 2.4.2) with
512 options "--binSize 10 --smoothLength 30" and ignoring duplicated reads. Normalization of reads

513 per kilobase per million reads (RPKM) was applied on separate chromosomes to render the
514 coverage comparable across samples and ploidy levels.

515 We determined chromosome aneuploidy in our samples sets based on sequencing coverage.
516 In **Fig. S6A** and **S10,** the normalized sequencing coverage was binned in contiguous 2,500bp
517 windows for each sample and for each chromosome, and the distribution of the windows' mean
518 coverage scores were displayed. Sequenced positions where more than 50% of the reads showed
519 a MAPQ lower than 50 were not considered in the analysis. To estimate chromosome copy number
520 differences, the window coverage was normalized by the median coverage of the chromosome
521 showing a stable disomic level across samples (chromosomes 21-25 respectively for **Fig. S6A** and
522 **S10**) and multiplied by two. The distribution of the mean window coverage was compared with
523 the R (*45*) Wilcoxon test and the p-values of the comparisons are reported in **Table S11**.

524

525 **Reference genomes**

526 The analyses regarding the use of the Sudanese *L. donovani* strain 1S2D (Ld1S) were performed
527 utilizing its PacBio genome assembly available from the NCBI website under the biosample (*46*)
528 accession n° SAMN07430226 and the bioproject (*46*) accession n° PRJNA396645. To produce
529 Ld1S genome annotation we combined different approaches. We ran the Companion pipeline (*47*)
530 without using transcript evidence and exploiting *L. major* Friedlin as a reference organism to
531 specify the models for gene finding and functional annotation transfer. The predicted gene
532 identifiers were renamed to match chromosomal localization. Then we complemented Companion
533 annotations with LdBPKv2 (*7*) homology-based predictions. We used NCBI-tblastn (version
534 2.2.28) (*48*) to search the amino acid sequence of LdBPKv2 genes with an e-value threshold of
535 0.01. To enable the screening of non-coding sequences we also ran gmap (version 2015-07-23)

536 with options "--npaths 30 --no-chimeras --nosplicing –prunelevel --min-trimmed-coverage 0.9 --
537 min-identity 0.9" searching Ld1S homologs of all LdBPKv2 genes. All the significant NCBI-
538 tblastn and gmap predictions that were not overlapping on the same strand with Companion
539 annotations were retained. BEDTools merge (version 2.25.0) was used to combine overlapping
540 predictions into single gene annotations spanning all of the combined predictions. Eventually
541 NCBI-blastn was used to search in the Ld1S intergenic space the *Leishmania major* Friedlin
542 UsnRNA, snoRNA, SLRNA and 7SL non-coding RNA classes. Overall, the Ld1S genome
543 encodes for 10,532 genes: 8,850 defined by the Companion pipeline, 104 defined by homology
544 with LdBPKv2 genes, and 1,578 defined by homology with *L. major* Friedlin ncRNA classes.

545 Ld1S gene functions were inferred adopting a hierarchical approach. Firstly, we transferred
546 the function of *L. donovani* strain BPK282A1 (LdBPK282A1) homolog genes available from
547 TriTrypDB (*49*) (downloaded the 05/07/2019). To this end, OrthoFinder (*50*) (version 2.2.7) was
548 used with the DIAMOND (*51*) search program to establish orthology relation between
549 LdBPK282A1 (downloaded from the Sanger FTP server
550 ftp://ftp.sanger.ac.uk/pub/project/pathogens/ on the 12/10/2018) and Ld1S gene annotations.
551 NCBI-blastn was run with options "-dust no -soft_masking false -evalue 10" to scan and try to
552 rescue the low complexity genes were OrthoFinder failed to determine homology. When "one-to-
553 many" homologs occurred (i.e. one Ld1S gene matching multiple significant LdBPK282A1
554 homologs), the function of all individual homologs was concatenated in a single, not-redundant
555 functional assignment. For the genes lacking functional annotation, or reported as "hypothetical
556 proteins" we assigned the function reported in the Companion GAF output file or the transcript
557 type if no description was available. Then, if no annotation was found other than "hypothetical
558 proteins", we sought to infer the gene function using HMMer (*52*) (version 3.1b2) to scan the gene
559 protein sequences against EGGNOG kinetoplastida database (*53*) (version 4.5)

560 (http://eggnogdb.embl.de/download/eggnog_4.5/data/kinNOG/), and ultimately combining the
561 annotated functions in EuPathDB (*54*) and UniProt (*55*) (downloaded the 20/09/2019). To extract
562 the UniProt function annotations we queried the Gene Ontology Annotation (GOA) (*56*) database
563 with the Uniprot identifiers of the LdBPK282A1 homologs of Ld1S genes.

564 In order to assign the Gene Ontology Identifiers (GO IDs) we combined the GOA-derived
565 identifiers with the ones available from the corresponding orthologs in target species: LdBPK, *L.*
566 *infantum*, *L. major*, *L. mexicana*, *Typanosoma brucei brucei 927* (Tbru) and *Typanosoma cruzi*
567 (Tcru). For each target species we retrieved both the "curated" and "computed" GO IDs from
568 TriTrypDB on the 11/09/2019. OrthoFinder with the DIAMOND search program was applied to
569 establish orthology between the genes in Ld1S and in target species. In "one-to-many" orthology
570 relations we concatenated all the non-redundant GO IDs from all the homologs. The GO IDs were
571 then assigned based on the hierarchy: LdBPK curated > LdBPK GOA > *L. infantum* curated > *L.*
572 *major* curated > *L. mexicana* curated > Tbru curated > Tcru curated > LdBPK computed > *L.*
573 *infantum* computed > *L. major* computed > *L. mexicana* computed > Tbru computed > Tcru
574 computed. The GO IDs were assigned if not present in any higher rank GO ID data set. The GO
575 IDs of snoRNAs, UsnRNA, SLRNA and 7SL classes defined by homology with *L. major* Friedlin
576 genes were manually attributed. Overall, we assigned biological process (BP), molecular function
577 (MF) and cellular component (CC) GO IDs to 5,246, 4,521 and 7,236 Ld1S genes.

578 All analyses regarding the 204 Indian sub-continent (ISC) isolates (*12*) used the Nepalese
579 PacBio *L. donovani* (LdBPKv2) genome and annotation as reference (*7*). LdBPKv2 gene function
580 was inferred adopting a hierarchical approach similar to the one described for Ld1S. Higher
581 priority was given to TriTrypDB, followed by Companion GAF (available from (*7*)) and
582 EGGNOG functional annotations.

583

584 **Repeat analysis**

585 We annotated the repetitive elements in the LdBPKv2 genome producing a comprehensive dataset
586 (**Table S3**). To build this dataset, we merged the repetitive elements identified by different
587 approaches in three distinct repeat sets and present with at least 5 copies in the genome.

588 The first set of repetitive elements accounts for the interspersed repeat elements predicted
589 with RepeatMasker (*57*) (version 4.0.8) with options "-e crossmatch -gff -xsmall -s", scanning the
590 genome to identify short simple/low complexity repetitive elements and interspersed repetitive
591 elements. RepeatMasker was run in combination with the Repbase (*58*) library to identify
592 *Leishmania*-specific and ancestral repeats, and with the output of RepeatModeler (version 1.0.11)
593 (*59*), a pipeline to identify transposable elements run with option "-engine ncbi". Redundant,
594 RepeatModeler-defined elements overlapping with Repbase elements were discarded.
595 RepeatMasker output was processed collapsing self-inclusive low-complexity motifs to the
596 shortest identical sub-motif (e.g. TAGTAG -> TAG) and merging together overlapping repeats
597 elements of the same type (**Table S3**).

598 The second set of repetitive elements consists in the DNA motifs enriched at the boundaries
599 of the CNV regions we detected in the core group of 191 ISC *L. donovani* field isolates, excluding
600 the "Yeti" strains (*12*). Such boundaries were defined as the genomic regions spanning 1kb and
601 flanking both sides of each CNVs. The CNVs defined in the core group include the 300bp genomic
602 bins with average read MAPQ score > 50, and showing a variation of normalized sequencing
603 coverage score > 1 in the sample set. The bins localized within 6kb were collapsed and their
604 sequencing scores averaged. Significant DNA motifs were detected using MEME (version 5.0.4)
605 (*60*) with option "-dna -revcomp -nmotifs 20 -evt 0.01 -mod anr" and specifying a custom

606 background model of order 3 with fasta-get-markov, a tool included in the MEME suite. The
607 sequences we used to build such background model were randomly extracted from the LdBPKv2
608 genome with BEDTools shuffle (*61*) (version 2.25.0) in order to resemble the ones scanned by
609 MEME in terms of number and size, but not including them. The DNA motifs predicted with
610 MEME were then mapped genome-wide with MAST using the options "-hit_list -comp -remcorr
611 -mt 0.0000001".

612 The third set of repetitive elements includes the DNA motifs enriched at the boundaries of
613 tandem repeat cluster (TRC) regions. TRCs are characterized by the tandem repetition of large
614 genomic segments. To detect TRCs in the LdBPKv2 genome we streamlined a pipeline including
615 several steps and relying on several bioinformatics tools. First, we detected inexact repetitive
616 elements on individual chromosomes. For this purpose we ran nucmer (version 3.23) (*62*) with
617 options "--maxmatch --nosimplify", and show-coords (version 3.23) (*62*) with options "-r -T -o -l
618 -d -I 80 -L 50". Possible overlapping predictions were collapsed with BEDTools merge (*61*). To
619 remove distantly related elements the resulting elements were re-clustered by sequence similarity
620 with cd-hit-est (version 4.6.8) with options "-d 0 -g 1". Then to refine the element boundaries we
621 compared all-versus-all sequences in each cluster with NCBI-blastn (version 2.2.28) (*48*) with
622 options "-dust yes -soft_masking true -evalue 0.01", retaining the elements longer than 50bp. Next,
623 we processed each cluster grouping the sequences based on a maximum distance cutoff of 2kb,
624 thus ensuring that the repetitive element occur in the same genomic area. Then the clusters found
625 to be intersecting each other were pooled together to generate the TRCs. Eventually the TRCs were
626 processed enforcing a minimum distance between an element and the one downstream of at least
627 100bp. The genomic space between the elements inside the TRC and the regions of 1kb flanking

628 each TRC were evaluated to identify enriched DNA motifs. Significant motifs were detected with
629 MEME and mapped genome-wide with MAST as already described for the CNV boundaries.
630
631 **CNV association analyses**
632 We evaluated weather LdBPKv2 CNVs are significantly associated with genes and intergenic
633 regions utilizing GAT (*63*), a tool for testing the significance of overlap of genomic intervals. In
634 this analysis we considered 10,000 simulations in which we assigned random positions to the
635 collapsed bin CNV regions. For each simulation we evaluated the number of nucleotides
636 overlapping the gene and intergenic regions, thus implementing the null models that the CNV set
637 of intervals is placed independently. The genomic regions accessible for simulation include (i) the
638 36 chromosomes of the LdBPKv2 genome, excluding the assembly gaps, (ii) the genomic positions
639 where more than 50% of the reads have a MAPQ < 50, and (iii) the regions shorter than 20kb not
640 containing any CNV. To account for possible non-annotated UTR elements, we expanded the gene
641 coordinates by 100bp on both sides, and consequently shrunk the intergenic regions.
642 GAT was also used to determine the association between the LdBPKv2 CNV boundaries
643 and the comprehensive repeat dataset. For each of the 10,000 simulations we evaluated the number
644 of CNV boundaries overlapping each repeat type. For this purpose, all low-complexity motifs were
645 collectively considered as a single class named "simpleRepeats". The CNV boundaries included a
646 1kb area flanking the CNV, plus the segments of 150bp spanning from the CNVs extremities and
647 moving toward their center. Eleven repeat types were found significantly associated to CNV
648 regions (**Fig. 1C**).
649
650 **Network analysis**

