Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani

Pablo Prieto Barja, Pascale Pescher, Giovanni Bussotti, Franck Dumetz, Hideo Imamura, Darek Kedra, Malgorzata Domagalska, Victor Chaumeau, Heinz Himmelbauer, Michel Pagès, et al.

Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani

Pablo Prieto Barja[1,6,$], Pascale Pescher[2,$], Giovanni Bussotti[3], Franck Dumetz[4], Hideo Imamura[4], Darek Kedra[1], Malgorzata Domagalska[4], Victor Chaumeau[5], Heinz Himmelbauer[1,7], Michel Pages[5], Yvon Sterkers[5], Jean-Claude Dujardin[4,8], Cedric Notredame[1,6,*] & Gerald Frank Späth[2,*]

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain; [2]Institut Pasteur, INSERM U1201, Unité de Parasitologie moléculaire et Signalisation, Paris, France; [3]HUB de Bioinformatique et Biostatistiques, Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France; [4]Institute of Tropical Medicine, Molecular Parasitology Unit, Antwerpen, Belgium; [5]University Montpellier, Faculty of Medicine, Laboratory of Parasitology–Mycology, CNRS 5290, IRD 224, University Montpellier (UMR "MiVEGEC") and Centre Hospitalier Universitaire, Montpellier, France; [6]Universitat Pompeu Fabra (UPF), Barcelona, Spain; [7]Institute of Biotechnology Muthgasse 18 1190 Vienna; [8]Universiteit Antwerpen, Department of Biomedical Sciences, Belgium

Running title: *L. donovani* haplotype selection

[*] Corresponding authors:

G. F. Späth, Tel: 01.40.61.38.58; E-mail: gerald.spaeth@pasteur.fr

C. Notredame, Tel: +34 933160271; Email: cedric.notredame@crg.eu

[$]these authors contributed equally to this work

Keywords: *Leishmania donovani*, aneuploidy, genome instability, dosage compensation, fitness, haplotypes

Summary

The parasite *Leishmania donovani* causes a fatal disease termed visceral leishmaniasis. The process through which the parasite adapts to environmental change remains largely unknown. Here we show that aneuploidy is integral for parasite adaptation and that karyotypic fluctuations allow for selection of beneficial haplotypes, which impact transcriptomic output and correlate with phenotypic variations in proliferation and infectivity. To avoid loss of diversity following karyotype and haplotype selection, *L. donovani* utilizes two mechanisms: polyclonal selection of beneficial haplotypes to create coexisting subpopulations that preserve the original diversity, and generation of new diversity as aneuploidy-prone chromosomes tolerate higher mutation rates. Our results reveal high aneuploidy turnover and haplotype selection as a unique evolutionary adaptation mechanism that *L. donovani* uses to preserve genetic diversity under strong selection. This unexplored process may function in other human diseases, including fungal infection and cancer, and stimulate innovative treatment options.

Introduction

Rapid pathogen adaptation to novel environments is a major threat to human health. In parasites evolving under constant host selection, fitness gains often cause increased pathogenicity[1]. Aneuploidy recently emerged as a key driver for evolutionary adaptation of fungal and protist pathogens[2–4], in which chromosome copy number variations induce phenotypic changes through modulation of transcript and protein abundance[5]. In addition to its effect on gene count, aneuploidy impacts cellular phenotypes through the selection of beneficial alleles. While this phenomenon has received considerable attention in human genetics[6] and cancer genome analysis[7,8,9], the role of aneuploidy on allelic selection in genome evolution and adaptation in pathogenic eukaryotes remains poorly understood in most organisms, in spite of recent attention[10,11].

We addressed this question in the protozoan parasite *Leishmania donovani*, an important human pathogen that causes fatal visceral leishmaniasis[12]. During its life cycle *Leishmania* undergoes a major developmental transition from insect-stage promastigotes to mammalian-stage amastigotes, which adapts these parasites for extra- and intracellular survival, respectively[13]. In addition to environmentally-induced stage differentiation, *Leishmania* can adapt to a variety of unpredictable fluctuations inside its human host, notably pharmacological interventions. Such environment-genotype interactions likely select parasites for higher fitness and have important consequences on disease outcome as demonstrated by the emergence of drug resistant clinical isolates[14,15]. However, the underlying genetic mechanisms that drive short-term *Leishmania* evolution are still largely unknown. In the absence of classical transcriptional gene regulation, these early-branching eukaryotes often control protein abundance via gene and chromosome amplification[10,16]. This process is fueled by frequent asymmetric chromosome allotments, which cause

constitutive mosaicism[17,18]. This model has recently received new support[19] with a large-scale deep sequencing study of *L. donovani* field isolates showing a correlation between polysomy variations and changes in heterozygosity. *Leishmania* therefore represents an ideal non-conventional system to investigate how karyotype variations and allelic selection drive parasite fitness gains and how transient and stable aneuploidies impact the long-term evolutionary trajectory.

Results

Karyotype and haplotype selection in field isolates

We first explored the impact of aneuploidy on allelic selection in *Leishmania* using the recently published sequencing data of 204 *L. donovani* field isolates[19]. These populations originate from independent evolutionary radiations that occurred after a population bottleneck caused by DDT vector control campaigns in the 1960s. This makes them a convenient benchmark to study genetic diversity dynamics. We conducted read-depth analyses on publicly available sequencing data sets[19] to explore ploidy variations across culture-adapted isolates (Fig. 1a). Even though a large number of intermediate read-depth values suggest frequent mosaicism (Fig. 1a, Supplementary Table 1), the relatively high frequency of discrete polysomy levels other than disomic levels (notably trisomy) indicates that a large fraction of isolates may be homogenous for single or co-occurring aneuploidies (e.g chr 5, 9, 23 and 26, Fig. 1b). While all chromosomes are predominantly disomic except for chr 31, the overall karyotypic diversity is very high with the most frequent karyotype representing only about 10% of the 204 isolates.

The frequent occurrence of karyotypic similarities among isolates suggests that the

97 corresponding aneuploidies may either have a common, monoclonal origin (i.e. one unique
98 founding duplication) or result from frequent convergent processes (i.e. selection of several
99 independent duplications of the same homologous chromosome). Relatively weak
100 correlations between aneuploidy and SNP variation[19] does not discriminate between these
101 two possibilities. While clonal expansion of a single founder may underlie the most common
102 aneuploidies (like chr 31, 5 or 26), their occurrence in karyotypically very diverse isolates is
103 also compatible with a polyclonal origin.

104 Under a polyclonal scenario, each culture-adapted isolate arises from the expansion
105 of mixed subpopulations sharing some independently generated polysomies. This model is
106 in good agreement with mosaicism, previously observed in cultured *Leishmania major*[18]. We
107 analyzed allelic frequency distributions to evaluate the likelihood of polyclonal origin
108 (Supplementary Fig. 1). In this analysis, well-defined peaks in the merged profiles of most
109 chromosome allelic frequency distributions indicate that most heterozygote alleles are
110 balanced similarly across isolates. For disomic profiles, we typically observed unimodal
111 allelic frequency distributions centered at 50% (Supplementary Fig. 1a - left column). These
112 observations are consistent with diploid asexual populations in which most alleles diffuse by
113 random drift.

114 Chromosomes with a read-depth level compatible with trisomy present us with a
115 more complex situation. In these chromosomes, the merged allelic frequency profiles
116 (Supplementary Fig. 1a – right column) revealed a continuum between the two extreme
117 cases of pure unimodal (e.g. chr 1, Fig. 1c, Supplementary Fig. 1d bottom scenario) and
118 pure bimodal (e.g. chr 5, Fig. 1c, Supplementary Fig. 1d other scenarios) allelic frequency
119 distributions. In the absence of mosaicism, trisomies usually cause bimodal allelic frequency

distributions. Unimodal distributions like chr 1 therefore reflect population mosaicism as a consequence of a duplication scenario with both homologous chromosomes equally likely to be duplicated across the entire population of cells constituting an isolate, thus maintaining the original allelic frequency equilibrium of the disomic state (Supplementary Fig 1d). The coexistence of haplotypically different trisomies within the same isolate clearly indicates mosaicism, which strongly suggests a polyclonal origin. In contrast, bimodal allelic frequency distributions centered on 33 and 66%, like that observed for chr 5, reflect the collective doubling in relative frequency of half of the alleles and are characteristic of homogenous duplication of the same homologous chromosome within each isolate population (thus defining an identical haplotype). While the integration of various mosaicism combinations by sequencing analysis could cause similar read-depths, it is unlikely that mosaicism would also cause bimodal allelic frequency distributions centered on 33 and 66%.

Even though they occur in a large fraction of the isolates, chr 5 trisomies are most likely the result of a convergent process as suggested by their polyphyletic dispersion on the phylogenetic tree of the 204 filed isolates[19]. Given such a convergent process, we asked if these trisomies were selected as karyotypes or as specific haplotype combinations. Haplotype selection should cause increased homogeneity across isolates. We therefore measured allelic dispersion across trisomic isolates. We estimated the relative loss of heterozygosity (LOH) between disomic and trisomic isolates for each chromosome. This analysis shows that chr 5 had one of the lowest allelic diversity (highest average LOH) and suggests frequent selection of the same homologous chromosome among the trisomic isolates.

