



**HAL**  
open science

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

Giovanni Bussotti, Laura Piel, Pascale Pescher, Malgorzata A. Domagalska, K. Shanmugha Rajan, Smadar Cohen-Chalamish, Tirza Doniger, Disha-Gajanan Hiregange, Peter J. Myler, Ron Unger, et al.

► **To cite this version:**

Giovanni Bussotti, Laura Piel, Pascale Pescher, Malgorzata A. Domagalska, K. Shanmugha Rajan, et al.. Genome instability drives epistatic adaptation in the human pathogen *Leishmania*. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118 (51), pp.e2113744118. 10.1073/pnas.2113744118 . pasteur-03512631

**HAL Id: pasteur-03512631**

**<https://pasteur.hal.science/pasteur-03512631>**

Submitted on 5 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



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

Giovanni Bussotti<sup>a,b</sup>, Laura Piel<sup>b,1</sup>, Pascale Pescher<sup>b</sup>, Malgorzata A. Domagalska<sup>c</sup>, K. Shanmugha Rajan<sup>d</sup>, Smadar Cohen-Chalamish<sup>d</sup>, Tirza Doniger<sup>d</sup>, Disha-Gajanan Hiregange<sup>e</sup>, Peter J. Myler<sup>f,g,h,i,j</sup>, Ron Unger<sup>d</sup>, Shulamit Michaeli<sup>d</sup>, and Gerald F. Späth<sup>b,2</sup>

<sup>a</sup>Bioinformatics and Biostatistics Hub C3BI, Institut Pasteur, Université de Paris, Unité de Service et de Recherche (USR) 3756 Institut Pasteur CNRS, Paris 75015, France; <sup>b</sup>Unité de Parasitologie moléculaire et Signalisation, Institut Pasteur, Université de Paris, INSERM U1201, Paris 75015, France; <sup>c</sup>Molecular Parasitology Unit, Institute of Tropical Medicine B-2000 Antwerp, Belgium; <sup>d</sup>The Mina and Everard Goodman Faculty of Life Sciences and Advanced Materials and Nanotechnology Institute, Bar-Ilan University, Ramat-Gan 52900, Israel; <sup>e</sup>Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot 7610001, Israel; <sup>f</sup>Seattle Structural Genomics Center for Infectious Disease, Seattle, WA 98195; <sup>g</sup>Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA 98195; <sup>h</sup>Department of Pediatrics, University of Washington, Seattle, WA 98195; <sup>i</sup>Department of Global Health, University of Washington, Seattle, WA 98195; and <sup>j</sup>Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA 98195

Edited by L. Sibley, Department of Molecular Microbiology, Washington University in St Louis, St. Louis, MO; received July 27, 2021; accepted October 28, 2021

**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*, which exploits frequent variations in chromosome and gene copy number to regulate expression levels. Using ecological genomics and experimental evolution approaches, we provide evidence that *Leishmania* adaptation relies on epistatic interactions between functionally associated gene copy number variations in pathways driving fitness gain in a given environment. We further uncover posttranscriptional regulation as a key mechanism that compensates for deleterious gene dosage effects and provides phenotypic robustness to genetically heterogeneous parasite populations. Finally, we correlate dynamic variations in small nuclear RNA (snRNA) gene dosage with changes in ribosomal RNA 2'-O-methylation and pseudouridylation, suggesting 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 selective pressure. The 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.**

genome instability | epistatic interactions | posttranscriptional regulation | evolutionary adaptation | fitness gain

**D**arwinian evolution plays a central, yet poorly understood, role in human disease. Iterative rounds of genetic mutation and environmental selection drive tumor development, microbial fitness, and therapeutic failure. Genome instability is a key source for genetic and phenotypic diversity, often defining disease outcome (1–4). However, the mechanism(s) by which genome instability is harnessed for fitness gain despite its potential deleterious effects remain largely elusive. Here we investigate this question in the protozoan parasite *Leishmania*, which causes a spectrum of severe diseases known as leishmaniasis that generate substantial human morbidity worldwide (5). Visceral leishmaniasis (also known as kala azar) is caused by *Leishmania donovani* or *Leishmania infantum* and is fatal if left untreated. Most *Leishmania* species show a digenic life cycle comprising two major developmental stages that infect two distinct hosts. The motile, extracellular promastigote form of *Leishmania* proliferates inside the digestive tract of phlebotomine sand flies, which transmit the highly virulent

metacyclic form of the parasite into mammalian hosts during a blood meal. There, following uptake by macrophages, *Leishmania* develops into nonmotile, intracellular amastigotes that proliferate inside fully acidified phagolysosomes and subvert host cell immuno-metabolism functions, thus causing the severe immuno-pathologies underlying leishmaniasis.

Genome instability is a hallmark of *Leishmania* biology since these parasites lack promoter-dependent gene regulation (6, 7) but exploit chromosome and gene copy number variations (CNVs) to regulate messenger RNA (mRNA) abundance by gene dosage (8–12). In the absence of confounding transcriptional control, *Leishmania* thus represents an ideal system to investigate the role of genome instability in fast-evolving eukaryotic cells. We applied an experimental evolution (EE) approach and assessed changes at genomic and transcriptomic levels during adaptation of animal-derived parasites to in vitro culture. As demonstrated in our previous study (11), fitness

## Significance

**Chromosome and gene copy number variations often correlate with the evolution of microbial and cancer drug resistance, thus causing important human mortality. How genome instability is harnessed to generate beneficial phenotypes and how deleterious gene dosage effects are compensated remain open questions. The protist pathogen *Leishmania* exploits genome instability to regulate expression via gene dosage changes. Using these parasites as a unique model system, we uncover complex epistatic interactions between gene copy number variations and compensatory transcriptomic responses as key processes that harness genome instability for adaptive evolution in *Leishmania*. Our results propose a model of eukaryotic fitness gain that may be broadly applicable to pathogenic fungi or tumor cells known to exploit genome instability for adaptation.**

Author contributions: G.B., P.P., M.A.D., K.S.R., and G.F.S. designed research; G.B., L.P., P.P., and K.S.R. performed research; G.B., L.P., P.P., K.S.R., S.C.-C., T.D., D.-G.H., P.J.M., R.U., S.M., and G.F.S. analyzed data; and G.B., P.P., K.S.R., and G.F.S. wrote the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission.

