PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

Leishmania allelic selection during experimental sand fly infection correlates with mutational signatures of oxidative DNA damage

Giovanni Bussotti[a,b,1], Blaise Li[a], Pascale Pescher[b], Barbora Vojtkova[c], Isabelle Louradour[b], Katerina Pruzinova[c], Jovana Sadlova[c], Petr Volf[c], and Gerald F. Späth[b,2]

Edited by Serap Aksoy, Yale University School of Public Health, New Haven, CT; received December 19, 2022; accepted January 13, 2023

Trypanosomatid pathogens are transmitted by blood-feeding insects, causing devastating human infections. These parasites show important phenotypic shifts that often impact parasite pathogenicity, tissue tropism, or drug susceptibility. The evolutionary mechanisms that allow for the selection of such adaptive phenotypes remain only poorly investigated. Here, we use *Leishmania donovani* as a trypanosomatid model pathogen to assess parasite evolutionary adaptation during experimental sand fly infection. Comparing the genome of the parasites before and after sand fly infection revealed a strong population bottleneck effect as judged by allele frequency analysis. Apart from random genetic drift caused by the bottleneck effect, our analyses revealed haplotype and allelic changes during sand fly infection that seem under natural selection given their convergence between independent biological replicates. Our analyses further uncovered signature mutations of oxidative DNA damage in the parasite genomes after sand fly infection, suggesting that *Leishmania* suffers from oxidative stress inside the insect digestive tract. Our results propose a model of *Leishmania* genomic adaptation during sand fly infection, with oxidative DNA damage and DNA repair processes likely driving haplotype and allelic selection. The experimental and computational framework presented here provides a useful blueprint to assess evolutionary adaptation of other eukaryotic pathogens inside their insect vectors, such as *Plasmodium* spp, *Trypanosoma brucei*, and *Trypanosoma cruzi*.

Leishmania | sand fly infection | allelic selection

Significance

During passage through their insect host, vector-borne pathogens show reduction in population heterogeneity, which can limit their adaptive potential. Using a comparative genomic approach, we demonstrate that the protozoan parasite *Leishmania* experiences such a bottleneck effect during experimental sand fly infection. We provide genetic evidence that these parasites accumulate new mutations and thus restore evolvability by exploiting the DNA damaging environment inside the sand fly midgut. We further uncover allelic selection as a key driver of genomic adaptation during *Leishmania* insect infection. Our study sets the stage for future investigations that aim to analyze the impact of reactive oxygen species (ROS) on *Leishmania* genomic adaptation.

Leishmaniases are vector-borne diseases that generate important human morbidity and mortality worldwide (1). These immunopathologies are characterized by a variety of clinical symptoms, including self-healing skin ulcers, disfiguring mucocutaneous lesions, and fatal hepatosplenomegaly. The etiological agents of leishmaniases are trypanosomatid pathogens of the genus *Leishmania* that exploit two major hosts, i) phlebotomine sand flies, where they proliferate extracellularly as motile, insect-stage parasites termed promastigotes and ii) vertebrate (mostly mammalian) hosts that are infected during a sand fly blood meal, where promastigotes develop into immotile, proliferating amastigotes inside acidified macrophage phagolysosomes.

In addition to stage differentiation, *Leishmania* uses an evolutionary strategy to adapt and increase fitness following the Darwinian paradigm of mutation and selection of the fittest (2). As all microbial pathogens, *Leishmania* is constantly under selection by fluctuations in the various host environments, e.g. changes in the sand fly microbiome, the host immunogenetic makeup or the presence of antiparasitic drugs. Unlike other pathogens that exploit stochastic changes in gene expression to generate selectable phenotypes (3, 4), *Leishmania* generates genetic and phenotypic variability through its intrinsic genome instability (5–8). Frequent changes in chromosome and gene copy number cause gene dosage-dependent changes in transcript abundance, which have been linked to experimentally induced drug resistance, changes in tropism, or fitness gain during culture adaptation (9–15). Aside this quantitative aspect on gene expression, we previously showed that gene dosage changes also allow for qualitative transcriptomic changes via haplotype selection, with chromosomal amplification being nonrandom as only observed for one of the two different haplotypes (14). This was observed both in vitro during culture adaptation of the Sudanese isolate *Leishmania donovani* Ld1S, but also during hamster infection in vivo, where transient trisomies seem to allow selection of tissue-specific allelic profiles in parasites recovered from spleen (14).

In contrast to mammalian infection, studies on *Leishmania* genomic adaptation inside its sand fly insect host are scarce and largely limited to karyotypic analyses and single-nucleotide polymorphism (SNP) profiling to monitor genetic hybridization events (11, 16–19). To date, no information is available on other forms of genomic adaptation

Author contributions: P.P., J.S., P.V., and G.F.S. designed research; P.P., B.V., and K.P. performed research; G.B., B.L., P.P., B.V., I.L., J.S., and G.F.S. analyzed data; and G.B., B.L., P.P., I.L., and G.F.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: Veracyte SAS, Immuno-oncology, F-13009 Marseille, France.

[2]To whom correspondence may be addressed. Email: gerald.spaeth@pasteur.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2220828120/-/DCSupplemental.

Published February 27, 2023.

during vector infection. Here, we address this important open question conducting experimental sand fly infection with bona fide amastigotes isolated from infected hamster spleen, and derived promastigotes with different karyotypic profiles. Applying our genome instability bioinformatic pipeline [GIP, (20)] on sand fly-recovered parasites revealed signature mutations of oxidative DNA damage that correlated with haplotype shuffling and allelic selection. These seem to drive parasite fitness gain during vector infection as judged by converging allele frequency shifts across independent infection experiments.

Materials and Methods

Animals and Ethics Statement. Two 7- to –8-wk-old female hamsters were used. Work on animals was performed in compliance with French and European regulations on care and protection of laboratory animals (EC Directive 2010/63, French Law 2013-118, February 6, 2013). All animal experiments were approved by the Ethics Committee and the Animal Welfare body of Institut Pasteur and by the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (project no. #19683).

Hamster Infection and Amastigote Isolation. Anesthetized hamsters H143 and H154 were inoculated by intracardiac injection with 5×10^7 infectious amastigotes obtained from infected hamster spleens (Fig. 1*A*). Hamster weight was monitored weekly, and the animals were killed with CO_2 before reaching the endpoint (20% weight loss). Spleens were collected and homogenized in phosphate-buffered saline (PBS) supplemented with 2.5 mg/mL saponin using the gentleMACS homogenizer with gentleMACS M tubes from Miltenyi Biotec. Amastigotes were purified as previously described (21) for DNA extraction, differentiation into promastigotes and further promastigote culture.

Parasites and Culture. *Leishmania donovani* strain 1S2D (MHOM/SD/62/1S-CL2D) was obtained from Henry Murray, Weill Cornell Medical College, New York, USA and maintained by serial passages in hamsters. The amastigote strains Ama143 and Ama154 were recovered from the spleens of two infected hamsters (H143 and H154) and differentiated into promastigotes in M199 complete medium (M199, 10% FBS, 25 mM HEPES; 100 μM adenine, 2 mM L-glutamine, 10 μg/mL folic acid, 13.7 μM hemin, 4.2 mM $NaHCO_3$, 1xRPMI1640 vitamins, 8 μM 6-biopterin, 100 units penicillin, and 100 μg/mL streptomycin, pH 7.4) at 26 °C. Promastigotes were then maintained in culture by dilution in fresh medium once they were reaching stationary phase. Promastigotes derived from Ama154 were maintained in culture until passage 2 (P2), while promastigotes derived from Ama143 were maintained in culture until P22 (Fig. 1*A*). This experimental lay out allowed us to i) investigate potential signals of stage-specific adaptation (comparing Ama154 with matching P2), ii) reveal signals of converging adaptation between independent strains (comparing P2 derived from Ama154 with P22 derived from Ama143), and iii) monitor the fate of aneuploidies during sand fly infection (comparing P22 input and P_P22-SF output strains). Ama, P2 and P22 were collected at exponential growth phase for DNA extraction and for sand fly feeding (see below).

Sand Fly Feeding and Parasites Recovery. Colonies of *Phlebotomus orientalis* were maintained under standard conditions as previously described (22). Three groups with 100 3- to 7-d-old females were fed with 10^6 parasites per ml suspensions of splenic amastigotes, promastigotes at passage 2, or at passage 22 (Fig. 1*A*). After 8 d postinfection, 10 female sand flies were dissected, and parasites