On these same subsets of trisomic and disomic chromosomes, we measured the segregation variation (SV). SV is defined as the average ratio between the frequency of the

144 dominant allele in each trisomic isolate and its corresponding average frequency in disomic
145 isolates. This index reflects the changes in segregation patterns between disomic and
146 trisomic states and is highest for chromosomes that show a unimodal distribution at a
147 disomic state and a bi-modal distribution at a trisomic state. As expected, we found chr 5 to
148 have one of the highest SV ratio. We also found a strong correlation between LOH and SV
149 across all chromosomes with the scatter plot displaying a clear continuum between chr 1
150 and chr 29 (Fig. 1d). This observation is in agreement with our hypothesis that high levels of
151 haplotype selection cause increased allelic homogeneity across isolates. These results
152 confirm that bimodal allelic frequency profiles are more likely induced by a selection process
153 rather than a monoclonal founding event. Monoclonal origin without haplotype selection
154 would cause a high SV and low LOH. Interestingly, distinctly bimodal chr 5 trisomies coexist
155 with unimodal trisomies of other chromosomes in a subset of the isolates (e.g. chr 12,
156 Supplementary Fig. 2). This finding indicates that various levels of haplotypic selection can
157 affect different chromosomes within the same parasite isolate. Altogether, the complex
158 karyotypic and haplotypic variations across field isolates clearly favors a polyclonal mosaic
159 origin as the most common scenario with diversity drawn from an equally diverse original
160 population, as postulated by Sterkers et al.[18].
161
162 **In vivo *karyotype fluctuations***
163 Polyclonal origin implies a pre-existing population-wide karyotypic mosaicism within each
164 individual isolate. Maintaining this degree of variability would require frequent polysomic
165 fluctuations. We measured this effect by monitoring chromosomal copy number when re-
166 passaging an aneuploid field isolate (strain LdBPK282) through hamsters. The rapid shift to
167 a predominant disomic karyotype in the field isolate analysed here (and an independent

168 isolate shown by Dumetz et al.[20]) demonstrates highly dynamic aneuploidy fluctuations (Fig.
169 2a). It is unclear, however, if the observed aneuploidies during culture adaptation arise from
170 expansion of existing subpopulations or constitute a reversible *de novo* phenomenon. We
171 therefore examined this question applying single cell DNA-FISH analysis on liver and spleen
172 amastigotes. We analyzed the experimental Sudanese *L. donovani* strain LD1S with
173 available DNA-FISH probes to monitor aneuploidies for chr 5, 17, 22 and 27. Hamsters were
174 infected with amastigotes isolated from infected hamster spleens that showed a disomic read
175 depth for all chromosomes except tetrasomic chr 31 (see Supplementary Fig. 3a). For all
176 chromosomes investigated, our analysis confirmed frequent *in situ* mosaic aneuploidies in
177 liver and spleen (Fig. 2b). This pre-existing diversity can facilitate population-wide
178 aneuploidies in response to environmental changes. This possibility is in line with the mosaic
179 origin of aneuploidies observed in culture[18].

180 ***In vitro* karyotype and haplotype selection**

181 While field isolate and *in vivo* analyses suggest that parasite adaptation relies on a
182 mechanism dedicated to accommodate frequent aneuploidies, these observations only
183 provide static snapshots of apparently highly dynamic fluctuations. To further elucidate this
184 process, we used again the experimental strain LD1S. In contrast to *L. donovani* clinical
185 isolates from the Indian sub-continent, this strain contains a larger number of heterozygous
186 sites (23014 compared to a median of 1143 per field isolate), which makes it an ideal model
187 for longitudinal monitoring of chromosome copy number variation and haplotype selection.
188 We used HTSeq to follow hamster-derived LD1S amastigotes during adaptation to *in vitro*
189 culture. Read depth analyses indicated that parasites rapidly establish stable trisomies for
190 chr 5, 9, 23, and 26 between *in vitro* passages p2 (ca. 20 generations) and p10 (ca. 60
191 generations) (Fig. 3a and Supplementary Fig. 3a). These karyotypic trajectories matched the

most common variations observed in the field isolates (Fig. 1a and b, Supplementary Fig. 1) and they were highly reproducible across two independent experiments[20]. Not all aneuploidies are stable and homogenous, however. For example, chr 20 underwent a transient trisomy between passages p2 and p20 (ca. 190 generations), while probable mosaic aneuploidies (Fig. 3a) occurred for chr 14 and 15 and were stably maintained as judged by their intermediate read-depth levels at p10 and p20.

Population mosaicism in *Leishmania* major was previously examined on cultured promastigotes using a combination of sub-cloning and DNA-FISH analysis[18]. We used systematic comparison between p20 and eight individual sub-clones to model the original population complexity (Fig. 3b and Supplementary Fig. 3b). Haplotype and karyotype comparisons suggest that the eight clones may have arisen from at least two independent founding individuals (Supplementary Fig. 3c). Indeed, aneuploidies provide a powerful insight into haplotype phasing, due to systematic frequency shifts associated with duplicated homolog chromosome[21]. In this case, direct haplotype comparisons (Fig. 4a, Supplementary Fig. 4) clearly show how the allelic nature of chr 5 and chr 9 trisomies sets clones 1 and 8 apart from the rest. For instance, the existence of two groups with differences of major alleles on chr 5 (1 and 8 vs the rest) indicates that these trisomies were established through duplications of different chr 5 homologs (i.e. same chr 5 homologue duplicated in clones 1 and 8 that is different from the other clones). The most parsimonious reconstruction simply requires two trisomies of independent origin and is consistent with a high aneuploidy turnover[18]. This explanation also applies to chromosome 9 and further supports the common origin of clones 1 and 8. Sequencing 8 subclones has limited statistical power, and our speculation of two original founders does rely on a parsimonious hypothesis. Yet, the existence of clearly distinct subpopulations featuring similar aneuploidies further supports the

mosaic origin scenario, under which individual strains represent complex mixtures of stable, karyotypically distinct subpopulations.

Between subpopulations, similar karyotypes and haplotypes likely reflect a form of convergence. At the karyotypic level, the most obvious indications of convergence occur on chr 5, 9 and 26, whose trisomies are shared across all clones despite their independent origins. Selection could also occur at the haplotypic level, with the clearest indication provided by chr 26 whose haplotype map shows a consistent selection of the same allelic combination across clones of different origin (Fig 4a). Unfortunately, the simultaneous haplotype/karyotype selection taking place on this chromosome makes it impossible to discriminate between purifying (lethality of one of the two possible trisomies) or positive (higher fitness of one trisomy) selection pressure being imposed on its haplotype. Chromosome 20 instead provides clear evidence for positive haplotype selection. This chromosome is disomic and heterozygous at p2, becomes trisomic at p10, but reverts back to disomy at p20 (Figs. 3a and 4b, Supplementary Fig. 5). After reversion, the allelic frequency profiles unexpectedly show a near-perfect homozygote disomy in 6 out of 8 clones (i.e. relatively flat allele frequency profile combined with a 60% decrease in the number of heterozygous sites, Fig. 4a and Supplementary Fig. 6). In chr 20, transient trisomy therefore provides an intermediate step to establish a homozygous disomy, whose selection contributes to positive fitness as previously seen on yeast[22].

Correlation of aneuploidies with phenotypic variations

Since *Leishmania* primarily lacks classical gene regulation mechanisms, it has long been speculated that polysomy should result in proportional transcriptional fluctuations[15] and that aneuploidy mediated gene expression changes could be a trait under selection. We used

HTSeq analysis to correlate read-depth levels between genome and transcriptome (Fig. 5a, Supplementary Fig. 7) and found a very high correlation for most chromosomes (r=0.72), with the notable exception of chr 31. In this chromosome that is predominantly tetrasomic across various *Leishmania* strains, transcriptomic output appears halved, suggesting some form of dosage compensation, which cancels the polysomic effect. We also found karyotypic fluctuations resulting from culture adaptation to correlate with rapidly increasing *in vitro* fitness as judged by the decreasing generation time (Fig. 5b, left panel). Concurrently with these variations the parasites showed decreased *in vivo* fitness as indicated by lower infectivity (Fig. 5b middle and right panels). To assess if these variations take place during the parasite life cycle, we separately sequence amastigotes obtained from spleen and liver from one infected hamster. We found significantly distinct allelic profile variations for chr 20, which showed a unimodal profile in liver and a bimodal profile in spleen (Fig. 5c, Supplementary Fig. 8), while read depth analysis indicated a disomic state in both tissues (Supplementary Fig. 9). These observations closely recapitulate our *in vitro* analyses of parasite clones and indicate that transient polysomies may contribute to parasite adaptation *in vivo*.