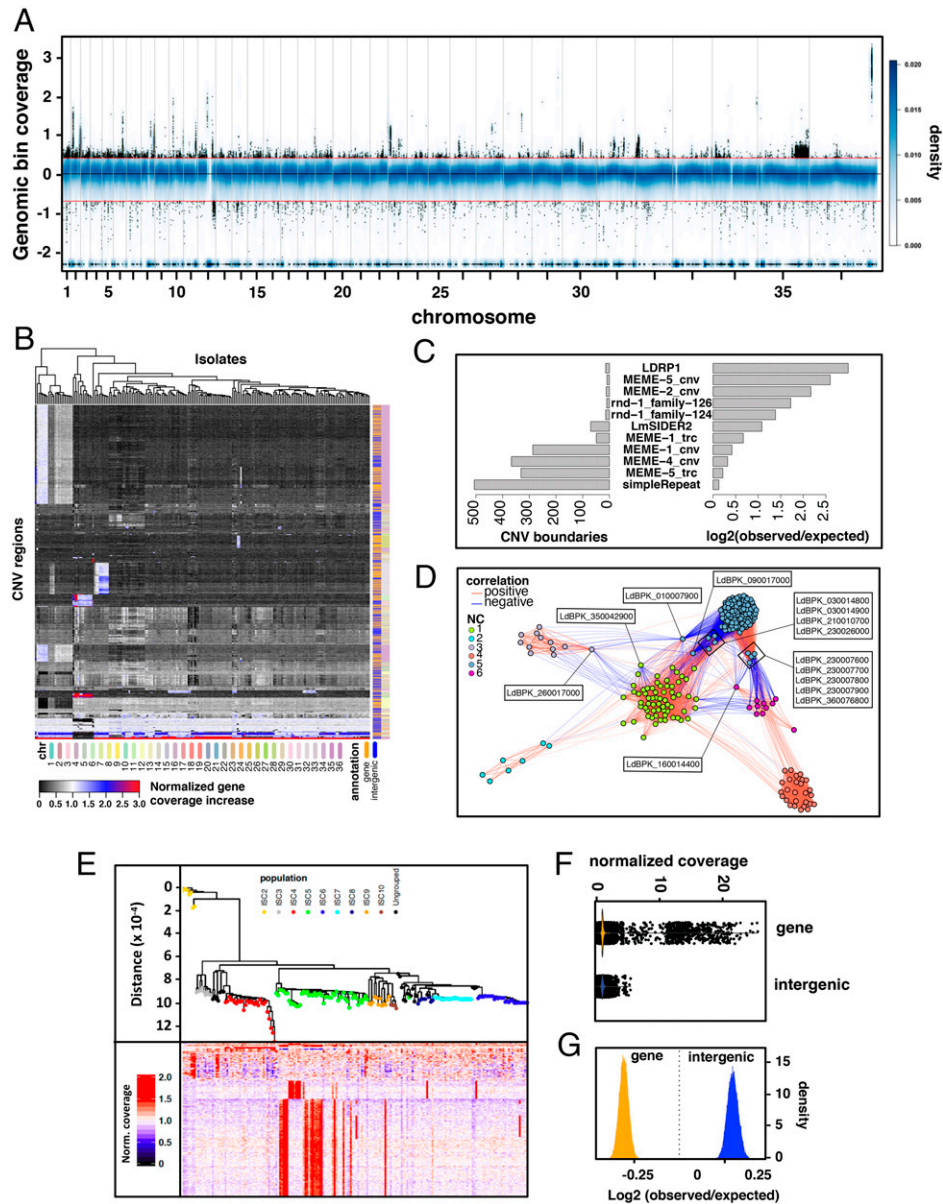
This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>Present address: Institut de Biologie Intégrative de la Cellule (I2BC) UMR 9198, Gif-sur-Yvette 91190, France.

<sup>2</sup>To whom correspondence may be addressed. Email: gerald.spaeth@pasteur.fr.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113744118/-DCSupplemental>.

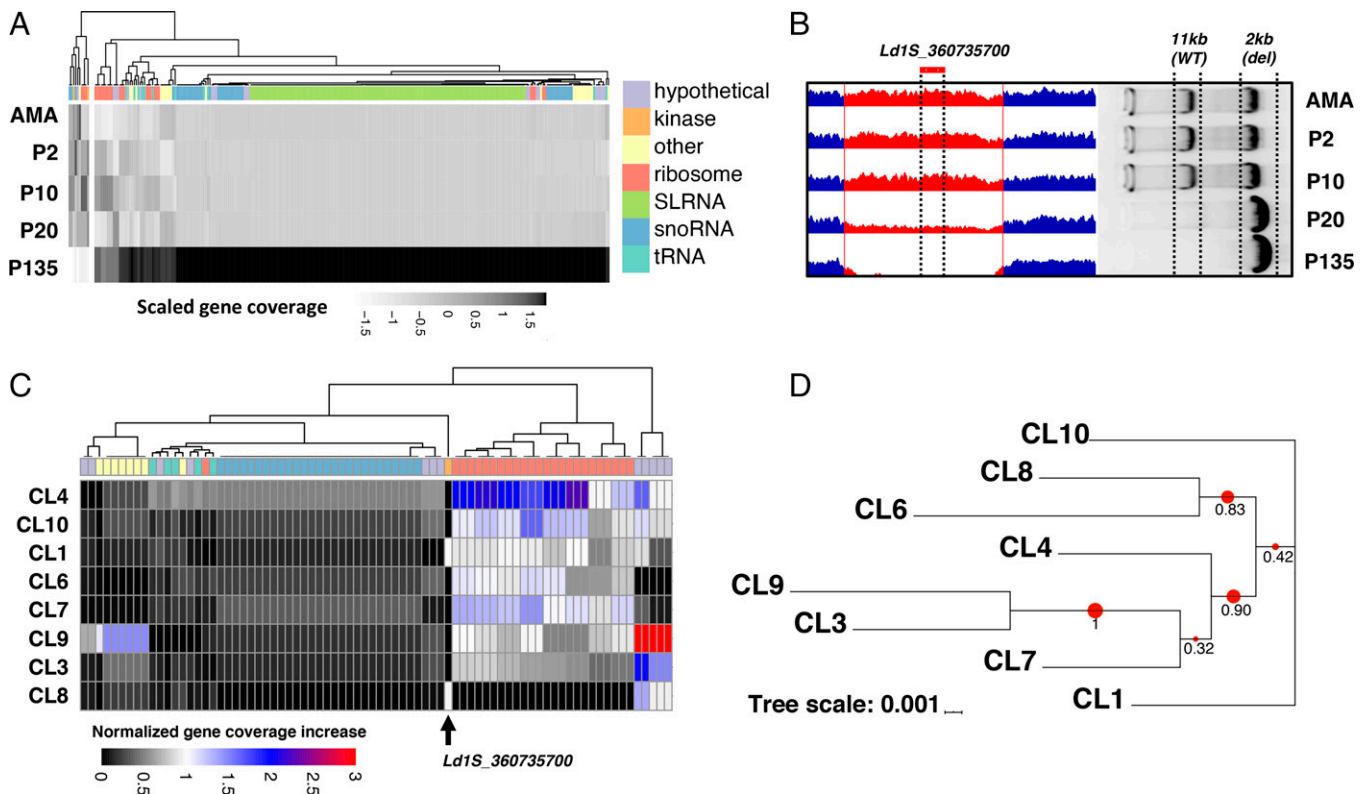
Published December 13, 2021.



