# Exploring the evolution and adaptive role of mosaic aneuploidy in a clonal *Leishmania donovani* population using high throughput single cell genome sequencing

Gabriel H. Negreira[1], Pieter Monsieurs[1], Hideo Imamura[1], Ilse Maes[1], Nada Kuk[2], Akila Yagoubat[2], Frederik Van den Broeck[1], Yvon Sterkers[2], Jean-Claude Dujardin[1,3], Malgorzata A. Domagalska[1]

Authors Affiliation:

[1] Institute of Tropical Medicine Antwerp, Molecular Parasitology Unit, Antwerp, Belgium

[2] Université Montpellier, UFR Médecine, Laboratoire de Parasitologie-Mycologie, Montpellier, France

[3] University of Antwerp, Department of Biomedical Sciences, Antwerp, Belgium.

Keywords: Leishmania, single cell genomic sequencing, mosaicism, aneuploidy

17

**Update: 25th May 2021**

A revision of the present preprint was released in BioRxiv on 11th May 2021 (https://www.biorxiv.org/content/10.1101/2021.05.11.443577v2). In the new version, we included two extra samples in our single-cell genome sequencing (SCGS) analysis – the BPK081 cl8 clone (a nearly euploid strain), and a population consisting of a mixture of four *L. donovani* strains which was used as control for high levels of mosaicism in aneuploidy and for estimation of doublets. We also upgraded the bioinformatics pipeline to determine single-cell karyotypes and performed new fluorescence in situ hybridization (FISH) analysis. The new findings observed especially in the BPK081 cl8 led to a reformulation of the text, a new hypothesis for the evolution of mosaicism and a general restructuring of the article. Therefore, the present preprint is obsolete. Please refer to the new preprint entitled "High throughput single cell genome sequencing gives insights in the generation and evolution of mosaic aneuploidy in *Leishmania donovani*" for more information.

**Abstract**

Maintenance of stable ploidy over continuous mitotic events is a paradigm for most higher eukaryotes. Defects in chromosome segregation and/or replication can lead to aneuploidy, a condition often considered deleterious. However, in *Leishmania,* a Protozoan parasite, aneuploidy is a constitutive feature, where variations of somies represent a mechanism of gene expression adaptation, possibly impacting phenotypes. Strikingly, clonal *Leishmania* populations display cell-to-cell somy variation, a phenomenon named mosaic aneuploidy (MA). However, until recently, no method was available for the determination of the complete karyotype of single *Leishmania* parasites. To overcome this limitation, we used here for the first time a high-throughput single-cell genomic sequencing (SCGS) method to estimate individual karyotypes of 1560 promastigote cells in a clonal population of *Leishmania donovani*. We identified 128 different karyotypes, of which 4 were dominant. A network analysis revealed that most karyotypes are linked to each other by changes in copy number of a single chromosome and allowed us to propose a hypothesis of MA evolution. Moreover, aneuploidy patterns that were previously described by Bulk Genome Sequencing as emerging during first contact of promastigotes populations with different drugs are already pre-existing in single karyotypes in the SCGS data, suggesting a (pre-)adaptive role of MA. Additionally, the degree of somy variation was chromosome-specific. The SCGS also revealed a small fraction of cells where one or more chromosomes were nullisomic. Together, these results demonstrate the power of SCGS to resolve sub-clonal karyotype heterogeneity in *Leishmania* and pave the way for understanding the role of MA in these parasites' adaptability.

40 **Introduction**

41 Historically, cell populations were analyzed in bulk assuming that this provides the
42 representative information on the biology and behavior of these cells in a given experimental
43 set up. The existence of heterogeneity between single cells, even in clonal populations, has
44 been acknowledged but for long it was not further explored due to lack of suitable
45 methodologies, and also because of the assumption that this cell-to-cell variation is random
46 and has no biological significance. In the last years, advances in single cell imaging, molecular
47 biology and systems biology enabled analysis and quantification of differences between single
48 cells, defining new cell types and their origin, and finally allowing the demonstration of
49 functional relevance of this variation[1]. This cell-to-cell variability of clonal populations is also of
50 essential importance for unicellular organisms, like *Staphylococcus aureus*[2] , or *Sacharomyces*
51 *cerevisiae*[3]. Similarly, for digenetic protozoan parasites, such as *Plasmodium spp*. or
52 *Trypanosoma cruzi*, cellular mosaicism was also demonstrated to be crucial for drug resistance
53 or tissue tropism during the vertebrate host infection[4].

54 *Leishmania* sp. are digenetic unicellular protozoan parasites responsible for a spectrum
55 of clinical forms of leishmaniasis worldwide and causing 0.7-1 million new cases per year[5]. As
56 other trypanosomatids, *Leishmania* belongs to the supergroup Excavata, one of the earliest
57 diverging branch in the Eukaryota domain[6]. Thus, several molecular mechanisms considered
58 canonical for eukaryotes are different in these parasites, including the genomic organization in
59 long polycistronic units, the near absence of transcription initiation regulation by RNA
60 polymerase II promoters with gene expression regulation happening mostly through post-
61 transcriptional mechanisms[7], and its remarkable genomic plasticity[8]. The genome of *Leishmania*
62 sp is ubiquitously aneuploid, and high levels of chromosomal somy variation, as well as local
63 gene copy number variation are found between all *Leishmania* species[9,10]. Moreover, these
64 variations are highly dynamic and change in response to new environment, such as drug
65 pressure, vertebrate host or insect vector[11]. Changes in somy, together with episomal gene
66 amplifications, and not variation in nucleotide sequence are the first genomic modifications
67 observed at populational level during the course of experimental selection of drug resistance,
68 suggesting that they are adaptive[12,13]. Importantly, somy alterations are reflected in the level
69 of transcriptome[11], and -for a same life stage- to a great degree of proteome derived from genes
70 located in polysomic chromosomes[14]. These findings further support the notion that *Leishmania*
71 exploits aneuploidy to alter gene dosage and to adapt to different environment.

72 Cell-to-cell variation in *Leishmania* has been so far reported in the form of mosaic
73 aneuploidy (MA). This phenomenon was demonstrated in several *Leishmania* species by means
74 of fluorescence in situ hybridization (FISH) performed on a few chromosomes, which showed
75 at least two somy states among individual cells[15]. As a consequence, thousands of different
76 karyotypes are expected to co-exist in a parasite population[16]. This heterogeneity of aneuploidy
77 provides a huge potential to generate diversity in *Leishmania* from a single parental cell, both
78 quantitatively through gene dosage, but also qualitatively through changes in heterozygosity[17].
79 Thus, it is hypothesized that mosaic aneuploidy constitutes a unique source of adaptability to
80 new environment for the whole population of parasites[16].

81 Although MA has been demonstrated by FISH, its extent to all over the 35-36
82 chromosomes of *Leishmania*, its dynamics in constant as well as new environment and its
83 potential role in adaptation to different environment remains to be determined. Accordingly,
84 pioneer FISH-based studies should be complemented and refined by single cell genome
85 sequencing (SCGS). In a previous study[18], we made a first step in that direction by combining
86 FACS-based sorting of single *Leishmania* cells with whole genome amplification (WGA) and
87 whole genomic sequencing (WGS). In this pilot study, we evaluated different WGA and
88 bioinformatic methods and detected 3 different karyotypes among 28 single cells of *L.*
89 *braziliensis*[18]. Here, we made one step beyond, applying a droplet-based platform for single cell
90 genomics, in order to undertake the first high-throughput study of MA in *Leishmania*.

91 **Material and Methods**

92 **Parasites**

93 *L. donovani* promastigotes of the strain MHOM/NP/03/BPK282/0 clone 4 (further called
94 BPK282, reference genome of *L. donovani*) were maintained at 26˚C on HOMEM medium
95 (Gibco, ThermoFisher) supplemented with 20% Fetal Bovine Serum, with regular passages done
96 every 7 days. The strain was submitted to SCGS 21 passages after cell cloning and analyzed by
97 FISH 23 passages later.

98 **Cell Suspension Preparation for SCGS**

99 BPK282 promastigotes at early stationary phase (day 5) were harvested by centrifugation
100 at 1000rcf for 5 minutes, washed twice with PBS1X (without calcium and magnesium) + 0,04%
101 BSA, diluted to $5x10^6$ parasites/mL and passed through a 5µm strainer to remove clumps of
102 cells. After straining, volume was adjusted with PBS1X + 0,04% BSA to achieve a final
103 concentration of $3x10^6$ parasites/mL. The absence of remaining cell duplets or clumps in the
104 cell suspension was confirmed by microscopy.

105 **Single Cell partitioning, barcoding, WGA and sequencing**

106 We used 4.2µL of the single-cell suspension as input to the 10X Chromium™ Single Cell CNV
107 Solution protocol (10X Genomics), targeting a total of 2000 sequenced cells according to the
108 manufacturer's instructions. Individual cells were portioned and incapsulated in a hydrogel
109 matrix using a microfluidic chip and the Chromium Controller (10X Genomics). During cell
110 partitioning, each cell was combined with the Cell Bead (CB) Polymer™ and the Cell Matrix™
111 reagents, forming the CBs. CBs were then incubated overnight at 21˚C and 1000RPM on a
112 thermomixer for homogeneous hardening of the CB polymer. The individually incapsulated cells
113 were lysed, releasing the genomic DNA (gDNA) inside the CB. In a second microfluidic chip, each
114 CB was individually joined with a Gel Bead (GB) containing multiple copies of one of the
115 ~750.000 unique 10X barcodes™, together with an enzyme mix used in downstream Whole
116 Genome Amplification (WGA) and ligation of 10X barcodes. The CB-GB joining was performed
117 with a partitioning oil, forming an emulsion where each droplet (GEMs) contain a single CB
118 linked to a unique GB. Inside the GEMs, the CB and GB were disrupted, and isothermal WGA
119 with random hexamers followed by ligation of the 10X barcode to the amplified gDNA
120 molecules was carried out. Then, GEMs were disrupted, amplified DNA molecules were pooled
121 and processed for Illumina sequencing with the addition of P5 and P7 adaptors and a sample

index. Sequencing of the library was performed at Genomics Core Leuven (Belgium), with a NextSeq High Output kit (Illumina) platform with 2 x 150 read length.