Long-term genetic implications on genetic diversity

Maintaining diversity under strong selection poses a dilemma for all microbial pathogens, since it requires a mechanism that permits some compromise between immediate survival through selection and long-term adaptation through the maintenance of genetic diversity. We therefore searched for traces of increased genetic diversity associated with frequent aneuploidies and found a clear signal in the 204 field isolates (Fig. 6). Integrating genetic

263 variation across each individual isolate shows that polysomy-prone chromosomes exhibit a
264 significantly higher level of heterozygous sites than their more stable counterparts. This
265 trend is especially strong for chr 31, which has the most stable, frequent and highest order
266 aneuploidy. This finding suggests that *L. donovani* might utilize aneuploidy to accumulate
267 mutations and increase its diversity. Further, these results confirm that even though transient
268 *in vivo* aneuploidies are difficult to detect and quantify, they are frequent enough to shape
269 the parasite genome and its genetic diversity.
270
271 **Discussion**
272 Utilizing recently published genomes of 204 field isolates[19] and conducting evolutionary
273 experiments, we demonstrated that karyotypic variation modulates transcript abundance and
274 thus likely generates considerable phenotypic variability with possible fitness gains
275 associated with tissue-specific haplotype selection *in situ*. The genomic landscape defined
276 by an exhaustive combination of all possible trisomies, disomies and monosomies along with
277 their haplotype variations is substantial and at least one order of magnitude larger than a
278 similar estimate recently made on the basis of aneuploidy variations[23]. Aneuploidy turnover
279 therefore provides an efficient alternative to sex-based haplotype selection. This mechanism
280 may be especially useful within the mammalian host, where the absence of sexual
281 reproduction creates an even heavier selective pressure on the parasite's survival prospects.
282 Rapid aneuploidy turnover combined with haplotype selection permits fast adaptation
283 but comes with a significant genetic cost. Since alleles with lowered frequencies are at a
284 higher risk of disappearing from the genetic pool[25], haplotype selection may induce loss of
285 heterozygosity. The transient trisomy of chr 20 and its *in vivo,* tissue-specific haplotypic
286 diversity illustrates the dilemma for the parasite between over-adaptation to any given

environment that may involve irreversible genetic tradeoffs and maintenance of sufficient genetic diversity for future adaptations. In this chromosome, the loss of heterozygosity could contribute to adaptation across conditions and tissues, however this same process may also yield an evolutionary dead end. The evolutionary success of *Leishmania* testifies of its ability to establish a delicate balance between these conflicting requirements for short- and long-term adaptation. Long-term survival for the parasite involves two complementary mechanisms to generate and maintain genetic diversity. The first one relies on the mosaic origin of selected populations. We demonstrated that culture adapted isolates do not have a single founding parent but result from the simultaneous selection of several individuals. This process maintains a high level of karyotypic and haplotypic diversity allowing polysomies to occur frequently and independently in genetically distinct individuals. The second source of diversity is a direct consequence of the relaxed selection that occurs after gene or chromosome duplication[26]. Our findings agree with these models and further suggest that aneuploidy-prone chromosomes have significantly higher mutation rates than their more stable counterparts. Thus, aneuploidy represents an additional powerful mechanism to increase genetic diversity in *L. donovani*.

Leishmania, like all known infectious agents, provides us with an excellent example of genetic selection driven by survival. But it does so in a very unusual way: while high mutation rates and genetic material exchange through sexual or analogous mechanisms mediate survival in most pathogens, *Leishmania* evolved an additional strategy based on genome instability to enhance parasite evolvability[27–29]. We show here that the parasite harnessed this process and efficiently combined it with haplotype selection for rapid and efficient genome adaptation under environmental pressure. This strategy has direct consequences for current protocols in *Leishmania* drug and biomarker discovery. First,

Leishmania genome instability limits all current and future drugs that directly target the parasite biology and select for resistance phenotypes[30,15,14]. New strategies for anti-leishmanial drug discovery should avoid direct parasite selection and rather target the parasites' dependence on host cell metabolism. Second, the massive genomic changes we observed during *Leishmania* culture adaptation challenges current protocols for biomarker discovery, which all rely on *in vitro* expansion of clinical isolates. In the light of our findings, we should modify these protocols to identify additional biomarkers hidden/diluted by culture-based methods especially when polysomy itself is considered a biomarker. Alternatively, we could establish culture-independent approaches that propagate field isolates in experimental animal models or apply direct tissue sequencing. It is clear that genome instability needs to be considered when investigating medically relevant aspects of the parasite's phenotype, such as tissue tropism, drug susceptibility and pathogenicity. Our findings establish a foundation for the future discovery of haplotypes with diagnostic and prognostic value.

Material and Methods

***L. donovani* strains, culture conditions and cell cloning.** Culture-adapted *L. donovani* field isolates maintained for more than 20 in vitro passages from the Indian Sub-Continent were used in our analyses to track evolutionary diversity maintained across populations[19]. Isolates were not sub-cloned and populations were used for all studies. Infectious *L. donovani* strain 1S2D (MHOM/SD/62/1S-CL2D, referred to here as LD1S) was obtained from Henry Murray (Weill Cornell Medical College, New York, USA) and maintained in hamsters by serial passaging (see below). All LD1S studies were performed with populations except when specified otherwise. For promastigote differentiation and culture, 1 x 10^7 LD1S amastigotes purified from hamster spleens were cultured at 26 °C in M199 media supplemented with 10 % FCS, 25 mM HEPES pH 6.9, 4 mM NaHCO 3 1 mM glutamine, 1 x RPMI 1640 vitamin mix, 0.2 µM folic acid, 100 µM adenine, 7.6 mM hemin, 8 µM biopterin, 50 U/ml of penicillin, and 50 µg/ml of streptomycin. Promastigotes were then maintained in culture by serial dilution in fresh medium once they reached stationary phase. At passages 2, 10 and 20 corresponding to approximately 20, 60 and 190 generations respectively, parasites in exponential growth phase were collected and adjusted to 2 x 10^8 parasites per tube for DNA extraction (see below). After 20 *in vitro* passages, serial dilutions of promastigotes were plated on M199 Agar plates and 8 clones were selected and amplified in promastigote medium.

Hamster infection and isolation of infectious amastigotes. We used female RjHan:AURA golden syrian hamsters between 4 to 6 weeks of age (Janvier Labs, France) and a weight between 60 to 90g. Housing of the animals and all experiments were

conducted in agreement with the project number 2013-0092 approved by the Institut Pasteur Ethics committee in accordance to the European legislation/guidelines EU 2010/63. We used a total of 13 hamsters for DNA-FISH analysis (1), DNA/RNA extraction from purified amastigotes (2), hamster infection (6), monitor tissue parasite burden (3), and comparative HTseq analysis of liver and spleen amastigotes (1). Given the nature of our animal experiments to compare infected versus non-infected animals and use infected animals as a source of biological material, our studies were not blinded or randomized. No animals were excluded from analysis. Anesthetized hamsters were inoculated by intra-cardiac injection with 5×10^7 amastigotes obtained from infected hamster spleens. Hamster weight was monitored and animals were euthanized with CO_2 after four months of infection. Spleens and livers were collected, weighed, and homogenized in PBS as described[31]. For amastigote purification, tissues were homogenized in 25 ml PBS supplemented with 2.5 mg/ml saponine using the gentleMACS homogenizer and gentleMACS M tubes from Miltenyi. Homogenates were cleared by centrifugation at 130*g* for 5 min at 4°C, the supernatants were collected, 1 ml of saponine (25 mg/ml) was added under gentle agitation, and parasites were harvested 5 min later by centrifugation at 2000*g* for 10 min at 4°C. After two washing steps with PBS, remaining host cell contaminants were removed by Percoll centrifugation. Briefly, parasite were resuspended in 6 ml of 45 % Percoll, and 3 ml were layered on a cushions of 2 ml of 90 % Percoll in 15 ml falcon tubes. After 35 min of centrifugation at 3500*g*, 15°C, amastigotes were recovered from the interface of the gradient and washed 3 times in PBS (centrifugation at 2000*g*, 10min, 4°C). Tissue-derived amastigotes were adjusted to 2×10^8 parasites per tube for RNA or DNA extraction or inoculated in culture medium for differentiation into promastigotes and further culture.

DNA FISH analysis. DNA probes for chromosomes 5, 17, 22, and 27 were prepared as described in[18]. Amastigotes were purified from infected hamster livers and spleens (previously inoculated with amastigotes obtained from infected hamster spleens) as described above, immobilized on glass slides, fixed in 4% paraformaldehyde and 4% acetic acid. Fluorescence *in situ* hybridization was performed according to Sterkers et al.[18]. *Leishmania* cells were viewed by phase contrast, and fluorescence was visualized using appropriate filters on a ZeissAxioplan 2 microscope with a 100 x objective. Digital images were captured using a Photometrics CoolSnap CDD camera (Roper Scientific) and processed with MetaView (Universal Imaging). Z-Stack image acquisitions (15 planes of 0.25 mm) were systematically performed for each cell analyzed using a Piezo controller, allowing to view the nucleus in all planes and to count the total number of labeled chromosomes. The ploidy was estimated on 100 labeled cells. We carefully considered the cell cycle state and excluded cells that were replicating their DNA as judged by the presence of large or duplicated kinetoplasts, which corresponded to less than1% and thus a negligible fraction. This low replication rate was confirmed by our own infection data analysis that allowed us to calculate a generation time of amastigotes *in vivo* of 10.4 days.

Genomic sequencing. All sequencing analysis were performed using Illumina short-read technology. The sequences for BPK282/0 promastigotes and amastigotes shown in Figure 2a were generated using the Nextera XT library preparation kit (Illumina), the KAPA Library Quantification Kits (KAPA Biosystems), and paired-end sequencing on a HTSeq 1500 sequencer (Illumina). The sequences of LD1S liver and spleen amastigotes (Figure 5c, Supplementary Figure 8 and 9) were carried out as a service by the Centro Nacional de

Análisis Genómico (CNAG, Barcelona) using the following protocol: 0.3-0.6 micrograms of sheared genomic DNA (Covaris E201) was end-repaired, adenylated and ligated to Illumina specific indexed paired-end adapters (adapter:insert molar ratio 40:1). The DNA library was size selected with AMPure XP beads in order to reach an insert size of 220-550bp and to remove unligated adapters. The final libraries were quantified with the Library Quantification Kit (Kapa Biosystems). On the base of the qPCR quantification the molar concentration sufficient for sequencing run was estimated. Each library was sequenced using TruSeq SBS Kit v3-HS (Illumina Inc.), in paired end mode, 2x101bp, in a fraction of one sequencing lane of an HTSeq2000 flowcell v3 (Illumina Inc.) according to standard Illumina operation procedures with a yield between 3,593 - 6,326 Gb for each sample. The rest of the LD1S *Leishmania* sequences were generated at the Centre for Genomic Regulation (CRG, Barcelona) using the following protocol: DNA was isolated using DNeasy blood and tissue kits from Qiagen, nucleic acid concentration was measured with a NanoDrop® spectrometer, and between 2 to 5 µg were used for sequencing and quality control. DNA sequencing was performed using an Illumina HiSeq 2000 platform and TruSeq v3 kits. One µg of nucleic acids was used for DNAseq while the rest of the material was used for quality control. All libraries were sequenced in 125 bp reads. A summary of the read numbers and read-length associated with each genome is provided in Supplementary Table 2.