**Fig. 1.** Genome-wide mapping of CNVs, their environmental selection, and epistatic interactions. (A) Genome-wide normalized coverage values in natural logarithm scale (y axis) across the 36 chromosomes (x axis) for 204 *L. donovani* field isolates from the Indian subcontinent (ISC) (13). The x axis reports the position of genomic windows along the chromosomes. The smoothed blue color represents the two-dimensional kernel density estimate of genomic bins. A sample of 50,000 genomic bins with normalized coverage  $\geq 1.5$  or  $\leq 0.5$  are materialized as black dots. The black horizontal line and the two red lines indicate normalized coverage values of 1.5, 1, and 0.5. (B) Heatmap generated for the 204 clinical isolates (columns) showing their phylogenetic relationship as a function of CNV regions (rows). The color scale reflects the deviation from the minimum bin coverage observed across all genomes. The two-colored columns on the right report the presence of annotated genes (orange) and intragenic regions (blue), and the chromosomal location of the CNVs is defined by the legend below the plot. (C) Association between CNVs and repetitive elements. The bar plots show the number of observed overlap instances between the boundaries of the CNV regions and repetitive elements (Left), and the  $\log_2$  ratio between the observed and expected overlap events over 10,000 randomizations (Right). (D) Gene CNV network analysis. The nodes represent gene CNVs while the edges indicate statistically significant positive (red) and negative (blue) correlations observed in the 204 field isolates. The nodes are colored according to the predicted network clusters (NC). The indicated gene identifiers refer to key nodes of the network, i.e., genes that connect positive or negative correlating clusters, such as a putative protein kinase (LdBPK\_160014000), a member of the 4F5 protein family (LdBPK\_010007900), the eukaryotic translation initiation factor 2 subunit alpha (LdBPK\_030014900), or the rRNA gene LdBPK\_210010700. (E) Phylogenetic tree based on SNVs ( $>90\%$  frequency) (Upper) for the ISC core population comprising 191 isolates (13) [not including the distant ISC1 strains (48); see also cladogram in *SI Appendix, Fig. S5*]. The heatmap (Lower) shows the normalized gene sequencing coverage across genomes (rows). To ease the visualization, gene amplifications with normalized coverage  $> 2$  are indicated as 2. (F) Violin plot showing the distributions of the normalized genomic coverage values of the collapsed CNV positions (dots) matching genic (orange) and intergenic (blue) regions. (G)  $\log_2$  ratio distributions of observed and expected nucleotide overlap between collapsed CNV regions and gene/intergenic annotations.

gain in this in vitro system is driven largely by the rate of parasite proliferation, which depends on accelerated cell cycle progression and increased transcription and translation efficiencies. Here, we uncover complex epistatic interactions between

CNVs and compensatory transcriptomic responses as key processes that harness genome instability for fitness gain in *Leishmania*. Our data may be broadly applicable to pathogenic fungi or cancer cells known to exploit genome instability for adaptation.



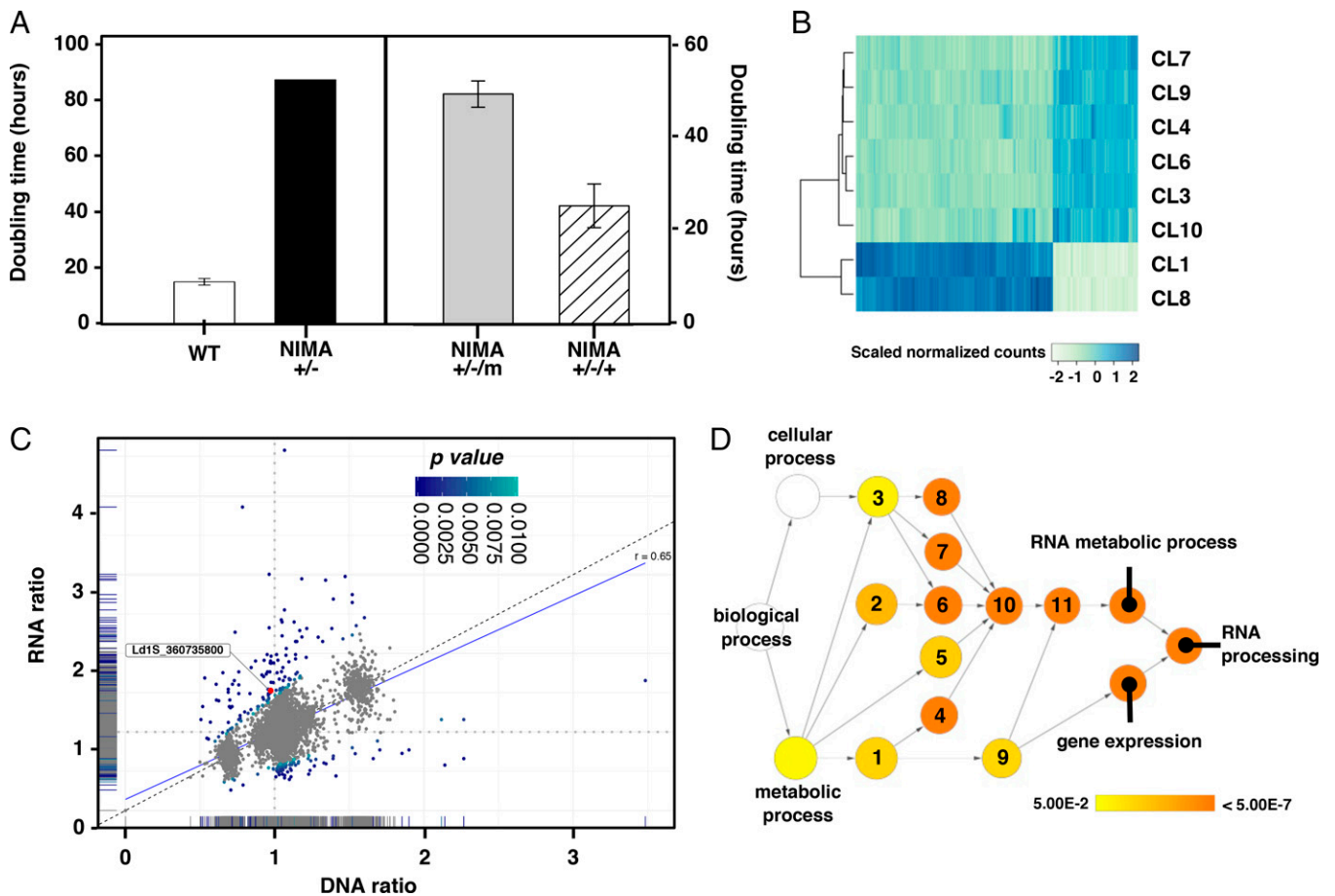
**Fig. 2.** Longitudinal analysis of gene CNVs during fitness gain in culture. (A) Heatmap generated by plotting gene read depth values (columns) across *L. donovani* amastigotes isolated from infected hamster spleen (AMA) and derived promastigotes evolving in continuous culture for 2, 10, 20, and 135 passages (P2, P10, P20, P135) (rows). The gray level reflects the scaled normalized gene coverage as indicated in the figure. The colored ribbon indicates the simplified gene annotation as shown in the legend to the right. (B) Screenshot of the IGV genome browser showing gradual loss of the NIMA-like kinase gene *Ld1S\_360735700* during culture adaptation between splenic amastigotes (AMA) and derived promastigotes at passages P2 and P135. The right panel shows the gel-electrophoretic analysis of PCR fragments obtained from lesion amastigotes (AMA) and derived promastigotes at the indicated culture passages that are diagnostic for the WT (11 kb) and the deleted NIMA-like kinase locus (2 kb) (see *SI Appendix, Fig. S7A* for a schematic overview of the PCR strategy). (C) Heatmap generated by plotting gene read depth variation (columns) across eight clones isolated from the P20 population (rows). The color code is defined in the legend and corresponds to the deviation from the minimum sequencing coverage measured for that gene in all clones. The colored ribbon indicates the simplified gene annotation as shown in the legend of A. The deletion of the NIMA-like kinase *Ld1S\_360735700* is indicated by the arrow. (D) Analysis of the phylogenetic relation of the P20 clones. The red dots and numbers indicate the node bootstrap support (1 = 100%). The genetic distance is indicated by the branch length and scaled as indicated in the figure.