651 To explore the interaction between gene CNVs we performed a network analysis. In the graph
652 shown in **Fig. 1D,** the edges indicate significant gene CNVs correlation with adjusted p-value
653 lower than 0.01. Nodes, representing the gene CNVs, are colored according to the MCL clustering
654 (R package version 1.0) (*64*) of the gene CNVs absolute correlation values. The clustering was
655 performed using respectively 2 and 3 for the expansion and inflation MCL parameters. The
656 centroid and the standard deviation were computed for each cluster. Nodes mapping more than 2
657 standard deviations away from the cluster centroid were labeled as outliers and not shown. The
658 network was produced using the R library igraph (R package version 1.1.2) (*65*), using the
659 "layout_with_fr" function, placing vertices on the plane using the force-directed layout algorithm
660 by Fruchterman and Reingold (*66*).
661
662 **Phyogenetic analysis**
663 The phylogenetic tree in **Fig. 1E** and the cladogram in **Fig. S5** include respectively the 191 ISC
664 core isolates and the full 204 ISC set (*12*) and were calculated with raxmlHPC-PTHREADS
665 (version 8.2.8) (*67*) with options '-m ASC_GTRGAMMA --asc-corr=lewis', with 10 starting trees
666 and 10,000 rapid bootstrap runs. The program was run providing single nucleotide variant (SNV)
667 positions returned by freebayes (version v1.0.1-2-g0cb2697) (*68*), and further filtered to produce
668 high-quality predictions. Freebayes was run with options "--no-indels --read-indel-limit 0 --no-
669 mnps --no-complex --read-mismatch-limit 3 --read-snp-limit 3 --hwe-priors-off --binomial-obs-
670 priors-off --allele-balance-priors-off --min-alternate-fraction 0.05 --min-base-quality 5 --min-
671 mapping-quality 50 --min-alternate-count 2 --pooled-continuous". Then the variants mapping to
672 the maxi-circle DNA or presenting multiple alternate alleles were discarded. Additionally, all
673 variants with an alternate allele frequency < 0.9 were removed. Next, SNV predictions where the
674 mean MAPQ of the reads supporting the alternate or the reference allele was < 20 were also

675 removed. We filtered SNVs with sequencing coverage above or below 4 median absolute
676 deviations (MADs) from the median chromosome coverage. Eventually, SNVs mapping inside
677 homopolymers were filtered as described (*41*). To ease the visualization, the branch representing
678 the reference genome was removed from the tree and the cladogram. The phylogenetic tree and
679 the associated gene CNV heatmap were displayed with the R library ggtree (version 2.0.1) (*69*). A
680 similar raxmlHPC-PTHREADS/freebayes approach was adopted to determine the evolutionary
681 relationship between the P20 clones (**Fig. 2D**). As compared to the analyses in **Fig. 1E** and **S5**, we
682 increased the number of starting trees (10,000), used 10,000 standard bootstrap iterations, loosened
683 the stringency of the alternate allele frequency filtering (< 0.1), and displayed the tree with iTol
684 (version 5) (*70*).
685
686 **RNAseq analysis**
687 RNA was prepared and RNA sequencing (RNAseq) performed using an Illumina Hiseq 2000
688 platform and TruSeq v3 kits as previously described (*10*) We assessed the statistical significance
689 of quantitative gene expression differences between WT clones and spontaneous NIMA-like
690 kinase null mutant clones (spo-KO) with DESeq2 (version 1.14.1) (*71*). Genes with an adjusted p-
691 value < 0.01 and a log2 fold change above 0.5 or below -0.5 were deemed differentially expressed.
692 Maxi-circle DNA genes were not considered in the analysis. The gene expression was quantified
693 by HTSeq (version 0.6.1) (*72*) while the reads were mapped as previously described (*41*).
694 To investigate post-transcriptionally regulated transcripts (**Fig. 3C**), we considered all the
695 genes excluding the ones with 0 RNAseq read counts and the ones encoded on the maxi-circle
696 DNAs. All HTSeq gene read counts estimated for RNAseq (RNA) and whole genome sequencing
697 (DNA) were incremented by 1 pseudocount. For RNA and DNA data, the mean read counts in

698 spo-KO clones was divided by the mean read counts in WT clones (representing the ratio scores
699 that are displayed respectively on the y- and the x-axes in **Fig. 3C**). Both the numerator and
700 denominator were normalized by the average library size (defined as the total number of reads
701 mapping on all genes). The double-ratio score was estimated as the ratio between the RNA and
702 DNA ratios. Genes with double-ratio score < 1 and p-value < 0.01 and were considered
703 "destabilized" (i.e. less RNA than expected considering the DNA amount). Genes with double-
704 ratio score > 1 and p-value < 0.01 were considered "stabilized" (i.e. more RNA than expected
705 considering the DNA amount).
706
707 **Gene Ontology (GO) term enrichment analysis**
708 Cytoscape (version 3.7.2) (*73*) was used in combination with *BiNGO* (version 3.0.3) (*74*) to detect
709 biological process GO terms enriched amongst the genes with significant double-ratio scores. In
710 **Fig. 3D** node color mapping is ranging from yellow to dark orange to represent increasing
711 significance levels, or lower adjusted p-values. White nodes are not significant. The BiNGO
712 analysis was run using the "go-basic.obo" ontology file (release 2019-10-07) (*75, 76*) available
713 from http://geneontology.org/docs/download-ontology/, and as a reference the set of 8,880 genes
714 reporting a mean of normalized counts taken over all samples > 0 ("baseMean" metric in the
715 DESeq2 analysis).
716
717 **PCR analyses**
718 The presence of a 10.3kb genomic deletion on chromosome 36 (coordinates 36:574,041 – 584,350)
719 comprising a conserved NIMA kinase homolog (LdBPK_361580.1) was monitored by PCR
720 amplification of a diagnostic fragment of 2kb. Briefly, genomic DNA was extracted from

721 amastigotes purified from hamster spleen (*77*) or logarithmic promastigotes in culture using the
722 DNeasy Blood and Tissue kit (Qiagen) according to manufacturer's instructions. DNA quality was
723 assessed by monitoring the 230/260 and 280/260 ratios and concentration was determined by
724 spectrophotometry measuring the absorbance at 260nm. Genomic DNA was subjected to PCR
725 analysis using LongAmp kit (New England Biolabs) according to the manufacturer's instructions.
726 Primers were designed targeting genomic sequence from either side of the deleted area (forward
727 primer (F): AGACAGACAGCAATGCGTGAT; reverse primer (R):
728 ACAGACGTCTGTCCGTGCTT) or the NIMA kinase open reading frame (F:
729 ATGTCGGGGGGTAACGCC; R: TCACTCCTTTGCGGGTAGAGCA). The PCR mix was
730 composed of 1x LongAmp Taq reaction buffer, 300µM dNTP each, 0.4µM of each primer, 10ng
731 of genomic DNA and 5 units of LongAmp Taq DNA polymerase. The thermal cycling conditions
732 were: initial denaturation at 94°C for 30s, followed by 30 cycles of 94°C for 30s, 60°C for 30s and
733 65°C for 10min. A final extension step was performed for 10min at 65°C.

734 PCR was performed for the CRISPR/cas9-mediated gene deletion assay. Based on the
735 sequence of the NIMA kinase target gene (LdBPK_361580.1), single-guide RNA (sgRNA) were
736 amplified using Expand High Fidelity polymerase (Roche) with 0.2mM dNTP each, 1x Expand
737 High Fidelity buffer with $MgCl_2$ (15 mM final concentration), 2µM each of primers G00 (sgRNA
738 scaffold) and either 5'-sgRNA primer
739 (gaaattaatacgactcactataggTAGAAAGAACGAAGTGATCGgttttagagctagaaatagc) or 3'-sgRNA
740 primer (gaaattaatacgactcactataggAAGAGGAAGAAGATCACCAAgttttagagctagaaatagc) in 20µl
741 final volume. PCR steps were: 30s at 98°C followed by 35 cycles of 10s at 98°C, 30s at 60°C, and
742 15s at 72°C. Sterilization of PCR products was performed at 94°C for 5min. PCR amplification
743 of the DNA donor was performed using 30ng of pT plasmid containing the blasticidin resistance

744 cassette as template (*78*). The PCR mix contained 0.2mM dNTP each, 1x Expand high Fidelity
745 buffer with $MgCl_2$ (3.375mM final concentration), 3% DMSO, 1 unit of Expand High Fidelity
746 polymerase and 2µM of each primer (F:
747 CTCTCCACCTCTGAATCTACTACGGTTTTCgtataatgcagacctgctgc; R:
748 CGCAAAGGGTGCCGCCACATCGCAAACGCGccaatttgagagacctgtgc) in 40µl final volume.
749 PCR steps were: 5min at 94°C followed by 40 cycles of 30s at 94°C, 30s at 65°C, 2min 15s at
750 72°C followed by a final elongation step for 7min at 72°C. Sterilization of PCR products was
751 performed at 94°C for 5 min.

752 For the heterozygote $NIMA^{+/-}$ null mutant, validation was performed by PCR amplification
753 of LdBPK_361580.1 open reading frame (NIMA_F and NIMA_R) using 10ng genomic DNA as
754 template and the amplification of the blasticidin resistance cassette (F: atgcctttgtctcaagaagaatc and
755 R: ttagccctcccacacataac) using the 11kb fragment surrounding the insertion locus (diagn_F and
756 diagn_R) as template.

757 For transgenic expression assay, the NIMA kinase gene LdBPK_361580.1 was PCR
758 amplified using the Expand high fidelity polymerase kit (Roche) and 30ng of genomic DNA of
759 promastigotes as template with primers F: ACCCTCGAGATGTCGGGGGGTAAC and R:
760 CGCCTTAAGTCACTCCTTTGCGGGTAGAG. The amplicon was first sub-cloned into pGEM-
761 T (Promega), validated by sequencing, and inserted into pXNG (kindly given by Stephen Beverley,
762 Washington University School of Medicine, St. Louis, MO, USA, (*79*)) to generate construct
763 pXNG-NIMA following digestion by XhoI and AflIII (*80*). For the validation of the pXNG
764 transfection in the heterozygote $NIMA^{+/-}$ null mutant, PCR assay was performed targeting the
765 nourseothricin resistance cassette (F: ACCGTCGACATGAAGATTTCGGTGAT; R:

766 CGGTCTAGATTAGGCGTCATCCTGT) after extraction of circular DNA from promastigotes at
767 logarithmic growth phase using the Nucleospin Plasmid kit (Macherey Nagel).
768
769 **Null mutant analysis**
770 Vector pTB007 was used, containing a T7 RNA polymerase gene, the humanized Streptococcus
771 pyogenes Cas9 gene, and homology regions allowing for stable integration into the beta-tubulin
772 locus (*78*). The plasmid was linearized with SbfI (NEB) and Hind-III (NEB) and dephosphorylated
773 using the Antarctic phosphatase (NEB) for 30min at 37°C. The 12kb-fragment was extracted and
774 purified from 0.8% agarose gel using the Wizard SV Gel and PCR clean-up system (Promega).
775 Transfection of P2 promastigotes from logarithmic growth phase was performed using a BioRad
776 Gene pulser (*81*). After centrifugation (1600*g*, 10min, room temperature), cells were resuspended
777 in Cytomix (0.15mM $Cacl_2$, 120mM KCl, 10mM KH_2PO_4, 5mM $MgCl_2$, 25mM HEPES (pH 7.5),
778 2mM EDTA (pH 7.6)) and 5.10^7 cells were electroporated by two pulses at 1,500V, 25µF, infinite
779 resistance. The next day, hygromycin B (Invitrogen) was added to select for transgenic cells
780 (30µg/ml final concentration). Virulence of the transgenic line (termed Ld1S PT007) was
781 maintained by two successive passages in hamsters as described (*77*) and recovered parasites were
782 grown in the presence of hygromycin B. To generate heterozygous $NIMA^{+/-}$ null mutants, Ld1S
783 PT007 promastigotes from logarithmic growth phase were transfected with the Biorad Gene Pulser
784 combining sterilized PCR constructs and sgRNA. After centrifugation (1600*g*, 10min, room
785 temperature), cells were resuspended in Cytomix and $3x10^7$ cells were electroporated using the
786 following conditions: 900V, 50µF, infinite resistance, two pulses. The next day, blasticidin S
787 hydrochloride (Sigma) was added to select for resistant cells (20 µg/ml final concentration) and
788 isolate heterozygous $NIMA^{+/-}$ null mutants.

