The effect of *in vitro* culture on unicellular eukaryotes: adaptation of *Trypanosoma brucei brucei* bloodstream forms results in gene copy-number changes.

[1]Julius Mulindwa, [2]Geofrey Ssentamu, [2]Enock Matovu [3]Kevin Kamanyi Marucha#, [4]Francisco Aresta-Branco, [3]Claudia Helbig, [3]Christine Clayton.

[1]Department of Biochemistry and Sports Science, College of Natural Sciences, Makerere University, P.O.Box 7062, Kampala, Uganda

[2]School of Biosecurity, Biotechnical and Laboratory Sciences, College of Veterinary Medicine, Animal Resources and Biosecurity, Makerere University, P.O. Box 7062, Kampala, Uganda

[3]Heidelberg University Centre for Molecular Biology (ZMBH), Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

[4]Division of Immune Diversity, Deutsche Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 280, D69120 Heidelberg, Germany.

#current address: Kisii University Medical Biochemistry, P.O. Box 408, 40200 Kisii, Kenya.

*to whom correspondence should be addressed. e-mail cclayton@zmbh.uni-heidelberg.de, tel +49 6221 546876

Abstract

Most researchers who study unicellular eukaryotes work with an extremely limited number of laboratory-adapted isolates that were obtained from the field decades ago, but the effects of passage in laboratory rodents, and adaptation to *in vitro* culture, have been little studied. For example, the vast majority of studies of *Trypanosoma brucei* biology have concentrated on just two strains, Lister 427 and EATRO1125, which were taken from the field over half a century ago and have since have undergone innumerable passages in rodents and culture. We here describe two new *Trypanosoma brucei brucei* strains. MAK65 and MAK98, which have undergone only 3 rodent passages since isolation from Ugandan cattle. Adaptation of these strains to culture was accompanied by changes in gene copy numbers, some of which were also evident when other lab-adapted strains, field isolates of *T. rhodesiense*, and the genome strain TREU927 were compared. Reproducible increases were seen for genes encoding histones, enzymes of mRNA processing and degradation, the cytosolic chaperone HSP70, and two proteins required for the DNA damage response. These results indicate that similar work with other eukaryotic pathogens would be worthwhile. Meanwhile, the two new *T. brucei* strains should be useful to researchers interested in trypanosome differentiation and pathogenicity. They have differing pathogenicities in mice and may also differ in their propensity for stumpy-form differentiation, as judged by morphology and mRNA expression. MAK65 grows better than MAK98 in bloodstream-form culture, and is uniformly diploid, whereas MAK98 is triploid for chromosome 5. Genome sequence exceeding 100-fold coverage is available for both strains.

Introduction

Trypanosoma brucei brucei and related parasites infect humans and various domestic animals, and can be transmitted mechanically and venereally as well as by their definitive host, the Tsetse fly. *T. brucei gambiense* is the subspecies that causes West African sleeping sickness, while *T. brucei rhodesiense* causes East African sleeping sickness. *T. brucei brucei* is found in cattle, but not humans; it differs from *T. brucei rhodesiense* only in the absence of the *SRA* gene, which enables *T. rhodesiense* to survive in human plasma [1, 2]. Within Tsetse flies, *T. brucei* multiply as procyclic forms in the midgut, before migrating to the salivary glands, where sexual reproduction occurs, with meiosis and gamete formation, followed by gamete fusion [3, 4].