## Results

**Leishmania Genomic Adaptation Is Governed by Epistatic Interactions between Gene CNVs.** We first assessed the level of CNV across 204 *L. donovani* clinical isolates from the Indian subcontinent (13). This collection includes a core group of 191 strains that are genetically highly homogenous as judged by the small number of single nucleotide variants (SNVs) (<2,500 total), which provided us with a useful benchmark to study the dynamics of CNVs across a large number of quasiclonal populations. DNA read depth analysis of these isolates revealed important CNVs in both coding and noncoding (nc) regions, with amplifications and deletions affecting respectively 14% and 4% of the genome (Fig. 1 *A* and *B* and *Datasets S1* and *S2*). Analyzing the statistical association of observed CNVs with repetitive sequence elements uncovered 11 DNA repeats, including the previously described SIDER (14) and LDRP1 elements (15). In addition, we describe simple/low complexity repeats (16) and eight repetitive sequence elements that may drive *Leishmania* genome instability through microhomology-mediated, break-induced replication as observed for human CNVs (17) (Fig. 1C, *SI Appendix, Fig. S1*, and *Dataset S3*). Gene dosage changes were not random but clearly under selection as judged by the reproducibility of genetic interactions across independent isolates and the enrichment of amplified genes in biological functions associated with proliferation and

thus fitness gain in culture. Statistically significant interactions were observed between positive (correlating) and negative (anticorrelating) read depth variations (Fig. 1D, *SI Appendix, Figs. S2* and *S3*, and *Datasets S4* and *S5*), including a highly connected network cluster (NC) containing 60 coamplified transfer RNA (tRNA) genes encoded on 16 different chromosomes (NC1, Fig. 1D, *SI Appendix, Fig. S4*, and *Datasets S6* and *S7*). Natural selection of gene CNVs is further supported by 1) their independent emergence across phylogenetically distinct strains providing evidence for evolutionary convergence, 2) the very high copy number observed for certain genes (up to 21-fold) suggesting strong, positive selection, and 3) the global enrichment of read depth variations in phenotypically silent, intergenic regions, suggesting purifying selection against deleterious effects caused by gene CNVs (Fig. 1 *E–G*, *SI Appendix, Fig. S5*, and *Datasets S8–S10*). Together, our data provide evidence that *Leishmania* genomic adaptation is governed by gene CNVs through highly dynamic, functional interactions that are under natural selection. These interactions define a form of epistasis at the gene (rather than nucleotide) level, with the phenotypic effect of a given gene amplification being dependent on coamplification of functionally related genes.

**Epistatic Adaptation Drives Polyclonal Fitness Gain In Vitro.** We next used an EE approach to directly assess the link between