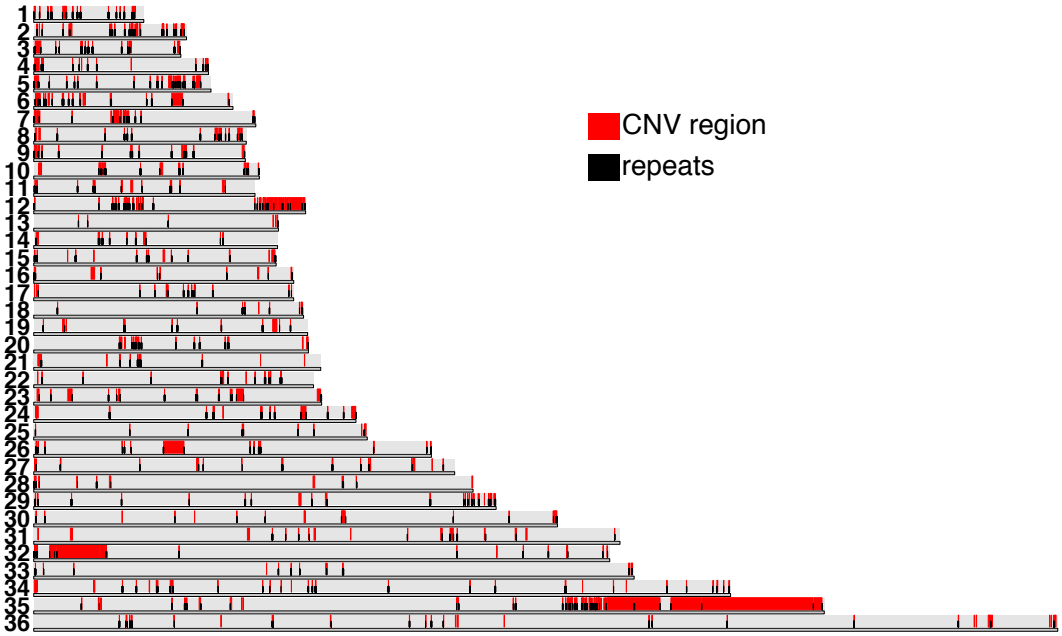
789 To generate the addback control, NIMA$^{+/-}$ promastigotes from logarithmic growth phase were
790 transfected in electroporation buffer containing 90mM sodium phosphate, 5mM potassium
791 chloride, 0.15mM calcium chloride, 50mM HEPES, pH 7.3 (*82*) using the program X-001 of the
792 Amaxa Nucleofector IIb (Lonza) with either 5µg pXNG Mock or pXNG-NIMA DNA. For
793 selection of transfected cells, nourseothricin (Sigma) was added 24 hours after electroporation at
794 100µg/ml final concentration.
795

796 **Methylation (Nm) and pseudouridine (Ψ) rRNA modification analyses**

797 We mapped the position of individual Nm sites in *L. donovani* rRNA based on Cryogenic electron
798 microscopy (Cryo-EM) (*40*) and mass-spectrometry (MS) (*83*) (**Table S20**). The snoRNAs
799 predicted to guide Nm site were identified by homology to *L. major* snoRNAs (*84*) (**Table S20**).
800 The relative methylation score (RMS score) of each individual Nm site was measured for both cell
801 culture P2 and P135 (**Table S20**). Three biological replicates of P2 and two biological replicates
802 of P135 were used for the analysis.
803 The ratio between the reads from Ψ-seq (*19*) at specific modification sites before and after CMC
804 treatment i.e. Ψ-fc (log2(+CMC/-CMC) of P135/P2 was calculated across the rRNA based on four
805 independent replicates. Only sites with log2 fold change >1.3 in all replicates were considered as
806 hypermodified.
807
808
809
810
811
812
813

814

815 **Fig. S1**



816

817 **Genomic distribution of CNV regions and DNA sequence repeats.** *Leishmania* chromosomes

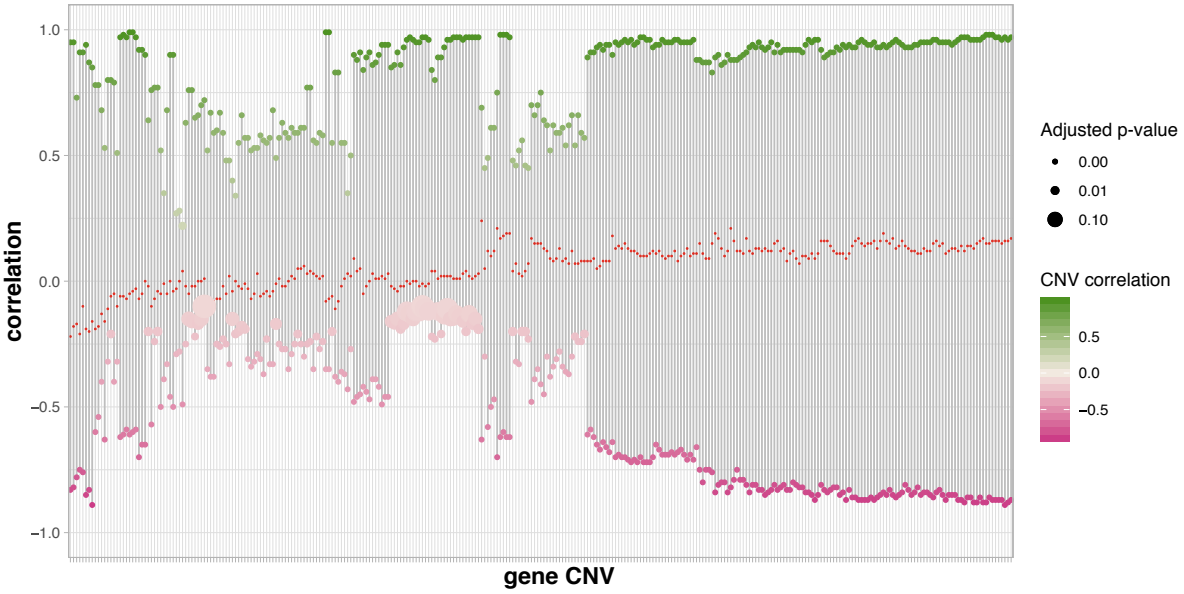
818 of the LdBPKv2 (*7*) are represented as stacked bars with karyoploteR (R package version 1.6.3)

819 (*85*). The collapsed CNV sequence bins are shown in red. The repeat elements significantly

820 associated with the CNV regions and mapping within their boundaries (1kb toward the outside and

821 150bp toward the inside of each collapsed CNV region) are shown in black.

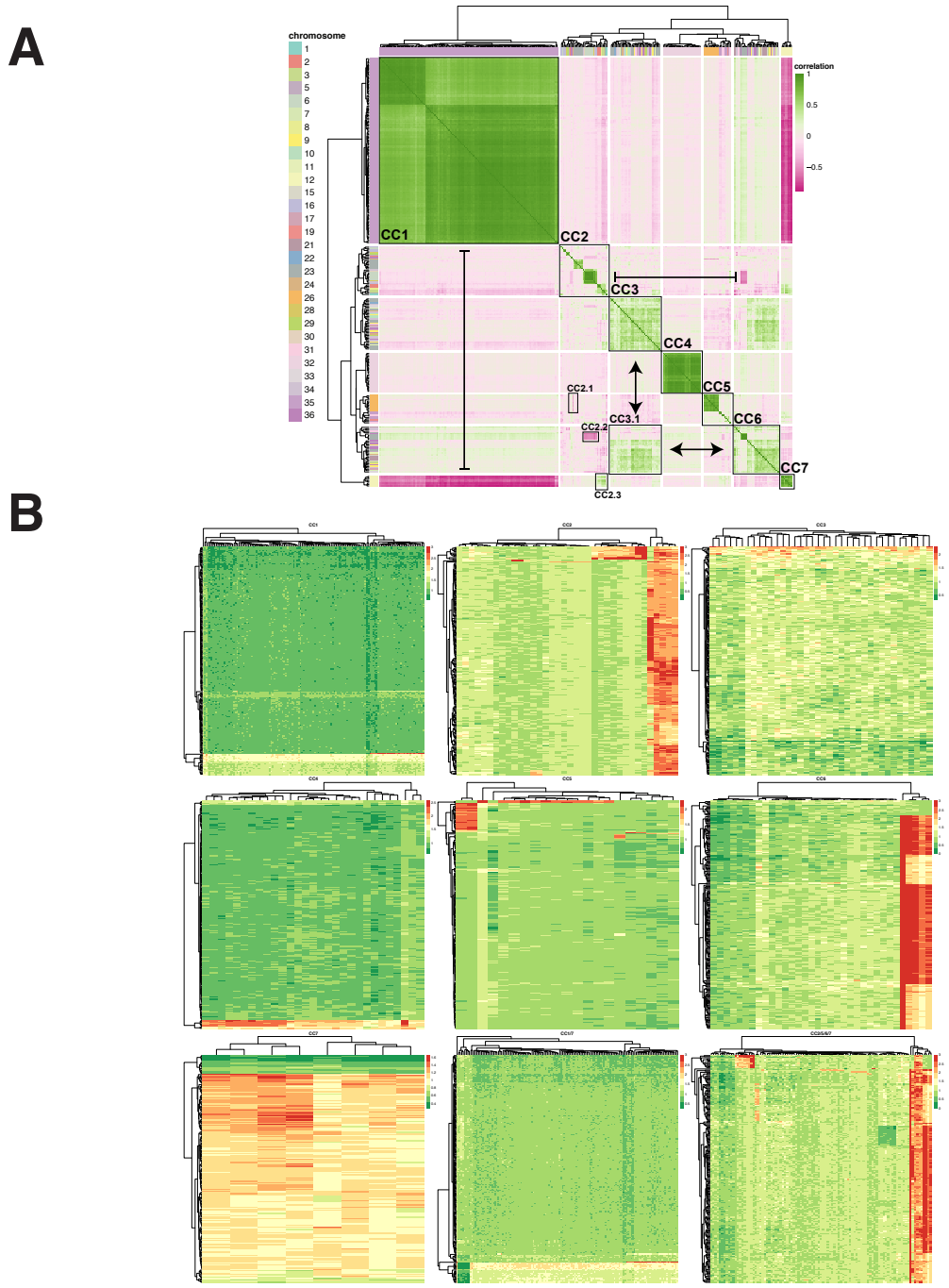
822 **Fig. S2**



823

824 **Distribution of correlation values measured between gene CNVs pairs.** Each column indicates
825 a different gene CNV. The *y*-axis reports the Pearson correlation value of the most positively (top,
826 green) and most negatively (bottom, pink) correlated gene CNVs. The color intensity and the size
827 of the dots match respectively to the correlation and the adjusted p-value scores. The red notches
828 in the center indicate the median correlation scores.