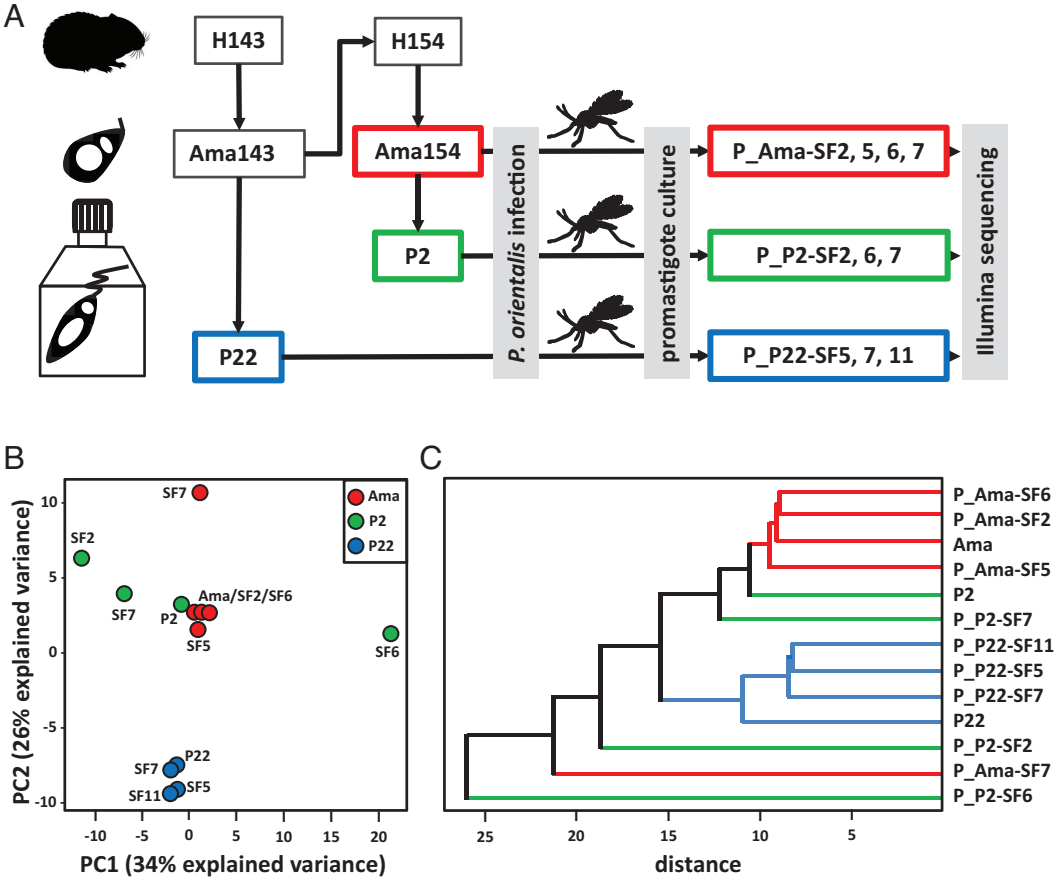

Fig. 1. Experimental layout and analysis of strain similarity. (*A*) Overview of *L. donovani* Ld1S isolates before and after experimental sand fly (SF) infection. Splenic amastigotes (Ama) were obtained from infected hamsters (H) 143 and 154, adapted to promastigote culture for two (P2) and twenty-two (P22) passages, and the indicated parasites were fed to 100 *P. orientalis* sand flies. Parasites were isolated 8 d postinfection and successful short-term cultures were established for 3 to 4 sand fly-derived promastigotes (SFn) (from a total of 10 dissected flies) per infection group. (*B*) Principal component analysis based on alternative ("alt") allele frequencies in the samples. The samples are represented by their projection along the two first components. (*C*) Dendrogram based on the Euclidean distances in the SNP "alt" allele frequencies space using the UPGMA (Unweighted Pair Group Method with Arithmetic mean) method.

recovered from each sand fly. Promastigotes from individual sand fly were then expanded in culture (max. two passages) and collected at exponential growth phase for DNA extraction. At least three promastigote cultures of individually infected sand flies were derived from each of the three experimental groups (Ama, P2, P22).

Genome Sequencing and Comparative Genomic Analyses. DNA was prepared and sequenced as detailed in *SI Appendix*. Read alignment, the quantification of chromosomes, genomic bins and genes, and part of the SNP analyses were performed with the GIP pipeline and the giptools toolkit (version 1.0.9) (20). All bioinformatics analyses were performed considering the *L. donovani* assembly GCA_002243465.1 of the 1S2D strain (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA396645, BioSample SAMN07430226) and the annotations we previously published (10). The genome sequencing datasets described in this study were deposited in the Sequence Read Archive (SRA) (23) and are available via the accession numbers reported in *SI Appendix*, Table S1. The mapping statistics are provided in *SI Appendix*, Table S2 and the GIP parameters are available in Dataset S1. A detailed description of the GIP tools and parameters used in this study is available in *SI Appendix*. Further SNP analyses were performed based on the filtered outputs of GIP using custom Python 3.10 code relying on the following libraries: Pandas (1.4.2) (24), Pysam (0.19.0) (25), Numpy (1.22.3) (26), Matplotlib (3.5.1) (27), Seaborn (0.11.2) (28), Biotite (0.32.0) (29), and Upsetplot (0.6.0) (30).

In the outputs of GIP, SNPs are characterized by a "ref" (reference) and an "alt" (alternative) allele, where the "ref" allele is the one present in the reference genome provided to GIP, and the "alt" allele is the one present in a sample. An "alt" allele frequency of 0 for a given sample indicates the allele's absence. "alt" alleles with a frequency below 0.1 were filtered out. See *SI Appendix, Computational Methods* for more details on SNP calling and filtering. The principal component analysis (PCA) (Fig. 1*B*) was based on "alt" allele frequency table of all samples. The dendrogram (Fig. 1*C*) was built using the UPGMA (Unweighted Pair Group Method with Arithmetic mean) method applied to pairwise Euclidean distances between "alt" allele frequencies of samples. "Upset plots" (Fig. 2*A*) were built from the sets of alleles with nonzero "alt" allele frequencies. Pairwise frequency scatterplots (Fig. 2*B*) represent "alt" allele frequencies of SNPs in sample pairs of. The red framed half of the plots are limited to those SNPs whose "alt" allele frequencies were strictly positive in all output samples and zero in the corresponding input. Frequency delta plots (Fig. 4*A* and *SI Appendix*, Figs. S6–S8) were built using a modified version of the giptools "SNP" module, computing and showing the "alt" allele frequency difference between input and output strains (frequency "delta"). The dots along the chromosomes were joined by lines to help visualize possible regions of

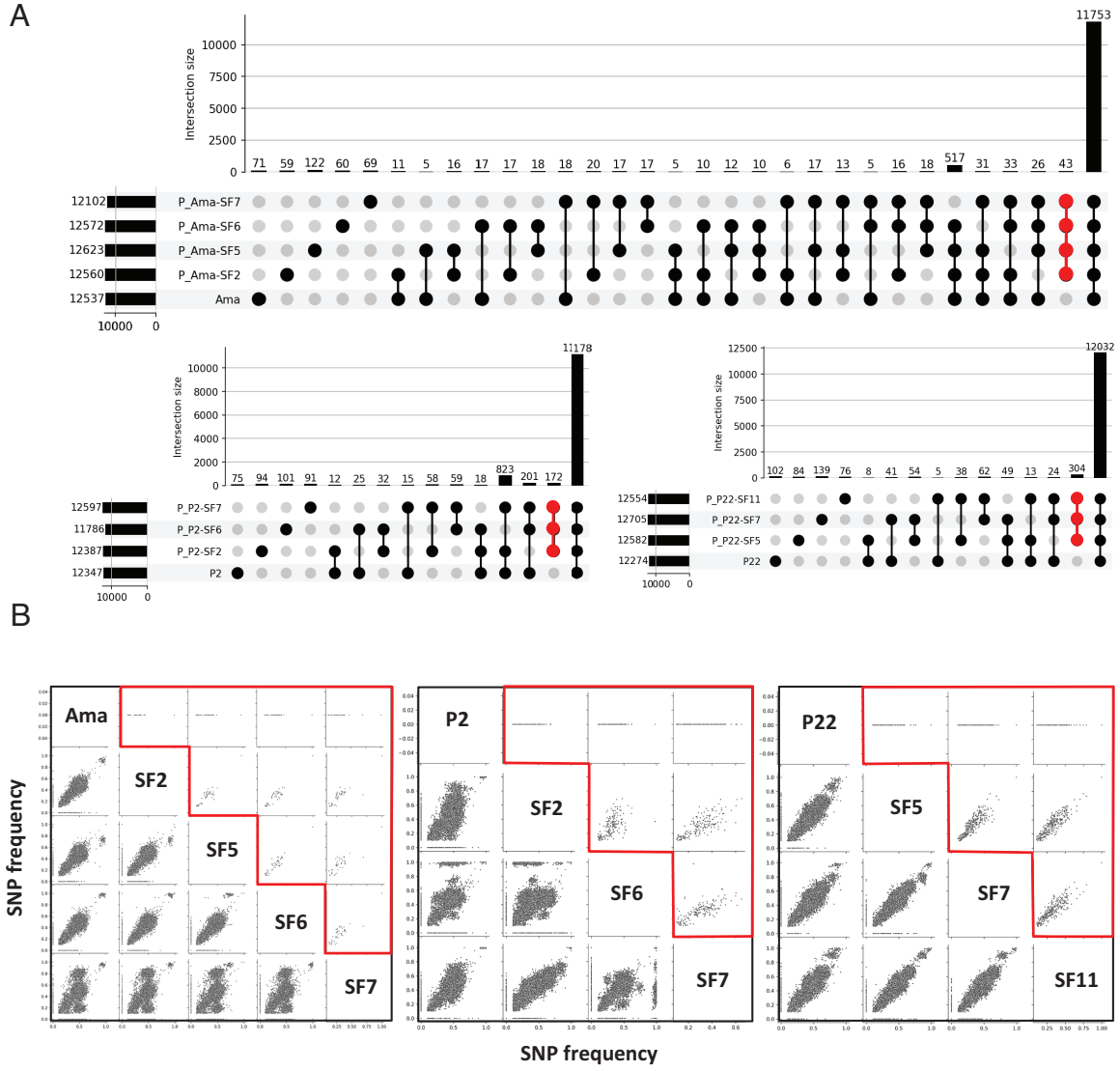

Fig. 2. Genome-wide SNP analyses. (*A*) *Upset* plots of SNPs found in the different input parasites (Ama, P2, P22) and their corresponding output isolates. The bars visualize the number of SNPs common to a given combination of samples. In red are highlighted the SNPs that are common to only the output isolates. (*B*) Comparison of SNP frequencies between input parasites and their corresponding output isolates. Highlighted in red is the comparative analysis of SNPs that are common to the output isolates but new compared to their respective input parasites.

similar frequency changes. The genomic positions of SNPs present in at least one of the samples are represented by a dot on the X-axis.