**CNV Calling and Somy Estimation**

Reads were associated to each sequenced cell based on their 10X-barcode sequence and mapped to a customized version of the reference *L. donovani* genome LdBPKv2[11] (available at ftp://ftp.sanger.ac.uk/pub/project/pathogens/Leishmania/donovani/LdBPKPAC2016beta/), where Ns were added to the ends of chromosomes 1 to 5 to reach the 500kb minimum size allowed by the Cell Ranger DNA™ pipeline (10X Genomics). The normalized read depth (NRD) within adjacent, non-overlapping 80kb bins was used to calculate copy number values (referred here as CNV values) in 80kb intervals, using the version 1.1 of the Cell Ranger DNA™ pipeline[19]. CNV values represent the integer NRD of each bin after scaling of NRD values by the baseline ploidy factor (S factor) of each cell determined by the scaling algorithm of the pipeline[20]. We used the arguments min-soft-avg-ploidy and max-soft-avg-ploidy set to 1.9 and 2.1, respectively, to encourage the scaling algorithm to establish cells baseline ploidy to 2. This was done to avoid overscaling artifacts generated by the scaling algorithm that were observed in some cells when these options were not used. CNV values were visualized in the Loupe scDNA Browser (10X Genomics)[21], with cells arranged in 512 clusters, the maximum number of clusters allowed by the software. These clusters were composed of 1 to 7 cells each that were grouped together based on CNV values similarities[22]. CNV values of each 80kb bin of all 512 clusters were exported as a csv file and the average CNV values of intrachromosomic bins were used to estimate chromosomal somy. Bins mapping to high copy number loci such as the H-locus (chromosome 23) and M-locus (chromosome 36)[23] were excluded from the calculation. Bin-to-bin variability was estimated by the normalized standard deviation of CNV values of intrachromosomal bins. Each cell cluster received a bin variability score based on the average bin-to-bin variability of all chromosomes. Clusters displaying a bin variability score higher than 0.05 were excluded. Remaining clusters were ungrouped, with somy values of each individual cell corresponding to the somy values estimated for their original clusters.

**Karyotypes identification and network analysis**

In order to identify different karyotypes present in the sequenced population, somy values of each cell were rounded to their closest integer. Cells displaying the same rounded somy values for all chromosomes were considered as having the same karyotype. Each unique

153 karyotype identified in the population received an identifier composed by the concatenated
154 rounded somy values of all chromosomes, from chromosome 1 to chromosome 36 (e.g.,
155 "222232223222222322222232232222423232"). Unique karyotypes were numerically named
156 according to their frequency in the sequenced population.

157 To perform a network analysis, a pairwise distance matrix was built based on the number
158 of different chromosomes between all karyotype and used to generate a randomized minimum
159 spanning tree with 100 randomizations, using the Pegas R package[24,25].

160 **DNA probes and fluorescence in situ hybridization (FISH)**

161 DNA probes were either cosmid (L549 specific of chromosome 1) or BAC (LB00822 and
162 LB00273 for chromosomes 5 and 22 respectively) clones that were kindly provided by Peter
163 Myler (Seattle Biomedical Research Institute) and Christiane Hertz-Fowler (Sanger Centre). DNA
164 was prepared using Qiagen Large-Construct Kit and labelled with tetramethyl-rhodamine-5-
165 dUTP (Roche Applied Sciences) by using the Nick Translation Mix (Roche Applied Sciences)
166 according to manufacturer instructions. Cells were fixed in 4% paraformaldehyde then air-dried
167 on microscope immunofluorescence slides, dehydrated in serial ethanol baths (50–100%) and
168 incubated in NP40 0,1 % for 5 min at RT. Around 100 ng of labelled DNA probe was diluted in
169 hybridization solution containing 50% formamide, 10% dextran sulfate, 2× SSPE, 250 $\mu g.mL^{-1}$
170 salmon sperm DNA. Slides were hybridized with a heat-denatured DNA probe under a sealed
171 rubber frame at 94°C for 2 min and then overnight at 37°C and sequentially washed in 50%
172 formamide/2× SSC at 37°C for 30 min, 2× SSC at 50°C for 10 min, 2× SSC at 60°C for 10 min, 4×
173 SSC at room temperature. Finally, slides were mounted in Vectashield (Vector Laboratories)
174 with DAPI. Fluorescence was visualized using appropriate filters on a Zeiss Axioplan 2
175 microscope with a 100× objective. Digital images were captured using a Photometrics CoolSnap
176 CCD camera (Roper Scientific) and processed with MetaView (Universal Imaging). Z-Stack image
177 acquisitions (15 planes of 0.25µm) were systematically performed for each cell analyzed using
178 a Piezo controller, allowing to view the nucleus in all planes and to count the total number of
179 labelled chromosomes. Around 200 cells [187-228] were analyzed per chromosome.

180 **Results**

181 In this study, we submitted the clonal population of BPK282 promastigotes to a high
182 throughput, droplet-based SCGS method (Chromium™ CNV Solution - 10X Genomics). In total,
183 148,7M reads were generated, of which 49,7M were mapped to 1703 cells after removal of
184 duplicated reads. The average effective coverage depth per cell was 0,14x, with only 2 cells
185 displaying a depth coverage higher than 1x (**Figure S1A)**. However, read mapping was evenly
186 distributed across chromosomes in general(**Figure S1B**), which allows somy estimation despite
187 low depth[18]. From 512 clusters, 452 passed the bin-to-bin variability filtering, representing a
188 total of 1560 cells. Most eliminated clusters displayed an average somy close to 1 (**Figure S1C**).

189 **Copy number variation of individual chromosomes**

190 The somies of the 36 *L. donovani* chromosomes in the 1560 filtered cells are depicted in
191 **Figure 1A**. Of these 36 chromosomes, 7 were predominantly trisomic (chromosomes 5, 9, 16,
192 23, 26, 33 and 35), while chromosome 31, the only chromosome that is extensively reported as
193 polysomic (often tetrasomic) in all *Leishmania* species studied so far[9,26], was also tetrasomic in
194 most cells in our SCGS data. The other chromosomes were mostly disomic. These observations
195 are very similar to the populational aneuploidy pattern estimated by BGS of the BPK282
196 reference strain (**Figure 1A, top heatmap, and Figure S2**). Notably, chromosome 13, which
197 display an intermediate somy value in the Bulk Genomic Sequence (BGS) data, was found as
198 disomic and trisomic at high proportions in the SCGS. Moreover, our SCGS data revealed that
199 high cell-to-cell somy variation was chromosome-specic. For instance, chromosomes 21 and 18
200 were disomic in 99,23% and 99,42% of the cells respectively, while chromosomes 13, and 35
201 were the most variable ones, with 33,54% and 25,24% of the cells displaying a copy number
202 divergent from the most frequent somy.

203 In the filtered data, we identified a cell where 4 chromosomes were estimated as nullisomic
204 (**Supplementary table 1 - kar89)**. The bam file of this cell shows that these chromosomes had
205 no sequenced reads mapping to them (**Figure 2A**, first plot), indicating that they were absent in
206 this cell. An investigation of the bam files of all cells, including the ones eliminated after data
207 filtering, revealed a total of 12 cells displaying at least one chromosome with no reads mapping
208 to it, of which 11 displayed a high intrachromosomal bin-to-bin variability of the mapped
209 chromosomes and were excluded during data filtering.

210 **Karyotype mosaicism**

211 A karyotype was defined as any unique combination of rounded somy values for all
212 chromosomes identified in the sequenced cells after data filtering. According to this definition,
213 128 different karyotypes were identified amid 1560 cells. The most frequent karyotype (kar1)
214 was found in 431 cells, representing 27,65% of the sequenced population; this karyotype is
215 equivalent to the average karyotype observed by BGS of BPK282. Four karyotypes are
216 dominant in the population, representing 62,5% of cells (**Figure 3A**). Noteworthy, the most
217 dominant karyotypes displayed little somy difference between each other. For instance, kar2,
218 3 and 4 differ from kar1 by changes in somy in a single chromosome, i.e. chromosomes 13, 35
219 and 5 respectively, while kar5 diverge from kar2 by a disomy in chromosome 35 (**Figure 3B**).
220 A network analysis showed that most karyotypes diverge from another by a difference in a
221 single chromosome, displaying an interesting pattern where most karyotypes can be linked
222 back to kar1 by single chromosome-change steps (**Figure 4**, black pies). Noteworthy, kar1 also
223 displays the biggest number of secondary karyotypes directly linked to it by somy changes in
224 single chromosomes, including the other 3 dominant karyotypes. Karyotypes which the closest
225 relatives were at two or more chromosome changes apart (red pies) were only present at
226 frequencies lower than 0,44%. Moreover, the only karyotype with nullisomic chromosomes
227 that passed the data filtering criteria (kar89) display several chromosomes with somies
228 different from most karyotypes, with the closest one (kar32) being 18 chromosomes somy
229 changes apart from it.

230 **Pre-existing karyotypes selected during early environmental adaptation.**

231 Previous studies have demonstrated through BGS that *in vitro Leishmania* populations
232 display changes in average aneuploidy patterns of the cell population as one of the first genomic
233 alterations when submitted to environmental changes, including hosts infection and drug
234 pressure[11–13,27]. However, it is still unknown if these variations in somies observed at
235 populational level are generated by selection of karyotypes pre-existing in subpopulations or
236 represent *de novo* alterations induced by the new environmental pressure. To address this
237 issue, we revisited previously published BGS data where the same BPK282 clonal promastigote
238 population was used as model in drug selection experiments. We evaluated in our SCGS data if
239 changes in aneuploidy patterns that emerge during adaptation to different drug pressures at
240 populational level are already present in single karyotypes in the BPK282 population.