829 **Fig. S3**



830

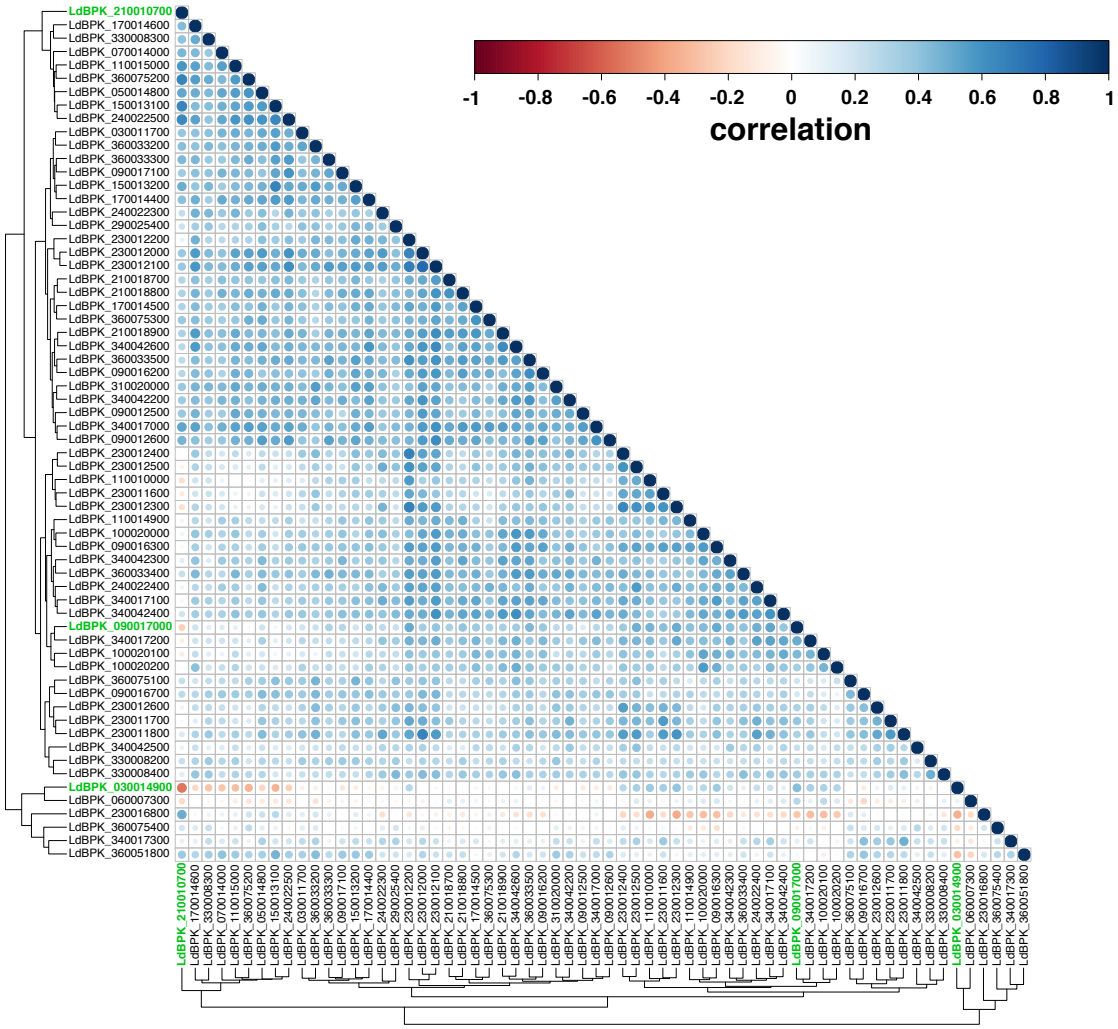
831 **All-vs-all gene CNV correlation.** (**A**) The heatmap shows gene CNVs on both axes. The color

832 range reflects the Pearson normalized coverage correlation value of gene CNVs across 204

833 samples. The side ribbon indicates the chromosomal localization of the gene CNV as defined in

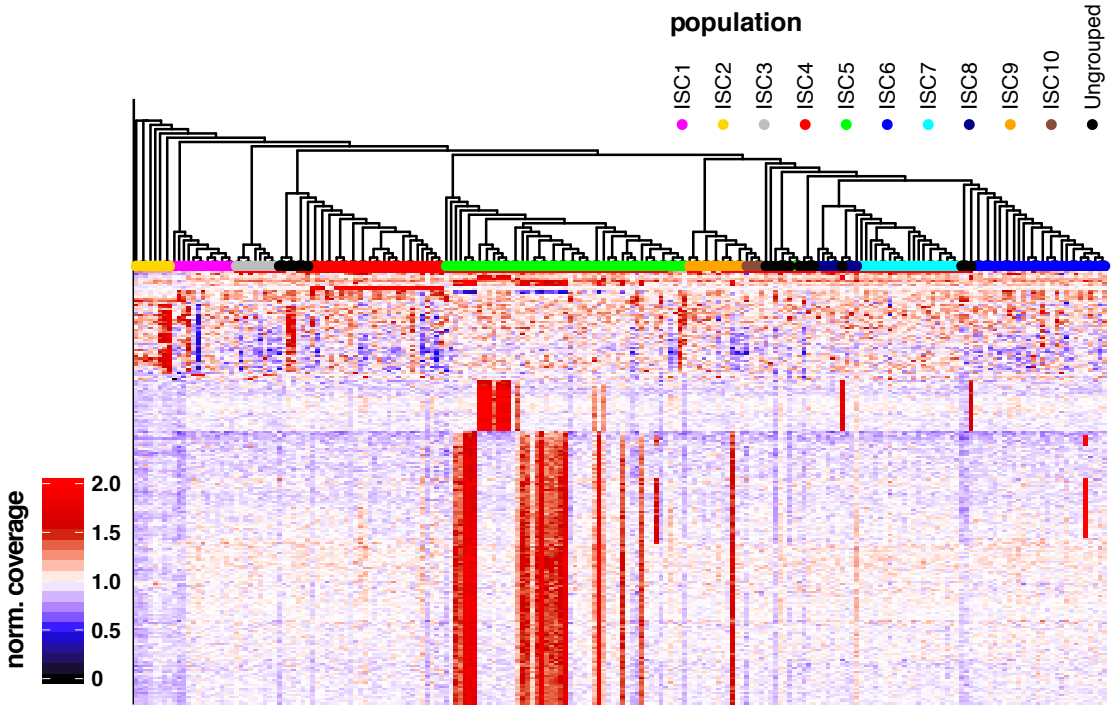
834 the legend. Black boxes highlight correlations clusters (CC). The arrows indicate higher-level
835 interactions between clusters. Pointy and flat arrowheads indicate respectively correlation and anti-
836 correlation between clusters. (**B**) The heatmaps show the normalized coverage values of the gene
837 CNVs (columns) belonging to the correlation clusters defined in panel A with respect to the 204
838 samples (rows). Green and red indicate respectively low and high coverage values as indicated in
839 the legends. To ease the visualization, all scores of > 3 were assigned to a value of 3. The two last
840 heatmaps show a combination of multiple CCs. In CC1/7 the columns include both CC1 and CC7
841 gene CNVs. In CC2/5/6/7 the columns include CC2, CC5, CC6, CC7 gene CNVs.

Fig. S4



tRNA gene CNV correlation map. Both the *x*- and *y*-axes report gene CNVs. In green are labeled three genes that are network-hubs as judged by their high connectivity (LdBPK_210010700, rRNA; LdBPK_090017000, tRNA; LdBPK_030014900, eukaryotic translation initiation factor 2 subunit alpha). The color range of the dots reflects normalized coverage correlation value of the gene CNVs. The level and direction of correlation are indicated by both dot size and color (red, negative correlation; blue, positive correlation).

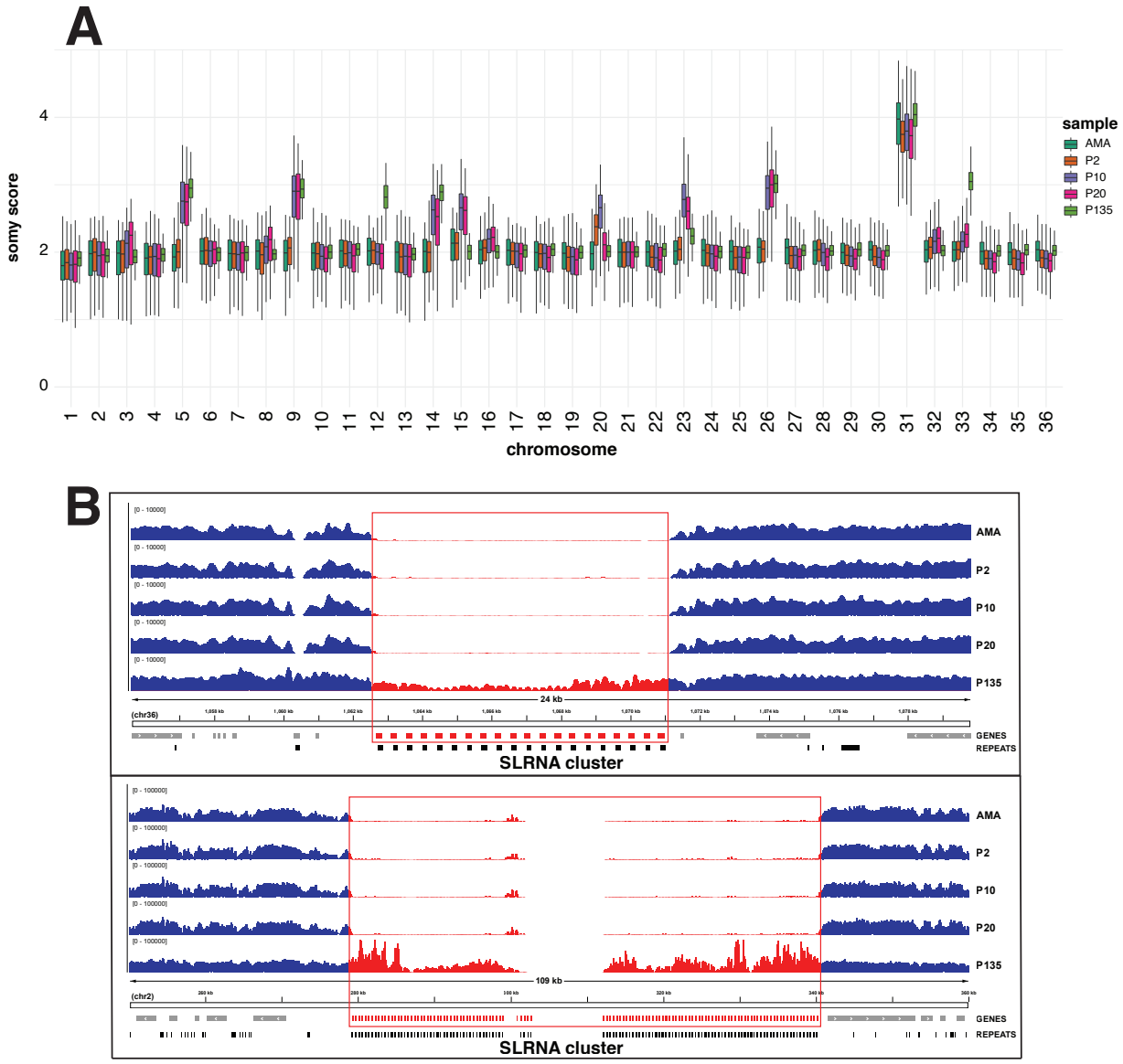
850 **Fig. S5**



851

852 **Polyphyletic distribution of gene CNVs.** Cladogram based on SNVs (>90% frequency) (upper
853 panel) and heatmap (lower panel) generated for the 204 ISC isolates. The cladogram topology
854 demonstrates the relationships between isolates. The branch length does not reflect genetic
855 distance. The heatmap has the same layout as in **Fig. 1E**.