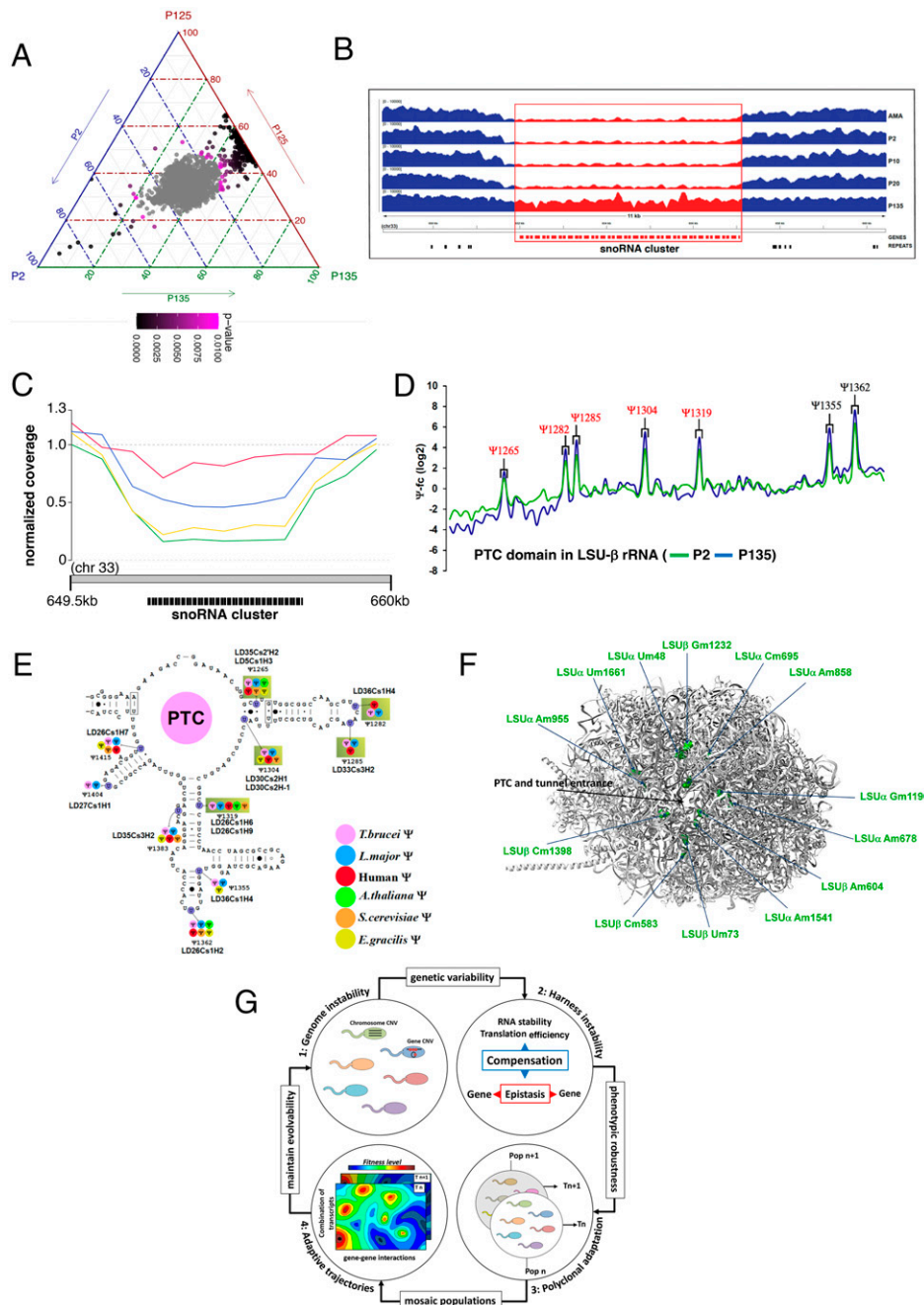


**Fig. 3.** Genetic and transcriptomic analyses of the NIMA-like kinase null mutant. (A) Growth analysis. WT and heterozygous NIMA<sup>+/-</sup> mutants generated by CRISPR/Cas9 gene editing (Left; see *SI Appendix, Fig. S7 D and E* for gene editing strategy and PCR validation of the mutant). Transgenic NIMA<sup>+/-</sup> parasites transfected with empty vector (NIMA<sup>+/-/m</sup>) or vector encoding for the NIMA-like kinase gene (NIMA<sup>+/-/+</sup>) (Right). The doubling time of parasites in logarithmic culture phase is shown. (B) Heatmap of the scaled normalized RNAseq counts of the genes differentially expressed in the P20 clones 1 and 8 (WT) with respect to the spontaneous NIMA null mutant (spo-KO) clones 3, 6, 4, 7, 9, and 10. A log<sub>2</sub> fold change > 0.5 with adjusted *P* value < 0.01 was considered significant. Darker levels of blue reflect higher expression levels as indicated by the legend. (C) Double-ratio scatter plot. The plot represents the ratio of the mean DNA (*x* axis) and RNA (*y* axis) sequencing read counts between the clones that lost the NIMA kinase gene (CL3-4-6-7-9-10) and the NIMA kinase WT clones (CL1 and CL8). Each dot represents an individual gene. The marginal distributions for DNA and RNA ratio values are displayed along the *x* and *y* axes. The color indicates the statistical significance level of the genes' double-ratio scores (i.e., RNA ratio divided by DNA ratio) as indicated in the legend. The NIMA-like kinase homolog Ld1S\_360735800 is labeled in red. The vertical and horizontal dotted lines indicate DNA and RNA ratio values of 1, while the diagonal dashed line specifies the bisector. The blue line represents a linear regression model built on the DNA and RNA ratio values and measuring a Pearson correlation value of 0.65. (D) Functional enrichment analysis of the biological process Gene Ontology terms for all genes in C showing a statistically significant double-ratio score (Dataset S18). The node color mapping is ranging from yellow to dark orange to represent increasing significance levels, or lower adjusted *P* values. White nodes are not significant. 1: organic substance metabolic process, 2: nitrogen compound metabolic process, 3: cellular metabolic process, 4: organic cyclic compound metabolic process, 5: primary metabolic process, 6: cellular nitrogen compound metabolic process, 7: heterocycle metabolic process, 8: cellular aromatic compound metabolic process, 9: macromolecule metabolic process, 10: nucleobase-containing compound metabolic process, 11: nucleic acid metabolic process.

epistatic interactions and fitness gain in hamster-derived *L. donovani* parasites during adaptation to in vitro culture (11) (*SI Appendix, Fig. S6A*; EE1). Following normalization for karyotypic variations (*SI Appendix, Fig. S6B* and Dataset S11), changes in read depth were monitored between passages 2 (P2, 2 wk in culture) and 135 (P135, 36 wk in culture), corresponding to ~20 and 3,800 generations. Our analysis revealed coamplification of coding and nc genes and gene clusters that are functionally linked to fitness gain in culture (i.e., accelerated cell proliferation), including genes encoding for ribosomal RNAs (rRNAs), tRNAs, small nuclear (snRNAs), small nucleolar RNA (snoRNA), spliced leader RNAs (SLRNAs), and ribosomal proteins (RPs; Fig. 2A, *SI Appendix, Fig. S6C*, and Datasets S12–S14). Epistatic adaptation to culture thus affects pathways linked to translation in both the LDBPK (Fig. 1) and Ld1S isolates, with amplification observed for homologous

genes encoding for various tRNAs, ribosomal RNAs (5S, 5.8S, 18S, 28S), or 60S acidic RPs (Dataset S10). Indeed, translation efficiency is one of the major rate-limiting steps for proliferation (18), with enhanced protein synthesis fulfilling the increased need for various molecular machines (DNA and RNA polymerases, ribosomes) in biosynthetically highly active, fast-growing cells. This functional link between a given fitness phenotype and its underlying epistatic network opens unexplored venues for the discovery of *Leishmania* pathways that govern adaptation to a given environment.

Unexpectedly, we discovered that gene depletion—as well as gene amplification—is also a major driver for environmental adaptation and fitness gain. We identified a genomic deletion of 11 kb containing a single gene encoding for a NIMA-related kinase gene (Ld1S\_360735700), which is gradually selected in the adapting promastigote population from a preexisting



**Fig. 4.** snoRNA genes are amplified in long-term adapted parasites and promote rRNA modification. (A) Ternary plot showing for each gene the relative abundance in culture passage P2, P125, and P135. The axes report the fraction of the normalized gene coverage in each sample, with each given point adding up to 100. Dots with color ranging from pink to black indicate significant gene CNVs ( $P$  value  $< 0.001$ ). (B) Recovery of the full snoRNA gene cluster during culture adaptation. The panel illustrates a genome browser representation of the sequencing depth measured in the samples at the indicated passage. Gene annotations and the predicted repetitive elements are indicated. (C) Line plot showing the normalized sequencing coverage (y axis) over the snoRNA cluster region on chromosome 33 (x axis) for the original amastigote sample used to derive the P2 strain (AMA, green) and three independent amastigote isolates (AMAH154, red; AMA07142, yellow; AMA1992, blue) obtained from different hamster infections. (D) Hyper-pseudouridylated ( $\Psi$ ) sites are present in the PTC functional domain of the ribosome. The relative change in  $\Psi$  levels was measured using  $\Psi$ -seq and is presented in Dataset S21. Representative line graph of the fold change in pseudouridylation levels ( $\Psi$ -fc,  $\log_2$ ) along the rRNA nucleotides (x axis) is presented for P2 (green line) and P135 (blue line) for the PTC domain in LSU- $\beta$  rRNA. The positions where the  $\Psi$  level is increased in four replicates are indicated in red. (E) The location of  $\Psi$  sites in the *L. donovani* Ld15 secondary structure, which is identical to *Leishmania major* and similar to *T. brucei* (25, 49). Hypermodified sites are highlighted in boxes. The snoRNA guiding on each  $\Psi$  is indicated. The color code for each  $\Psi$  site is indicative of the organism where it was already reported. (F) Hypermethylated (Nm) sites are located around the functional domains of the ribosome. The complete stoichiometry of each Nm site was measured by RiboMeth-seq in P2 and P135 *L. donovani* strains as presented in Dataset S20. The hypermodified Nm sites are highlighted in green, and their identity is indicated on the three-dimensional structure of the *L. donovani* large subunit ribosome based on the previously deposited cryo-EM coordinates [Protein Data Bank (50) accession 6AZ3] (26). (G) Model of *Leishmania* polyclonal adaptation. 1) *Leishmania* intrinsic genome instability generates constant genetic variability. 2) Epistatic interactions between gene CNVs and compensatory responses at the level of RNA stability and translation efficiency eliminate toxic gene dosage effects and harness genome instability. 3) This mechanism generates phenotypic robustness while at the same time maintaining genetic variability, thus allowing for polyclonal adaptation. 4) The genetic mosaic structure allows for distinct adaptive trajectories (T) inside and in-between adapting populations, a process that conserves genetic heterogeneity and thus evolvability of the population despite constant selection.