The *Trypanosoma brucei* genome consists of eleven megabase-length chromosomes, which are generally diploid, and a variable number of "minichromosomes". The parasite escapes the immune response by antigenic variation, expressing a single Variant Surface Glycoprotein (VSG). The expressed *VSG* gene is located at a telomere, and can be changed either through transcriptional switching or, more commonly, by genetic rearrangement. Every parasite has at least 1000 alternative *VSG* genes or pseudogenes, which are located in sub-telomeric arrays and on the minichromosomes [5]. DNA contents vary up to 30% between *T. brucei* isolates [6]. Although some of this can be attributed to differing minichromosome contents [6], the lengths of the megabase chromosomes also differ substantially both within, and between, strains [6-8]. Variations in the numbers of *VSG* genes are to be expected, but there are also other differences. Some genes are arranged in multi-copy arrays, which facilitates high expression but also leaves the genes prone to homologous recombination. For example, the beta- and alpha-tubulin genes are present in an alternating array, and one study found fewer copies in *T. brucei gambiense* than in *T. brucei rhodesiense* or *T. brucei brucei* [6]. The chromosome copy number might be expected to be made more uniform by meiosis and mating, but although classical Mendelian inheritance is seen [9], triploid progeny also appear to be relatively common [10, 11]; also, the sexual stage is not obligatory.

In the past few decades, nearly all studies of *T. brucei* molecular and cellular biology have used just two strains: Lister 427, which was probably originally isolated from a cow in 1956 (see http://tryps.rockefeller.edu/DocumentsGlobal/lineage_Lister427.pdf), and EATRO1125, which was isolated from a bushbuck in 1966 (see

http://tryps.rockefeller.edu/DocumentsGlobal/lineage_antat1.pdf). These strains have been passaged innumerable times in rodents or culture. This selects for an accelerated growth rate, which presumably reflect changes in metabolism and cell-cycle regulation. Stumpy forms are growth-arrested bloodstream forms which are pre-adapted for differentiating into procyclic forms. Prolonged passage clearly selects for a diminished ability to enter cell-cycle arrest, and therefore loss of the "stumpy form" life-cycle stage. Since some genes that are required for survival as procyclic forms are not needed in the bloodstream, prolonged culture or rodent passage as bloodstream forms can also result in loss of the ability to differentiate into proliferation-competent procyclic forms. This has, for example, occurred for the Lister 427 bloodstream forms currently used for genetic manipulation. Although several *T. brucei* strains, including EATRO1125, have been maintained in such a way as to preserve their differentiation capacity, we do not know whether their differentiation pathways are identical to those found in natural populations.

Another problem is that many cultured cell lines that have the same name have been maintained for decades in separate labs. As a consequence, the parasites that we study now are likely to show considerable differences in gene copy numbers, sequence and regulation compared with their ancestors and also between labs. This will have been exacerbated by multiple cloning steps that have occurred during the selection of lines suitable for genetic manipulation. Analysis of gene copy numbers indeed suggested that in comparison with recently-isolated *T. rhodesiense*, common "lab" strains of *T. brucei* had varying expansions in multicopy gene arrays encoding proteins required for rapid cell division, and also differed from each other [12].

In this paper we set out to establish new *T. b. brucei* strains that have undergone minimal passage since field isolation in order to expand the repertoire of trypanosomes available for lab investigation. We describe the infection and culture characteristics of two new Ugandan *T. b. brucei* strains that have very different disease profiles in mice. Available stocks of these parasites have undergone only 3 mouse passages since their isolation from cattle. These lines should be very useful for analyses of factors governing *T. brucei* disease course, tissue distribution, and differentiation. In addition, we aimed to find out what happens when *T. brucei* are adapted to bloodstream-form culture. The growth of the new trypanosomes was slower than of standard lab lines but genome changes were rapid, with gene copy number changes after only a few weeks.

Results and discussion

In vivo growth of two new *T. b. brucei* isolates

T. b. brucei MAK65 was isolated from a cow in Banya parish, Apac district on Feb 1st, 2016, while *T. b. brucei* MAK98 was isolated from the same place on July 30th, 2016. The identity as *T. brucei* was confirmed by the presence of a 480bp band PCR of the rRNA internal transcribed spacer, and absence of the SRA gene [13]. The strains were shown to be different by microsatellite typing [14] (S1 text). The frozen cow blood was passaged once through mice in order to make stabilates for further use.