Detection of loci with strong allele frequency changes between input and output samples relied on the following procedure. First, for a given "output" sample, the frequency "delta" was computed as above. A positive delta value indicates an increase in the frequency of the "alt" allele in the output with respect to the input (Datasets S5–S7). The absolute value of delta was computed (abs_delta), and its minimum across output samples of an experiment (min_abs_delta) was used to characterize a given locus. The 95th percentile of the distribution of min_abs_delta was used as a threshold above which a locus was considered as displaying strong allele frequency shifts (possibly convergent or divergent) (*SI Appendix*, Fig. S9). The same genomic regions were also considered in the "rainbow plots" (Fig. 4*B*), and the same threshold was used to select the loci counted in the right heatmap of Fig. 5*E*.

Estimates of the fixation index F_{ST} for individual SNPs were computed based on "alt" allele frequencies for a given set of samples considering each individual sample as a subpopulation (S), and the ensemble of all samples as the total population (T). The following procedure was used: Let p_i and q_i be the frequencies of the "ref" and "alt" alleles in sample i. The frequencies p (and q) for the total population (T) correspond to the average of p_i (and q_i) across subpopulations (S). The heterozygosity for a subpopulation is computed as $2p_iq_i$ (denoted as H_i), and that for the total population as 2pq (denoted as H_T). H_s was defined as the heterozygosity score averaged across all subpopulations (i.e. mean H_i) and the F_{ST} was computed by $(H_T - H_S) / H_T$. This procedure was applied for input–output pairs to estimate the degree of differentiation between inputs and outputs, and for the set of outputs to estimate the degree of differentiation between outputs. In Fig. 5*A*, a given SNP is represented by as many dots as there are output samples. These dots share the same X-axis coordinate, the F_{ST} estimated between outputs. The Y-axis coordinate of a given dot is the F_{ST} between the input and the corresponding output sample.

Results

Experimental Sand Fly Infection Affects the Genetic Structure of the Parasite Population. *L. donovani* Ld1S splenic amastigotes (Ama) obtained from infected hamsters and promastigotes derived from independent Ama strains and maintained for two passages (P2, early culture adaptation) and 22 passages (P22, late culture adaptation) were fed to *P. orientalis* sand flies (Fig. 1*A*). The use of different promastigote passages derived from independent Ama strains allowed us to assess converging signals of genomic adaptation and to monitor the fate of in vitro established aneuploidies during sand fly infection. Eight days after infection, promastigote cultures were established from 3 to 4 individual sand flies for each experimental group. Cultures were expanded for a maximum of two passages, which we previously showed to be too short to have a significant impact on genomic adaptation (14, 31). Following whole-genome sequencing analysis, we first assessed the genomic distance between the parental input parasites (Ama, P2, P22) and corresponding promastigote (P) sand fly (SF) output isolates (P_Ama-SF2, 5, 6, 7; P_P2-SF2, 6, 7; P_P22-SF5, 7, 11) using allele frequencies at SNPs. Surprisingly, when considering the first principal components determined by a principal component analysis (PCA) based on these frequencies, only P22 input clustered with the corresponding outputs. This suggests that sand fly infection significantly changes the genetic structure of the parasite population (Fig. 1*B*). This result was confirmed by a clustering analysis based on pairwise Euclidean distances in the SNP allele frequency space. In the resulting dendrogram, only the P22 samples form a coherent cluster (Fig. 1*C*). Our results are best explained by a bottleneck effect caused by sand fly infection (32–34), which reduces the genetic complexity of the P_Ama-SF and P_P2-SF output parasites compared with the genetically more heterogenous input strains, thereby affecting population structure and allele frequencies. Such bottleneck effects may be less pronounced in the P_P22-SF isolates, given that long-term culture of the P22 input parasites already reduced genetic heterogeneity by selecting for common aneuploidies and haplotypes that are beneficial for in vitro growth (14). We next investigated parasite genomic signals that arise during sand fly infection. We tried to distinguish between their random nature due to the bottleneck effect or their nonrandom nature due to natural selection by assessing their reproducibility in independent biological replicates.

Karyotypic Changes during Sand Fly Infection. We previously showed that copy number variations (CNVs) of both chromosomes and genes play an important role during promastigote fitness gain in culture (10, 14, 31). Here, we assess the link of these forms of genome instability to promastigote differentiation, proliferation, and genetic adaptation during sand fly infection. We applied read depth analysis on the genome sequences of the various input and corresponding output parasites (Fig. 1) using our genome instability pipeline (GIP) and the giptools analysis package (20). For all experimental groups, each output parasite population can be considered as an independent biological replicate as they are derived from independent sand flies. Only minor differences between input and output parasites were observed thus confirming previous observations that sand fly infection has little influence on the karyotypic profile (11) (*SI Appendix*, Fig. S1*A* and Datasets S1–S3). Small somy fluctuations observed in individual Ama- and P2-SF isolates can be attributed to the bottleneck effect described in Fig. 1. In contrast, we observed convergence of somy changes in the three independent P22-SF isolates for most of the chromosomes that underwent amplification during long-term culture adaptation (10, 14, 35) (*SI Appendix*, Fig. S1*A*, *Lower*).

Similar to our karyotypic analysis, only minor changes were observed for gene copy number variations (CNVs) between the Ama, P2 and P22 input parasites and their respective sand fly isolates (*SI Appendix*, Figs. S1*B* and S2 and Dataset S4). For example, we observed gene amplification in individual SF isolates for the HSP83 gene cluster in P_Ama-SF5 (1.5-fold), for the snoRNA LD36Cs-1C2_c1084 in P_P2-SF6 (1.7-fold). As they were only observed in individual samples (*SI Appendix*, Fig. S1*B*, see also Dataset S4), these CNVs do not seem to be the result of selection but are likely to result from the bottleneck effect occurring during the sand fly infection.

Taken together, our analyses indicate that changes in chromosome and gene copy number are not major drivers of promastigote adaptation at least during experimental sand fly infection, which most likely is explained by the short time of infection of only 8 d. Even though no karyotypic selection was observed for the Ama and P2 experimental groups under our experimental conditions, karyotypic adaptation may occur during prolonged promastigote proliferation following repeated blood feeding (36–38).

Sand Fly Infection Causes Important Changes in the SNP Profile. We next investigated the nature and frequency of single-nucleotide polymorphisms (SNPs) in input and output populations. SNPs are substitutions of individual nucleotides at specific positions in the genome, which can give rise to gene variants (alleles) affecting gene function. SNPs are key drivers of evolutionary adaptation that can shape the phenotypic landscape through purifying and positive selection of respectively detrimental or beneficial alleles. Changes in SNP frequency in a population can be caused by bottleneck events, allelic selection, aneuploidy, or recombination events. To distinguish between these possibilities, we investigated the SNP distribution before and after sand fly infection and assessed the

reproducibility of frequency changes across the independent SF isolates for a given experimental group.

We detected over 12,000 heterozygous SNPs across all input parasite populations (Ama, P2, P22) and their corresponding SF isolates compared with the reference genome (*Methods*) (Fig. 2*A* and *SI Appendix*, Fig. S3 and Datasets S5–S7). While the majority of SNPs were shared between the various parental input and SF output parasites (>85%), a small number of SNPs were detected only in the three input parasite populations, which may have been lost in the output strains due to the bottleneck event described above (Fig. 1). We also observed specific SNPs in the output strains, a significant number of which was shared between the biological replicates of a given experimental group (Fig. 2*A*, red dots). The emergence of these variants and their convergence in the output strains suggest that these SNPs may either represent mutations that are selected during sand fly infection, or again are the result of the bottleneck effect that can randomly enrich for low-frequency alleles, making them