mutant that was detected in splenic amastigotes (Fig. 2B and Fig. S7 A and B). Clonal analysis of the P20 population revealed the presence of the spontaneous knockout (spo-KO) in six out of eight individual clones (Fig. 2C). The only partial penetration of spo-KO cells after several months in culture indicates that the fitness gain provided by the absence of the NIMA kinase may be very small, a fact we confirmed by establishing growth curves for wild-type (WT) clone CL1 and spo-KO clone CL6, which was indeed unable to identify a statistically significant difference in proliferation (SI Appendix, Fig. S7C). The different spo-KO clones were clearly not the descendants of a single founder cell but were of independent, polyclonal origin as judged by their distinct gene CNV profiles (Fig. 2C and Dataset S15), and their polyphyletic clustering based on SNVs compared to WT clones (Fig. 2D and Dataset S16). This evolutionary convergence strongly supports natural selection of the deletion during culture adaptation and suggests a potential role of the deleted NIMA kinase in growth restriction, which we further assessed by gene editing.

**Toxic Gene Dosage Effects Are Compensated at Posttranscriptional Levels.** Unlike the spo-KO clones, CRISPR/Cas9-generated, NIMA knockout mutants (cri-KO) (SI Appendix, Fig. S7 C and D) were not viable, while heterozygous mutants showed a strong growth defect, which was partially rescued by episomal overexpression of the NIMA kinase gene (Fig. 3A). This paradoxical result suggests that spo-KO cells must have evolved mechanisms that can compensate for the loss of this essential gene. Read depth analyses of the spo-KO and WT clones ruled out genetic compensation (Dataset S15). In contrast, RNA sequencing (RNAseq) analysis showed highly reproducible, compensatory transcript profiles in the six independent spo-KO clones (Fig. 3B and Dataset S17). Our analysis revealed reduced stability in spo-KO clones of 23 transcripts implicated in flagellar biogenesis (SI Appendix, Fig. S8), which correlated with reduced flagellar activity (Videos S1 and S2). Loss of motility may represent a fitness tradeoff, providing energy required for accelerated in vitro growth. Likewise, the spontaneous loss of the NIMA-like kinase gene may be the result of an evolutionary process that selects against motility, given the implication of the NIMA-related kinase Cnk2p in regulating flagellar length in the protist *Chlamydomonas* (19) and the localization of the *Trypanosoma brucei* NIMA-related kinase TbNRKC to the flagellum basal body (20). However, selection for the spontaneous NIMA deletion mutant was not observed in three independent evolutionary experiments conducted with different amastigote isolates (SI Appendix, Fig. S6A; EE1-3, SI Appendix, Fig. S9A), suggesting that compensation for the essential NIMA functions may be complex and thus rather a rare event.

On the other hand, we identified 350 transcripts with increased abundance (Dataset S17, second sheet), which either result from increased gene dosage or mRNA posttranscriptional stabilization. Direct comparison of DNA and RNA read depth variations allowed us to distinguish between these two possibilities and identified a set of transcripts whose expression changes between WT and spo-KO did not correlate with gene dosage. In absence of transcriptional regulation in *Leishmania* (6), the abundance of these transcripts is likely regulated by differential RNA turnover. Increased abundance was observed in spo-KO clones for the mRNA of another NIMA-related kinase (Ld1S\_360735800) encoded adjacent to the deleted region, suggesting a direct, posttranscriptional compensation of kinase functions (Fig. 3C and Dataset S18). Likewise, we observed increased stability for functionally related small ncRNAs, including 43 snoRNAs, several rRNAs, and tRNAs but also various metabolic enzymes (e.g., calpain-like cysteine peptidases, a glycosomal phosphoenolpyruvate carboxykinase or a

histone methylation protein DOT1) (Fig. 3D and Dataset S18), suggesting ribosomal biogenesis, translational regulation, and various metabolic processes as yet additional levels of nongenomic adaptation. Together, our data identify gene deletion and compensatory, posttranscriptional responses as drivers of *Leishmania* fitness gain.

**The ncRNome as a Driver of *Leishmania* Fitness Gain.** The selective stabilization of snoRNAs during early culture adaptation (P2-P20) suggests these ncRNAs as key drivers in *Leishmania* fitness gain. We confirmed this possibility in long-term adapted parasites that were continuously cultured for ~3,800 generations (P135). Read depth analysis of P2, its derived P135 population, and an independent population evolved until P125 (SI Appendix, Fig. S6A; EE4) revealed selective amplification of snoRNAs, as judged by the shift observed in the shown ternary plot (Fig. 4A, SI Appendix, Fig. S9B, and Dataset S19). Surprisingly, rather than amplification of individual snoRNA genes, increased read depth was caused by the recovery of a single locus on chromosome 33 containing a cluster of 15 snoRNA genes in the P135 population, which was depleted in the original amastigote population (Fig. 4B). The restoration of this locus between P20 and P135 is further proof that snoRNAs are under positive selection during culture adaptation. Rather than intrachromosomal expansion of gene copies, the recovery of this locus is more likely due to the selection of a small subpopulation that may preexist in the original amastigote isolate and penetrate the culture between P20 and P135, similar to what we observed for the NIMA kinase spo-KO. This scenario is indeed supported by analyzing CNVs in the snoRNA cluster across independent amastigotes isolates, which revealed significant differences in snoRNA gene copy number, as indicated by the observed variations in read depth (Fig. 4C and SI Appendix, Fig. S10). Thus, while short-term adaptation in *Leishmania* is governed by control of RNA stability, long-term adaptation occurs through more cost-efficient gene dosage regulation. Increased snoRNA abundance likely satisfies a quantitative need for ribosomal biogenesis in our fast-growing cultures but could also affect the nature and quality of ribosomes (21, 22).