To characterize the two isolates, we first infected 8-week-old swiss mice with 1000 parasites each. Examples of the parasitaemias are shown in Figure 1A and B. Results suggested that MAK65 was slightly less virulent than MAK98. We also infected rats with 5000 parasites, after a single previous mouse passage to avoid artefacts due to differences in stabilate viability. Parasitaemias were again lower for MAK65. We harvested the parasites 5 or 6 days after infection (Fig 1C). We had hoped that the populations harvested earlier in infection would show low expression of stumpy-form markers, and would therefore be suitable for characterization of long-slender-form transcriptomes. On the contrary, the proportion of stumpy forms (as judged by PAD1 staining [15]) was highest for MAK65 parasites at low parasitaemia (Fig 1C, D). Conversely, in the rat with 50-times higher MAK98 parasitaemia, no stumpy forms were detected (MAK98B) even though cell proliferation had clearly slowed. These results hint that the two isolates might have differing tendencies to make stumpy

127 forms - or that stumpy forms are preferentially generated in locations other than the blood. A
128 lower threshold for stumpy formation in MAK65 might explain its lower virulence in mice [16].

129 Transcriptome analysis (E-MTAB-9320, S1 Table) revealed that parasites of the MAK65A
130 population had highest expression of mRNA encoding the procyclic-form surface protein
131 GPEET procyclin as well as other procyclic-specific mRNAs such as those encoding trans-
132 sialidase and enzymes of the citric acid cycle. Interestingly, although culture 98B did not
133 show PAD1 staining, the population with the lowest expression of stumpy-form markers was
134 98A. Comparison of the 65A and 98A transcriptomes with those of pure EATRO1125 long
135 slender and stumpy forms [17] confirmed that MAK65A indeed had a more stumpy-form like
136 expression pattern than 98A (Figure 1E). Although these were by no means pure
137 populations, the correlation was better than that obtained for *in vitro* stumpy-form
138 differentiation of EATRO1125 [18] (S1 Fig. A). We had previously suggested that poly(A)
139 selection also selects against long mRNAs [19]. Sequencing of mRNA prepared either by
140 poly(A) selection, or by rRNA depletion (using RNase H and complementary
141 oligonucleotides) firmly confirmed this (S1 Fig. B).

142 We also sequenced genomic DNA from the two new strains. The initial populations that were
143 sequenced were MAK65 from rats, and MAK98 cultured for 2 days (see below). We obtained
144 Oxford nanopore reads (E-MTAB-9318) as well as 72-nt paired-end Illumina reads at over
145 100-fold coverage (E-MTAB-9759). In combination these might allow *de novo* genome
146 assembly and should certainly be sufficient to enable assembly based on existing scaffolds.
147 To minimize manipulation, we did not clone these cells at any stage. Although *T. brucei*
148 populations are to some extent clonal [20, 21], we cannot rule out the possibility of mixed
149 infections, and some within-population variation is also to be expected.

150 ***In vitro* growth of the two new *T. b. brucei* isolates**

151 To assess the abilities of the two new strains to grow *in vitro*, we placed trypanosomes
152 directly from mouse blood stabilates into HMI-9 medium and followed cell numbers, diluting
153 them regularly to prevent densities did not exceed 1.5x 10^6/ml. Both isolates grew rapidly
154 from stabilate but after 1-2 days, the growth slowed markedly and became intermittent
155 (Figure 2A-D) with average doubling times of 30-50 hours. This suggests that the parasites
156 had run out a nutrient that is available only *in vivo*, but were able subsequently to adapt to
157 growth without it. As expected, there were clear deleterious effects when the density
158 accidentally exceeded 2x 10^6/ml. To find out whether the growth in culture was
159 reproducible, new stabilates were thawed and placed into culture. The MAK65 cells again
160 adapted to culture quite readily, with an initial division time of about 17h which shortened to
161 about 11 h after 5 days (Figure 2E). During culture for a further 2 weeks, growth remained
162 somewhat erratic, with clear deleterious effects when the density accidentally exceeded 1.0
163 x10^6/ml, but the division times progressively decreased to about 7h (S2 Fig. A; cultures
164 MAK65cA, MAK65cB). Similar observations were made for two additional cultures initiated
165 from a new stabilate (S2 Fig. B; cultures MAK65cC, MAK65cD).