241 In the 3 revisited drug resistance selection studies, the same pattern was observed. First
242 rounds of selection led to changes in the average populational aneuploidy profile that were
243 equivalent to single karyotypes present in our SCGS data (**Figure 5**). Authors[13] reported a
244 reduction in the somy of chromosomes 13 and 35 in a BPK282 population exposed to 3µM and
245 6µM of Miltefosine. The observed populational aneuploidy pattern is similar to kar9, found in
246 24 cells in our SCGS data, where chromosomes 13 and 35 are disomic (**Figure 5A**). A similar
247 observation happened when BPK282 promastigotes were exposed to 2µM and 4µM of
248 Paramomycin in another study[27], where the somy changes are also similar to kar9 (**Figure 5B**).
249 Finally, *in vitro* SbIII selection experiments with BPK282 led to changes in populational average
250 somies which were similar to karyotypes 3, 4, or 6 among 3 replicates exposed to 48µM of the
251 drug [12] (**Figure 5C**). However, in these 3 studies, higher drug concentrations led to changes in
252 the average populational somies that are not represented by single karyotypes in our data.

253 **Validation of the SCGS results with FISH**

254 As a validation to our SCGS data, we also evaluated the somy of 3 chromosomes in 187-228
255 BPK282 cells using FISH. We observed very similar somies distributions in both methods (**Figure**
256 **6**). For instance, chromosome 5, the most variable of the 3 assessed chromosomes, was
257 estimated as trisomic in 78,5% of the cells in the SCGS data, and 74,5% by FISH, while it was
258 disomic in 20,9% in SCGS and 22,6% in FISH. Chromosome 1 was estimated as disomic in 94,8%
259 of the cells in the SCGS and 96,2% in FISH, with a smaller fraction of cells being monosomic
260 (5,1% in SCGS and 3,2% in FISH). Chromosome 22 had slightly more divergent values between
261 both methods, being disomic in 99,5% in SCGS and 92,1% in FISH, with a fraction of 7,9% of cells
262 that were found as monosomic in FISH while monosomy for chromosome 21 was found in only
263 0,19% of the cells in SCGS. In general, the similarities between the distributions found in both
264 methods supported the reliability of SCGS.

265 **Discussion**

266 The parallel whole genomic sequencing of 1560 *Leishmania* promastigotes reported here
267 allowed us to reveal for the first time a complex mosaicism of complete karyotypes in
268 *Leishmania* with unprecedent resolution. By estimating the somy of all 36 chromosomes in each
269 sequenced single parasite, we could determine which chromosomes were more prone to cell-
270 to-cell somy variability, which were the frequencies of each identified karyotype in the clonal
271 population and the divergencies between these karyotypes. Our data was further supported by
272 FISH, which is a well establish method for the quantification of individual chromosome copy
273 number in single *Leishmania* cells[28]. Thus, this work represents the first high throughput SCGS
274 study of MA in *Leishmania*.

275 **Performance and challenges of the SCGS method**

276 In general, the whole genome amplification (WGA) method used in the Chromium™ Single
277 Cell CNV solution seems to generate an even genome coverage, which has been demonstrated
278 as a more relevant aspect than coverage depth for accurate somy estimation[18,29]. However, a
279 low percentage of cells displaying a higher read count variability might reflect on inaccurate
280 CNV values, and therefore, unreliable karyotype determination. The Cell Ranger DNA™ pipeline
281 uses an hierarchical clustering algorithm to combine the read counts of cells with similar CNV
282 values to enhance the resolution and reliability of CNV values determination[22], thus greatly
283 reducing the number of faulty karyotypes. Yet, cells with slightly different karyotypes might be
284 clustered together due to inaccurate single-cell CNV values estimation, leading to a cluster somy
285 pattern that does not reflect a true biological karyotype. Since we noticed that artificial
286 karyotypes were caused by chromosomes with high variation in intrachromosomal CNV values,
287 we applied an addition data filtering step where we eliminated clusters displaying a high
288 intrachromosomic bin-to-bin variability. This led to the removal of 56 clusters (143 cells),
289 corresponding to 51 karyotypes. Most of the eliminated karyotype displayed somy patterns that
290 were highly distinct from the ones found in the remaining karyotypes after data filtering,
291 probably due to inaccurate somy estimation caused by the high intrachromosomic bin-to-bin
292 variability.

293 Most eliminated clusters had the majority of CNV values assigned to 1. Since CNVs are
294 calculated by the relative read depth of the bins, the Cell Ranger DNA™ pipeline uses an
295 algorithm that determines the baseline ploidy of each cell by scaling the normalized read depth

296 values of each bin by a factor S that satisfies the condition where all normalized read depth
297 values are integers multiple of S[20]. Candidate values for S are heuristically determined following
298 a rule where if the cell displays the same copy number in most of the genome, S is set to a value
299 that leads to an average ploidy close to 2, otherwise, in the presence of high variable CNV
300 values, S is determined as the lowest value that satisfies the condition mentioned above. Thus,
301 it is likely that the high variation of local CNV values present in these clusters, reflecting in a
302 high intrachromosomic bin-to-bin variability, led to an unprecise determination of the S factor
303 by the scaling algorithm, leading to an average ploidy in these clusters close 1. Indeed, the
304 software cannot identify haploid cells since it is based on a normalized read depth quantification
305 method. Haploid cells could be identified by the evaluation of heterozygous SNPs in single cells.
306 However, the read depth per cell of our data does not allow such analysis.

307 The Cell Ranger DNA™ pipeline was developed to handle mammal genomes[30]. The nuclear
308 genome of *Leishmania* is about 100 times smaller than human's, and even the biggest *L.*
309 *donovani* chromosome (Chr36, 2,768Mbp[31]) is still 17 times smaller than the smallest human
310 chromosome (Chr21, 46,7Mbp). With the 80kb bin size determined by the pipeline in this data,
311 small chromosomes such as Chr1, Chr2 and Chr3 are represented by 3 to 5 bins. However, this
312 does not seem to introduce a variability bias to small chromosomes in general, since the
313 observed cell-to-cell somy variability does not correlate to chromosome size. For instance, Chr4,
314 Chr5 and Chr6 are all represented by 6 bins, and while Chr5 was found as triploid and diploid in
315 high proportions for both states, Chr4 and Chr6 were among the less variable chromosomes
316 (**Figure 1B**). Conversely, Chr34 and Chr35 had 24 and 27 bins respectively, and while the former
317 was found as disomic in more than 95% of the cells, the later was estimated as disomic in 26,9%
318 and trisomic in 71,4% of the cells. There is the possibility, however, that the very low number
319 of bins for chromosome 1 (3 bins) introduced a bias towards monosomy.

320 On the other hand, the reduced size of *Leishmania* genome means that higher coverage
321 depth per cell are achieved with relatively lower total sequence depth. The average 29.191
322 mapped deduplicated reads per cell resulted in an average coverage depth/cell of 0,14x. For
323 human cells, the recommended 750.000 reads/cell is expected to yield a coverage depth of
324 ~0,05x/cell, allowing CNVs assessment at 2mb resolution only[32]. This means that for each bin,
325 a higher number of reads are expected in *Leishmania*, increasing CNV estimation resolution.
326 Therefore, the higher coverage depth/cell seem to compensate the lower number of bins per
327 chromosome in *Leishmania* genome.

328 **SCGS data supports errors in chromosome replication as main drivers of mosaicism**

329 In a previous FISH-based study, authors reported asymmetrical chromosome allotments
330 (ACA) in dividing nuclei as the cause of MA[15]. Remarkably, in all cells displaying ACA, the total
331 copy number of the evaluated chromosome in the daughter cells were always odd ("3+2" or
332 "2+1"). This observation suggests that mosaicism in aneuploidy is generated by defects in
333 replication of chromosomes during cell division, as discussed by the authors, and not by errors
334 in chromosome segregation, where the expected number of copies between daughter cells
335 would be even. In our SCGS data, a pairwise comparison between all karyotypes that diverged
336 from another by a single chromosome revealed that complementary karyotypes that could
337 represent mis-segregation event are very rare. Of 127 possible pairs, only 7 pairs could be
338 explained by a mis-segregation event (**Figure 7 and Supplementary table 2**). The other 120 pairs
339 displayed a "2+1", "2+3", "3+4" or "4+5" pattern in the divergent chromosome, supporting the
340 hypothesis that such karyotypes were probably generated by under- or over-replication of
341 single chromosomes during mitosis.

342 One of the effects of ACA events is the possibility of some cells losing all copies of a
343 chromosome (nullisomy). Among all 1703 cells sequenced, including the ones that were
344 removed for further analysis, 12 had one or more chromosome which was nullisomic. This
345 represents ~0,7% of the total, which is a relatively high frequency considering that an *in vitro*
346 *Leishmania* culture can consist of millions of cells. The fact that almost all cells with an absent
347 chromosome displayed a karyotype that was not similar to any other in the sequenced
348 population (**Figure S4**) is possibly due to the high bin-to-bin variability found in these cells,
349 which hampers CNV estimation. The noisy coverage of these cells could be an indicative of DNA
350 degradation after cell death, potentially caused by the lack of one or more chromosomes,
351 suggesting that nullisomy is lethal for these parasites. In this case, the relatively high rate of
352 cells with absent chromosomes is an indicative that ACA events that lead to nullisomy happen
353 at high frequencies. Indeed, the FISH analysis of dividing nuclei found that for 2 chromosomes
354 (chromosomes 2 and 22), around 1% of the evaluated parasites were displaying a "1+0"
355 distribution of chromosomes between sister nuclei in *L. major*[15].