In the following, we assessed the possibility of such fitness-adapted ribosomes in *Leishmania*, especially since snoRNA genes guiding 2'-O-methylation (Nm) and pseudouridine ( $\Psi$ ) modifications were extensively amplified. Amplification of snoRNAs should lead to increased modifications on sites that are accessible for modification. The mapping of Nm via RiboMeth-seq (23) revealed an increased level of modification for 18 sites by at least twofold (Dataset S20), while the mapping of pseudouridylation by  $\Psi$ -seq (24, 25) showed increase for 5  $\Psi$  sites during adaptation from P2 to P135 (Fig. 4D and Dataset S21). Interestingly, mapping these sites on the resolved cryogenic electron microscopy (cryo-EM) structure of the *L. donovani* large ribosomal subunit (26) reveals their localization around the peptidyl-transferase center (PTC) and mRNA entrance tunnel, whereas the hyper-modified  $\Psi$  sites are located in the PTC itself (Fig. 4 E and F and SI Appendix, Fig. S11). Together, our data infer a complex model of *Leishmania* fitness gain where epistatic interactions between gene amplifications and compensatory responses at posttranscriptional levels harness *Leishmania* genome instability for polyclonal adaptation (see model Fig. 4G and details in the legend).

## Discussion

A common strategy in microbial evolutionary adaptation is known as “bet-hedging,” where the fitness of a population in changing environments is increased by stochastic fluctuations in gene expression that are regulated at epigenetic levels (27–31). The protozoan pathogen *Leishmania* largely lacks transcriptional

regulation, raising the question of how these parasites generate variability in transcript abundance and phenotype required for adaptation (6, 7). Our data uncover an alternative mechanism of bet-hedging that has evolved based on the unique biology of *Leishmania*.

First, we provide evidence that *Leishmania* genome instability may be driven by 11 repetitive DNA elements associated with genome-wide amplifications and deletions, both of which are under positive and purifying selection. *Leishmania* thus compensates for the absence of stochastic gene regulation by the generation of stochastic gene CNVs, which can cause dosage-dependent changes in transcript abundance (8–12). Second, we reveal an important role for posttranscriptional regulation in *Leishmania* fitness gain, which 1) compensates for the deleterious deletion of a NIMA kinase by selectively stabilizing the transcripts of an orthologous kinase, 2) allows for gene dosage-independent increase in the abundance of ncRNAs (SLRNAs, snoRNAs) required for proliferation, and 3) provides phenotypic robustness to genetically heterogeneous populations as documented by the converging transcript profiles of independent NIMA spo-KO mutants. This adaptation process guards against toxic gene dosage effects and simultaneously increases the phenotypic landscape available to *Leishmania* for adaptation via gene deletions and compensatory transcriptional responses. Significantly, the NIMA ortholog as well as the SLRNA and snoRNA loci are amplified during long-term culture (Fig. 4A and B and *SI Appendix*, Figs. S6C and S9B), revealing a two-step adaptation process reminiscent to yeast (32), implicating first, a posttranscriptional mechanism via transient changes in RNA stability followed by a second genomic mechanism via selection for stable CNVs.

The dynamic changes in snoRNA stability and gene copy number observed in our EE system identifies this class of ncRNAs as an unexpected driver of *Leishmania* fitness gain. snoRNAs guide rRNA modification and processing as well as modification of snRNAs and additional ncRNAs (33, 34). There is currently no information on the effect of individual rRNA  $\Psi$  and Nm modifications on translation. It has been shown that ribosomes with different RP composition are specifically affected in translation of only a subset of proteins (35). Conceivably, changes in rRNA conformation and interaction due to modification may have similar effects. Indeed, we recently demonstrated that individual  $\Psi$  modifications affect protein binding (33, 36) and similarly may impact RP binding, generating specialized ribosomes with different translation properties. In addition,  $\Psi$  modification also affects rRNA structure, which was shown to change translation efficiency and fidelity (22, 37). Based on our results, it is interesting to speculate that differential snoRNA expression and subsequent differential rRNA modification generates fitness-adapted ribosomes with unique translation properties. Such specialized ribosomes may represent an additional layer of regulation capable of counteracting toxic gene dosage effects, providing phenotypic robustness, and adapting the proteome profile to a given

environment, much like it was observed in cancer cells or during differentiation of stem cells (38, 39).

Our findings have important clinical implications for *Leishmania* infection. *Leishmania* adapts to various environmental cues, notably the presence of antileishmanial drugs. In contrast to high frequency amplifications observed during experimental drug treatment in culture (40–42), treatment failure and drug resistance observed during natural infection may evolve through multilocus epistatic interactions such as those described here, which can balance the fitness tradeoff between drug resistance and infectivity (43). Therefore, our data define biological networks, rather than individual genes, as biomarkers with potential diagnostic or prognostic value. Conceivably, the epistatic mechanisms we uncover in *Leishmania* can be of broader relevance to other human pathologies caused by fast-evolving eukaryotic cells exploiting genome instability for polyclonal adaptation, such as cancer cells. While single nucleotide, epistatic interactions are recognized as important drivers of tumor development (44, 45), the role of epistatic interactions between structural mutations and between the genome and transcriptome in drug-resistant cancer cells remains to be elucidated (46).

In conclusion, our results propose a model of *Leishmania* fitness gain (Fig. 4F), where polyclonal adaptation of mosaic populations is driven by epistatic interactions that 1) buffer the detrimental effects of genome instability, 2) coordinate expression of functionally related genes, and 3) generate beneficial phenotypes for adaptation. This mechanism of fitness gain avoids genetic death by maintaining heterogeneity in competing parasite populations under environmental selection and may be generally applicable to other eukaryotic systems that adapt through genome instability.

## Materials and Methods

The extended supporting materials and methods provide detailed information on the 1) *L. donovani* strains, culture conditions, and cell cloning, 2) nucleic acid extraction and deep sequencing, 3) whole genome sequencing data analysis, 4) reference genomes, 5) repeat analysis, 6) CNV association analyses, 7) network analysis, 8) phylogenetic analysis, 9) RNAseq analysis, 10) Gene Ontology term enrichment analysis, 11) PCR analyses, 12) null mutant analysis, and 13) methylation (Nm) and pseudouridine ( $\Psi$ ) rRNA modification analyses.

**Data Availability.** Reads were deposited in the Sequence Read Archive database (47) and are publicly available under accession no. PRJNA605972. All data are available in the main text or the *SI Appendix*. Previously published data were used for this work (11, 13).

**ACKNOWLEDGMENTS.** This study was supported by a seeding grant from the Institut Pasteur International Department to the LeishShield Consortium, the EU H2020 project LeishShield-MATI-REP-778298-1, the Fondation pour la Recherche Médicale (Grant 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 41131ZD. We thank Cedric Notredame and Jean-Claude Dujardin for critical reading of the manuscript.