166 For MAK98, over the first 24h the division time was 8h (Figure 2F; this is the DNA that was
167 used for genomic sequencing). Cells were then divided into 10 replicate cultures with
168 different starting densities varying from 6000/ml to 2 x10^5/ml. Over the next 24h the division
169 time was 17±3h irrespective of the starting density (Figure 2G) and the maximum density
170 that was subsequently obtained was 1.5 x10^6/ml. Several additional attempts to culture
171 MAK98 from stabilates failed. Although this might be an artefact - a property of the particular
172 stabilates used - the subsequent slow growth of MAK98 in culture suggests that it is
173 intrinsically less culture-adaptable, or that perhaps only a small subset of the population is
174 adaptable. For prolonged growth, we first continued the MAK98 culture shown in Figure 2F,
175 giving culture MAK98 culture A; this grew erratically with an overall division time of 13h
176 despite never exceeding densities of 5x 10^5/ml (S2 Fig. C). Later, we initiated a new culture
177 from a frozen stock made from the initial two-day culture; this resulted in culture MAK98
178 culture B, which after a further week achieved a division time of 11h (S2 Fig. D).

179 Very surprisingly, neither strain survived in medium containing methyl cellulose, which we
180 use routinely for passage of EATRO1125 bloodstream forms. Results of preliminary
181 experiments suggested that it was possible to obtain procyclic forms by cultivating the cells
182 with cis aconitate at 37°C for 24h, then placing them in procyclic-form medium (SDM79) at
183 27°C, but this has not been investigated in detail.

184 For future genetic manipulation attempts, the MAK65 line appears preferable because it is
185 uniformly diploid, and appears to be more stable in culture than MAK98.

186 ***In vitro* growth affects gene ploidy**

187 To find out how culture affects the genome, we sequenced MAK65 that had been cultured
188 for 5 weeks (65 culture A, 65 culture B, S1 Fig. A, E-MTAB-10457) and 2 weeks (65 culture
189 C, 65 culture D, S1 Fig. B, E-MTAB-10457); and MAK98 that had been cultured for 7 weeks
190 (MAK98 culture A, S1 Fig. C, E-MTAB-10466) and 3 weeks (MAK98 culture B, S1 Fig. D, E-
191 MTAB-10457).

192 Given the uncloned nature of the starting population we, expected to see selection of
193 parasites with better abilities to grow in culture, and perhaps of new mutations. Once the
194 starting genomes have been assembled, it will be possible to do detailed analyses of single
195 nucleotide polymorphisms that were selected. In the meantime, however, we measured
196 variations in gene copy number. We had previously found differences in numbers of multi-
197 copy genes between lab-adapted trypanosomes and *T. b. rhodesiense* isolated from patients
198 [12, 19], but these analyses were constrained by the limited number of trypanosome
199 populations available and in particular, the absence of data for un-passaged versions of the
200 lab-adapted strains. Copy-number calculations were based on the assumption that most
201 genes are diploid. Only central chromosome regions containing conserved regions were
202 considered and for multicopy genes, only one representative was considered (S2 Table).