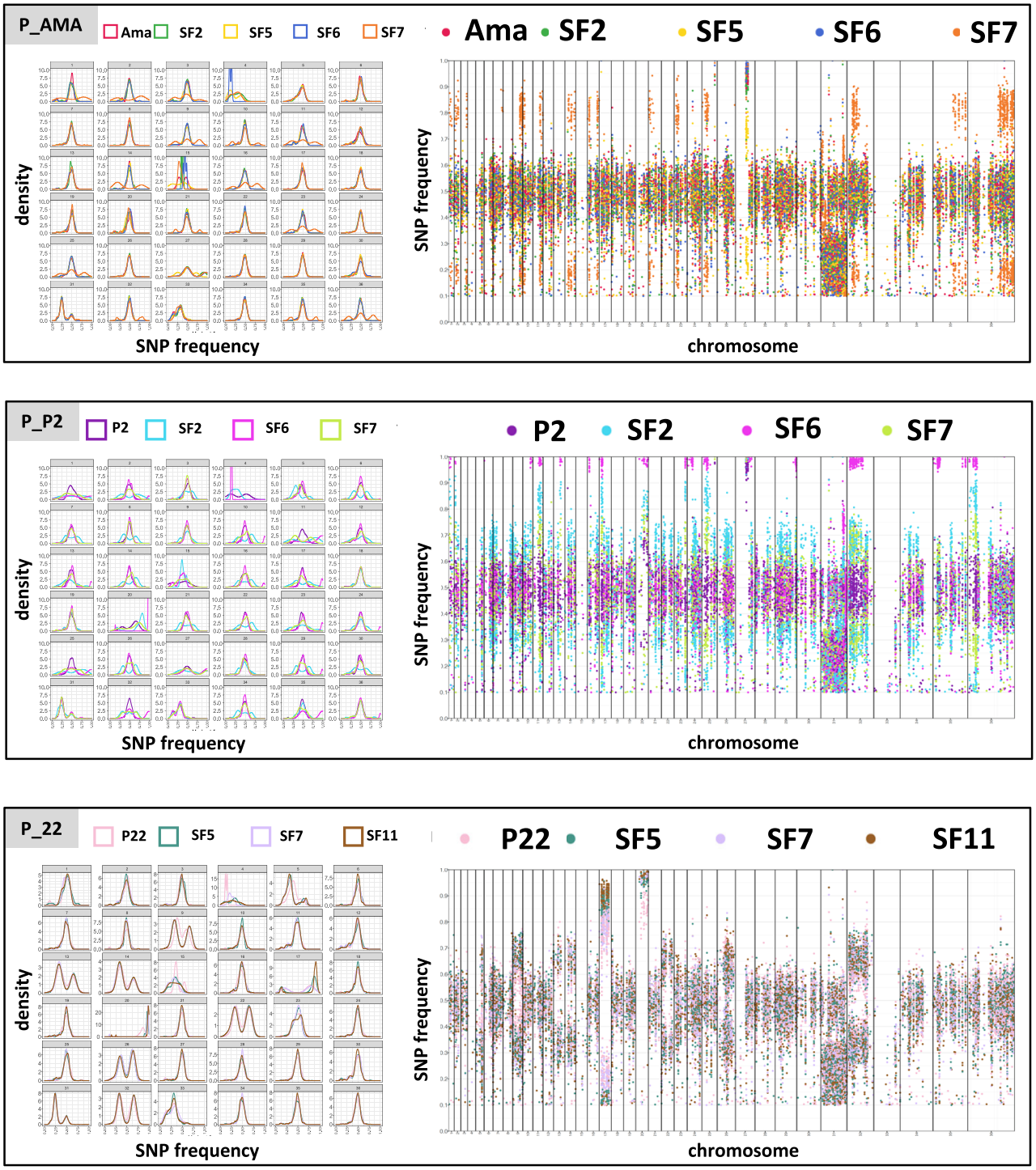

Fig. 3. SNP frequency distribution. The plots on the *Left* represent on the X-axis the variant allele frequency and on the Y-axis the corresponding estimated probability density. Different chromosomes are reported in different panels. The scatterplots on the *Right* display individual SNPs as dots. The X-axis indicates the genomic position while the Y-axis indicates the variant allele frequency. The different chromosomes are displayed one after the other, and their boundaries are visualized as vertical lines. The lines (on the *Left*) and the individual SNPs (on the *Right*) are colored according to the samples as indicated in the graph.

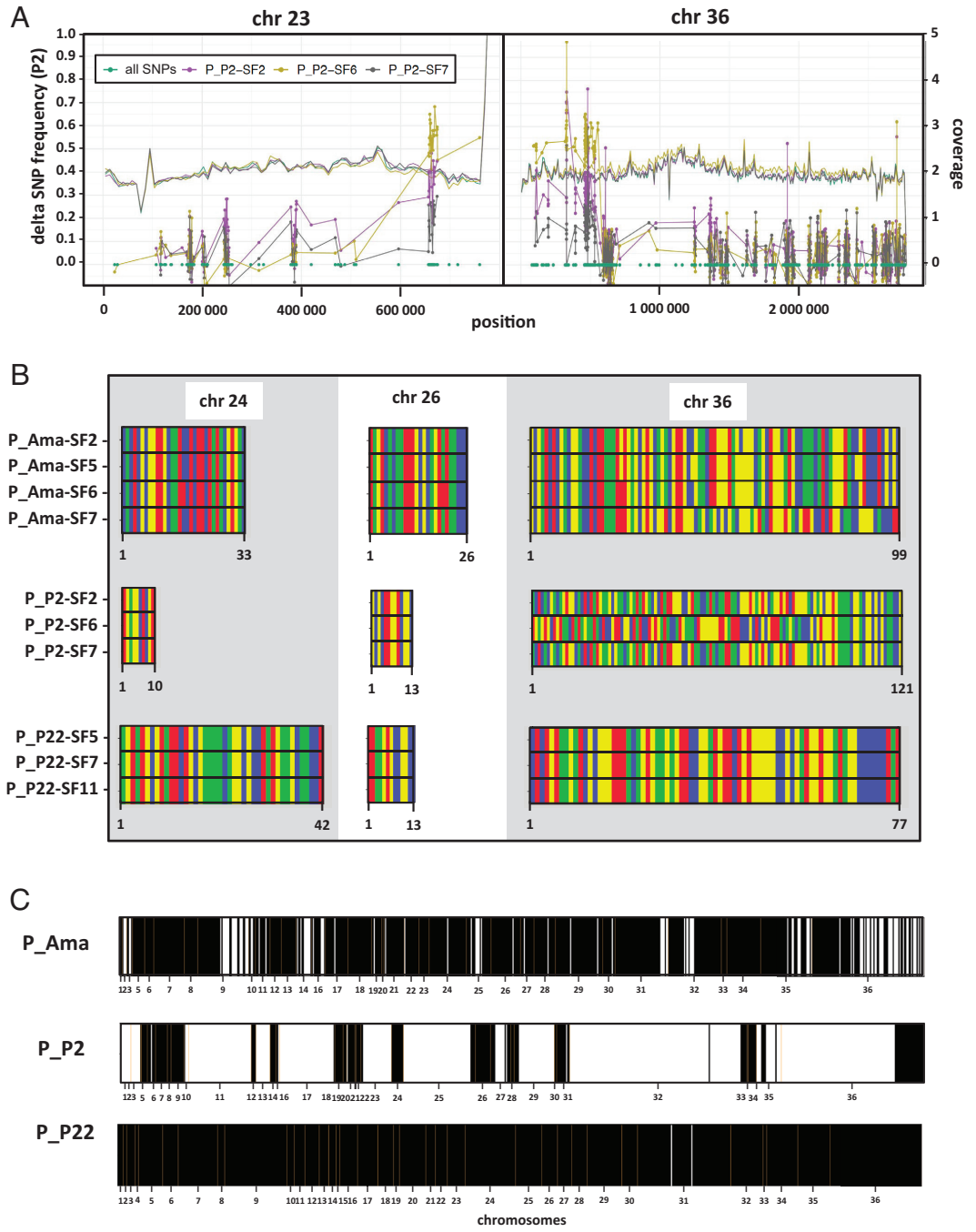

Fig. 4. Convergent SNP frequency shifts and haplotype shuffling. (*A*) Differential allele frequency analysis. SNP frequency changes ("delta") with respect to the input are represented for the output samples (Y-axis), and the positions of SNPs are represented as dots along the chromosome coordinates (X-axis). For a given output sample, dots for successive SNPs are connected by lines, in order to better visualize possible "hotspots" for SNP frequency increases. The genomic sequencing coverage for a given output sample is indicated as a continuous line and is reported on the *Right* Y-axis. Only chromosomes 23 and 36 for samples corresponding to the P2 input are represented (see *SI Appendix*, Figs. S6–S8 for all panels). (*B*) Rainbow plots. SNPs of the output samples that showed a strong frequency increase in either the reference allele or the "alt" allele compared with their corresponding input samples are represented. A frequency increase is considered "strong" if it belongs to the top 5% percent (*SI Appendix*, Fig. S9) in terms of the "minimum absolute delta" (*Methods*). To visualize haplotype selection, the bars are colored according to the allele that shows the frequency increase (A: blue, C: green, G: yellow, T: red). Only chromosomes 24, 26, and 36 are represented. The number of SNPs per chromosome fulfilling our criteria is indicated below the X-axis. (*C*) Genome-wide assessment of haplotype selection. The rainbow plots for a given experimental group were scored for convergent (black bar) or divergent (white bar) alleles. Chromosome boundaries (orange line) and numbers are indicated. The number of SNPs represented for a given chromosome may vary between experiments (input–output sets). Some chromosomes may even have no SNPs passing our selection criteria. Therefore, the rainbow plots for different experiments cannot be directly compared.

detectable in the output strains while they were below the detection threshold in the input strains.