856 **Fig. S6**



857

858 **Chromosome and gene copy number comparison of splenic amastigotes (AMA) and derived**

859 **promastigotes at different culture passages P2, 10, 20 and 135**. (**A**) The box plots demonstrate

860 the normalized sequencing coverage distributions of each chromosome. Bases where more than

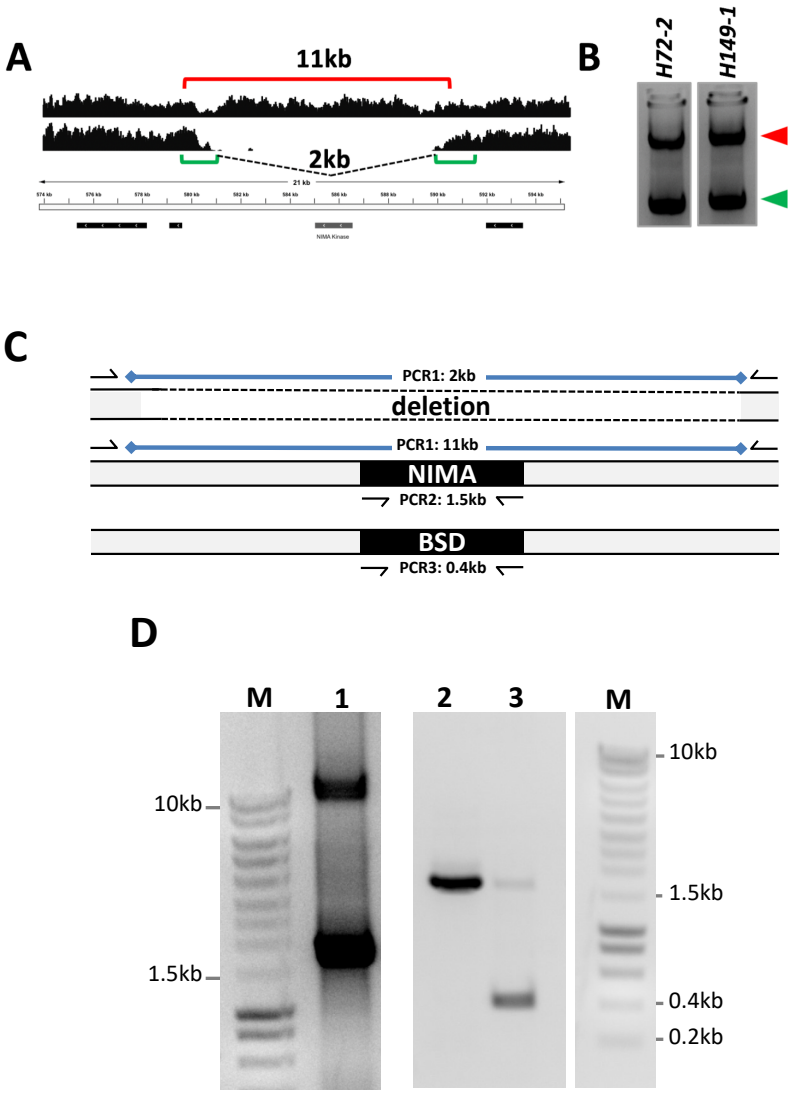
861 50% of the reads have a MAPQ score lower than 50 were not considered. The somy score measured

862 on the *y*-axis is an estimation of the chromosome copy number (see methods). Chromosome 21

863 showed a steady disomic level across this sample set and thus was used to normalize the read depth

864 of the other chromosomes. (**B**) Examples of detected CNV gene clusters. Each panel illustrates a
865 genome browser representation of the sequencing depth measured in the samples (see methods).
866 Gene annotations and the predicted repetitive elements are indicated.

867 **Fig. S7**



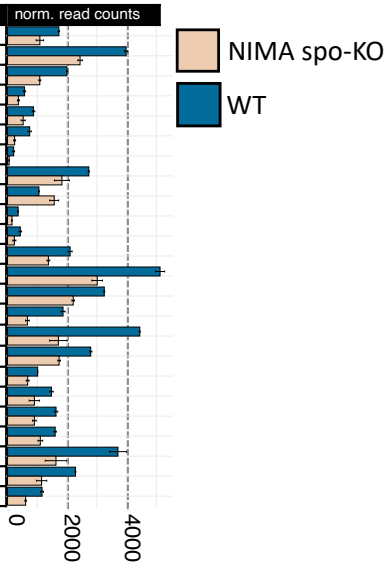
868

869 **Generation of heterozygous NIMA null mutants using CRISPR/Cas9 gene editing.** (**A**)
870 Schematic representation of the NIMA locus, indicating the 11kb and 2kb PCR fragments
871 diagnostic for the WT and deleted locus, respectively. (**B**) Gel electrophoretic separation of PCR
872 fragments obtained from two independent splenic isolates purified from hamsters H72-2 and
873 H149-1. The arrows heads indicate the 11kb (red) and 2kb (green) PCR fragments. (**C**) Schematic
874 representation of the PCR amplification strategy. Given the size of the genomic fragment under
875 investigation and the repetitive sequence surrounding the deletion site, a nested PCR protocol was

876 applied first amplifying the target locus (PCR1) before reamplification of the open reading frames
877 for the NIMA-related kinase gene (NIMA, PCR2) or the blasticidin resistance gene (BSD, PCR3).
878 The PCR primers are indicated by the arrows. (**D**) Gel electrophoretic separation of PCR fragments
879 obtained from a culture-adapted parasite population at *in vitro* passage 3 before gene editing using
880 primer pair for PCR1 (lane 1), and of a representative NIMA+/- heterozygous clone using primer
881 pairs PCR2 and PCR3 (lanes 2 and 3). The molecular weights of standard DNA marker fragments
882 (M) are indicated.

883 **Fig. S8**

gene identifier	annotation
Ld1S_070847100	paraflagellar rod protein, putative
Ld1S_080869200	intraflagellar transport protein 122B, putative
Ld1S_160120000	Flagellar calcium-binding protein, putative
Ld1S_160120100	flagellar calcium-binding protein, putative
Ld1S_160120200	flagellar calcium-binding protein, putative
Ld1S_160125800	paraflagellar rod protein 2C
Ld1S_160126200	paraflagellar rod protein 2C
Ld1S_160128500	Flagellar Member 3
Ld1S_270358700	Flagella connector protein 1
Ld1S_290432600	paraflagellar rod protein 1D, putative
Ld1S_290433000	paraflagellar rod protein 1D, putative
Ld1S_320533400	Flagellum attachment zone protein 9
Ld1S_320534400	dynein light chain, flagellar outer arm, putative
Ld1S_320535200	intraflagellar transport protein 140, putative
Ld1S_320549600	Paraflagellar Rod Proteome Component 9, putative
Ld1S_330583300_pseudogene	paraflagellar rod component, putative
Ld1S_330583400	intraflagellar transport protein 57,55, putative
Ld1S_330604900	Flagellum attachment zone protein 12
Ld1S_330611700	Flagellar Member 8 fragment
Ld1S_350689100	Flagellar Member 5
Ld1S_360750100	Flagellar Member 1
Ld1S_360766000	paraflagellar rod component, putative
Ld1S_360771700	paraflagellar rod component, putative
Ld1S_360783000	paraflagellar rod component, putative

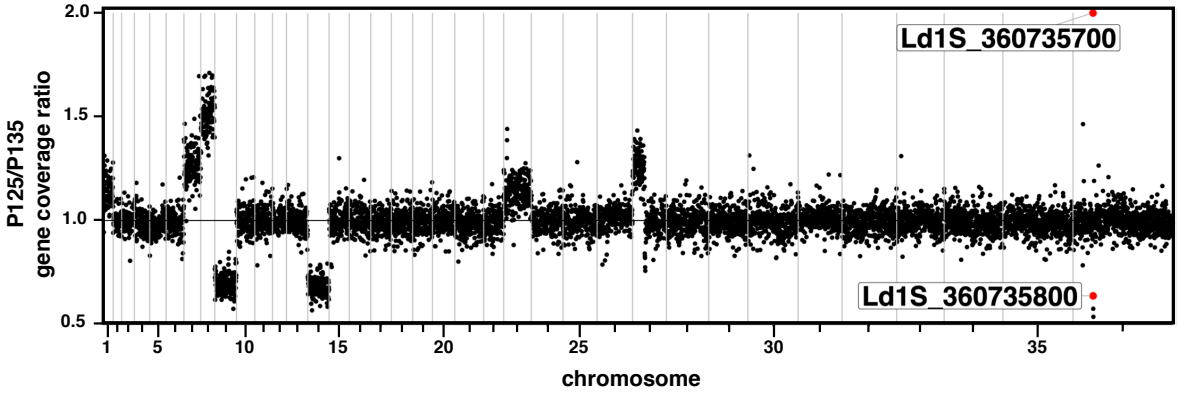


884

885 **Functional analysis of genes differentially expressed between spo-KO and WT clones.** Bar
886 plot (right panel) showing the normalized mean RNAseq read counts of the genes related to
887 flagellar function that were differentially expressed between WT (mean of two clones) and the six
888 NIMA-kinase spo-KO clones. For each entry, the gene identifier and the annotated function are
889 shown (left panel). The error bars indicate the standard deviations from the mean.

890

891 **Fig. S9**

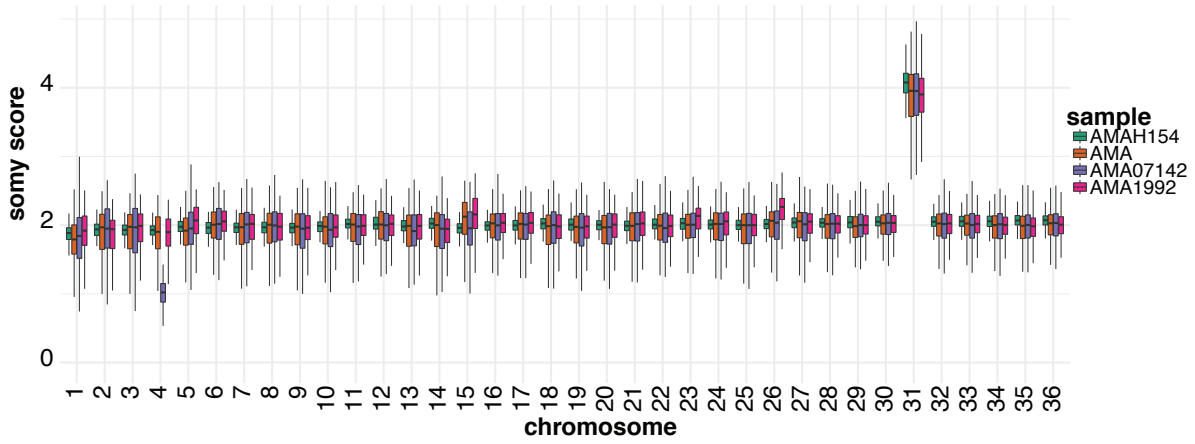


892
893

894 **Different parasite populations show distinct evolutionary trajectories.** (**A**) Genome-wide gene
895 coverage ratio between sample P125 and P135. Each dot represents a different gene. Vertical lines
896 mark chromosome boundaries. The deleted NIMA-kinase gene (Ld1S_360735700) and its
897 homologue (Ld1S_360735800) are indicated by the red dot. To ease plot readability, all scores of
898 > 2 were assigned to a value of 2.