203 First, we re-analysed the genomes of trypanosomes we maintain routinely in the laboratory,
204 comparing them with the previous previously-characterized *T. b. rhodesiense* strains, and
205 the reference "genome" strain, TREU927. We analysed data from three long-term-cultured
206 bloodstream-form lines (two Lister 427, one EATRO1125) and one Lister 427 procyclic-form
207 culture; all had been cloned at least ten years previously after genetic manipulation, and had
208 been published intermittently since. All were predominantly diploid. Figure 3 shows the copy
209 number distribution for the majority of genes - those with 0.5 - 2 copies per haploid genome;
210 and copy-numbers across the genome are shown in Figure 4. The *T. rhodesiense* genomes
211 were the most cleanly diploid among this group (Figure 3 and Figure 4F,G). These
212 trypanosomes were all derived from patients in a single sleeping-sickness focus and had
213 undergone minimal rodent passage [12]. The "genome" reference strain TREU927 was
214 initially chosen because the genome size is relatively small relative to other strains, and
215 because it is differentiation-competent. The source of the DNA for TREU927 sequencing is
216 not clear from the publication: it may have been derived from procyclic forms that had
217 recently-differentiated from blood parasites [7], but in that period isolation from rodent blood
218 was more common. Interestingly, although all TREU927 chromosomes were diploid, the
219 copy-number distribution was quite broad, suggesting population heterogeneity with low
220 levels of various gene duplications and deletions (Figure 3, Figure 4A). We discovered that
221 our EATRO1125 culture is trisomic for small regions of chromosomes 1 and 7, and
222 monosomic for a small segment for chromosome 11 (Figure 4B). These discrepancies most
223 likely represent intra-chromosomal duplications and deletions. We were also surprised to
224 discover that at the time of sequencing, our standard procyclic-form tet-repressor-expressing
225 Lister 427 line was trisomic for chromosome 6 (Figure 4B). Moreover, the "2T1" Lister 427
226 line, which contains a "landing pad" that allows targeting of plasmids to a specific rRNA
227 spacer [22] showed numerous copy-number differences from Lister 427 carrying pHD1313
228 [23], including a possible trisomy of part of chromosome 5 (Figure 4C). Such partial trisomies
229 might compromise attempts at homozygous gene deletion. They highlight the fact that

230 strains with the same name, but grown for protracted periods in different laboratories, may
231 have diverged considerably.

232 We next compared the genomes of the new strains with the previous ones. For both strains,
233 most genes were clearly diploid (Figure 3, Figure 5 A, B, D and E) except that MAK98 was
234 triploid for chromosome 5 (Figure B, E). This might be a remnant of mating, but could also
235 have arisen in the infected cow. It was interesting that the *T. rhodesiense* strains from
236 human patients showed somewhat more copy-number heterogeneity than the two cattle-
237 derived *T. brucei*. The difference is unlikely to reflect geographical origin since the *T.*
238 *rhodesiense* isolates were from Lwala, in Kaberamaido district, which is only about 100 km
239 from Banya (Apac district) and the dates of isolation were less than 5 years apart. The
240 difference is also unlikely to have been caused by growth in different hosts since cattle are a
241 reservoir for *T. rhodesiense,* and *T. rhodesiense* almost certainly undergoes genetic
242 exchange with *T. brucei* [24].

243 Surprisingly, just a week of culture was sufficient to broaden the ploidy distribution somewhat
244 for both MAK65 and MAK98 (Figure 5 A, B). This indicates selection of parasite
245 subpopulations with either deletions, or duplications of different chromosome regions.
246 Nevertheless, gene copy numbers of all MAK65 cultures looked broadly similar to the source
247 population; an example is shown in Figure 5D and the remainder are in S3 Fig. One of the
248 MAK98 cultures - MAK98B - also looked similar to the starting population although some
249 changes were evident (Figure 5F). In contrast, the MAK98A culture had an exceptionally
250 broad gene copy-number distribution profile, suggesting that an unusual number of genes
251 was present in more than 2 copies in the population (Figure 5B). Many small segments had
252 changed in copy number throughout the genome (Figure 5G), with some chromosome
253 segments present in triploid, tetraploid or more copies. We do not know whether the
254 additional gene copies are internal chromosome duplications, or extra small chromosomes.
255 However this result, combined with the slow growth rate, suggests that while adapting to
256 culture, the parasites suffered defects in chromosome replication and/or segregation.