Read depth Sequencing Analysis. All DNA sequences were mapped against *L. donovani* BPK282A1 from TriTrypDBv7. Mapping and post processing was carried out using a combination of BWA mem (v0.7.8)[32] along with Samtools (v1.3)[33] in order to refine read information and clean mapped sequences. Alignments were further refined using GATK (v2.8) IndelRealigner and MarkDuplicates[34]. Samtools depth command was used to estimate

read-depth for every base, so as to determine median read-depth for each chromosome. In order to estimate chromosome polysomy level in each sample, chromosome 34 - which showed a stable disomic level across samples - was used to normalize other chromosomes read depth. This measure was calibrated to determine polysomy levels, under the assumption that the most frequent polysomy levels correspond to disomy. Intermediate polysomy levels were classified as trisomic or tetrasomic events by using cutoffs on the normalized chromosome read depth values. Chromosomes with values ranging between 2.6 and 3.4 were classified as trisomic, and values between 3.6 and 4.4 as tetrasomic. These calibrated values were used to select subsets of chromosomes having comparable aneuploidy levels. (Accession number: PRJEA61817)

Aneuploidy co-occurrence. Aneuploidy co-occurrence analysis was carried out by measuring the Pearson correlations between pairs of chromosomes, with each chromosome being represented as vector of normalized depth values, each value being drawn from an isolate. P-Values were measured so as to identify pairs having the highest levels of co-variation - as opposed to constant or unrelated polysomy levels.

Variant calling and allele frequency analysis. Pileups were generated across the reference genome for all the samples. Samtools mpileup and its multiallelic caller were then used to detect variants separately in all samples. In order to filter out sites potentially affected by episomal amplifications, regions with a read-depth higher than 1.6 x the median depth of each chromosome were excluded from the analysis. The resulting sites were then used to draw allele frequency statistics, and produce allele frequency profiles. Variants were then filtered using a quality threshold of 15 using a Read Position Bias (RPB) filter of 0.1 and

alleles with frequencies comprised between 0.1 and 0.9 were retained. This approach, that was explicitly designed to prioritize sensitivity over specificity, yields about 10 times more sites than a more conservative approach (28071 vs 2418 sites [19,20]). The expected rate of false positive associated with this procedure can be shown to be in the order of 3% (-10*log10(1/31.6)). This amount of noise that would be incompatible with biomarker discovery does not compromise the current analysis focused on tracing homologue chromosome fate in aneuploidies.

Merged allele frequency profiles. The allele frequencies associated with variants called as explained in the previous section were then used to build merged allele frequency distribution profiles for each chromosome across isolates. Profiles of homologue chromosomes classified with the same polysomy level - as explained in read depth analysis - were added up to yield a set of merged profiles (Supplementary Figure 1). The stacking of pre-computed profiles presents the advantage of maintaining the individual segregation profiles and can therefore be used to estimate profile homogeneity across samples as indicated by well-defined peaks that reflect conserved allelic frequencies across isolates. Most of these profiles exhibit the near canonical profiles expected for allele distribution in a diploid, triploid or tetraploid organism and therefore confirm the suitability of the noise to signal ratio of our site calling procedure.

Segregation variations and Loss of Heterozygosity dispersion plots. Segregation variations (SV) between disomic and trisomic chromosomes were estimated by considering for each heterozygous site the log ratio between the frequency of the most major (major allele) allele in the considered isolate, and the average frequency of the same allele in the

disomic population.

$$\text{SV(Chromosome N)} = Avg\left(\sum_{\substack{Het\\Sites}}^{S}\sum_{\substack{Trisomic\\Isolates}}^{t}\log\left(\frac{Frequency(MA(s,t),t)}{AvgFrequency(MA(s,t),D)}\right)\right)$$

With

- *s* heterozygote site in an isolate
- *t* isolate trisomic for chromosome N
- *T* all the isolates trisomic for chromosome N
- *d* isolate disomic for chromosome N
- D all the isolates disomic for chromosome N
- *MA(s,t)* the Major Allele on site *s* in the isolate *t* trisomic for *Chromosome N*
- *Frequency (MA(s,t),t)* is the frequency of *MA(s,t)* in isolate *t*
- *AvgFrequency (MA(s,t),D)* the average frequency of *MA(s,t)* across all isolates *D* disomic for Chromosome N

This procedure amounts to considering the disomic average as a null hypothesis with respect to the fate of the corresponding alleles in their trisomic states. When doing so the nature of the allele is not considered and two isolates having different major alleles on the same site can nonetheless contribute the same value to the distribution of ratios, as plotted in Supplementary Figure S1b. These ratios were then averaged for each chromosome and used as a measure for Segregation variation. On Supplementary Fig S1, profiles are ordered according to this ratio that tends to reflect the disomic tendency in the trisomics.

Loss of heterozygosity (LOH) was estimated by considering for each trisomic heterozygous site the log odd ratio between the average frequency of the major allele across

trisomic isolates and the corresponding frequency across disomic isolates. These ratios were then averaged for each chromosome and used as a loss of heterozygosity measure using the following formula

$$\text{LOH(Chromosome N)} = \text{Avg}\left(\sum_{\substack{Het\\Sites}}^{S} \log\left(\frac{AvgFrequency\left(MA(s,T),T\right)}{AvgFrequency\left(MA(s,T),D\right)}\right)\right)$$

With

- *t* isolate trisomic for chromosome N
- *T* all the isolates trisomic for chromosome N
- *d* isolate disomic for chromosome N
- D all the isolates disomic for chromosome N
- *MA(s,T)* the allele *s* that has the highest average frequency across all the *T* isolates trisomic for *Chromosome N*
- *AvgFrequency(MA(s,T),T)* is the average frequency of *MA(s,T)* across all the *T* isolates trisomic for *Chromosome N*
- *AvgFrequency(MA(s,T),D)* is the average frequency of *MA(s,T)* across all the *D* isolates disomic for *Chromosome N*

This measure reflects the tendency of a given allele to dominate in frequency the overall trisomic population. It makes it possible to discriminate between a scenario where the two original disomic alleles are equally likely to have increased frequencies in each isolate (i.e. random selection of the duplicated homologous chromosome), and situations where the same allele is consistently amplified across all isolates (i.e. clonal amplification of the same trisomy or consistent selection of the same homologous chromosome amplification). The comparison between SV and LOH across field isolates is shown in Fig1d, with a p-value

518 calculated as the t-test statistic based on Pearson's product moment correlation coefficient
519 using cor.test() function in R.
520
521 **Haplotype phasing.** Haplotype phasing was carried out by considering the effect of
522 aneuploidy on the original frequencies following the protocol originally defined in [21]. In this
523 analysis, alleles undergoing similar shifts are considered as being part of the same
524 haplotype block. While heterozygous SNPs are not dense enough in the 204 filed isolates for
525 this approach to be validated, we took advantage of the LD1S hybrid origin to verify that
526 SNPs occurring with the insert distances of the pair-ended sequencing had similar
527 segregation shifts after duplication and could therefore considered to be genetically linked,
528 thus ruling out recombination as a major confounding factor. This finding is in agreement
529 with the original study of very low recombination rates in the 204 filed isolates. Our analysis
530 indicates that 50% of the SNP pairs within insert distance of one another have allelic
531 frequencies less that 10% point apart in the bimodal trisomics (Fig S11), as compared with
532 23% when randomizing pair ends in order to test for pair connectivity and allele linkage.
533
534 **Transcriptome sequencing.** RNA was prepared using RNeasy mini plus blood and tissue
535 kits from Qiagen according to manufacturer instructions. Nucleic acid concentrations were
536 measured with a NanoDrop® spectrometer. RNA sequencing was performed using an
537 Illumina Hiseq 2000 platform and TruSeq v3 kits. Four µg of nucleic acids were used for
538 RNAseq and the rest of the material was used for the quality control. RNA libraries were
539 sequenced on a single flow cell single stranded 51 bp read. Sequencing reads have been
540 submitted to the European Nucleotide Archive (ENA) and are available under the submission
541 number PRJEB15282. The amount and the length of reads obtained from each analysis are

542 summarized in the Supplementary Table 2.

543

544 **Relative expression estimation**. Gene copy numbers and expression levels were
545 estimated from DNA and RNA read counts respectively, mapping to annotated ORFs using
546 reference gene annotation from TriTrypDBv7. The actual read-counts were measured with
547 BedTools (v2.19)[35] using the BAM files obtained after mapping. In order to normalize for
548 gene length, read counts were estimated in Reads Per Kilobase Mapped (RPKM)[36] across
549 all genes and samples. These RPKM measurements were then used to compare expression
550 levels across samples.