356 **Chromosome Somy Variability**

357 Although pioneer FISH studies demonstrated that somy mosaicism was more prominent in
358 some chromosomes than others, the complete somy landscape of all chromosomes in a

359 *Leishmania* population was still unknown at single-cell level, since only 11 chromosomes were
360 evaluated by FISH hitherto. Our SCGS data revealed that some chromosomes display a
361 remarkable lack of cell-to-cell variability in the sequenced population, while others are more
362 prone to mosaicism. Several of the less variable chromosomes also display low inter-strain
363 variation between *L. donovani* isolates. For instance, chromosomes 10, 17, 18, 19, 21, 24, 25,
364 27, 28, 30, 34 and 36 were previously reported as disomic by BGS in more than 95% of 204
365 isolates in a previous study[33], and were disomic in more than 98% of the cells in our SCGS data,
366 suggesting a pressure to maintain these chromosomes as disomic, at least under standard *in*
367 *vitro* conditions. Conversely, chromosomes 5, 13, 33 and 35, the most variable in the SCGS, are
368 also present as disomic or trisomic at high proportions between these different isolates. This
369 indicates that, in a given environment, somy variability is restricted to a specific group of
370 chromosomes. One possibility is that gene contents of some chromosomes are more tolerant
371 to dosage imbalance than others, suggesting that selective pressure maintain somy stability in
372 some chromosomes while allow more plasticity in others. Moreover, the WGS of 204 *L.*
373 *donovani* isolates revealed that only chromosome 34 was consistently disomic in all strains[33].
374 However, BPK282/0 cl4 parasites displays a trisomy for this chromosome when exposed to high
375 SbIII concentrations[12], suggesting that every chromosome has the potential to became
376 polysomic depending on environmental pressures. The fact that, in our data, the karyotypes
377 that dominate the population are similar to each other while karyotypes with several somy
378 changes happen at low frequencies also suggests that selective pressure play a role in somy
379 variability.

380 **Hypothesis on the evolution of mosaic aneuploidy in the clonal BPK282 population**

381 By performing a pairwise comparison of the identified karyotypes in BPK282, we revealed a
382 network structure where most karyotypes are linked to each other by single somy changes. This
383 network allows proposing a hypothesis for the evolution of mosaicism during the 20 passages
384 that followed cellular cloning. Kar1, the most frequent in the sequenced cells, is also the one
385 displaying the highest number of karyotypes directly linked to it (**Figure 4**). Kar1 shows the same
386 aneuploidy pattern as the 'average' karyotype assessed by BGS in the uncloned BPK282/0 line[11],
387 suggesting that this was also the dominant karyotype in the parental population. The high
388 frequency of this karyotype in the parental population naturally increases the chance of it being
389 randomly picked when starting a clonal population. Altogether, these data suggest that kar1
390 was the potentially founder karyotype of the BPK282 clonal population derived from this

391 uncloned BPK282/0 line. Kar1 is likely well adapted to *in vitro* condition and it further spread
392 during clonal propagation. According to the network, kar1 would have generated -through
393 changes in somy of single chromosomes- a series of 'primary' derived karyotypes, (i) some
394 diverging early and being quite successful like kar2-4 and (ii) others having diverged later and/or
395 being less fit like the 19 minor karyotypes present around kar1. In this sense, there would be
396 secondary, tertiary waves of karyotype diversification, like kar5 emerging from kar2 and kar15
397 emerging from kar5. Time lapse single cell sequencing would be required to test this
398 hypothesis.

399 **Drug pressure and aneuploidy**

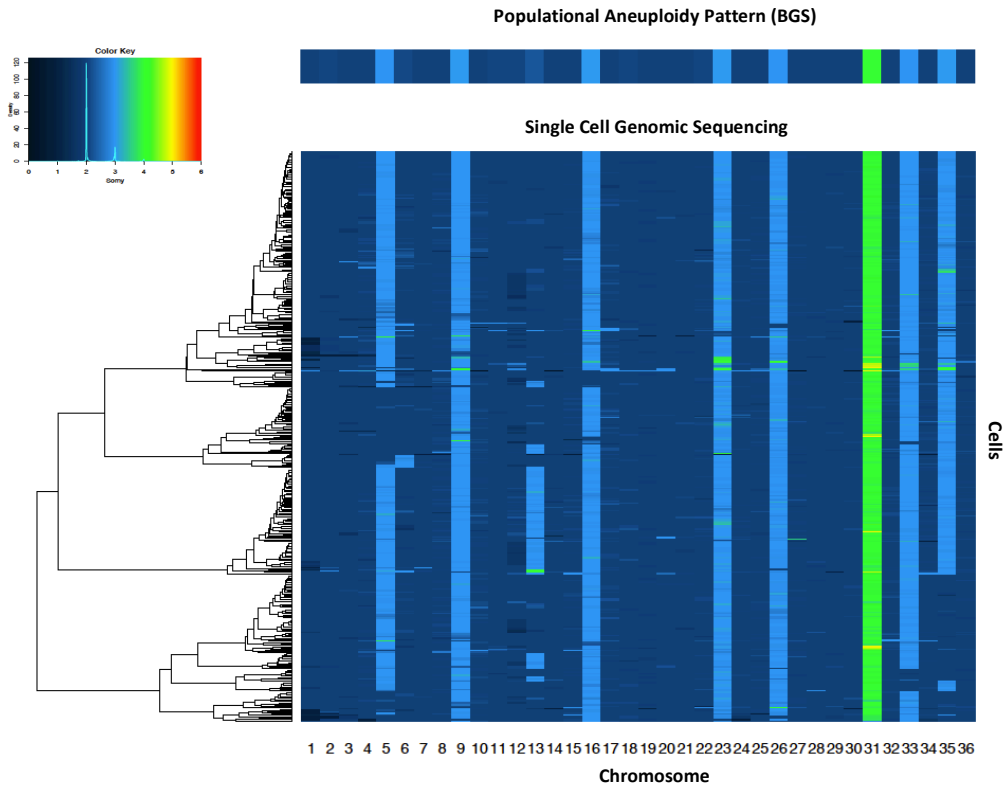
400 In cancer cells and in human pathogenic yeast, genetic diversity created by mosaic
401 aneuploidy has been linked to drug resistance[34,35]. In *Leishmania,* several *in vitro* studies report
402 whole chromosome somy changes as one of the first genetic changes in populations under drug
403 pressure[12,13,27,36,37]. However, it is still unknown if such variation in population average
404 aneuploidy is a reflect of a positive selection of a sub-population displaying a pre-existing
405 karyotype or is acquired *de novo* as a response to the pressure. The fact that the BPK282
406 population was previously used in *in vitro* drug selection experiments with at least 3 drugs[12,13,27]
407 allowed us to revisit the data of these experiments and compare the observed populational
408 aneuploidy pattern changes with single karyotypes identified by SCGS in the population in order
409 to address this question. In this context, we observed that the first contact of the parasites with
410 the drugs lead to the emergence of new 'populational karyotypes' that are similar to karyotypes
411 which were found in single-cells in the SCGS data (**Figure 5**). Thus, it seems that early stages of
412 aneuploidy changes are caused by selection of pre-existing karyotypes. However, further
413 exposure to higher concentrations led to somy changes that does not correspond to any
414 karyotype in our data. It is possible that these changes reflect selection of other, rarer
415 karyotypes that occur at frequencies lower than the detection limit of the droplet-based
416 method applied here. Nonetheless, since *Leishmania* is able to constitutively generate mosaic
417 aneuploidy[38], it is also possible that new karyotypes are generated *de novo* as a response to the
418 increasing challenge imposed by higher drugs concentration. Clonal lineages tracking methods,
419 such as DNA barcodes[39], should allow further insights in this matter.

420 **Conclusions**

421 In summary, this work represents the first description of complete karyotypes of individual
422 *Leishmania* parasites and open an unprecedent technological milestone to study MA in
423 trypanosomatids. Future work should focus on the emergence and evolution of MA during
424 clonal expansion by sequencing cells during very early stages of clonal evolution at different
425 timepoints, as well as following the dynamics of karyotype changes during different stages of
426 adaptation to drug pressure. Combining SCGS with single-cell transcriptomics could also allow
427 to understand better the impact of gene dosage imbalance on transcription with a single cell
428 resolution. Thus, high throughput single-cell sequencing methods represent a remarkable tool
429 to understand key aspects of *Leishmania* biology and adaptability.

430 **Figures:**

**A)**




**B)**

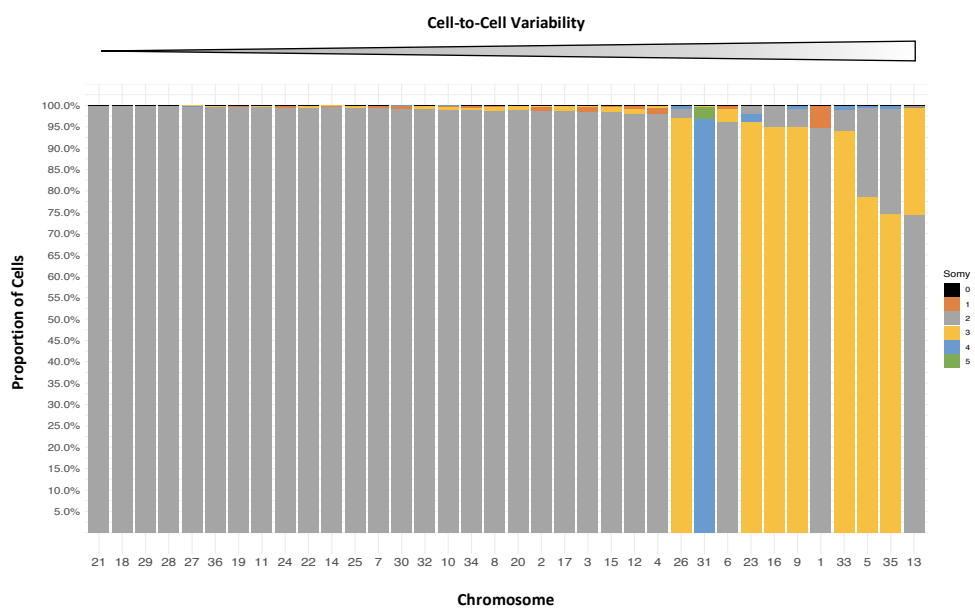



**Figure 1** – **A)** Somy values of all 36 chromosomes of all 1560 cells in the BPK282 clonal population. The color key inset displays the distribution of somy values in the SCGS data. A heatmap based on BGS data of BPK282 is displayed on the top. **B)** Proportion of somies found for each chromosome. Chromosomes are arranged, from left to right, by cell-to-cell variability, defined as the proportion of cells displaying a somy different from the most frequent one. For each column, somy bars are stacked by the frequency they happen in the population.