257 ***In vitro* growth selects for increased copies of a few specific genes**

258 Finally, we looked to see whether there were any consistent changes in copy numbers after
259 culture adaptation. First, we examined changes that happened in the MAK65 cultures and
260 MAK98 culture B. (We did not further consider MAK98 culture A because it appeared to be
261 severely compromised, but calculations are included in S2 Table.) It was immediately clear
262 that some genes were reproducibly affected. Table 1 is a list of all genes showing an
263 increase of at least one copy (per diploid genome) in all considered cultures, relative to the
264 source blood population, and Table 2 shows decreases; examples are also plotted in S3 Fig.
265 F-H. Increases were seen for histones, several paraflagellar rod proteins, the chaperones
266 HSP70 (major cytosolic isoform) and mitochondrial HSP60, two mitochondrial RNA editing
267 proteins and two translation factors, and the cyclin F box proteins CFB1 and CFB2. (These
268 share sequence; in TREU927 there are actually 5 genes encoding CFB1, followed by only
269 one encoding CFB2). In contrast, decreases were seen for rRNA genes and PIP39. The
270 change in rRNA genes is surprising since rRNA is needed for rapid growth, but PIP39
271 promotes differentiation to stumpy forms [25] so there could be selection for decreased
272 expression. A detailed alignment of the region that includes *HSP70* is shown in S4 Fig.. It
273 compares one MAK65 culture (culture A) with the starting population. The *HSP70* gene
274 (Tb927.11.11330) is arranged as tandem repeats, but is present only once in the reference
275 genome, because the (approximately 8) repeats were eliminated during genome assembly.
276 The alignment for the starting population reveals differences between the homologous
277 chromosomes (changes, relative to 927, seen in only half of the reads) and additional
278 heterogeneity. Comparing *HSP70* reads with those over the surrounding region, it is clear
279 that in the culture, amplification occurred within the tandem repeat, without affecting
280 neighboring genes.

If the changes that we saw in MAK65 and MAK98 are important for adaptation to culture, then they should also be apparent in more established lab strains. S2 Table, sheet 2 includes all genes showing reproducible copy-number changes in culture-adapted parasites relative to field isolates (see Table Legend for our definition of "reproducible"). The status of the TREU927 parasites is questionable in the comparisons - they had not been cultured as bloodstream forms but had undergone at least 40 rodent passages (see https://tryps.rockefeller.edu/DocumentsGlobal/lineage_TREU927.pdf). In some cases, genes that appeared to be slightly sub-diploid in the starting populations became slightly more than diploid in the selected cells; the significance of this is unclear (see Legend to S2 Table), but it could indicate haploidy in a small proportion of cells. Interpretation of the results is also complicated by the fact that selection for duplication for one gene may well also affect the surrounding region. Nevertheless, changes in some genes were reproducible. It is not surprising that tandemly repeated gene arrays were commonly affected, since the arrangement facilitates recombination; but there was clear specificity because many repeated genes showed no changes (S2 Table, sheet 3). Most cultures showed increases in genes encoding histone H3 (Figure 6A) and histone H2A (Figure 6B). Others included AUK2 kinase (Figure 6C), which is involved in DNA repair [26]; marginal increases for PUF9 (Figure 6D), which stabilizes S-phase mRNAs [27]; and a putative monooxygenase (Figure 6E). Most cultures had increases in *HSP70* gene copy numbers (though some reads could come from homologous genes) (Figure 6F) and the gene encoding the splicing factor TSR1 (Figure 6G) [28]. The numbers of genes encoding CFB1 or CFB2 were also reproducibly increased (Figure 6H); from the read counts (also of the 3'-untranslated regions) it would appear that the entire region had been amplified and it is not clear which gene was selected for. The function of CFB1 is unknown [29], while CFB2 binds to, and stabilizes the mRNA encoding the variant surface glycoprotein [30]. Gene copy numbers for REC8/SCC1, needed for sister chromatid adhesion [31]), two flagellar proteins, an RNA polymerase III subunit, and various other proteins of unknown function were also significantly increased. Among the few significant decreases were genes encoding a pteridine transporter (Figure 6I) and, rather oddly, the translation factor eIF3c (Figure 6I). Overall the increases seemed to be biased towards genes likely to be required for rapid proliferation or stress tolerance.