551

552 **Data availability.** All Sequencing reads have been submitted to the European Nucleotide
553 Archive (ENA) and are available under the Accession Number PRJEB15282. The amount
554 and the length of reads obtained from each analysis are summarized in the Supplementary
555 Table 2.

556

557 **Code availability**

558 Custom scripts and code necessary to generate figures and files can be found in the
559 following repository: https://github.com/pprieto/genomevoleish

560

561 **Biography**

562 1. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449,** 835–842 (2007).
563 2. Torres, E. M. *et al.* Effects of aneuploidy on cellular physiology and cell division in
564 haploid yeast. *Science* **317,** 916–924 (2007).
565 3. Pavelka, N. *et al.* Aneuploidy confers quantitative proteome changes and phenotypic

566 variation in budding yeast. *Nature* **468,** 321–325 (2010).
567 4. Pavelka, N., Rancati, G. & Li, R. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular
568 adaptation and cancer. *Curr. Opin. Cell Biol.* **22,** 809–815 (2010).
569 5. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of
570 aneuploidy provides increased fitness during the evolution of antifungal drug resistance.
571 *PLoS Genet.* **5,** e1000705 (2009).
572 6. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from
573 haplotype structure. *Nature* **419,** 832–837 (2002).
574 7. Staaf, J. *et al.* Landscape of somatic allelic imbalances and copy number alterations in
575 human lung carcinoma. *Int. J. Cancer* **132,** 2020–2031 (2013).
576 8. Scheinfeldt, L. B. & Tishkoff, S. A. Recent human adaptation: genomic approaches,
577 interpretation and insights. *Nat. Rev. Genet.* **14,** 692–702 (2013).
578 9. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative
579 breast cancer. *Nat. Genet.* **48,** 1119–1130 (2016).
580 10. Sterkers, Y., Crobu, L., Lachaud, L., Pagès, M. & Bastien, P. Parasexuality and mosaic
581 aneuploidy in Leishmania: alternative genetics. *Trends Parasitol.* **30,** 429–435 (2014).
582 11. Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the
583 Indian subcontinent. *Elife* **5,** (2016).
584 12. Alvar, J. *et al.* Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*
585 **7,** e35671 (2012).
586 13. Zilberstein, D. & Shapira, M. The role of pH and temperature in the development of
587 Leishmania parasites. *Annu. Rev. Microbiol.* **48,** 449–470 (1994).
588 14. Leprohon, P., Fernandez-Prada, C., Gazanion, É., Monte-Neto, R. & Ouellette, M. Drug
589 resistance analysis by next generation sequencing in Leishmania. *Int. J. Parasitol.*

590 *Drugs Drug Resist.* **5,** 26–35 (2015).
591 15. Downing, T. *et al.* Whole genome sequencing of multiple Leishmania donovani clinical
592 isolates provides insights into population structure and mechanisms of drug resistance.
593 *Genome Res.* **21,** 2143–2156 (2011).
594 16. Dujardin, J.-C., Mannaert, A., Durrant, C. & Cotton, J. A. Mosaic aneuploidy in
595 Leishmania: the perspective of whole genome sequencing. *Trends Parasitol.* **30,** 554–
596 555 (2014).
597 17. Rogers, M. B. *et al.* Chromosome and gene copy number variation allow major
598 structural change between species and strains of Leishmania. *Genome Res.* **21,** 2129–
599 2142 (2011).
600 18. Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pagès, M. FISH analysis reveals
601 aneuploidy and continual generation of chromosomal mosaicism in Leishmania major.
602 *Cell. Microbiol.* **13,** 274–283 (2011).
603 19. Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the
604 Indian subcontinent. *Elife* **5,** (2016).
605 20. Dumetz, F. *et al.* Modulation of Aneuploidy in Leishmania donovani during Adaptation to
606 Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *MBio* **8,**
607 (2017).
608 21. Legrand, M. *et al.* Haplotype mapping of a diploid non-meiotic organism using existing
609 and induced aneuploidies. *PLoS Genet.* **4,** e1 (2008).
610 22. Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress.
611 *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21010–21015 (2012).
612 23. Sterkers, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: mosaic
613 aneuploidy in Leishmania. *Mol. Microbiol.* **86,** 15–23 (2012).

24. Rougeron, V., De Meeûs, T., Kako Ouraga, S., Hide, M. & Bañuls, A.-L. 'Everything you always wanted to know about sex (but were afraid to ask)' in Leishmania after two decades of laboratory and field analyses. *PLoS Pathog.* **6,** e1001004 (2010).
25. Lande, R. Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* **30,** 314–334 (1976).
26. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290,** 1151–1155 (2000).
27. Victoir, K. & Dujardin, J.-C. How to succeed in parasitic life without sex? Asking Leishmania. *Trends Parasitol.* **18,** 81–85 (2002).
28. Bastien, P., Blaineau, C. & Pages, M. Leishmania: sex, lies and karyotype. *Parasitol. Today* **8,** 174–177 (1992).
29. Dujardin, J.-C. *et al.* Clonal propagation and the fast generation of karyotype diversity: An in vitro Leishmania model. *Parasitology* **134,** 33–39 (2007).
30. Gazanion, É., Fernández-Prada, C., Papadopoulou, B., Leprohon, P. & Ouellette, M. Cos-Seq for high-throughput identification of drug target and resistance mechanisms in the protozoan parasite Leishmania. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E3012–21 (2016).
31. Pescher, P., Blisnick, T., Bastin, P. & Späth, G. F. Quantitative proteome profiling informs on phenotypic traits that adapt Leishmania donovani for axenic and intracellular proliferation. *Cell. Microbiol.* **13,** 978–991 (2011).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

34. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
36. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).

Authors Contribution

P.P.B., P.P, C.N. and G.F.S worked on all aspects of work, contributed to the design of the project and wrote the article, G.B contributed to *in silico* analysis, F.D. performed hamster infection experiment with field isolate, H.I. and J.C.D. contributed to the field isolates analyses and revised the paper, D.K. helped analyzing the sequencing data, H.H. was responsible for the genomic sequencing of the *in vitro* clones, V.C, P.B. and Y.S. contributed the DNA-FISH analysis.

Work on samples from the Indian sub-continent was supported by the EU FP7 (Kaladrug-R, contract 222895), the Belgian Science Policy Office (TRIT, P7/41), the Department of Economy, Science and Innovation in Flanders (ITM-SOFIB) and the Flemish Fund for Scientific Research (G.0.B81.12). We thank Life Science Editors for editing assistance.

Funding

Plan Nacional [BFU2011-28575 to C.N., P.P.B., D.K.]; Center for Genomic Regulation (CRG); Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013–2017' [SEV-2012–0208]; Center for Genomic Regulation (CRG) (C.N, P.P.B., D.K). Grant from the Institut Pasteur International Department strategic fund to the LeiSHield consortium (P.P.B, P.P, C.N., G.F.S., G.B., Y.S.). Agence Nationale de la Recherche (ANR) within the frame of the "Investissements d'avenir" program (ANR-11-LABX-0024-01 "PARAFRAP") (V.C.,P.B., Y.S.). The EU FP7 (Kaladrug-R, contract 222895), the Belgian Science Policy Office (TRIT, P7/41), the Department of Economy, Science and Innovation in Flanders (ITM-SOFIB) and the Flemish Fund for Scientific Research (G.0.B81.12) (J.C.D., F.D., H.I., M.D.).

669 **Competing interests**

670 The authors declare no competing financial interests.

671

672 **Figures**

673 **Figure 1:** Genome instability, aneuploidy co-occurrence and haplotype selection in 204 *L.*
674 *donovani* field isolates. Polysomy analysis based on read-depth comparison and derived
675 allele frequency analysis on specific aneuploidies is shown. *(a) Polysomy analysis.*
676 Polysomy level was estimated for each chromosome and sample by read-depth analysis.
677 The distribution of chromosome copy number is shown across all field isolates. *(b) Co-*
678 *occurrence analysis.* Pearson correlation measured for each pair of chromosomes was
679 plotted while considering the median read depth measured in every field isolate. Empty
680 positions correspond to either uncorrelated pairs or insignificant correlations (significance
681 level of 0.05). One main cluster appears containing typically trisomic chromosomes. *(c)*
682 *Allele frequency profiles.* Merged allele frequency distributions are shown for chromosomes
683 1 and 5 for both disomic (right) and trisomic (left) states. *(d) Segregation variation vs allele*
684 *diversity analysis.* This panel summarizes the relationship between segregation variations
685 (SV, reflecting bi-modality of the merged frequency profiles) and loss of heterozygosity
686 (LOH, reflecting allelic diversity) when comparing trisomic and disomic isolates across 204
687 field isolates and the 36 chromosomes of the genome. For most chromosomes, both
688 quantities are highly correlated as indicated by the blue regression line.