431

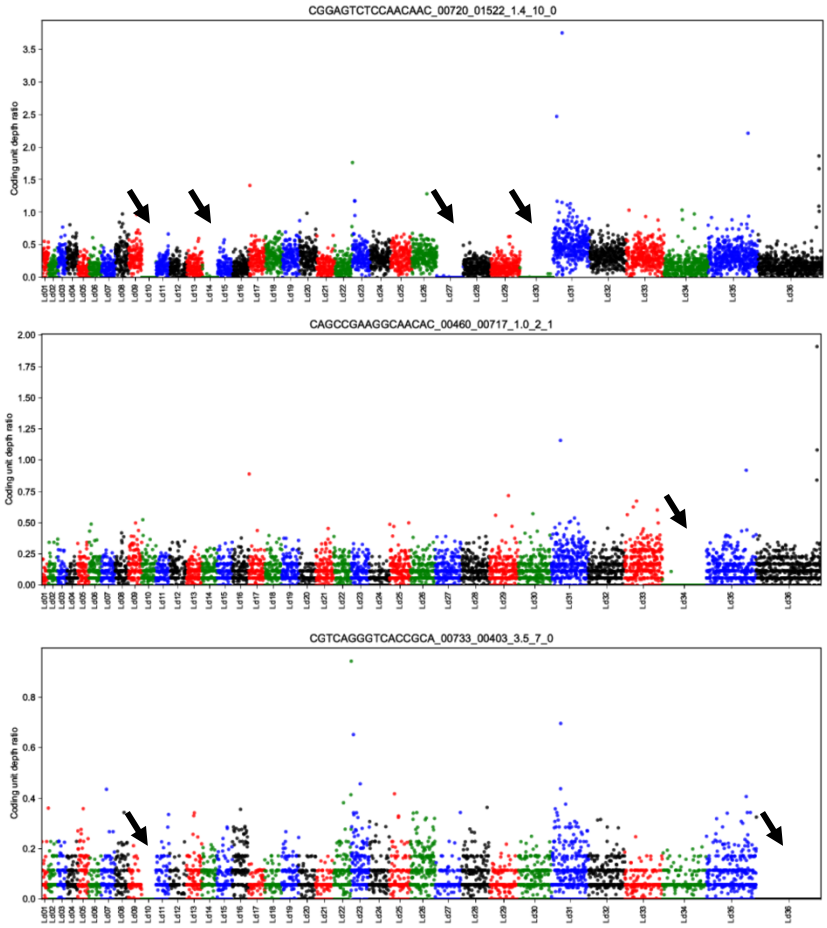
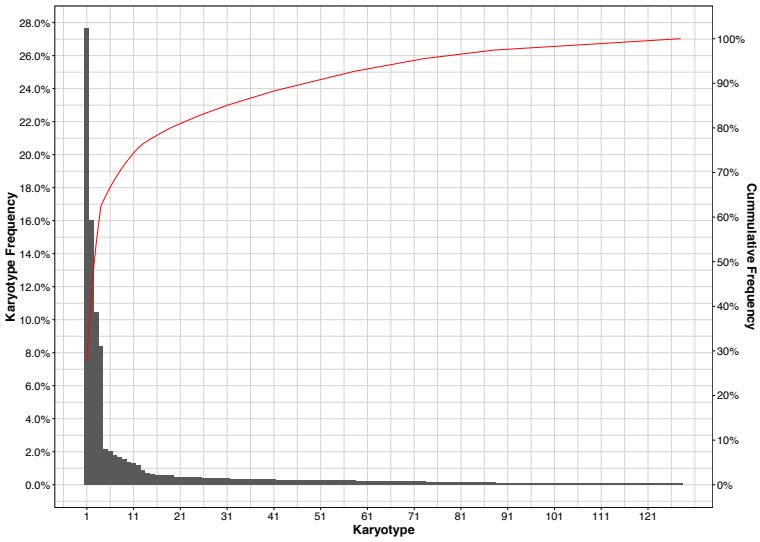



**Figure 2** – Three examples of individual cells, identified by their 10X barcode sequence (top of each plot), where one or more chromosome is nullisomic. Manhattan plot displays average depth per 5kb. Nullisomic chromosomes are indicated with a black arrow. When present, single reads mapping to nullisomic chromosomes are reads composed by repetitive nucleotides that can be mapped to multiple regions of the genome.

**A)**




**B)**

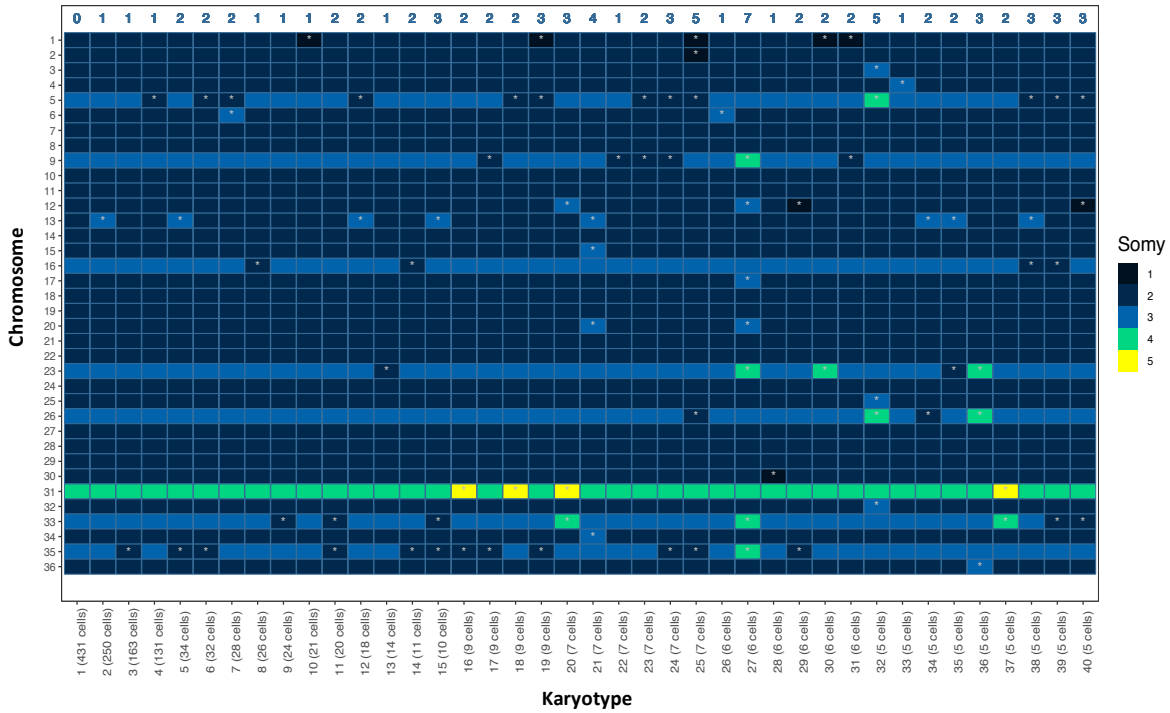



**Figure 3** – Distribution and profile of different karyotypes identified among 1560 BPK282/0 cl4 promastigotes. **A)** Frequency distribution of all the 128 different karyotypes identified in the SCGS data. Red line indicates the cumulative frequency in the secondary axis. **B)** Heatmap displaying the somy values off each chromosome in the 40 most frequent karyotypes. Karyotypes are ordered from left to right and numbered according to their frequency in the population. Chromosomes in which somies diverge from the most frequent karyotype (Karyotype 1) are marked with a white asterisk. Blue numbers in the top part of the graph indicate the number of chromosomes with somy different from the first karyotype.