Outlook

Our results show that cultured lines from the same original trypanosome isolate can have differences in karyotype, and that gene copy numbers and even chromosome ploidies in cell lines that are in routine use are very likely to differ between labs and clones, even if the parasites had the same origin. Perhaps we should not be surprised, therefore, if some results are not consistent between laboratories. It is perhaps also worth checking ploidies when thinking about gene knock-outs - although Crispr-Cas approaches may to some extent abrogate that problem.

Initial growth of the new parasites in culture was relatively slow, but usually began to speed up after a few weeks. We do not know what is "missing" from the culture medium, relative to mammalian blood and tissue fluids. The observed slow proliferation could reflect slow overall metabolism in all of the parasites, but could also reflect a heterogeneous population, some growing quite fast, and others dying. The extensive gene copy number changes seen after only a few weeks of culture adaptation might be adaptive, but could also reflect errors in DNA replication and chromosome segregation. Our observations on copy number are the tip of the iceberg: a survey of just a single ~10 kb region revealed selection for smaller insertions, deletions and point mutations (S4 Fig). More detailed studies of multiple cloned parasite lines would be necessary to examine this.

It was intriguing that the trypanosomes used for TREU927 genome sequencing appeared to have more genome heterogeneity than either the well-established cultured trypanosomes, or the field isolates. Perhaps this reflected extensive rodent passage, with different selective pressures relative to culture? Examination of historical stocks might clarify this. It would also be really interesting to compare the genome of the original Lister 427 line [32] with those in

334 routine use today. The results clearly raise the question of how much other cultured - or
335 merely rodent-adapted - organisms differ from their original counterparts, and to what extent
336 the parasites that are studied in the laboratory are representative of those in the field.

337 We here provide two new *T. brucei* strains for biological studies, with differences in rodent
338 pathogenicity and the ability to grow in culture. Full genome assembly will enable a much
339 more thorough comparison between the uncultured and cultured cell populations, as well as
340 an assessment of the *VSG* gene repertoires of parasites subject to selection in the field. We
341 did not clone the isolated parasites, or the cultures, so it would be useful to compare several
342 cloned lines both genetically and biologically. The very different courses of infection in mice,
343 as well as differences in the apparent threshold for PAD1 expression, suggest that there
344 might be differences between MAK65 and MAK98 in tissue distribution. Results so far
345 suggest that although MAK65 appears to have a lower threshold for stumpy formation in
346 mice, and is consequently less virulent, it grows more readily in culture and is less prone to
347 genome changes than MAK98. The MAK65 line would therefore be ideal in order to find out
348 whether the numerous observations made concerning differentiation of the lab-adapted
349 EATRO1125 trypanosome line are also valid for trypanosomes that have not undergone
350 prolonged lab adaptation. Can the new lines make stumpy forms *in vitro*? Can they
351 differentiate directly from long slender forms? In either case, which signals are important?
352 Do they require cis-aconitate for differentiation? Do they infect Tsetse more readily than lines
353 in routine use? How do they respond to stress? Are the tissue distributions different and if
354 so, how does this impact pathogenicity and differentiation? We hope that labs with
355 appropriate expertise will investigate these questions.