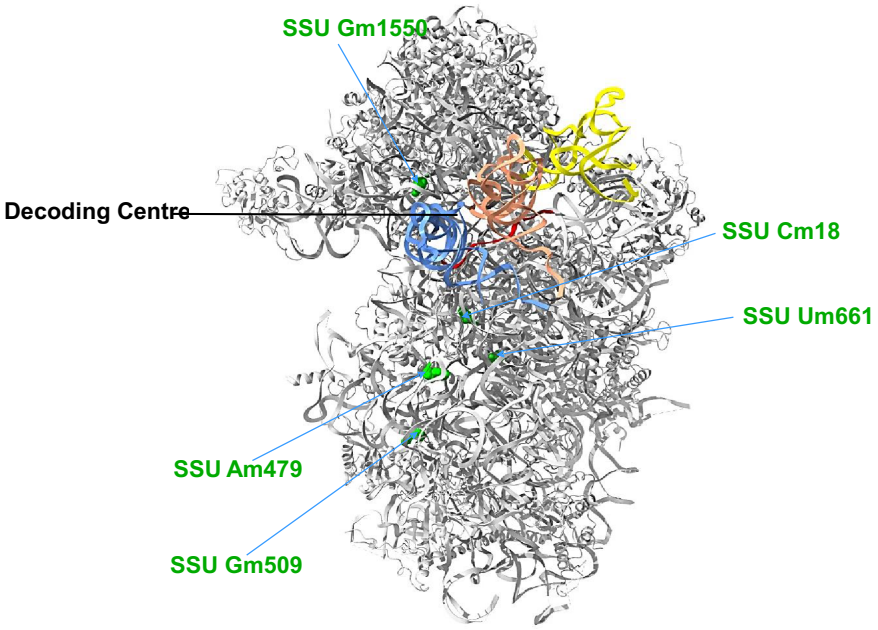
899

Fig. S10



Chromosome ploidy analysis. The plot shows the karyotypic heterogeneity of four independent amastigote isolates (AMAH154, AMA, AMA07142, AMA1992). Same layout as in **Fig. S6A**, but the stable disomic chromosomes used for normalization was 25.

907 **Fig. S11**

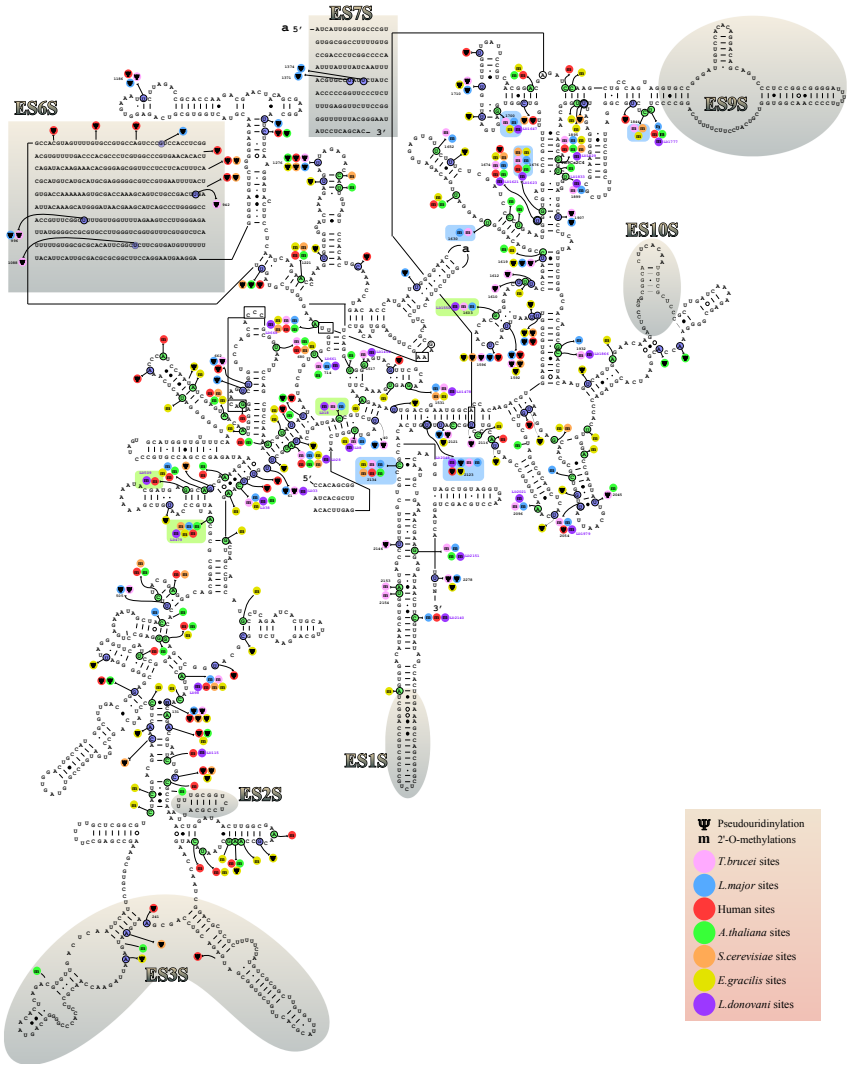


908

909 **3D-structure of the ribosome small subunit (SSU).** The hypermodified Nm sites are shown in
910 green. The location of the decoding center is indicated. A model of a tRNA positioned in the
911 ribosome A (blue), P (orange) and E (yellow) slots is shown. The model in red indicates the
912 mRNA position of the mRNA being translated. The representation is based on the published *L.*
913 *donovani* cryo-EM coordinates deposited with Protein Data Bank (PDB) (*39*) accessions 6AZ3
914 and 6AZ1 for the large and small subunits, respectively (*40*).

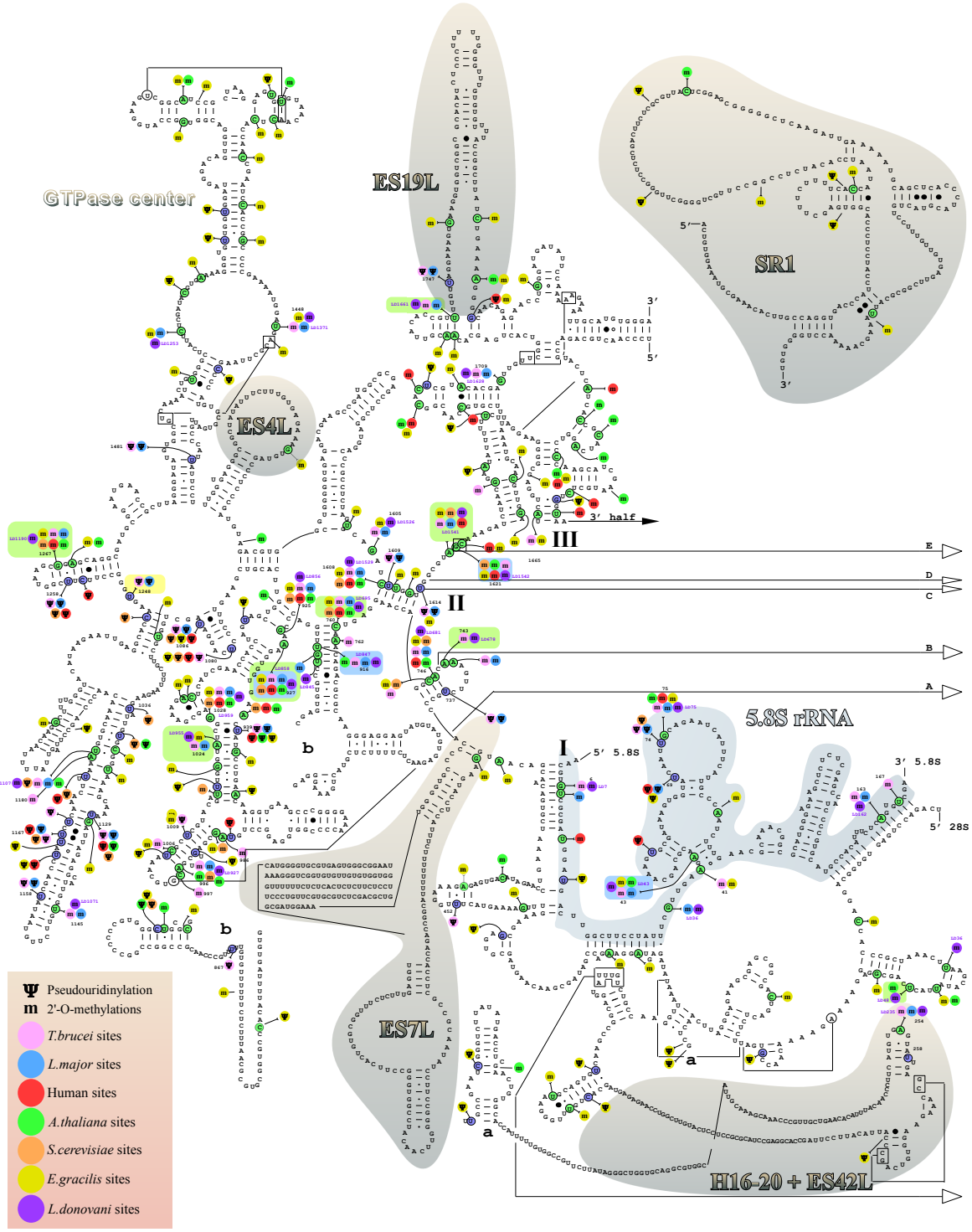
915

Fig. S12



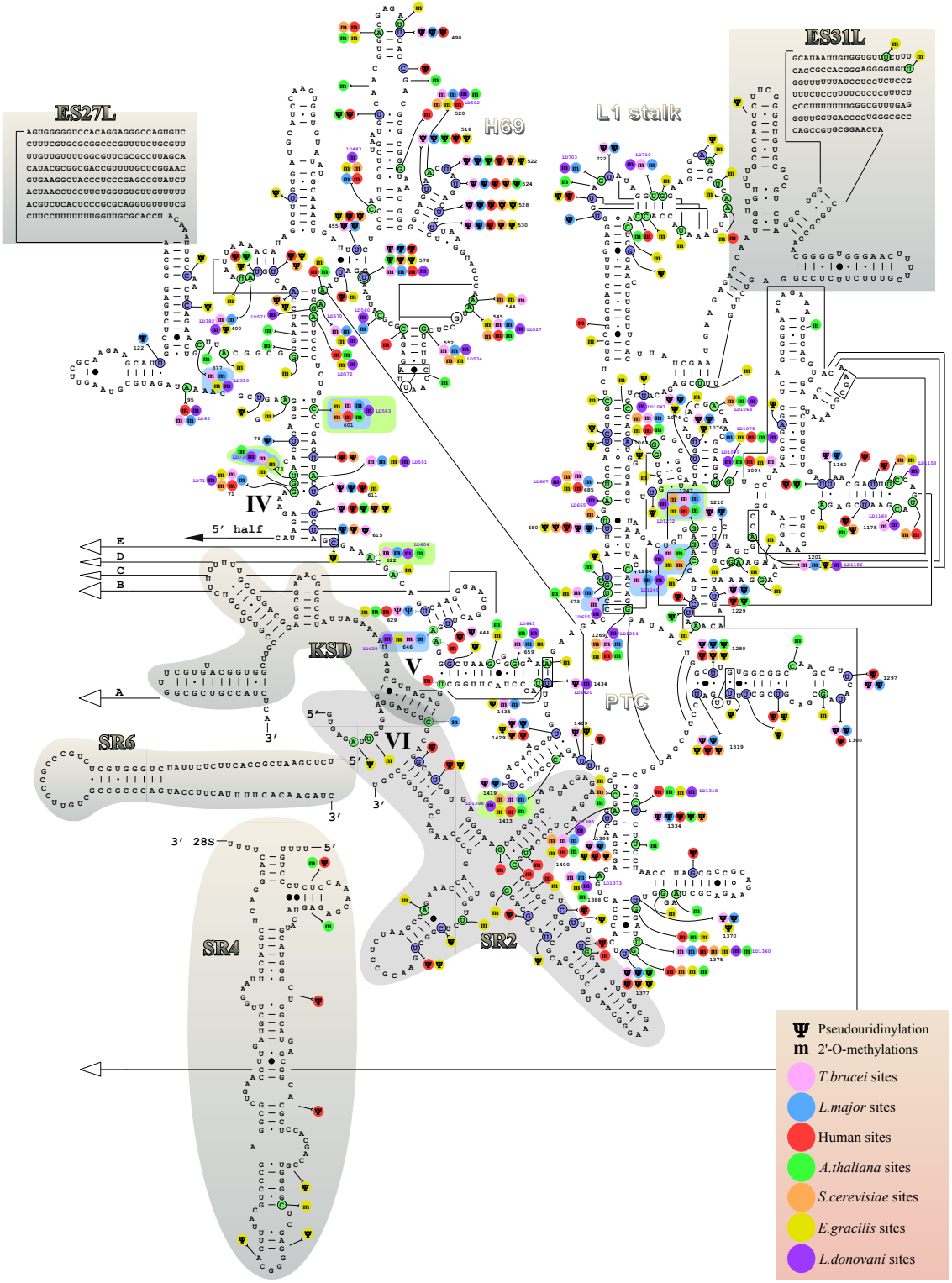
Localization of Nms and Ψs in the secondary structure of trypanosome rRNA SSU. The plot shows the position of Nm and Ψ sites in seven species: *Trypanosoma (T.) brucei*, *L. major*, *Homo sapiens*, *Arabidopsis (A.) thaliana*, *Saccharomyces (S.) cerevisiae*, *Euglena (E.) gracilis* and *L. donovani* (*2–7*). The location of Nm and Ψ sites are indicated for *T. brucei* (black number). The gene identifier of the *L. donovani* snoRNA genes driving the modification are shown in purple. Hypermodified Nm sites are indicated for *Leishmania* (green box) and *T. brucei* bloodstream form (blue box).