1. E. Darmon, D. R. Leach, Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* **78**, 1–39 (2014).
2. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
3. D. Hughes, D. I. Andersson, Evolutionary consequences of drug resistance: Shared principles across diverse targets and organisms. *Nat. Rev. Genet.* **16**, 459–471 (2015).
4. N. McGranahan, C. Swanton, Clonal heterogeneity and tumor evolution: Past, present, and the future. *Cell* **168**, 613–628 (2017).
5. Anonymous, Leishmaniasis in high-burden countries: An epidemiological update based on data reported in 2014. *Wkly. Epidemiol. Rec.* **91**, 287–296 (2016).
6. C. E. Clayton, Gene expression in Kinetoplastids. *Curr. Opin. Microbiol.* **32**, 46–51 (2016).
7. S. Michaeli, Trans-splicing in trypanosomes: Machinery and its impact on the parasite transcriptome. *Future Microbiol.* **6**, 459–474 (2011).
8. F. Dumetz et al., Modulation of aneuploidy in *Leishmania donovani* during adaptation to different *in vitro* and *in vivo* environments and its impact on gene expression. *MBio* **8**, e00599-17 (2017).
9. S. A. Iantorno et al., Gene expression in *Leishmania* is regulated predominantly by gene dosage. *MBio* **5**, e01393-17 (2017).
10. P. Leprohon et al., Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Res.* **37**, 1387–1399 (2009).
11. P. Prieto Barja et al., Haplotype selection as an adaptive mechanism in the protozoan pathogen *Leishmania donovani*. *Nat. Ecol. Evol.* **1**, 1961–1969 (2017).
12. M. B. Rogers et al., Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* **21**, 2129–2142 (2011).
13. H. Imamura et al., Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife* **5**, e12613 (2016).



14. M. Smith, F. Bringaud, B. Papadopolou, Organization and evolution of two SIDER retroposon subfamilies and their impact on the Leishmania genome. *BMC Genomics* **10**, 240 (2009).
15. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
16. J. Ellis, J. Crampton, Characterisation of a simple, highly repetitive DNA sequence from the parasite *Leishmania donovani*. *Mol. Biochem. Parasitol.* **29**, 9–17 (1988).
17. P. J. Hastings, J. R. Lupski, S. M. Rosenberg, G. Ira, Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551–564 (2009).
18. C. Waldron, F. Lacroute, Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J. Bacteriol.* **122**, 855–865 (1975).
19. B. A. Bradley, L. M. Quarmby, A NIMA-related kinase, Cnk2p, regulates both flagellar length and cell size in *Chlamydomonas*. *J. Cell Sci.* **118**, 3317–3326 (2005).
20. L. C. Pradel, M. Bonhivers, N. Landrein, D. R. Robinson, NIMA-related kinase TbNRKC is involved in basal body separation in *Trypanosoma brucei*. *J. Cell Sci.* **119**, 1852–1863 (2006).
21. V. Chikne *et al.*, Small nucleolar RNAs controlling rRNA processing in *Trypanosoma brucei*. *Nucleic Acids Res.* **47**, 2609–2629 (2019).
22. X. H. Liang, Q. Liu, M. J. Fournier, rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell* **28**, 965–977 (2007).
23. U. Birkedal *et al.*, Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.* **54**, 451–455 (2015).
24. S. Schwartz *et al.*, Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
25. V. Chikne *et al.*, A pseudouridylation switch in rRNA is implicated in ribosome function during the life cycle of *Trypanosoma brucei*. *Sci. Rep.* **6**, 25296 (2016).
26. M. Shalev-Benami *et al.*, Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. *Nat. Commun.* **8**, 1589 (2017).
27. M. Ackermann *et al.*, Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987–990 (2008).
28. J. K. Graham, M. L. Smith, A. M. Simons, Experimental evolution of bet hedging under manipulated environmental uncertainty in *Neurospora crassa*. *Proc. R. Soc. B Biol. Sci.* **281**, 2014.0706 (2014).
29. I. M. Rouzine, A. D. Weinberger, L. S. Weinberger, An evolutionary role for HIV latency in enhancing viral transmission. *Cell* **160**, 1002–1012 (2015).
30. M. K. Stewart, B. T. Cookson, Non-genetic diversity shapes infectious capacity and host resistance. *Trends Microbiol.* **20**, 461–466 (2012).
31. J. Seger, H. J. Brockmann, "What is bet-hedging?" in *Oxford Surveys in Evolutionary Biology*, P. H. Harvey and L. Partridge, Eds. (Oxford University Press, Oxford, UK, 1987), vol. **4**, pp. 182–211.
32. A. H. Yona *et al.*, Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21010–21015 (2012).
33. K. S. Rajan *et al.*, Pseudouridines on *Trypanosoma brucei* spliceosomal small nuclear RNAs and their implication for RNA and protein interactions. *Nucleic Acids Res.* **47**, 7633–7647 (2019).
34. K. S. Rajan, V. Chikne, K. Decker, H. Waldman Ben-Asher, S. Michaeli, Unique aspects of rRNA biogenesis in trypanosomatids. *Trends Parasitol.* **35**, 778–794 (2019).
35. Z. Shi *et al.*, Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* **67**, 71–83.e7 (2017).
36. K. S. Rajan *et al.*, Pseudouridines on *Trypanosoma brucei* mRNAs are developmentally regulated: Implications to mRNA stability and protein binding. *Mol. Microbiol.* **116**, 808–826 (2021).
37. T. H. King, B. Liu, R. R. McCully, M. J. Fournier, Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol. Cell* **11**, 425–435 (2003).
38. M. Penzo *et al.*, Human ribosomes from cells with reduced dyskerin levels are intrinsically altered in translation. *FASEB J.* **29**, 3472–3482 (2015).
39. C. Bellodi *et al.*, H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell Rep.* **3**, 1493–1502 (2013).
40. F. Ariey *et al.*, A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55 (2014).
41. J. B. Koenderink, R. A. Kavishe, S. R. Rijpma, F. G. Russel, The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26**, 440–446 (2010).
42. A. Selmecki, A. Forche, J. Berman, Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* **313**, 367–370 (2006).
43. A. H. Melnyk, A. Wong, R. Kassen, The fitness costs of antibiotic resistance mutations. *Evol. Appl.* **8**, 273–283 (2015).
44. A. Ashworth, C. J. Lord, J. S. Reis-Filho, Genetic interactions in cancer progression and treatment. *Cell* **145**, 30–38 (2011).
45. J. van de Haar *et al.*, Identifying epistasis in cancer genomes: A delicate affair. *Cell* **177**, 1375–1383 (2019).
46. C. Kim *et al.*, Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* **173**, 879–893.e13 (2018).
47. R. Leinonen, H. Sugawara, M. Shumway, International Nucleotide Sequence Database Collaboration, The sequence read archive. *Nucleic Acids Res.* **39** (suppl. 1), D19–D21 (2011).
48. B. Cuyppers *et al.*, Integrated genomic and metabolomic profiling of ISC1, an emerging *Leishmania donovani* population in the Indian subcontinent. *Infect. Genet. Evol.* **62**, 170–178 (2018).
49. D. Eliaz *et al.*, Genome-wide analysis of small nucleolar RNAs of *Leishmania major* reveals a rich repertoire of RNAs involved in modification and processing of rRNA. *RNA Biol.* **12**, 1222–1255 (2015).
50. H. M. Berman *et al.*, The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).