432

433

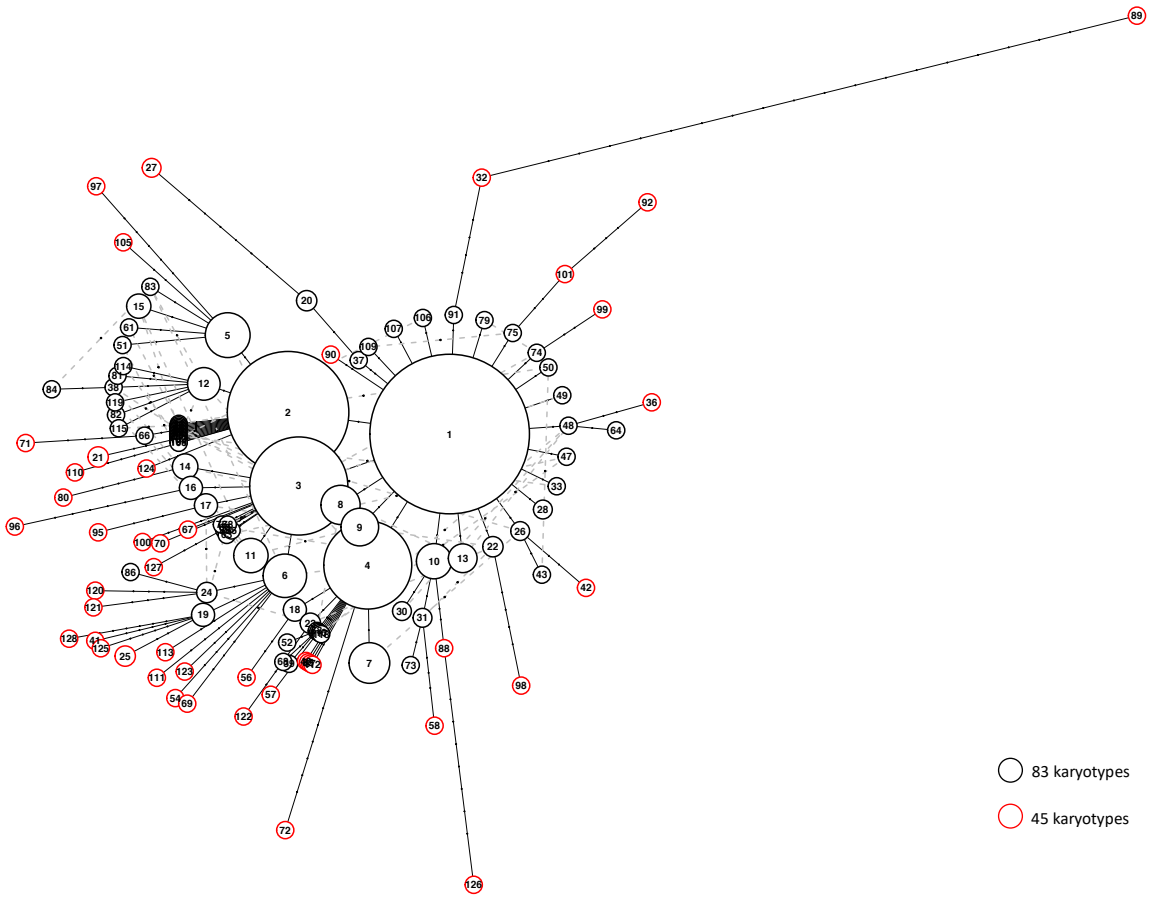



**Figure 4** – Haplotype network between all 128 karyotypes. The dots in the link lines represent the number of different chromosomes between 2 karyotypes. Dashed grey lines indicate alternative links between karyotypes that diverge by a single chromosome (black pies). Red pies highlights karyotypes where the closest related karyotype display different somy in at least 2 chromosomes. The size of each pie is proportional to the number of cells displaying a given karyotype. The smallest pies represent a total of 7 or less cells.

434

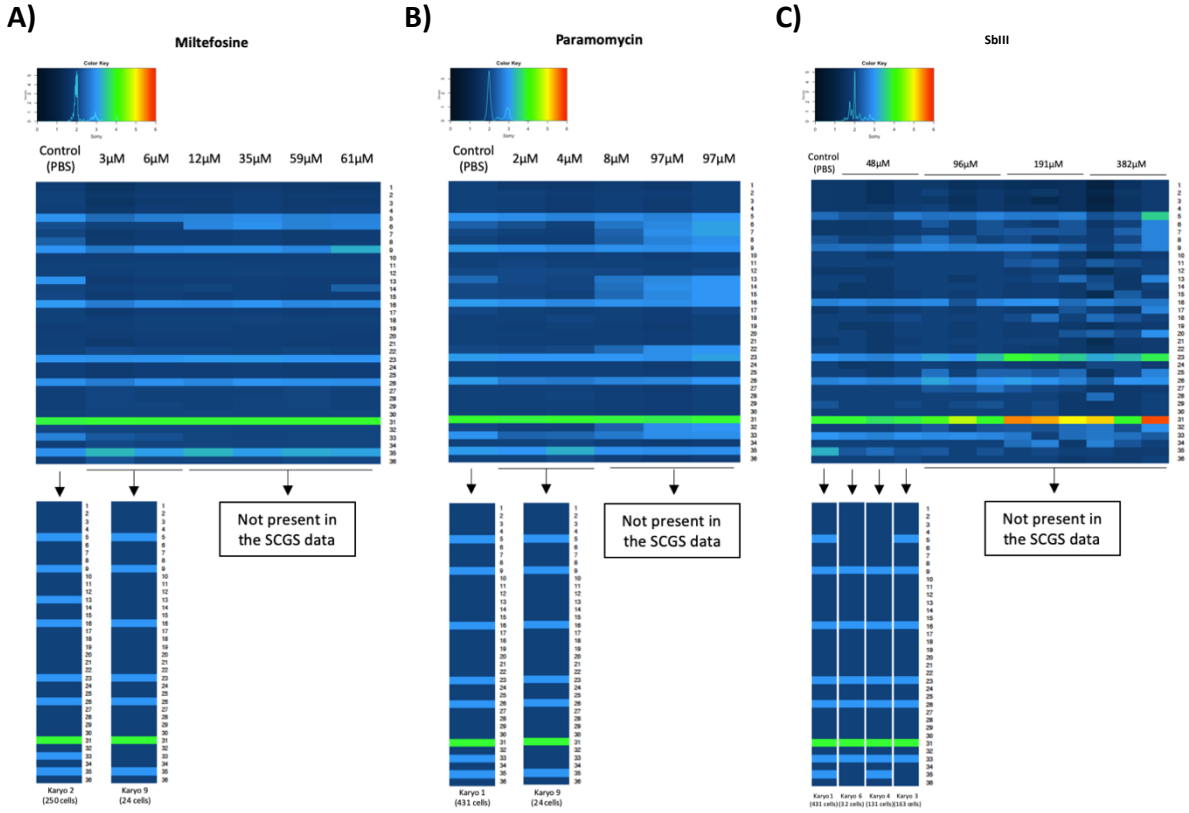



**Figure 5** – Comparison between previously published somy estimation by BGS of BPK282/0cl4 strain during early adaptation to drug pressure (upper heatmaps) and individual karyotypes found in the SCGS data (bottom heatmaps). BGS somy values were published elsewhere for Miltefosine[13], Paramomycin[26] and SbIII[12].

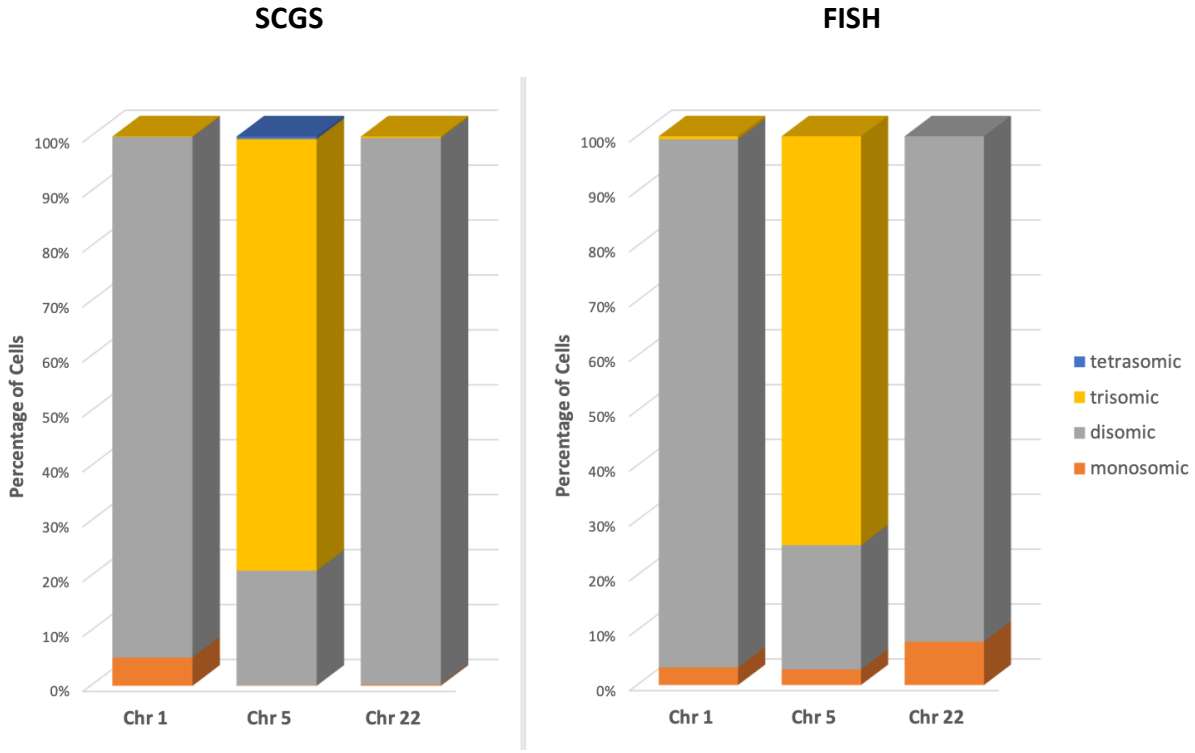



**Figure 6** - Validation of the SCGS data using FISH. The proportions of cells displaying monosomic, disomic, trisomic and tetrasomic chromosomes 1, 5 and 22 were evaluated by FISH and compared to the same proportions found in the SCGS data.

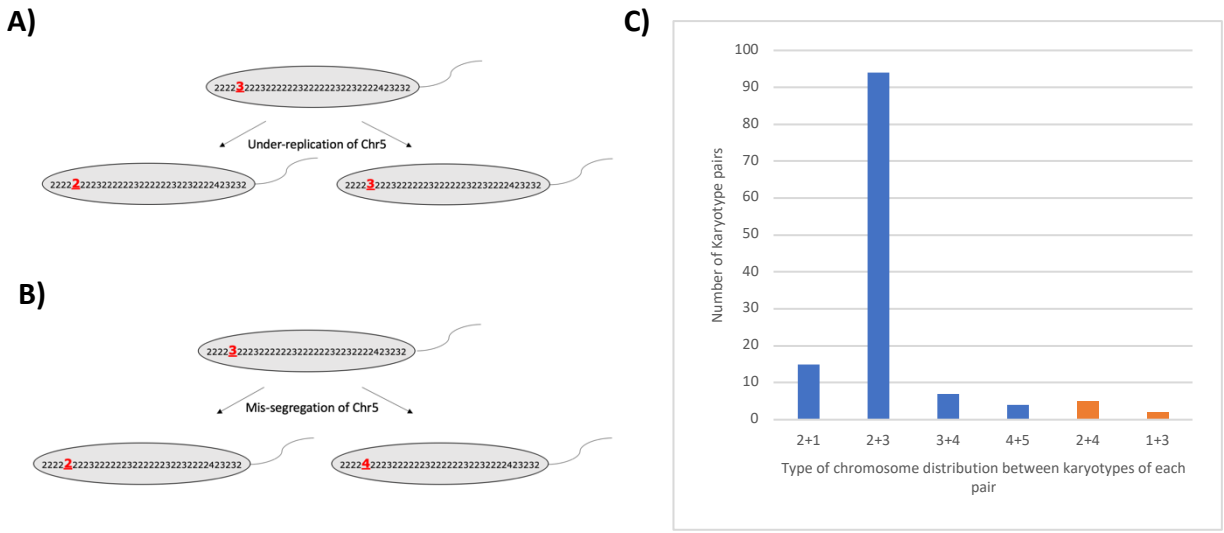



**Figure 7** – Karyotypes found in the SCGS data support errors in chromosome replication during mitosis as main drivers of aneuploidy mosaicism. **A)** An example of an under-replication event that leads to two karyotype pairs displaying a "2+3" (odd) somy distribution in chromosome 5. **B)** A mis-segregation event in chromosome 5 would generate two karyotypes that display a "2+4" (even) somy distribution in the daughter cells. **C)** Karyotypes diverging by a single chromosome from the SCGS data were compared in pairs. The pattern of somies of the divergent chromosome in each pair is represented in the x axis. Odd patterns, representing putative under or over-replication events, are depicted in blue, while even patterns, a putative indication of mis-segregation events, are represented in orange.