Plotting SNP frequencies of the input parasites against their respective SF isolates resulted in a largely diagonal pattern suggesting similar SNP frequencies (Fig. 2*B*). Nevertheless, we observed significant changes in SNP frequencies for a number of output isolates, including P_Ama-SF7, P_P2-SF2 and P_P2-SF6 (and to a lesser extend also P_P2-SF7), all of which showed frequency shifts to both higher and lower values compared with their corresponding input parasites. These shifts are neither explained by chromosome nor gene amplifications as judged by the largely constant read depth observed in the Ama and P2 experimental groups (*SI Appendix*, Fig. S1). Unlike the Ama and P2 infection groups, no significant differences were observed for the

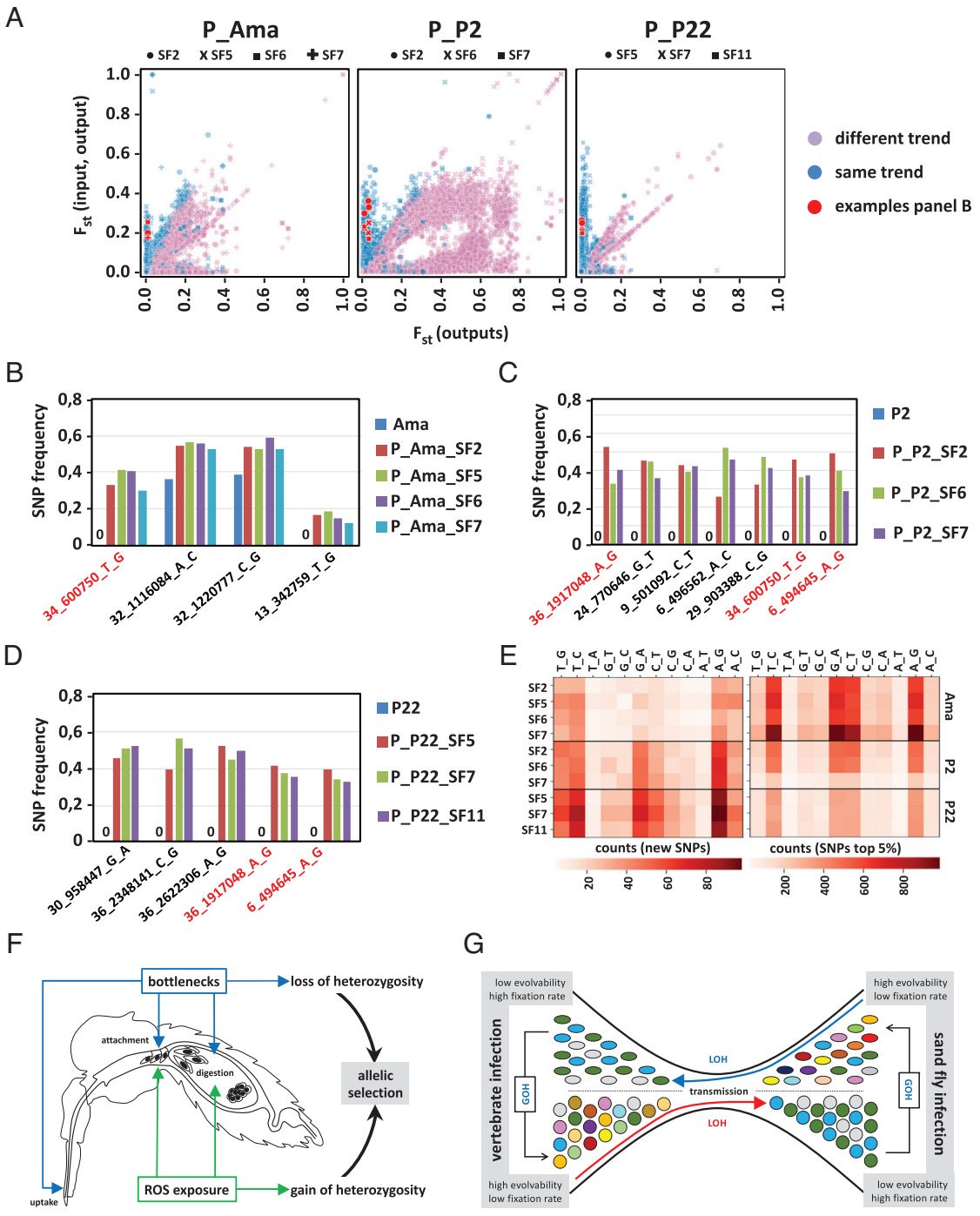

Fig. 5. Allelic selection and oxidative DNA damage during sand fly infection. (*A*) Analysis of the fixation index (F_{ST}). Each dot represents a SNP identified in the indicated output samples. The Y-axis coordinates show the difference in frequency between input and output, with increasing differences indicated by higher coordinates. The X-axis coordinates show the difference in SNP frequency between the individual output strains of a given experimental group, with similar SNP frequency shifts across the individual output stains indicated by lower coordinates. Alleles under selection are thus identified by both a low X-axis coordinate and high Y-axis coordinates. "Same trend" refers to those SNPs showing allele frequency changes in the same direction for all outputs with respect to the input (blue dots). The red dots correspond to the SNPs further analyzed in more detail in panels *B*–*D*. (*B*–*D*) Examples of allelic selection. Plots showing the frequencies of the indicated, nonsynonymous SNPs in the Ama (*B*), P2 (*C*), and P22 (*D*) experimental groups. SNPs that are shared between experimental groups are indicated in red. (*E*) Mutation distribution analysis. Heatmaps showing the distribution of the type of mutation observed in the indicated output isolates compared with the corresponding input parasites. The color intensity indicates the count for the indicated the type of mutation according to the shown legend. The *Left* heatmap represents the number of SNPs that are new in the output sample with respect to its corresponding input. The right heatmap represents the number of SNPs whose abs_delta indicator was above the 95th percentile in terms of min_abs_delta, that is, those that display high frequency changes in all outputs (but not necessarily all in the same direction with respect to the input, see *Methods* and *SI Appendix*, Fig. S9). (*F*) Model of *Leishmania* genomic adaptation during sand fly infection. Sand fly infection can affect the *Leishmania* genetic diversity in two ways: i) various bottlenecks (blue) that cause a loss of heterozygosity in the parasite population due to random sampling during uptake, blood meal digestion, or parasites attachment to the midgut epithelium following excretion of the digested blood meal; ii) ROS exposure (green) and oxidative DNA damage that cause a gain of heterozygosity through the generation of new SNPs. Both the bottleneck events and oxidative DNA damage contribute to allelic selection of potentially beneficial SNPs (*Discussion*). (*G*) Model of *Leishmania* evolutionary dynamics. Alternating cycles of i) loss of heterozygosity (LOH, symbolized by low color diversity) caused by parasite transmission (bottleneck represented by the space constriction) between the vertebrate and insect hosts (red and blue arrows indicating the direction of transmission) and ii) gain of heterozygosity (GOH, symbolized by high color diversity) can drive a high fixation rate of beneficial mutations, while maintaining high evolvability. Round shapes, amastigotes; oval shapes, promastigotes.

P22 input parasites and the derived P_P22-SF isolates as indicated by the diagonal SNP distribution. Comparison of only the SNPs that are new with respect to the parental input and shared by the SF isolates further sustains their convergent selection during sand fly infection. Those SNPs attain overall similar frequencies in the SF isolates as shown in Fig. 2*B* (red frames).

In conclusion, we observed important shifts in SNP frequency during sand fly infection that were not linked to chromosome or gene amplification. These shifts could be explained by transient changes in ploidy, genetic recombination events, or nucleotide mutation during sand fly infection – possibilities we further explored by a more in-depth analysis of the SNP signals.

Haplotype Selection Occurs in the Absence of Stable Aneuploidy during Sand Fly Infection. We next analyzed the distribution of SNP frequencies, chromosome by chromosome, for the various samples (Fig. 3) to test if sand fly infection is associated with parasite selection for specific haplotypes. As expected from the presence of two haplotypes in *L. donovani* Ld1S (14) and the largely disomic state of the Ama input parasites, most heterozygous SNPs in the Ama sample are present at a frequency of 50% (or 0.5), giving rise to a single peak (Fig. 3, *Upper Left*; see *SI Appendix*, Fig. S4 for the individual frequency distribution plots). The density profiles in P_Ama-SF7 shows important differences to the Ama input and the other Ama-SF outputs, including bimodal distribution for several chromosomes with peaks at 20% and 80% frequencies (i.e. chromosomes 1, 2, 14), indicating strong selection of one over the other haplotype. In the absence of any observed karyotypic change (*SI Appendix*, Fig. S1*A*), such frequency shifts are likely explained by transient trisomy and strong haplotype selection as we previously reported (14). The same is observed for many chromosomes of the P_P2-SF isolates that – unlike the P2 parental parasites – show a bimodal SNP frequency distribution at 33% and 66% diagnostic of haplotype selected trisomies (14), even though they remain disomic (Fig. 3, *Middle Left* and *SI Appendix*, Fig. S1*A*). A case of transient trisomy is indeed observe for chr 20 that is amplified in the P2 input parasites showing a 33/66% frequency distribution (*SI Appendix*, Fig. S1 and Fig. 3), while all P_P2SF output isolates return to a disomic state, but now showing bimodal SNP frequency for the two haplotypes, with an extreme distribution close to 10/90% observed for P_P2-SF6. In contrast to the Ama and P2 infection groups, the P_P22-SF isolates largely reproduce the profiles of their P22 parental parasites (Fig. 3, *Lower Left*). Of note, the partial trisomies observed for chr 5 and 9 (*SI Appendix*, Fig. S1*A*) were further selected to full trisomies in all P_P22-SF isolates as indicated by the 33/66% frequency distribution (Fig. 3, *Lower Left*; see *SI Appendix*, Fig. S4 for the individual SNP frequency distribution), suggesting that one of the haplotypes may provide an unknown fitness advantage to promastigotes proliferating in the sand fly midgut.