Fig. S13



Localization of Nms and Ψs in the secondary structure of trypanosome rRNA large subunit α (LSUα). Same layout as **Fig. S12**.

Fig. S14



Localization of Nms and Ψs in the secondary structure of trypanosome rRNA large subunit β (LSUβ). Same layout as **Fig S12**.

936 **Data S1. (separate file)**
937 **Table S1: Mapping statistics of the dataset for the 204 clinical isolates.**

938 **Data S2. (separate file)**
939 **Table S2: Position and normalized coverage values of the collapsed CNV regions.**

940 *Footnote: The values reported for each sample indicate the normalized collapsed CNV coverage.*

941 **Data S3. (separate file)**
942 **Table S3: Repetitive elements.**

943 *Footnotes: The columns "GC%" and "N%" indicate respectively GC (guanine-cytosine) and the*
944 *undetermined bases content. The "longest ORF" column indicates the length of the amino acid*
945 *sequence of the longest open reading frame computed with EMBOSS getorf (version 6.6.0) (86)*
946 *with option "-find 1". (*) The observed count. (**) The expected count based on the sampled*
947 *CNV boundaries. (***) The value at the 5% percentile of random samples. (****) The value at*
948 *the 95% percentile of random samples. (*****) The standard deviation of random samples.*
949 *(******) The fold enrichment, given by the ratio observed / expected. (*******) Log2 of the fold*
950 *enrichment value. (********) The p-value of enrichment/depletion. (*********) The multiple-*
951 *testing corrected p-value.*

952 **Data S4. (separate file)**
953 **Table S4: Most positive and most negative LdBPK gene coverage correlations.**

954 **Data S5. (separate file)**
955 **Table S5: Correlation map data.**

956 *Footnote: The first three columns and the first three rows indicate respectively LdBPKv1 gene*
957 *identifiers, gene annotations and LdBPKv2 gene identifiers.*

958 **Data S6. (separate file)**
959 **Table S6: Correlation network data.**

960 *Footnote: The first three columns and the first three rows indicate respectively LdBPKv1 gene*
961 *identifiers, gene annotations and LdBPKv2 gene identifiers.*

962 **Data S7. (separate file)**
963 **Table S7: Network hub genes.**

964 **Data S8. (separate file)**
965 **Table S8: Collapsed CNVs showing highest read depth in the 204 *L. donovani* field isolates.**

966 *Footnote: (*) Maximum measured coverage of the top 30 collapsed CNV detected in genic areas.*

967 **Data S9. (separate file)**
968 **Table S9: Genes amplified in phylogenetically distinct clades.**

969 *Footnote: The values reported for each sample indicate the normalized gene coverage.*

970 **Data S10. (separate file)**
971 **Table S10: Normalized gene coverage values across ISC samples.**

972 *Footnote: The values reported for each sample indicate the normalized gene coverage.*

973 *(*) Difference between maximum and minimum normalized gene coverage.*

974 **Data S11. (separate file)**
975 **Table S11: Polysomy level analyses of two sets of Ld1S samples: (i) AMA to P135 isolates;**
976 **(ii) four AMA isolates.**

977 **Data S12. (separate file)**
978 **Table S12: Ld1S annotation.**

979 *Footnote: (*) Gene Ontology (GO) identifiers of the biological process (BP), molecular function*
980 *(MF) and cellular component (CC). (**) LdBPKv1 homologs predicted with OrthoFinder. Genes*
981 *with no significant match were searched with NCBI-blastn. One-to-many homologs are reported*
982 *with comma separated gene identifiers.*

983 **Data S13. (separate file)**
984 **Table S13: Definitions of gene clusters in two sets of Ld1S samples: (i) AMA-P135 isolates;**
985 **(ii) P20 clones isolates.**

986 **Data S14. (separate file)**
987 **Table S14: Gene CNVs and cluster CNVs detected during Ld1S culture adaptation.**
988 *Footnotes: (*) The AMA, P2, P10, P20 and P135 columns indicate the normalized gene coverage*
989 *in the respective samples. (**) "increasing" gene CNVs have a P135 normalized coverage*
990 *increase of at least 0.5 with respect to AMA. "decreasing" gene CNVs have a P135 normalized*
991 *coverage decrease of at least 0.5 with respect to AMA. All the other gene CNVs are labeled as*
992 *"transient". (***) The number of genes in the gene cluster (clstr). Individual (non-cluster) genes*
993 *have by default a value of 1. (****) Gene Ontology (GO) identifiers of the biological process (BP),*
994 *molecular function (MF) and cellular component (CC). (*****) The genes belonging to each*
995 *cluster CNV are reported as separate entries sharing the same attributes.*

996 **Data S15. (separate file)**
997 **Table S15: Gene CNVs and cluster CNVs detected in the eight Ld1S P20 clones.**
998 *Footnotes: (*) The CL1, CL3, CL4, CL6, CL7, CL8, CL9 and CL10 columns indicate the*
999 *normalized gene coverage in the respective samples. (**) The number of genes in the gene cluster*
1000 *(clstr). Individual (non-cluster) genes have by default a value of 1. (***) Gene Ontology (GO)*
1001 *identifiers of the biological process (BP), molecular function (MF) and cellular component (CC).*
1002 *(****) The genes belonging to each cluster CNV are reported as separate entries sharing the same*
1003 *attributes.*

1004 **Data S16. (separate file)**
1005 **Table S16: Single nucleotide variants (SNVs) in Ld1S P20 clones.**

1006 *Footnotes: (*) Number of alternate observations, (**) Total read depth at the locus, (***)*
1007 *Alternate allele quality sum in phred, (****) Reference allele quality sum in phred, (*****) Mean*
1008 *mapping quality of observed alternate alleles, (******) Mean mapping quality of observed*
1009 *reference alleles, (*******) Reference base and alternate allele, (********) context sequence of*
1010 *the variant (+/- 5bp), (*********) SnpEff predicted variant effect.*
1011

1012 **Data S17. (separate file)**
1013 **Table S17: Data of the differential expression analysis.**

1014 *Footnotes: DESeq2 was utilized to test differential gene expression between the NIMA spo-KO*
1015 *and the WT clones groups. Significant genes (adjusted p-value < 0.01) with a log2 fold change >*
1016 *0.5 were considered "enriched" in the NIMA spo-KO group, while those with a log2 fold change*
1017 *< -0.5 were considered "depleted" in the NIMA spo-KO group. Maxi-circle genes were not*
1018 *considered in the analysis. (*) Mean of normalized counts, taken over all samples, (**) log2 fold*
1019 *change between the groups, (***) standard error of the log2 fold change estimate, (****) Wald*
1020 *statistic, (*****) Wald test p-value, (******) Benjamini-Hochberg adjusted p-value.*

1021 **Data S18. (separate file)**
1022 **Table S18: DNA and RNA double-ratio data.**

1023 *Footnotes: (*) Whole genome sequencing (DNA) ratio: the mean read counts in spo-KO clones is*
1024 *divided by the mean DNA read counts in WT clones. Both the numerator and denominator are*
1025 *normalized by the average library size. (**) RNAseq (RNA) ratio: the mean read counts in spo-*
1026 *KO clones is divided by the mean RNA read counts in WT clones. Both the numerator and*
1027 *denominator are normalized by the average library size. (***) Double-ratio score: the ratio*
1028 *between the RNA ratio and the DNA ratio. (****) P-values of the double-ratio scores. (*****)*

1029 *Genes with double-ratio score p-value < 0.01 are considered "destabilized" or "stabilized" for*
1030 *double-ratio scores < or > 1, respectively. (******) Gene Ontology (GO) identifiers of the*
1031 *biological process (BP), molecular function (MF) and cellular component (CC).*

1032 **Data S19. (separate file)**
1033 **Table S19: P125/P135 gene coverage ratio values.**

1034 *Footnote: (*) The P125 and P135 columns indicate the gene coverage normalized by the median*
1035 *gene coverage in the respective samples.*

1036 **Data S20. (separate file)**
1037 **Table S20: The complete stoichiometry of Nm sites in *L. donovani* rRNA.**

1038 *Footnote: Data are shown as mean ± SEM (standard error of mean). The fold change (FC) of*
1039 *RMS of individual Nms is presented and the hypermodified sites showing FC > 2 are highlighted.*
1040 *Note that we have not identified all the snoRNAs guiding the existing Nm modification.*

1041 **Data S21. (separate file)**
1042 **Table S21: The relative fold-change of Ψs in *L. donovani* rRNA between P135 and P2**
1043 **parasites.**
1044
1045 *Footnote: (*) Four independent replicates (R). Only sites with log2 fold change (FC) >1.3 in all*
1046 *replicates are considered as hypermodified sites and are highlighted in pink.*