440 **Supplementary Material**

441

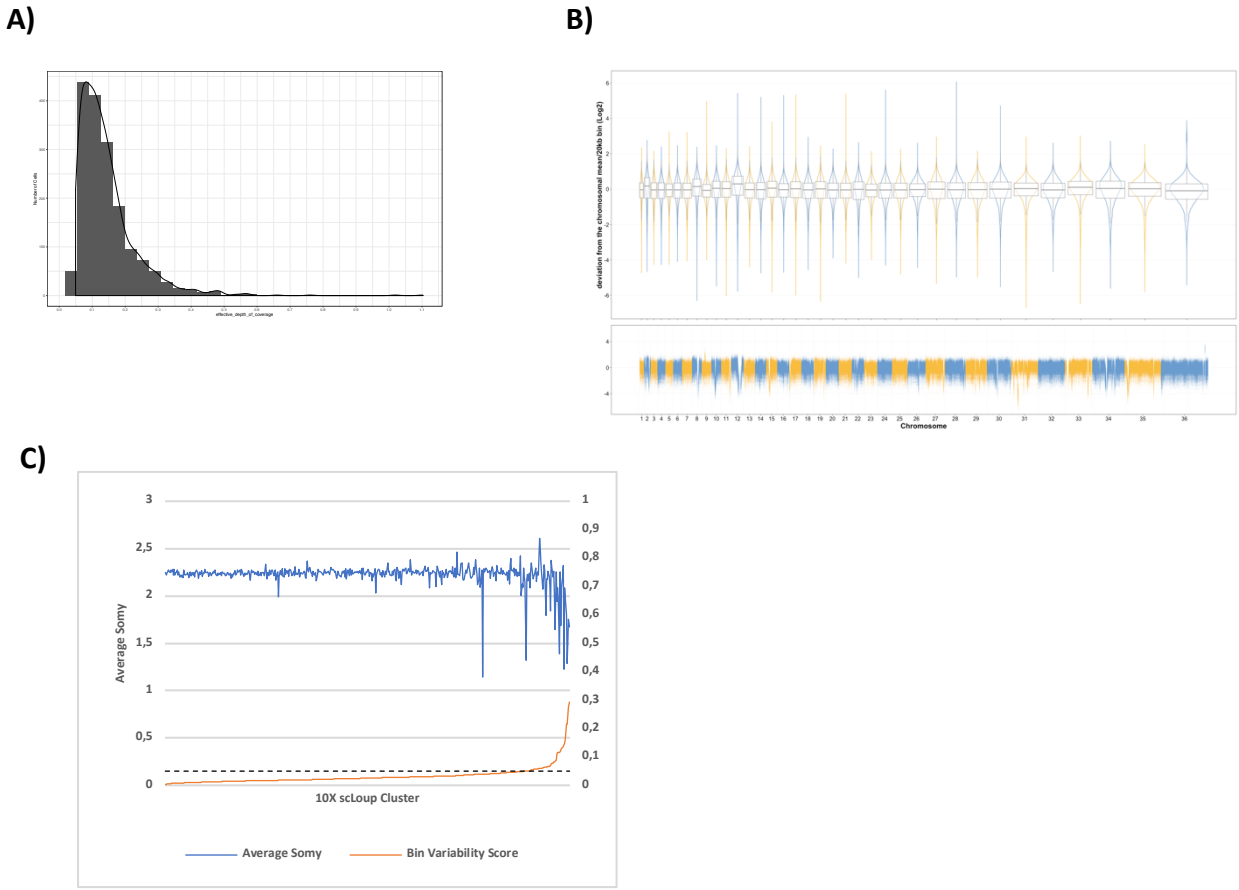



**Figure S1 – A)** Distribution of the effective coverage depth per cell. **B)** Violin plot showing the read depth deviation from the mean of intrachromosomal 20kb bins. For each cell, the mean depth/20kb of each chromosome was calculated and the ration between the depth of each intrachromosomal 20kb bin and the average was log2 transformed. Values close to 0 represent bins with depth similar to the chromosomal average. A box plot indicates the median (line in the center), the upper and lower quartiles (boxes) and the maximum and minimum of each chromosome (lines). A Manhattan plot on the bottom shows individual deviation values. **C)** Average somy and bin variability score of all 512 clusters. The black dashed line represents the threshold set to eliminate clusters with high bin-to-bin variability.

442

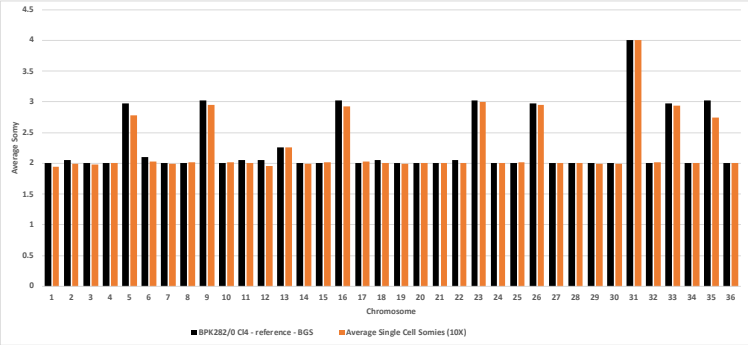



**Figure S2** - Comparison of the average combined somy values of all 1560 BPK282/0 cl4 promastigotes assessed by SCGS (orange bars) with the average somies of a population of the reference BPK282 strain estimated by BGS (black bars).

443

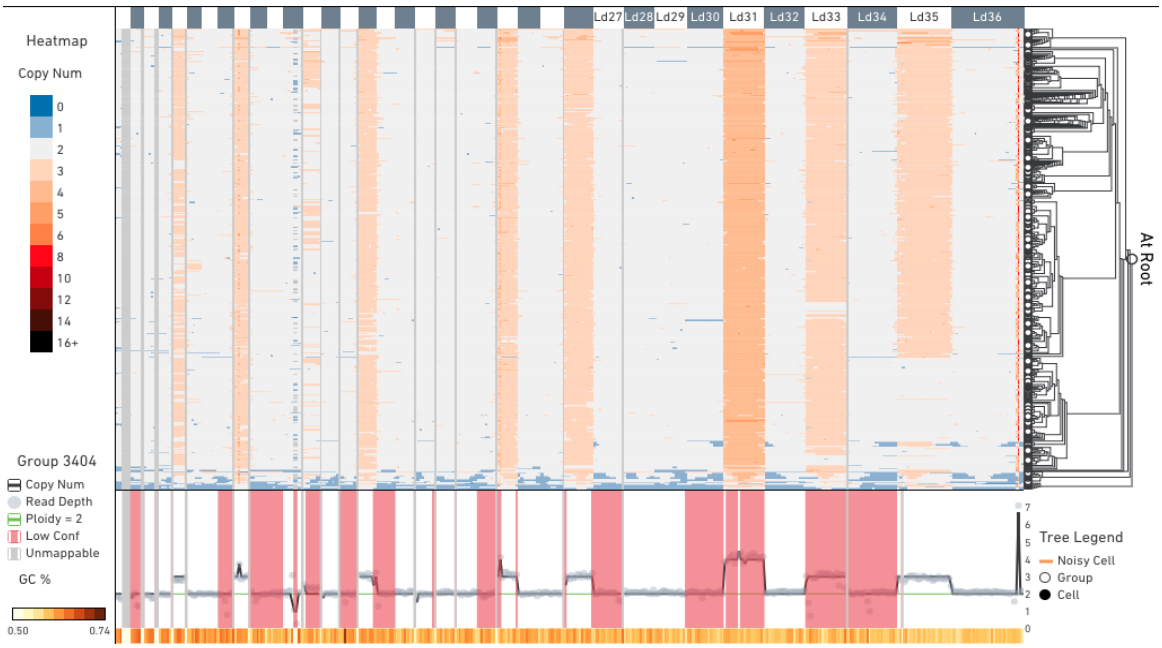



**Figure S3** - CNV values calling by 10X Cell Ranger DNA pipeline visualized in 10X Loupe scDNA Browser tool. The 1703 cells are arranged in 512 clusters (horizontal lines). Bars on the top indicate the position of each chromosome, while lines and grey dots in the bottom represent the CNVs and reads/megabase, respectively, of all clusters combined. High copy number regions in chromosome 23 and 36 represent, respectively, the H-locus and the M-locus.