Subchromosomal Regions Show SNP Frequency Shifts. We next investigated in more detail the changes in SNP frequency, with the aim to identify genomic regions that may be under selection. Plotting variant frequency against chromosomal location (Fig. 3, *Right* and *SI Appendix*, Fig. S5) revealed a patchy SNP distribution in all chromosomes of the input parasites Ama, P2, P22, with short zones of heterozygosity separated by zones devoid of SNPs. Like the Ama input sample, most P_Ama-SF isolates show a frequency distribution around 50% (Fig. 3, *Upper Right*) as expected from their disomic state (*SI Appendix*, Fig. S1*A*) and the largely balanced distribution of the two haplotypes (Fig. 3, *Upper Left*). In contrast, the isolate P_Ama-SF7 shows important shifts of individual SNP patches toward a 20/80% frequency distribution for chr 1, 2, 3, 7, 9, 10, 11, 14, 16, 21, 23, 25, 32, 35, and 36, consistently with the bimodal and trimodal SNP frequency profiles observed for these chromosomes (*Left*). Likewise, two SNP patches on chr 31 show a frequency shift from 25 to 40% and 65%.

We observed many subchromosomal SNP frequency shifts for all P_P2-SF isolates, with P_P2-SF6 attaining frequencies of close to 100% for SNP patches on chromosomes 1, 2, 10, 11, 13, 16, 17, 20, 23, 24, 25, 29, 32, 35, and 36 (Fig. 3, *Middle Right*). For many of these SNP patches, we also observed a significant increase in frequency in the other two SF isolates, ranging from 70% in P_P2-SF7 to 85% in P_P2-SF2. This striking convergence between independent SF isolates both in SNP localization and the direction of frequency shift identifies regions with alleles that may be under selection during sand fly infection. In contrast to P_Ama- and P_P2-SF isolates, no major SNP frequency shifts were observed in the P_P22-SF isolates compared with the P22 input parasites, except for SF7 that shows minor frequency shifts for chr 17 and 20 (Fig. 3, *Lower Right*).

In conclusion, the observed SNP frequency shifts and their potential convergence in both location and frequency change reveals an additional form of *Leishmania* genome instability during sand fly infection that may be caused by haplotype shuffling and allelic selection.

Haplotype Shuffling during Sand Fly Infection. We next more closely investigated the regions that showed converging allele frequency shifts. Plotting the difference in frequency between parental input and the SF output isolates for a given allele position revealed that only a subset of SNPs undergo frequency shifts (*SI Appendix*, Figs. S6–S8 for the entire panel of all experimental groups). Some shifts were observed across all SF isolates of a given experimental group, as exemplified by the two chromosomes 23 and 36 of P2 input and output parasites (Fig. 4*A*).

SNP frequency shifts largely involved short DNA regions and were not caused by copy number variation as judged by the constant read coverage between input and output parasites plotted in the same graph (Fig. 4*A*).

We then investigated to what extent these frequency shifts affected the otherwise balanced haplotypes. As indicated above (Fig. 3) and published previously (14), the *L. donovani* Ld1S genome shows two distinct chromosomal haplotypes that are defined by heterozygous SNPs. The convergent SNP frequency shifts we observed thus can be produced by both alleles that represent a given heterozygous SNP. We therefore assessed if the observed frequency shifts occur at random, or if one allele is preferred over the other and thus may be under selection. Painting the dominant allele with distinct colors for the four possible nucleotide bases revealed a surprising degree of allelic selection, with many chromosomes showing a highly reproducible and converging SNP pattern either along stretches of a given chromosome (e.g. chr 36 of the P_Ama and P_P2 SF output parasites, Fig. 4*B*) or the entire chromosomes (e.g., chr 24, 26, or 36 in the P_P22-SF output parasites, Fig. 4*B*). Plotting the allelic choice (i.e., selection of identical or different alleles between output isolates) for each convergent SNP position at genome-wide level shows that allelic selection affects all chromosomes in the Ama and P22 experimental groups, but only a subset of chromosomes in the P2 experimental group (Fig. 4*C*).

In conclusion, our analyses reveal that allelic selection operates at the subchromosomal level affecting only certain genomic regions. This can occur via shuffling of existing haplotypes generating new haplotype combinations, which may facilitate the selection of beneficial alleles during *Leishmania* sand fly infection.

Allele Frequency Changes Indicate Oxidative DNA Damage. We analyzed convergent selection of individual alleles by calculating the fixation index (F_{ST}) that assesses population divergence. F_{ST} values were calculated for each experimental group of input and output parasites (*Methods*). On the shown scatterplots (Fig. 5*A*), a SNP is represented by one dot per output sample. The Y-axis coordinates of those dots correspond to the input–output F_{ST} values for each of these output samples. On the X-axis, the dots share the same within-outputs F_{ST} value. Alleles under convergent selection are identified by high F_{ST} values between input and output (revealing changes in allele frequency) and low F_{ST} values between outputs (revealing converging trends in all output strains, indicating selection). We indeed observed converging allele frequency shifts for the three experimental groups Ama, P2 and P22, suggesting that these SNPs are under natural selection (Fig. 5*A*). We then investigated the potential biological impact of allelic selection by analyzing converging SNP frequency shifts in coding sequence that caused nonsynonymous mutations (Fig. 5 *B*–*D* and Datasets S5–S7). For most of these examples, the parental input parasites showed a zero-frequency value, indicating that either this SNP does not exist in the input or has been filtered out by our computational pipeline due to quality issues linked to read mapping or further filtering criteria (in particular, SNPs with alternate allele frequency below 0.1 were filtered out, and so may appear as having a frequency of 0 in subsequent analyses). Strikingly, some of these SNPs not only converged between the SF output isolates of a given experimental group, but even between experimental groups. For example, the SNP 34_600750_T_G (defining chr_position_reference_variant) that causes an Asn221His amino acid (aa) change in a Rab-like GTPase activating protein on chromosome 34 (Ld1S_340628300) is observed in all P_Ama-SF and P_P2-SF isolates. Likewise, convergent SNPs between the P_P2 and P_P22 experimental groups have been observed for the ras-like protein Ld1S_360772200 on chr 36 (SNP 36_1917048_A_G, aa change Asn166Ser) and the hypothetical protein Ld1S_060842800 (SNP 6_494645_A_G, aa change Asn15Ser). How these mutations may affect protein structure, function, and interactions, and if they establish beneficial parasite phenotypes remains to be elucidated in future studies.

Converging frequency increases in the SF output isolates were also observed for synonymous SNPs that have no overt effect on gene function (Datasets S5–S7). Such synonymous SNPs may just be "hitchhiking" with nonsynonymous SNPs under selection. Alternatively, synonymous SNPs may alter the codon profile and thus may be selected for changes in translation efficiency, a possibility that is supported by our recent observation that *Leishmania* fitness gain correlates with tRNA gene amplification (10). Finally, convergent SNPs were even observed in noncoding regions (Datasets S5–S7). These mutations may again be "hitchhiking" or could be selected for by beneficial changes in DNA topology or by changes in gene expression if affecting regulatory sequences in 3′ and 5′ Untranslated regions (UTRs) that for example control RNA turn-over (39, 40).

We next assessed the possible mechanism that triggers the observed allele frequency shifts. The observed mutations may be the result of parasite exposure to DNA-damaging agents, such as exogenous reactive oxygen species (ROS), known to regulate antimicrobial immunity and microbiome homeostasis in the insect gut (41), or endogenous ROS that can be generated in response to stress (42). Oxidative DNA damage increases genome instability and introduces signature mutations due to mismatches caused by DNA repair intermediates (43). To investigate signatures of ROS-mediated mutation, we quantified genome-wide the type of nucleotide changes that caused allele frequency shifts in the SF output isolates compared with the corresponding input parasites. For all SF isolates, the induced nucleotide changes mainly involved A ⇔ G and T ⇔ C transitions, which represent the majority of observed SNP changes (Fig. 5*E* and Dataset S8). Such transitions are indeed diagnostic for oxidative DNA damage and can be caused by i) oxidation of guanine to 7,8-dihydro-8-oxoguanine (8-oxo-G), ii) oxidation of cytosine to 5-hydroxycytosine (5-OH-C) or cytosine to 5-hydroxyuracil (5-OH-U), and iii) oxidation of adenine to hypoxanthine, all of which generate repair intermediates with altered base pairing properties (44).

In conclusion, our data establish an association between sand fly infection and *Leishmania* DNA damage likely caused by exposure to midgut oxidants that can increase parasite genetic heterogeneity and allows for the selection of beneficial alleles (Fig. 5*F*).

Discussion

Here, we combined our *Leishmania* in vitro evolution system (10, 14, 35) with experimental sand fly infection to assess the impact of the insect midgut environment on parasite genomic adaptation. We identified vector infection as an important genetic bottleneck that causes loss of heterozygosity (LOH) and revealed haplotype shuffling and allelic selection that are likely driven by oxidative DNA damage (Fig. 5*F*). Our findings generate an important insight into mechanisms underlying *Leishmania* genetic diversity and evolvability and open a series of questions on how *Leishmania* genomic adaptation drives parasite fitness inside its insect vector and what factors may cause oxidative DNA damage.

During their passage through their insect host, vector-borne pathogens (notably viruses) show a strong reduction in population heterogeneity, due to the limited number of microbes that are ingested or establish productive infection (45, 46). In evolutionary biology, such reduction in population size due to environmental factors is referred to as a bottleneck, during which random sampling can have profound effects on the frequency of genotypes or alleles – a phenomenon known as genetic drift (47). In contrast to this stochastic process, changes in allele frequencies can also be nonrandom and driven by natural selection for traits that are beneficial in a given environment (2). Dissociating random from nonrandom allele frequency changes represents a key challenge in evolutionary studies. This holds true for our analysis on *Leishmania* genomic adaptation in the sand fly, which revealed a bottleneck event as judged by PCA, nonreproducible karyotypic changes and the loss of SNPs following sand fly passage, indicating that certain parasite subpopulations were eliminated from the infecting population. This bottleneck is likely caused by i) the small number of parasites ingested during a sand fly blood meal, ii) the potential elimination of parasites by toxic by-products during blood meal digestion, or iii) their failure to attach to the midgut that leads to parasites excretion via defecation of the digested blood meal (32, 37, 48–51) (Fig. 5*F*).