689

690 **Figure 2:** *In vivo* aneuploidy dynamics. The three panels document the dynamic nature and
691 the mosaicism of *L. donovani* aneuploidy *in situ* during hamster infection. *(a)* Aneuploidy
692 pattern of *L. donovani* field isolate BPK282/0 after over 20 passages in culture (*in vitro*, iv)

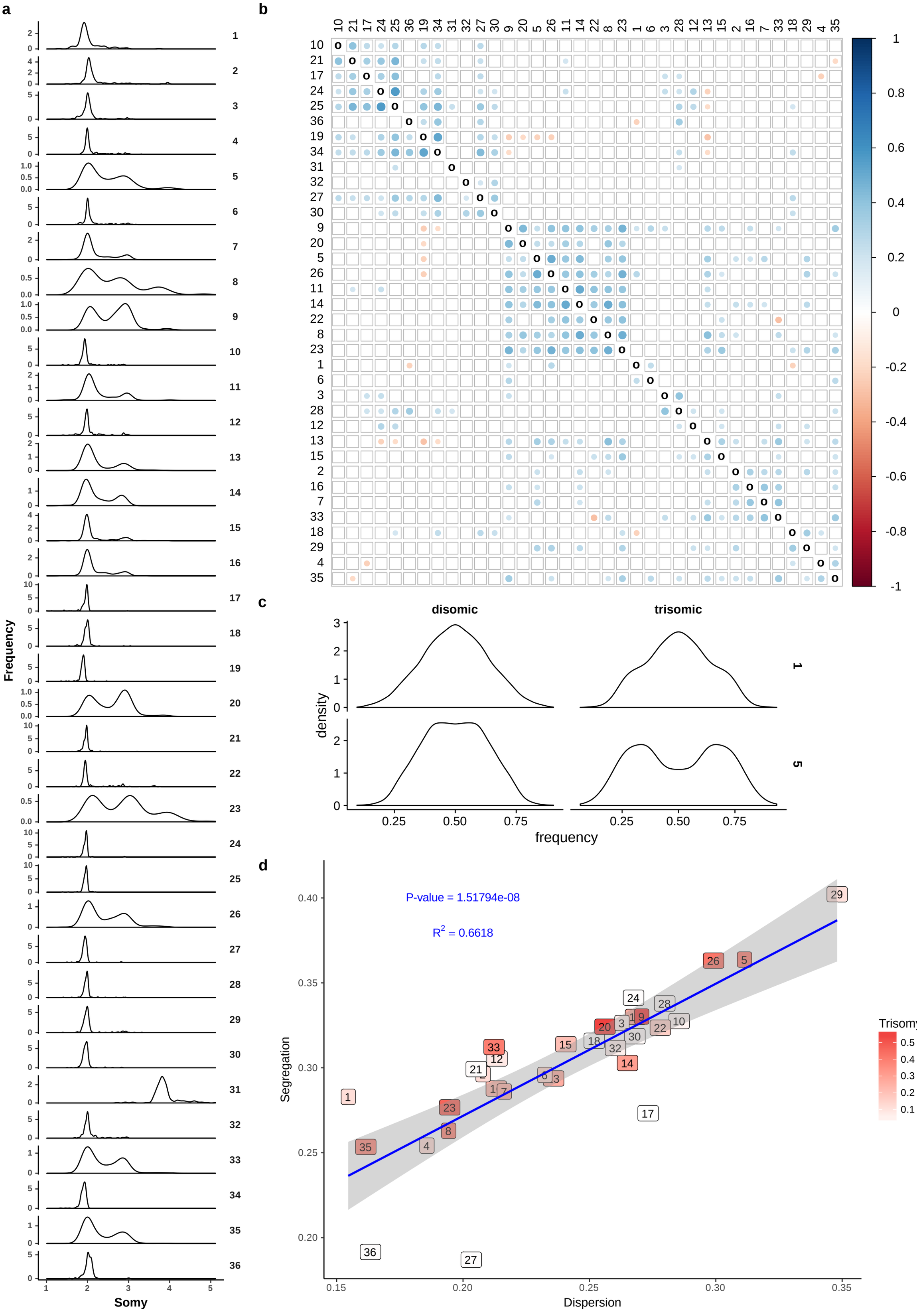
693 and following three consecutive passages of these p>20 parasites in the hamster (H3). Heat
694 maps show median normalized read depths of the 36 chromosomes in both samples and an
695 estimation of the polysomy level of each chromosome: blue, two copies; green, three copies;
696 yellow four copies. *(b) DNA-FISH analysis. L. donovani* strain LD1S amastigotes purified
697 from infected hamster liver (L) and spleen (S) were analyzed with fluorescent labeled probes
698 specific for chromosomes (chr) 5, 17, 22, and 27 and signals were analyzed by microscopy.
699 The signal for chr 5 in spleen-derived amastigotes is shown as an example (phase, phase
700 contrast; FISH, fluorescent signal from DNA-FISH analysis; overlay, merged image of DNA-
701 FISH and nuclear signal obtained with DAPI stain). The bar corresponds to 2 μm. The %
702 chromosome number was calculated counting 100 individual cells per condition (lower
703 panel). The polysomy level is indicated by the bar filling, with white for monosomy, gray for
704 disomy and black for trisomy.
705
706 **Figure 3.** *In vitro* aneuploidy dynamics. This figure shows read-depth variations in *L.*
707 *donovani* LD1S genes (8760 genes) across chromosomes at different passages during
708 culture adaptation (4 samples) and in the 8 sub-clones derived from the parasite population
709 at passage 20. Chromosome gene read-depth distributions are shown in boxplots depicting
710 upper and lower quartiles and the median. *(a) Read depth variation during culture*
711 *adaptation.* Variations in read-depth for *L. donovani* LD1S amastigotes isolated from infected
712 hamster spleen (sp-ama) and derived promastigotes at passage 2, 10, and 20 (p2, p10,
713 p20). Only aneuploid chromosomes are shown (see full panel in Supplementary Fig. 3a).
714 Normalized read depth for every gene within each chromosome is displayed using a
715 standard box-plot representation with the central line representing the median of the
716 distribution. *(b) Read depth variation on individual sub-clones.* Variations on these same

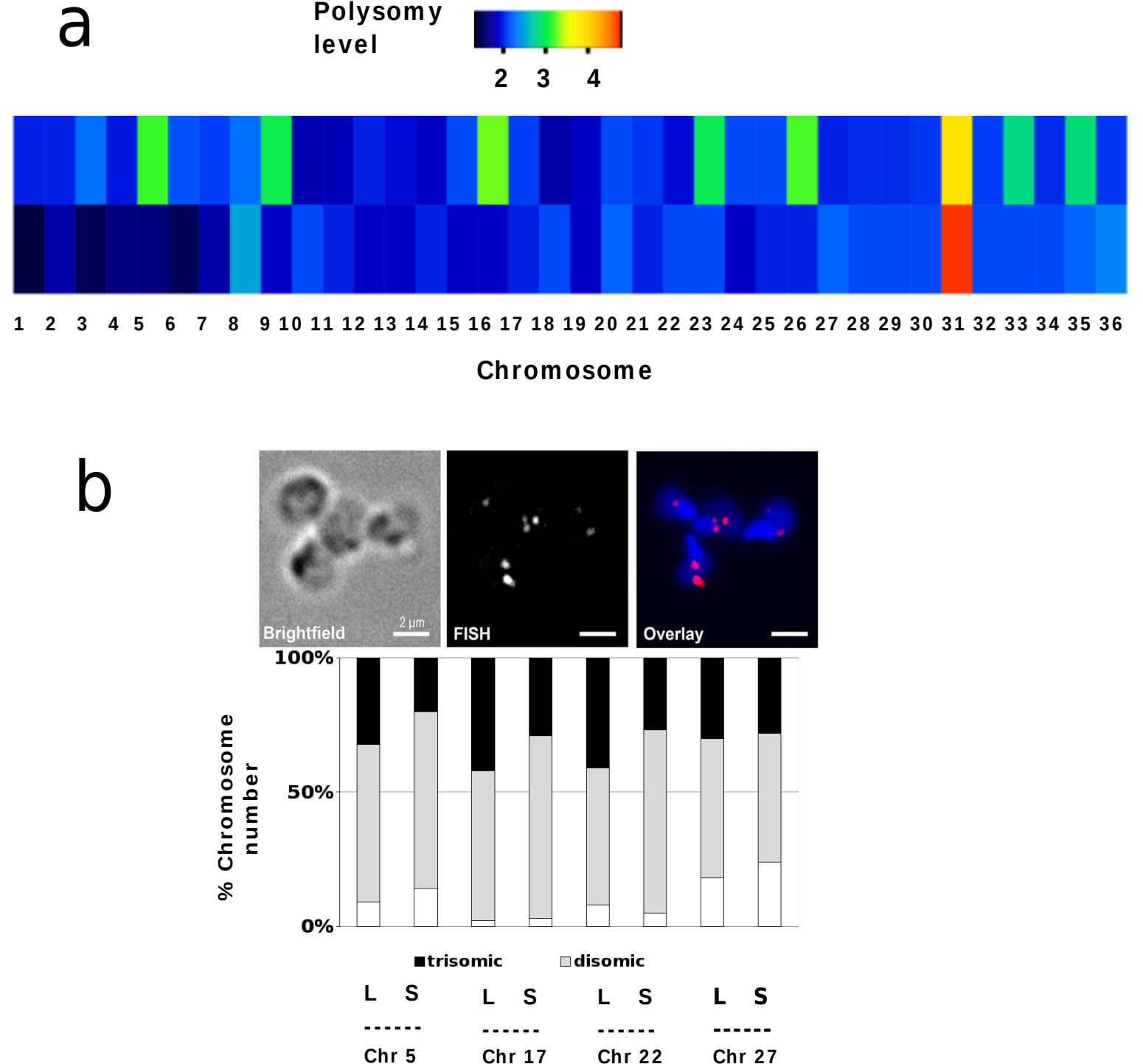
717 chromosomes as shown in panel A are shown for 8 individual parasite cultures subcloned
718 from p20 parasites (sub-clones are indicated by CL, for full panel see Supplementary Fig.
719 3b).
720
721 **Figure 4:** Fluctuations of allele frequency during culture adaptation. These panels document
722 haplotype selection by following the allele frequency distribution in 8 sub-clones and the
723 original p20 parasite population (n=9 samples in total) across . *(a) Haplotype selection in*
724 *clones.* Variable sites in chromosomes undergoing amplification were either colored
725 according to the major alleles (i.e. highest frequency) or left in gray for balanced
726 heterozygote sites. Each painted pattern represents a chromosome, either from the original
727 p20 population (first line) or one of the 8 derived sub-clones. Allele frequency distributions
728 for each sample are shown on the left of each panel. The color code of the distributions
729 corresponds to the read depth, with disomic-to-trisomic-to-tetrasomic transitions indicated by
730 a continuum colored from red to green. The number of heterozygous sites for each
731 chromosome and sample are shown by the histograms (right side of each panel). Samples
732 with low counts as observed for chromosomes 15 and 20 represent loss of heterozygosity
733 occurring during culture adaptation. *(b) Haplotype selection during culture adaptation.* Allele
734 frequency distributions corresponding to hamster-derived splenic amastigotes (sp-ama) and
735 derived promastigotes during culture adaptation at passage p2, p10 and p20 are shown for
736 selected chromosomes. All chromosomes undergoing polysomy variations are displayed in
737 Supplementary Fig. 3.
738
739 **Figure 5:** Aneuploidies and haplotype selection correlate with phenotype and fitness. This
740 figure shows the relationship between aneuploidy and haplotype selection with