1047
1048

1049 **References**

1050 1. E. Darmon, D. R. Leach, Bacterial genome instability. *Microbiol Mol Biol Rev* **78**, 1-39
1051 (2014).
1052 2. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674
1053 (2011).
1054 3. D. Hughes, D. I. Andersson, Evolutionary consequences of drug resistance: shared
1055 principles across diverse targets and organisms. *Nat Rev Genet* **16**, 459-471 (2015).
1056 4. N. McGranahan, C. Swanton, Clonal Heterogeneity and Tumor Evolution: Past, Present,
1057 and the Future. *Cell* **168**, 613-628 (2017).
1058 5. C. E. Clayton, in *Curr. Opin. Microbiol.* (2016), vol. 32, pp. 46-51.
1059 6. S. Michaeli, Trans-splicing in trypanosomes: machinery and its impact on the parasite
1060 transcriptome. *Future Microbiol* **6**, 459-474 (2011).
1061 7. F. Dumetz *et al.*, in *MBio*. (2017), vol. 8.
1062 8. S. A. Iantorno *et al.*, Gene Expression in Leishmania Is Regulated Predominantly by Gene
1063 Dosage. *mBio* **8**, (2017).
1064 9. P. Leprohon *et al.*, in *Nucleic Acids Res.* (2009), vol. 37, pp. 1387-1399.
1065 10. P. Prieto Barja *et al.*, in *Nat Ecol Evol*. (2017), vol. 1, pp. 1961-1969.
1066 11. M. B. Rogers *et al.*, in *Genome Res.* (2011), vol. 21, pp. 2129-2142.
1067 12. H. Imamura *et al.*, in *Elife*. (2016), vol. 5.
1068 13. P. J. Hastings, J. R. Lupski, S. M. Rosenberg, G. Ira, Mechanisms of change in gene copy
1069 number. *Nat Rev Genet* **10**, 551-564 (2009).
1070 14. C. Waldron, F. Lacroute, Effect of growth rate on the amounts of ribosomal and transfer
1071 ribonucleic acids in yeast. *J Bacteriol* **122**, 855-865 (1975).
1072 15. V. Chikne *et al.*, Small nucleolar RNAs controlling rRNA processing in Trypanosoma
1073 brucei. *Nucleic Acids Res* **47**, 2609-2629 (2019).
1074 16. X. H. Liang, Q. Liu, M. J. Fournier, rRNA modifications in an intersubunit bridge of the
1075 ribosome strongly affect both ribosome biogenesis and activity. *Mol Cell* **28**, 965-977
1076 (2007).
1077 17. U. Birkedal *et al.*, Profiling of ribose methylations in RNA by high-throughput sequencing.
1078 *Angew Chem Int Ed Engl* **54**, 451-455 (2015).
1079 18. S. Schwartz *et al.*, Transcriptome-wide mapping reveals widespread dynamic-regulated
1080 pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).
1081 19. V. Chikne *et al.*, A pseudouridylation switch in rRNA is implicated in ribosome function
1082 during the life cycle of Trypanosoma brucei. *Sci Rep* **6**, 25296 (2016).
1083 20. M. Ackermann *et al.*, Self-destructive cooperation mediated by phenotypic noise. *Nature*
1084 **454**, 987-990 (2008).
1085 21. J. K. Graham, M. L. Smith, A. M. Simons, Experimental evolution of bet hedging under
1086 manipulated environmental uncertainty in Neurospora crassa. *Proc Biol Sci* **281**, (2014).
1087 22. I. M. Rouzine, A. D. Weinberger, L. S. Weinberger, An evolutionary role for HIV latency
1088 in enhancing viral transmission. *Cell* **160**, 1002-1012 (2015).
1089 23. M. K. Stewart, B. T. Cookson, Non-genetic diversity shapes infectious capacity and host
1090 resistance. *Trends Microbiol* **20**, 461-466 (2012).
1091 24. J. Seger, H. J. Brockmann. (Oxford, UK: Oxford University Press, 1987), vol. 4, pp. 182–
1092 211.
1093 25. A. H. Yona *et al.*, Chromosomal duplication is a transient evolutionary solution to stress.
1094 *Proc Natl Acad Sci U S A* **109**, 21010-21015 (2012).

1095 26. K. S. Rajan *et al.*, Pseudouridines on Trypanosoma brucei spliceosomal small nuclear
1096 RNAs and their implication for RNA and protein interactions. *Nucleic Acids Res* **47**, 7633-
1097 7647 (2019).
1098 27. K. S. Rajan, V. Chikne, K. Decker, H. Waldman Ben-Asher, S. Michaeli, Unique Aspects
1099 of rRNA Biogenesis in Trypanosomatids. *Trends Parasitol* **35**, 778-794 (2019).
1100 28. M. Penzo *et al.*, Human ribosomes from cells with reduced dyskerin levels are intrinsically
1101 altered in translation. *FASEB J* **29**, 3472-3482 (2015).
1102 29. C. Bellodi *et al.*, H/ACA small RNA dysfunctions in disease reveal key roles for noncoding
1103 RNA modifications in hematopoietic stem cell differentiation. *Cell Rep* **3**, 1493-1502
1104 (2013).
1105 30. F. Ariey *et al.*, A molecular marker of artemisinin-resistant Plasmodium falciparum
1106 malaria. *Nature* **505**, 50-55 (2014).
1107 31. J. B. Koenderink, R. A. Kavishe, S. R. Rijpma, F. G. Russel, The ABCs of multidrug
1108 resistance in malaria. *Trends Parasitol* **26**, 440-446 (2010).
1109 32. A. Selmecki, A. Forche, J. Berman, Aneuploidy and isochromosome formation in drug-
1110 resistant Candida albicans. *Science* **313**, 367-370 (2006).
1111 33. A. H. Melnyk, A. Wong, R. Kassen, The fitness costs of antibiotic resistance mutations.
1112 *Evol Appl* **8**, 273-283 (2015).
1113 34. A. Ashworth, C. J. Lord, J. S. Reis-Filho, Genetic interactions in cancer progression and
1114 treatment. *Cell* **145**, 30-38 (2011).
1115 35. J. van de Haar *et al.*, Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **177**,
1116 1375-1383 (2019).
1117 36. C. Kim *et al.*, Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by
1118 Single-Cell Sequencing. *Cell* **173**, 879-893 e813 (2018).
1119 37. R. Leinonen, H. Sugawara, M. Shumway, I. N. S. D. Collaboration, in *Nucleic Acids Res.*
1120 (2011), vol. 39, pp. D19-21.
1121 38. B. Cuypers *et al.*, Integrated genomic and metabolomic profiling of ISC1, an emerging
1122 Leishmania donovani population in the Indian subcontinent. *Infect Genet Evol* **62**, 170-178
1123 (2018).
1124 39. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
1125 40. M. Shalev-Benami *et al.*, Atomic resolution snapshot of Leishmania ribosome inhibition
1126 by the aminoglycoside paromomycin. *Nat Commun* **8**, 1589 (2017).
1127 41. G. Bussotti *et al.*, Leishmania Genome Dynamics during Environmental Adaptation
1128 Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype
1129 Instability, and Telomeric Amplification. *MBio* **9**, (2018).
1130 42. W. Li, A. Godzik, in *Bioinformatics*. (2006), vol. 22, pp. 1658-1659.
1131 43. J. T. Robinson *et al.*, in *Nat. Biotechnol.* (2011), vol. 29, pp. 24-26.
1132 44. F. Ramírez *et al.*, in *Nucleic Acids Res.* (2016), vol. 44, pp. W160-165.
1133 45. *R Core Team*.
1134 46. T. Barrett *et al.*, BioProject and BioSample databases at NCBI: facilitating capture and
1135 organization of metadata. *Nucleic Acids Res* **40**, D57-63 (2012).
1136 47. S. Steinbiss *et al.*, Companion: a web server for annotation and analysis of parasite
1137 genomes. *Nucleic Acids Res* **44**, W29-34 (2016).
1138 48. C. Camacho *et al.*, in *BMC Bioinformatics*. (2009), vol. 10, pp. 421.
1139 49. M. Aslett *et al.*, TriTrypDB: a functional genomic resource for the Trypanosomatidae.
1140 *Nucleic Acids Res* **38**, D457-462 (2010).

1141 50. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative
1142 genomics. *Genome Biol* **20**, 238 (2019).
1143 51. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND.
1144 *Nat Methods* **12**, 59-60 (2015).
1145 52. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).
1146 53. J. Huerta-Cepas *et al.*, in *Nucleic Acids Res.* (2016), vol. 44, pp. D286-293.
1147 54. C. Aurrecoechea *et al.*, EuPathDB: the eukaryotic pathogen genomics database resource.
1148 *Nucleic Acids Res* **45**, D581-D591 (2017).
1149 55. UniProt, in *Nucleic Acids Res.* (2012), vol. 40, pp. D71-75.
1150 56. E. Camon *et al.*, The Gene Ontology Annotation (GOA) Database: sharing knowledge in
1151 Uniprot with Gene Ontology. *Nucleic Acids Res* **32**, D262-266 (2004).
1152 57. A. Smit, Hubley, R & Green, P. (2013-2015).
1153 58. J. Jurka *et al.*, in *Cytogenet. Genome Res.* (2005), vol. 110, pp. 462-467.
1154 59. J. M. Flynn *et al.*, RepeatModeler2: automated genomic discovery of transposable element
1155 families. *bioRxiv [856591]*, (2019).
1156 60. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*
1157 *Res* **37**, W202-208 (2009).
1158 61. A. R. Quinlan, I. M. Hall, in *Bioinformatics*. (2010), vol. 26, pp. 841-842.
1159 62. S. Kurtz *et al.*, in *Genome Biol.* (2004), vol. 5, pp. R12.
1160 63. A. Heger, C. Webber, M. Goodson, C. P. Ponting, G. Lunter, GAT: a simulation framework
1161 for testing the association of genomic intervals. *Bioinformatics* **29**, 2046-2048 (2013).
1162 64. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale
1163 detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).
1164 65. G. C. a. T. Nepusz, The igraph software package for complex network research.
1165 *InterJournal* **Complex Systems**, 1695 (2006).
1166 66. T. M. J. R. Fruchterman, Edward M., Graph Drawing by Force-Directed Placement.
1167 *Software – Practice & Experience* **21**, 1129–1164 (1991).
1168 67. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1169 large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
1170 68. E. Garrison, G. Marth, in *arXiv preprint arXiv:1207.3907 [q-bio.GN]*. (2012).
1171 69. G. Yu, T. T. Lam, H. Zhu, Y. Guan, Two Methods for Mapping and Visualizing Associated
1172 Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041-3043 (2018).
1173 70. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: recent updates and new
1174 developments. *Nucleic Acids Res* **47**, W256-W259 (2019).
1175 71. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
1176 RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
1177 72. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput
1178 sequencing data. *Bioinformatics* **31**, 166-169 (2015).
1179 73. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of
1180 biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
1181 74. S. Maere, K. Heymans, M. Kuiper, BiNGO: a Cytoscape plugin to assess
1182 overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**,
1183 3448-3449 (2005).
1184 75. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology
1185 Consortium. *Nat Genet* **25**, 25-29 (2000).

1186 76. C. The Gene Ontology, The Gene Ontology Resource: 20 years and still GOing strong.
1187 *Nucleic Acids Res* **47**, D330-D338 (2019).
1188 77. P. Pescher, T. Blisnick, P. Bastin, G. F. Spath, Quantitative proteome profiling informs on
1189 phenotypic traits that adapt Leishmania donovani for axenic and intracellular proliferation.
1190 *Cell Microbiol* **13**, 978-991 (2011).
1191 78. T. Beneke *et al.*, A CRISPR Cas9 high-throughput genome editing toolkit for
1192 kinetoplastids. *R Soc Open Sci* **4**, 170095 (2017).
1193 79. S. M. Murta, T. J. Vickers, D. A. Scott, S. M. Beverley, Methylene tetrahydrofolate
1194 dehydrogenase/cyclohydrolase and the synthesis of 10-CHO-THF are essential in
1195 Leishmania major. *Mol Microbiol* **71**, 1386-1401 (2009).
1196 80. M. Dacher *et al.*, Probing druggability and biological function of essential proteins in
1197 Leishmania combining facilitated null mutant and plasmid shuffle analyses. *Mol Microbiol*
1198 **93**, 146-166 (2014).
1199 81. J. H. LeBowitz, Transfection experiments with Leishmania. *Methods Cell Biol* **45**, 65-78
1200 (1994).
1201 82. G. Schumann Burkard, P. Jutzi, I. Roditi, Genome-wide RNAi screens in bloodstream form
1202 trypanosomes identify drug transporters. *Mol Biochem Parasitol* **175**, 91-94 (2011).
1203 83. H. Nakayama *et al.*, Method for Direct Mass-Spectrometry-Based Identification of
1204 Monomethylated RNA Nucleoside Positional Isomers and Its Application to the Analysis
1205 of Leishmania rRNA. *Anal Chem* **91**, 15634-15643 (2019).
1206 84. D. Eliaz *et al.*, Genome-wide analysis of small nucleolar RNAs of Leishmania major
1207 reveals a rich repertoire of RNAs involved in modification and processing of rRNA. *RNA*
1208 *Biol* **12**, 1222-1255 (2015).
1209 85. B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes
1210 displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017).
1211 86. P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open
1212 Software Suite. *Trends Genet* **16**, 276-277 (2000).
1213