Despite random genomic changes caused by these bottlenecks, the comparison of independent biological replicates allowed us to also reveal nonrandom and thus selected genomic signals at nucleotide level. We identified several SNP frequency shifts that converged not only between biological replicates within, but even in-between experimental groups, providing strongest support for natural selection. Even though the phenotype associated with these mutations eludes us, the functional annotation of the mutated genes allows to speculate on the mechanism of fitness gain. For example, two of the converging alleles affected distinct members of the Ras GTPase family and thus a signaling pathway known to regulate eukaryotic cell growth, division, and differentiation (52). Indeed, Ras GTPases have been linked to cell proliferation and differentiation in *Trypanosoma cruzi* (53), microtubule biogenesis

and cytokinesis in *Trypanosoma brucei* (54) and their survival during tse tse fly infection (55). Significantly, in *L. donovani*, a non-synonymous SNP in the Ras-like small GTPase-RagC (LdBPK_366140.1) affects tissue tropism (15), indicating that a single SNP may produce important phenotypic change during infection. Hence, the convergent amino acid changes we have observed during sand fly infection in *L. donovani* Ras GTPases may drive fitness gain by increasing the parasite's reproductive capacity inside the vector gut, for example by enhancing proliferation or promoting parasite attachment, possibilities that need to be investigated in future studies for example using CRISPR/Cas9 gene-edited parasites.

The sand fly environment seems surprisingly conducive for *Leishmania* allele frequency changes and allelic selection considering the short, 8-d sand fly infection period. This could be a consequence of the bottleneck and resulting founder effects as the reduction of population heterogeneity allows for fast penetration of rare variants, even if they have only a minor beneficial effect (56). On the other hand, such LOH has been linked to reduced evolvability (57), which could prove fatal once these parasites encounter subsequent bottlenecks during transmission to the vertebrate host (Fig. 5*G*). The observed SNP dynamics thus calls for an active mechanism that can replenish *Leishmania* genetic heterogeneity and fuel natural selection. We propose oxidative DNA damage as such a key mechanism as judged by the highly homogeneous, diagnostic mutation signature we observed across all analyzed output genomes, i.e. transitions between A ⇔ G and T ⇔ C (43, 44).

Oxidative DNA damage can be caused by reactive oxygen species (ROS), such as superoxide anion (O_2^-), hydroxyl radical (OH^-) and hydrogen peroxide (H_2O_2) (43), which are a constant threat to DNA integrity and have been identified as key drivers of genome instability in various systems (43, 57). ROS can be generated from various external and internal sources. In the context of *Leishmania* sand fly infection, ROS may be produced intrinsically inside the parasite as an enzymatic by-product of the amastigote-to-promastigote differentiation process that is linked to important structural changes and retooling of parasite metabolism (58). Likewise, the transition from an anaerobic, intracellular lifestyle (amastigotes) to an aerobic extracellular lifestyle (promastigotes) is associated with a shift in energy production that may result in ROS production (59).

On the other hand, ROS may be produced by external sources inside the sand fly gut. Indeed, ROS were shown in *Drosophila* to participate in the immune protection against pathogenic bacteria (60–62), and are produced in *Anopheles* (*A.*) *gambiae* and in *A. aquasalis* following infection with the parasite *Plasmodium falciparum* (63) or *Plasmodium vivax* (64), respectively. Even though *Leishmania* infection itself seems not to elicit detectable ROS inside the infected sand fly gut (65), a possible role of basal-level ROS in affecting *Leishmania* genome integrity is supported by indirect evidence. First, the sand fly genome encodes for a member of the NAPDH oxidase family responsible for the generation of ROS (termed dual oxidase, DuOx) and several ROS detoxification enzymes, including superoxide dismutase and catalase (https://vectorbase.org/vectorbase/app/record/dataset/TMPTX_llonJacobina). Second, sand flies can produce immune protective ROS as shown in *Lutzomyia longipalpis* orally infected with the pathogenic bacteria *Serratia marcescens* (65). Finally, the same study showed that silencing of the antioxidant enzyme catalase in *L. longipalpis* flies and the oral administration of ROS to the infected flies reduces the number of *L. mexicana* parasites in the gut. The mutation signature observed in our experimental system lends further support to the exposure of *Leishmania* to ROS inside the sand fly gut. Whether such external ROS is directly produced by the insect gut cells, or indirectly by the endosymbiont bacteria composing the microbiota or toxic by-products of bloodmeal digestion and hemoglobin catabolism (50, 66, 67) remains to be established.

Independent of the source of toxic oxidants, it seems that *Leishmania* may benefit from oxidative DNA damage to counteract the LOH caused by parasite transmission and to gain heterozygosity that can promote parasite evolvability (Fig. 5*G*). Given the exposure of intracellular amastigotes to oxidative stress inside host macrophages, a similar scenario of ROS-driven gain-of-heterozygosity (GOH) may also apply to mammalian infection. However, such a strategy bears the risk to accumulate irreversible and damaging mutations, which is especially the case in asexual populations. This phenomenon – known as "Muller's Ratchet" – can only be alleviated by sex or recombination to allow efficient restoration of the wild-type sequence (68, 69). As a matter of fact, *Leishmania* shows a cryptic sexual cycle capable of producing hybrid genotypes during sand fly infection, which can be induced in vitro by culture exposure to ROS (70). *Leishmania* ROS-induced hybridization thus counteracts "Muller's Ratchet" and enables ROS-dependent GOH.

In conclusion, our results suggest that *Leishmania* may benefit from the DNA damaging environment in the sand fly midgut to increase its genetic heterogeneity and thus to further expand the already vast adaptive landscape of these parasites defined by other forms of genome instability, including chromosome and gene copy number variations (10, 14, 35). Based on these observations, we propose a model of *Leishmania* evolutionary adaptation where alternating cycles of LOH and GOH in both the sand fly and mammalian hosts, respectively, drive a high fixation rate of beneficial mutations, while maintaining high evolvability (Fig. 5*G*). Conceivably, the bottleneck events during host-vector transmission can have detrimental effects at the level of an individual parasite population. However, this LOH–GOH cycle may be overall beneficial on the level of the "metaparasite population" represented by the thousands of infected mammals and sand flies in a given endemic area. Our study sets the stage for future investigations that aim to assess the quantity and quality of DNA damage during sand fly infection using single-cell sequencing technologies, and to analyze the impact of ROS-generating and -detoxifying pathways on *Leishmania* genomic adaptation during sand fly infection by applying CRISPR/Cas9-mediated gene editing on both the vector and the parasite (70, 71).

Data, Materials, and Software Availability. Genome sequences data have been deposited in Sequence Read Archive (SRA) [PRJNA605972 (72) and PRJNA748346 (73)].

ACKNOWLEDGMENTS. We thank Pascal Campagne, Lucy Glover, and Shaden Kamhawi for helpful discussions. This study was supported by a seeding grant from the Institut Pasteur International Department to the LeiSHield Consortium and the EU H2020 project LeiSHield-MATI–REP-778298-1. Prague group is supported by European Regional Development Fund (project CZ.02.1.01/0.0/0.0/16_019/0000759). I.L. is financed by a Roux-Cantarini fellowship (S-FB14002-74A). We thank the Centro Nacional de Análisis Genómico (CNAG) for whole genome library preparations and sequencing.

Author affiliations: [a]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France; [b]Institut Pasteur, Université Paris Cité, INSERM U1201, Unité de Parasitologie moléculaire et Signalisation, F-75015 Paris, France; and [c]Department of Parasitology, Faculty of Science, Charles University, Prague 2, 128 44, Czech Republic