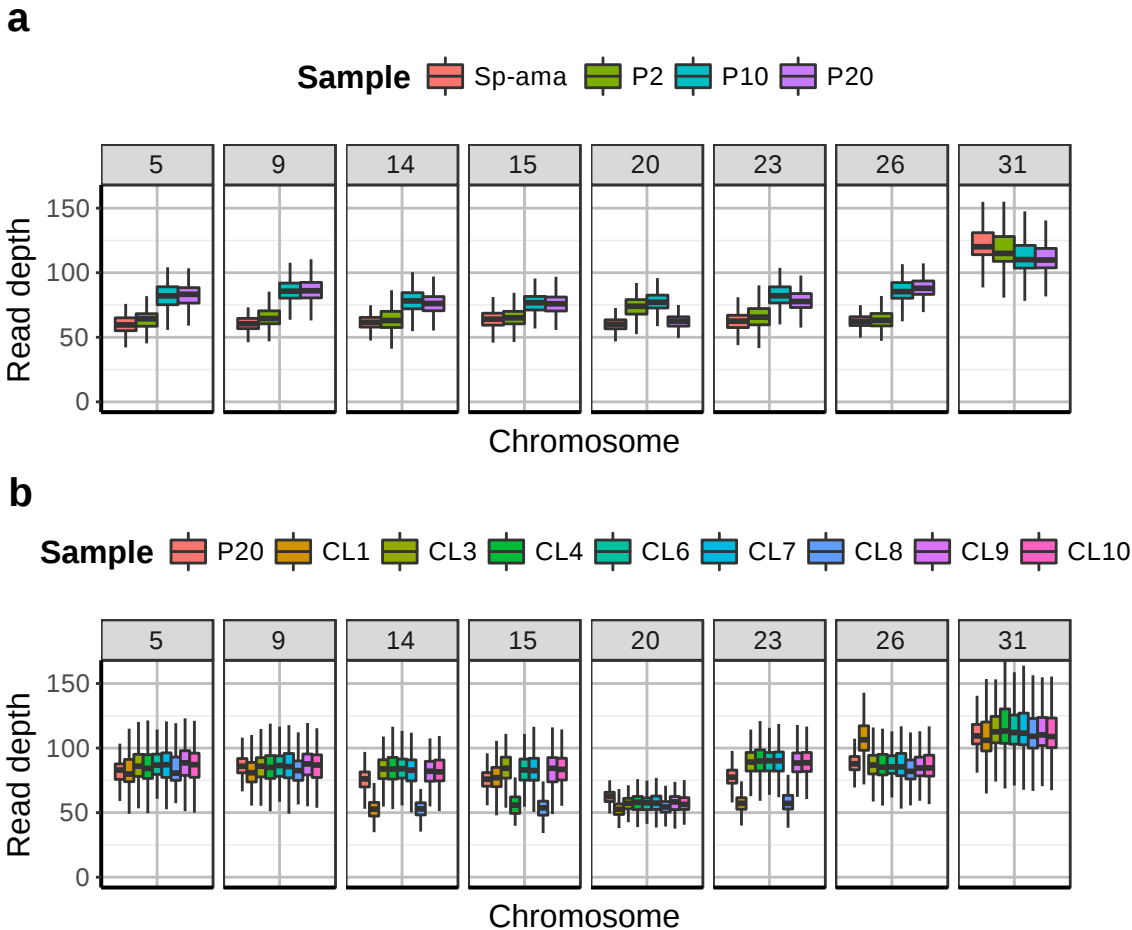
741 transcriptomic output, *in vitro* and *in vivo* growth, and tissue-specific parasite adaptation. *(a)*
742 *Correlation of aneuploidy and transcript output.* Each dot corresponds to a chromosome with
743 the X-axis reflecting the transcriptome median read-depth and the Y-axis the corresponding
744 genomic median read-depth, as estimated for the eight subclones. A subset of
745 chromosomes was colored as indicated by the legend. *(b) Parasite growth phenotype* in vitro
746 *and* in vivo*.* LD1S promastigotes during culture adaptation at passages p2, p10, and p20
747 were assessed for *in vitro* growth to determine the generation time (left panel). Three
748 independent triplicate experiments were performed and results of one representative
749 experiment with standard deviation represented by the error bars is shown. *In vivo* parasite
750 growth was assessed in infected hamster spleens and livers by limiting dilution assay
751 (middle panel). The plot represents the mean parasite burden +/- SD of three infected
752 hamsters. Pathogenicity was monitored following hamster weight as a function of time (right
753 panel) for three non-infected and three infected hamsters. *(c) Tissue-specific haplotype*
754 *selection.* Amastigotes were isolated from liver and spleen of one infected hamster, purified
755 genomic DNA was subjected to HTseq analysis, and allele profiles were established by
756 plotting allele density versus frequency.
757
758
759 **Figure 6:** Aneuploidy influence on heterozygosity. Each chromosome is represented by its
760 number on the graph and colored according to aneuploidy frequency after culture
761 adaptation, as measured in the 204 *L. donovani* field isolates. The two lines (blue and green)
762 show apparent regimens of mutation rates. The green line fits a linear model for
763 chromosomes with an aneuploidy frequency higher than 25%, while the blue line indicates
764 chromosomes that are lower in frequency. Chromosome 31 is excluded from the fitted

765 models as it stands out as the only chromosome with aneuploidy in all samples.

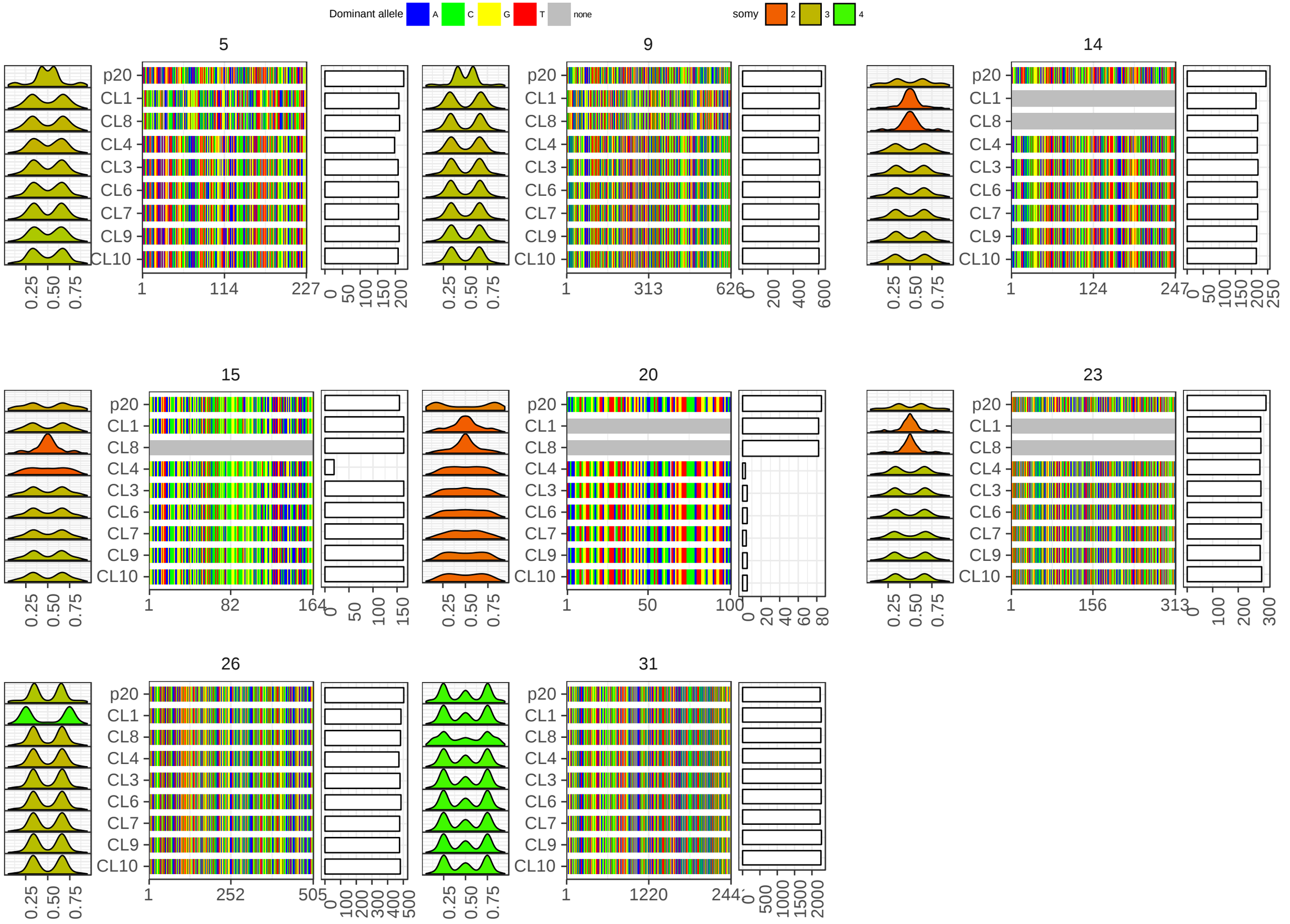
766


a
Frequency
Somy
b
10 21 17 24 25 36 19 34 31 32 27 30 9 20 5 26 11 14 22 8 23 1 6 3 28 12 13 15 2 16 7 33 18 29 4 35
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
c
disomic
trisomic
density
frequency
d
P-value = 1.51794e-08
R^2 = 0.6618
Segregation
Dispersion
Trisomy frequency
0.5
0.4
0.3
0.2
0.1