444

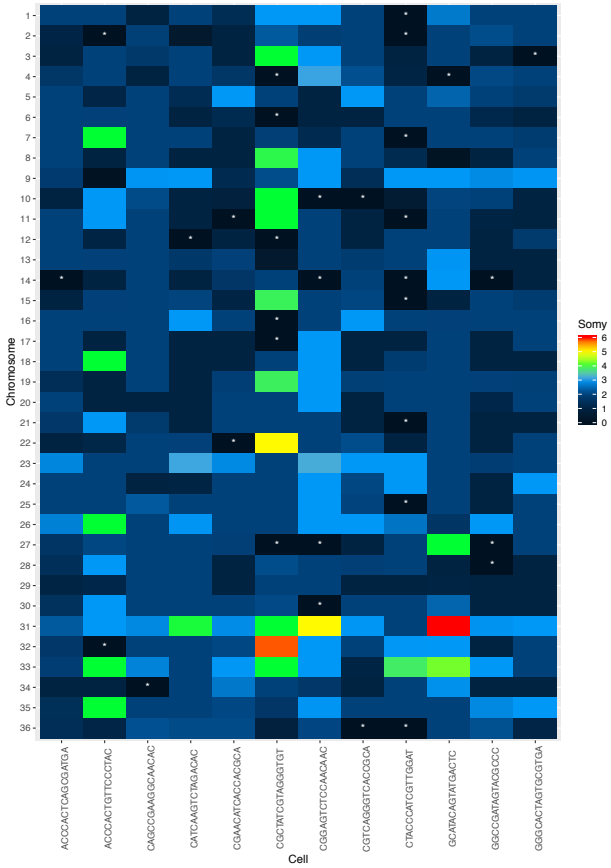



**Figure S4** – Heatmap showing the estimated somy of all cells where at least one chromosome was missing, including cells removed from by data filtering. Nullisomic chromosomes are marked with a white asterisk.

445 **Bibliography**


446 1. Pelkmans, L. Cell Biology. Using cell-to-cell variability--a new era in molecular biology.
447 *Science* **336**, 425–6 (2012).

448 2. García-Betancur, J. C. & Lopez, D. Cell Heterogeneity in Staphylococcal Communities.
449 *Journal of Molecular Biology* (2019). doi:10.1016/j.jmb.2019.06.011

450 3. Gilchrist, C. & Stelkens, R. Aneuploidy in yeast: Segregation error or adaptation
451 mechanism? *Yeast* **36**, 525–539 (2019).

452 4. Seco-Hidalgo, V., Osuna, A. & Pablos, L. M. De. To bet or not to bet: deciphering cell to
453 cell variation in protozoan infections. *Trends Parasitol.* **31**, 350–356 (2015).

454 5. Burza, S., Croft, S. L. & Boelaert, M. Leishmaniasis. *Lancet* **392**, 951–970 (2018).

455 6. Adl, S. M. *et al.* The revised classification of eukaryotes HHS Public Access. *J. Eukaryot.*
456 *Microbiol Microbiol* **59**, 429–493 (2012).

457 7. Clayton, C. Regulation of gene expression in trypanosomatids: Living with polycistronic
458 transcription. *Open Biol.* **9**, (2019).

459 8. Reis-Cunha, J. L., Valdivia, H. O. & Bartholomeu, D. C. Gene and Chromosomal Copy
460 Number Variations as an Adaptive Mechanism Towards a Parasitic Lifestyle in
461 Trypanosomatids. *Curr. Genomics* **19**, 87–97 (2017).

462 9. Mannaert, A., Downing, T., Imamura, H. & Dujardin, J. C. Adaptive mechanisms in
463 pathogens: Universal aneuploidy in Leishmania. *Trends Parasitol.* **28**, 370–376 (2012).

464 10. Rogers, M. B. *et al.* Chromosome and gene copy number variation allow major structural
465 change between species and strains of Leishmania. *Genome Res.* **21**, 2129–2142 (2011).

466 11. Dumetz, F. *et al.* Modulation of aneuploidy in leishmania donovani during adaptation to
467 different in vitro and in vivo environments and its impact on gene expression. *MBio* **8**, 1–
468 14 (2017).

469 12. Dumetz, F. *et al.* Molecular Preadaptation to Antimony Resistance in *Leishmania*
470 *donovani* on the Indian Subcontinent. *mSphere* **3**, e00548-17 (2018).

471 13. Shaw, C. *et al.* In vitro selection of miltefosine resistance in promastigotes of Leishmania
472 donovani from Nepal: Genomic and metabolomic characterization. *Mol. Microbiol.* **99**,
473 1134–1148 (2016).

474 14. Cuypers, B. A systems biology approach for a comprehensive understanding of molecular
475 adaptation in *Leishmania donovani*. (University of Antwerp, 2018).

476 15. Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pagès, M. FISH analysis reveals
477 aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*.
478 *Cell. Microbiol.* **13**, 274–283 (2011).

479 16. Sterkers, Y. *et al.* Novel insights into genome plasticity in Eukaryotes: Mosaic aneuploidy
480 in Leishmania. *Mol. Microbiol.* **86**, 15–23 (2012).

481 17. Barja, P. P. *et al.* Haplotype selection as an adaptive mechanism in the protozoan
482 pathogen *Leishmania donovani*. *Nat. Ecol. Evol.* **1**, 1961–1969 (2017).

483 18. Imamura, H. *et al.* Evaluation of whole genome amplification and bioinformatic methods
484 for the characterization of Leishmania genomes at a single cell level. *bioRxiv* (2020).

485 19. What is Cell Ranger? - Software - Single Cell Gene Expression - Official 10x Genomics
486 Support. Available at: https://support.10xgenomics.com/single-cell-
487 dna/software/pipelines/latest/what-is-cell-ranger-dna. (Accessed: 29th January 2020)

488 20. Single cell copy number calling -Software -Single Cell CNV -Official 10x Genomics Support.
489 Available at: https://support.10xgenomics.com/single-cell-
490 dna/software/pipelines/latest/algorithms/cnv_calling. (Accessed: 27th February 2020)

491 21. What is Loupe scDNA Browser? -Software -Single Cell CNV -Official 10x Genomics
492 Support. Available at: https://support.10xgenomics.com/single-cell-
493 dna/software/visualization/latest/what-is-loupe-scdna-browser. (Accessed: 29th
494 January 2020)

495 22. Clustering Cells -Software -Single Cell CNV -Official 10x Genomics Support. Available at:
496 https://support.10xgenomics.com/single-cell-
497 dna/software/pipelines/latest/algorithms/clustering. (Accessed: 6th February 2020)

498 23. Downing, T. *et al.* Whole genome sequencing of multiple Leishmania donovani clinical
499 isolates provides insights into population structure and mechanisms of drug resistance.
500 *Genome Res.* **21**, 2143–2156 (2011).

501 24. Paradis, E. Analysis of haplotype networks: The randomized minimum spanning tree
502 method. *Methods Ecol. Evol.* **9**, 1308–1317 (2018).

503 25. Paradis, E. Pegas: An R package for population genetics with an integrated-modular
504 approach. *Bioinformatics* **26**, 419–420 (2010).

505 26. Patino, L. H., Muñoz, M., Muskus, C., Méndez, C. & Ramírez, J. D. Intraspecific Genomic
506 Divergence and Minor Structural Variations in Leishmania ( Viannia ) panamensis. 1–19
507 (2020). doi:10.3390/genes11030252

508 27. Shaw, C. *et al.* Genomic and metabolomic polymorphism among experimentally selected
509 paromomycin-resistant Leishmania donovani strains. *Antimicrob. Agents Chemother.*
510 (2019). doi:10.1128/AAC.00904-19

511 28. Sterkers, Y., Crobu, L., Lachaud, L., Pagès, M. & Bastien, P. Parasexuality and mosaic
512 aneuploidy in Leishmania: Alternative genetics. *Trends Parasitol.* **30**, 429–435 (2014).

513 29. Hou, Y. *et al.* Comparison of variations detection between whole-genome amplification
514 methods used in single-cell resequencing. *Gigascience* **4**, 1–16 (2015).

515 30. Can I run non-human samples? – 10X Genomics. Available at:
516 https://kb.10xgenomics.com/hc/en-us/articles/360005055712-Can-I-run-non-human-
517 samples-. (Accessed: 5th February 2020)

518 31. Camacho, E. *et al.* Complete assembly of the *Leishmania donovani* (HU3 strain) genome
519 and transcriptome annotation. *Sci. Rep.* **9**, 1–15 (2019).

520 32. How much of a single cell's genome is amplified? – 10X Genomics. Available at:
521 https://kb.10xgenomics.com/hc/en-us/articles/360005108931-How-much-of-a-single-
522 cell-s-genome-is-amplified-. (Accessed: 6th February 2020)

523 33. Imamura, H. *et al.* Evolutionary genomics of epidemic visceral leishmaniasis in the Indian
524 subcontinent. *Elife* **5**, 1–39 (2016).

525 34. Sotillo, R., Schvartzman, J. M., Socci, N. D. & Benezra, R. Mad2-induced chromosome
526 instability leads to lung tumour relapse after oncogene withdrawal. *Nature* **464**, 436–440
527 (2010).

528 35. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of
529 aneuploidy provides increased fitness during the evolution of antifungal drug resistance.
530 *PLoS Genet.* **5**, 1–16 (2009).

531 36. Rastrojo, A. *et al.* Genomic and transcriptomic alterations in *Leishmania donovani* lines

532 experimentally resistant to antileishmanial drugs. *Int. J. Parasitol. Drugs Drug Resist.* **8**,
533 246–264 (2018).

534 37. Patino, L. H. *et al.* Major changes in chromosomal somy, gene expression and gene
535 dosage driven by SbIII in Leishmania braziliensis and Leishmania panamensis. *Sci. Rep.* **9**,
536 9485 (2019).

537 38. Lachaud, L. *et al.* Constitutive mosaic aneuploidy is a unique genetic feature widespread
538 in the Leishmania genus. *Microbes Infect.* **16**, 61–66 (2014).

539 39. Blundell, J. R. & Levy, S. F. Beyond genome sequencing: lineage tracking with barcodes
540 to study the dynamics of evolution, infection, and cancer. *Genomics* **104**, 417–430 (2014).

541