1. WHO, Leishmaniasis in high-burden countries: An epidemiological update based on data reported in 2014. *Wkly. Epidemiol. Rec.* **91**, 287–296 (2016).
2. C. Darwin, L. Kebler, *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life* (J. Murray, London, 1859).
3. J. K. Graham, M. L. Smith, A. M. Simons, Experimental evolution of bet hedging under manipulated environmental uncertainty in Neurospora crassa. *Proc. Biol. Sci.* **281**, 20140706 (2014).
4. M. K. Stewart, B. T. Cookson, Non-genetic diversity shapes infectious capacity and host resistance. *Trends Microbiol.* **20**, 461–466 (2012).
5. T. Downing *et al.*, Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **21**, 2143–2156 (2011).
6. 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).
7. Y. Sterkers *et al.*, Novel insights into genome plasticity in Eukaryotes: Mosaic aneuploidy in Leishmania. *Mol. Microbiol.* **86**, 15–23 (2012).
8. J. M. Ubeda *et al.*, Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite Leishmania. *PLoS Biol.* **12**, e1001868 (2014).
9. M. C. Brotherton *et al.*, Proteomic and genomic analyses of antimony resistant Leishmania infantum mutant. *PLoS One* **8**, e81899 (2013).
10. G. Bussotti *et al.*, Genome instability drives epistatic adaptation in the human pathogen Leishmania. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2113744118 (2021).
11. 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).
12. S. A. Iantorno *et al.*, Gene expression in leishmania is regulated predominantly by gene dosage. *MBio* **8**, e01393–17 (2017).
13. 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).
14. P. Prieto Barja *et al.*, Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. *Nat. Ecol. Evol.* **1**, 1961–1969 (2017).
15. W. W. Zhang *et al.*, Genetic analysis of Leishmania donovani tropism using a naturally attenuated cutaneous strain. *PLoS Pathog.* **10**, e1004244 (2014).
16. N. S. Akopyants *et al.*, Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. *Science* **324**, 265–268 (2009).
17. E. Inbar *et al.*, The mating competence of geographically diverse Leishmania major strains in their natural and unnatural sand fly vectors. *PLoS Genet.* **9**, e1003672 (2013).
18. E. Inbar *et al.*, Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in Leishmania. *PLoS Genet.* **15**, e1008042. (2019).
19. A. Romano *et al.*, Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16808–16813 (2014).
20. G. F. Spath, G. Bussotti, GIP: An open-source computational pipeline for mapping genomic instability from protists to cancer cells. *Nucleic Acids Res.* **50**, e36 (2022).
21. P. Pescher *et al.*, Quantitative proteome profiling informs on phenotypic traits that adapt Leishmania donovani for axenic and intracellular proliferation. *Cell Microbiol.* **13**, 978–991 (2011).
22. P. Volf, V. Volfova, Establishment and maintenance of sand fly colonies. *J. Vector Ecol.* **36**, S1–S9 (2011).
23. R. Leinonen *et al.*, The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
24. W. McKinney, "Data structures for statistical computing in Python" in *Proc. of the 9th Python in Science Conf* (Austin Texas: SciPy, 2008), (2010).
25. J. K. Bonfield *et al.*, HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **10**, giab007 (2021).
26. C. R. Harris *et al.*, Array programming with NumPy. *Nature* **585**, 357–362 (2020).
27. J. D. Hunter, Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
28. M. L. Waskom, Seaborn: Statistical data visualization. *J. Open Source Software* **6**, 3021 (2021).
29. P. Kunzmann, K. Hamacher, Biotite: A unifying open source computational biology framework in Python. *BMC Bioinform.* **19**, 346 (2018).
30. A. Lex *et al.*, UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983–1992 (2014).
31. G. Bussotti *et al.*, Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *MBio* **9**, e01399–18 (2018).
32. A. Dostalova, P. Volf, Leishmania development in sand flies: Parasite-vector interactions overview. *Parasit Vectors* **5**, 276 (2012).
33. S. M. Gossage, M. E. Rogers, P. A. Bates, Two separate growth phases during the development of Leishmania in sand flies: Implications for understanding the life cycle. *Int. J. Parasitol.* **33**, 1027–1034 (2003).
34. E. Inbar *et al.*, The transcriptome of Leishmania major developmental stages in their natural sand fly vector. *MBio.* **8**, e00029–17 (2017).
35. L. Piel *et al.*, Experimental evolution links post-transcriptional regulation to Leishmania fitness gain. *PLoS Pathog.* **18**, e1010375 (2022).
36. J. Sadlova *et al.*, Virulent and attenuated lines of Leishmania major: DNA karyotypes and differences in metalloproteinase GP63. *Folia Parasitol.* **53**, 81–90 (2006).
37. T. D. Serafim *et al.*, Leishmaniasis: The act of transmission. *Trends Parasitol.* **37**, 976–987 (2021).
38. T. D. Serafim *et al.*, Sequential blood meals promote Leishmania replication and reverse metacyclogenesis augmenting vector infectivity. *Nat. Microbiol.* **3**, 548–555 (2018).
39. R. Aly *et al.*, A regulatory role for the 5′ and 3′ untranslated regions in differential expression of hsp83 in Leishmania. *Nucleic Acids Res.* **22**, 2922–2929 (1994).
40. C. E. Clayton, Gene expression in Kinetoplastids. *Curr. Opin. Microbiol.* **32**, 46–51 (2016).
41. E. L. Telleria *et al.*, Leishmania, microbiota and sand fly immunity. *Parasitology* **145**, 1336–1353 (2018).
42. T. Finkel, N. J. Holbrook, Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
43. A. R. Poetsch, The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Comput. Struct. Biotechnol. J.* **18**, 207–219 (2020).
44. M. S. Cooke *et al.*, Oxidative DNA damage: Mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–214 (2003).
45. N. L. Forrester *et al.*, Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLoS Pathog.* **8**, e1002897 (2012).
46. S. C. Weaver *et al.*, Population bottlenecks and founder effects: Implications for mosquito-borne arboviral emergence. *Nat. Rev. Microbiol.* **19**, 184–195 (2021).
47. J. B. Reece *et al.*, *Genetic Drift* (Campbell Biology, ed. 10, 2011), pp. 488–490.
48. J. S. P. Doehl *et al.*, Skin parasite landscape determines host infectiousness in visceral leishmaniasis. *Nat. Commun.* **8**, 57 (2017).
49. J. Myskova *et al.*, Characterization of a midgut mucin-like glycoconjugate of Lutzomyia longipalpis with a potential role in Leishmania attachment. *Parasit Vectors* **9**, 413 (2016).
50. K. Pruzinova *et al.*, Leishmania mortality in sand fly blood meal is not species-specific and does not result from direct effect of proteinases. *Parasit Vectors* **11**, 37 (2018).
51. V. Seblova *et al.*, Phlebotomus orientalis sand flies from two geographically distant Ethiopian localities: Biology, genetic analyses and susceptibility to Leishmania donovani. *PLoS Negl. Trop. Dis.* **7**, e2187 (2013).
52. D. Bar-Sagi, A. Hall, Ras and Rho GTPases: A family reunion. *Cell* **103**, 227–238 (2000).
53. G. R. dos Santos *et al.*, The GTPase TcRjl of the human pathogen Trypanosoma cruzi is involved in the cell growth and differentiation. *Biochem. Biophys. Res. Commun.* **419**, 38–42 (2012).
54. H. P. Price *et al.*, The small GTPase ARL2 is required for cytokinesis in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **173**, 123–131 (2010).
55. S. K. Natesan *et al.*, The trypanosome Rab-related proteins RabX1 and RabX2 play no role in intracellular trafficking but may be involved in fly infectivity. *PLoS One* **4**, e7217 (2009).
56. T. Wein, T. Dagan, The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria. *Genome Biol. Evol.* **11**, 3283–3290 (2019).
57. A. Sallmyr, J. Fan, F. V. Rassool, Genomic instability in myeloid malignancies: Increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair. *Cancer Lett.* **270**, 1–9 (2008).
58. D. Rosenzweig *et al.*, Retooling Leishmania metabolism: From sand fly gut to human macrophage. *FASEB J.* **22**, 590–602 (2008).
59. P. Barrera *et al.*, Natural sesquiterpene lactones induce oxidative stress in Leishmania mexicana. *Evid. Based Complement. Alternat. Med.* **2013**, 163404 (2013).
60. S. Bai *et al.*, Regulatory mechanisms of microbial homeostasis in insect gut. *Insect Sci.* **28**, 286–301 (2021).
61. E. M. Ha *et al.*, A direct role for dual oxidase in Drosophila gut immunity. *Science* **310**, 847–850 (2005).
62. E. M. Ha *et al.*, An antioxidant system required for host protection against gut infection in Drosophila. *Dev. Cell* **8**, 125–132 (2005).
63. A. Molina-Cruz *et al.*, Reactive oxygen species modulate Anopheles gambiae immunity against bacteria and Plasmodium. *J. Biol. Chem.* **283**, 3217–3223 (2008).
64. A. C. Bahia *et al.*, The role of reactive oxygen species in Anopheles aquasalis response to Plasmodium vivax infection. *PLoS One* **8**, e57014 (2013).
65. H. Diaz-Albiter *et al.*, Reactive oxygen species-mediated immunity against Leishmania mexicana and Serratia marcescens in the sand phlebotomine fly Lutzomyia longipalpis. *J. Biol. Chem.* **287**, 23995–24003 (2012).
66. A. Das, M. Kamran, N. Ali, HO-3867 induces ROS-dependent stress response and apoptotic cell death in Leishmania donovani. *Front. Cell Infect. Microbiol.* **11**, 774899 (2021).
67. A. Mandal *et al.*, Deprivation of L-Arginine induces oxidative stress mediated apoptosis in Leishmania donovani promastigotes: Contribution of the polyamine pathway. *PLoS Negl. Trop. Dis.* **10**, e0004373 (2016).
68. J. Maynard-Smith, *The Evolution of Sex* (Cambridge University Press, Cambridge [England], New York, 1978).
69. H. J. Muller, The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9 (1964).
70. I. Louradour *et al.*, CRISPR/Cas9 Mutagenesis in phlebotomus papatasi: The immune deficiency pathway impacts vector competence for Leishmania major. *MBio* **10**, e01941-19 (2019).
71. I. Louradour *et al.*, Stress conditions promote Leishmania hybridization in vitro marked by expression of the ancestral gamete fusogen HAP2 as revealed by single-cell RNA-seq. *Elife* **11**, e73488 (2022).
72. G. Bussotti, G. F. Späth, Transcriptomic compensation as sources for Leishmania genetic adaptation. NCBI SRA database, https://www.ncbi.nlm.nih.gov/sra/PRJNA605972
73. G. Bussotti, G. F. Späth, Evolutionary genomic adaptation of Leishmania donovani parasites in sandflies. NCBI SRA database, https://www.ncbi.nlm.nih.gov/sra/PRJNA748346