a
Polysomy level
2 3 4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
Chromosome
b
Brightfield
2 µm
FISH
Overlay
% Chromosome number
100%
50%
0%
trisomic
disomic
L S
Chr 5
Chr 17
Chr 22
Chr 27

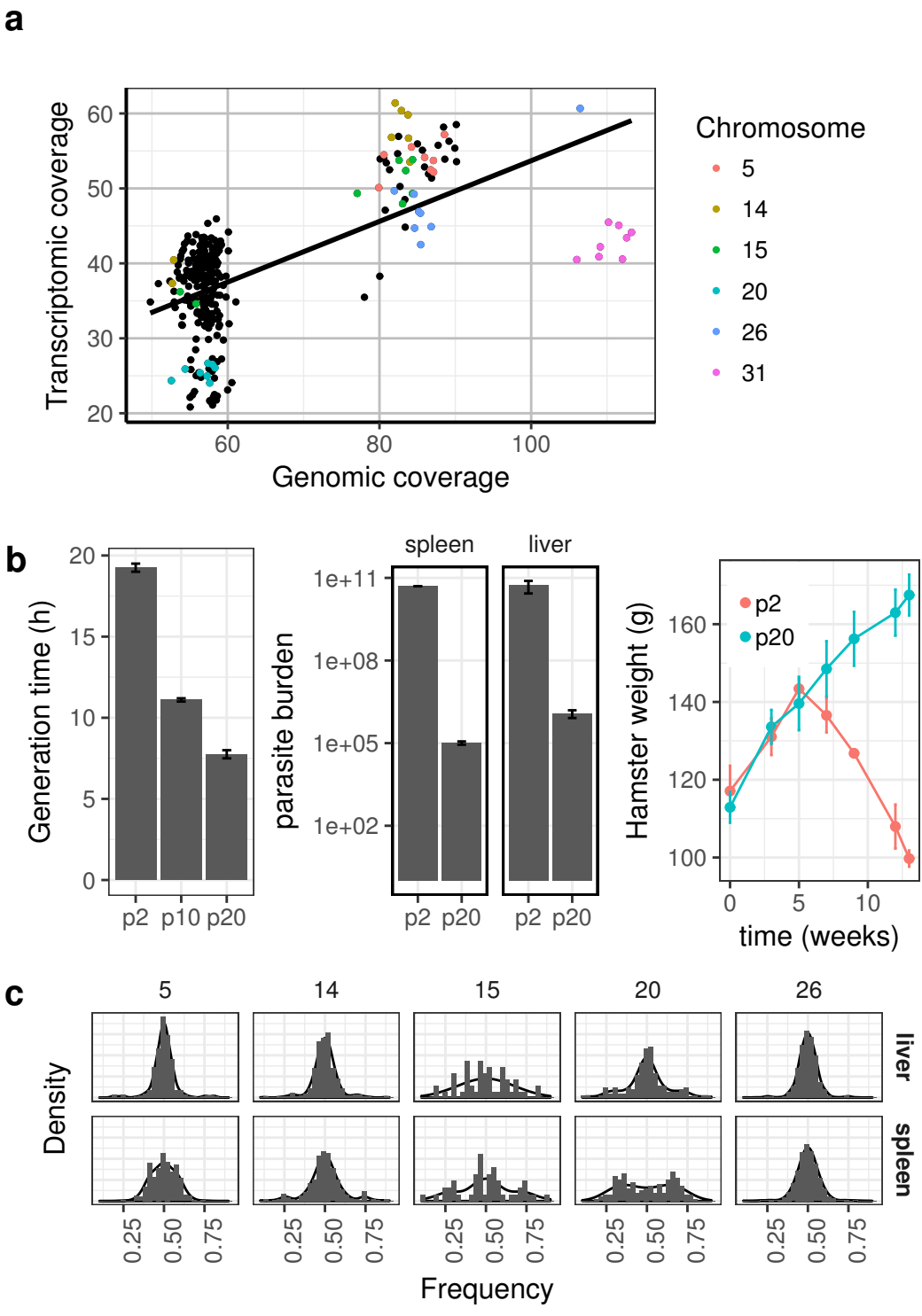

a
Sample
Sp-ama
P2
P10
P20
5
9
14
15
20
23
26
31
Read depth
150
100
50
0
Chromosome
b
Sample
P20
CL1
CL3
CL4
CL6
CL7
CL8
CL9
CL10
5
9
14
15
20
23
26
31
Read depth
150
100
50
0
Chromosome

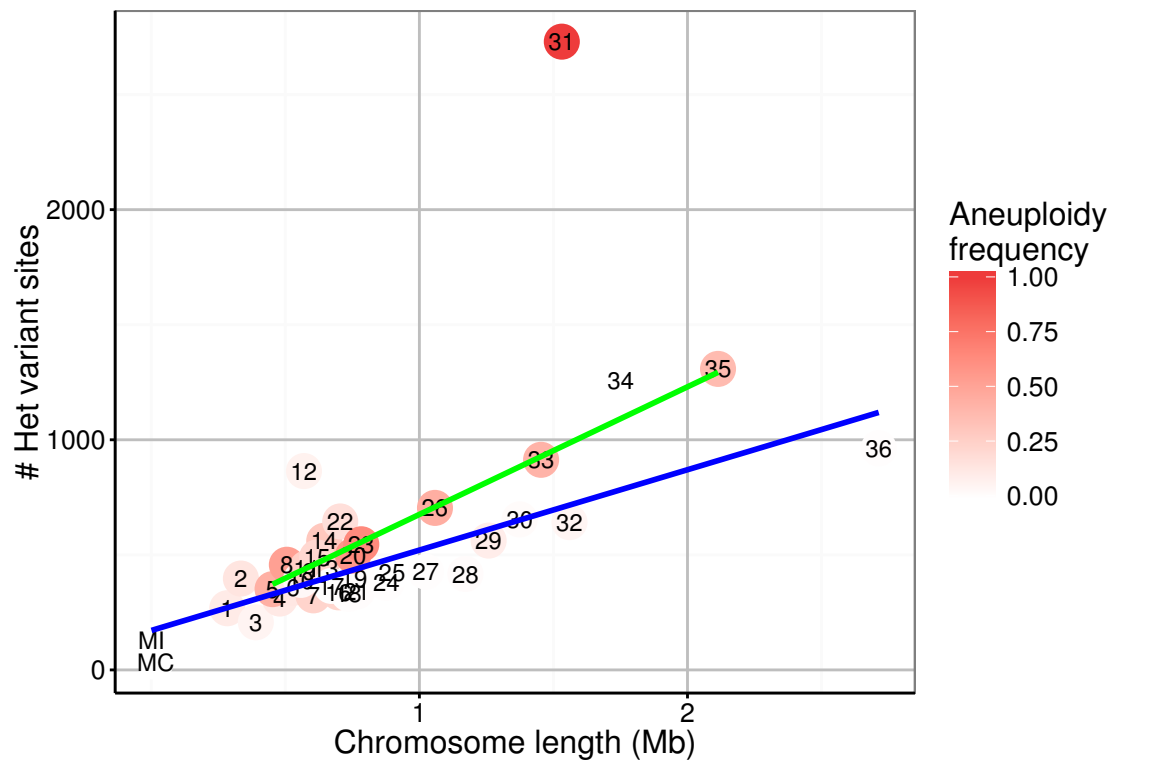
a


Dominant allele
A
C
G
T
none
somy
2
3
4
5
9
14
15
20
23
26
31
p20
CL1
CL8
CL4
CL3
CL6
CL7
CL9
CL10
0.25
0.50
0.75

b

Sample
Splenic amastigote
p2
p10
p20
5
9
14
15
20
23
26
31
Density
Frequency


a
Transcriptomic coverage
Genomic coverage
Chromosome
5
14
15
20
26
31
b
Generation time (h)
p2
p10
p20
parasite burden
spleen
liver
Hamster weight (g)
time (weeks)
c
Density
Frequency
liver
spleen


31
34
35
33
36
12
26
22
32
30
29
14
2
27
28
MI
MC
Het variant sites
2000
1000
0
1
2
Chromosome length (Mb)
Aneuploidy frequency
1.00
0.75
0.50
0.25
0.00