356

357

358 **Materials and Methods**

359 **Trypanosome samples and infections**

360 Peripheral blood (3-4ml) from cattle was collected into an EDTA tube (BD Vacutainer). An
361 aliquot of the whole blood (600µl) was cryopreserved, a drop (10µl) was spotted on
362 Whatman paper for PCR diagnosis. To determine the genus of the Trypanosome isolates,
363 PCR was carried out on the Internal transcribed spacer (ITS-1) as described by Njiru et al
364 2004 [13].
365 To follow mouse infections, a stabilate was thawed and injected into a mouse. Once these
366 parasitaemias had attained about 5 x 10^7 trypanosomes/ml, 2-week-old inbred Swiss white
367 mice were infected with 500 parasites each. Parasites were counted by diluting 10µL of tail
368 blood into 1mL of phosphate-saline-glucose, then counting in a haemocytometer. Thin blood
369 smears were also prepared for immunofluorescence staining by fixed in methanol and
370 permeabilizing with 0.2%TritonX-100. Mice were euthanized if obvious terminal symptoms
371 were observed.
372 For RNA preparation, rats were infected with 5000 parasites each. Parasites were followed
373 using wet blood films. After the times shown in Figure 1, blood (3-4mL) was collected by
374 cardiac puncture and approximately 2.5ml drawn into a Paxgene tube. The Paxgene blood
375 was incubated at room temperature for one hour and thereafter centrifuged at 5000g for
376 10min. The supernatant was discarded, the pellet washed once with nuclease free water by
377 centrifuging at 5000g for 10min. The pellet was then resuspended in 1ml of Trifast reagent
378 (Peqlab, GmbH) and transferred to 1.5ml microfuge tube for RNA preparation.
379 Trypanosomes were cultured as described in [33], starting with frozen mouse blood. PAD1
380 staining was done at 4°C overnight and DNA was counterstained with DAPI. Stained slides
381 were blinded and evaluated by an independent observer.

382 **Sequence analysis**

383 Ribosomal RNA was depleted from the total RNA by hybridisation with antisense
384 oligonucleotide and digestion with RNase H as described in [34].
385 Sequencing of genomic DNA and RNA was done using standard Illumina kits, and some
386 genomic DNA was also sequenced using an Oxford nanopore device. Reads were aligned to
387 the genome, then those aligning to open reading frames, annotated 3'-untranslated regions,
388 and functional non-coding RNAs of the TREU927 genome were counted using the
389 tryprnaseq [35]. The transcriptome data were analysed using DeSeqU1 [36].
390 To obtain gene copy numbers, all genomic DNA reads were allowed to align 20 times. We
391 then selected unique genes, plus, for genes present in more than one copy, only a single
392 representative copy. We then calculated reads per million per kilobase for this set of open
393 reading frames. The result was a continuous distribution, with a strong peak. For each
394 dataset we chose the modal RPKM value to represent a single-copy gene, then adjusted it
395 slightly to get a symmetrical distribution as shown in S2 Table. We divided the other values
396 by this to attain gene copy numbers.
397

398 **Data availability**

399 The genome sequences are available with accession numbers E-MTAB-9318 and E-MTAB-
400 9759 for original strains, and E-MTAB-10457 and E-MTAB-10466 for cultures. The
401 transcriptome results are deposited as E-MTAB-9320.

402

403 **Ethical approval**

404 The sample collection and work with experimental animals was approved by the Makerere
405 University Animal use committee. The rodent experiment approval is SBLS/HDRC/19/012.
406

407 **Acknowledgements**

408 We thank Keith Matthews for antibody to PAD1, and George Cross and Annette McLeod for
409 useful comments concerning the origin of the TREU927 genome DNA. This work was
410 partially funded by Deutsche Forschungsgemeinschaft grant number CI112/28-1 to CC and
411 JM.

412

413 **References**

414 1. MacLeod A, Tait A, Turner CM. The population genetics of *Trypanosoma brucei* and
415 the origin of human infectivity. Philos Trans R Soc Lond B Biol Sci.
416 2001;356(1411):1035-44. Epub 2001/08/23. doi: 10.1098/rstb.2001.0892. PubMed
417 PMID: 11516381; PubMed Central PMCID: PMCPMC1088498.
418 2. Sistrom M, Evans B, Benoit J, Balmer O, Aksoy S, Caccone A. De Novo Genome
419 Assembly Shows Genome Wide Similarity between Trypanosoma brucei brucei and
420 Trypanosoma brucei rhodesiense. PLoS One. 2016;11(2):e0147660. Epub
421 2016/02/26. doi: 10.1371/journal.pone.0147660. PubMed PMID: 26910229; PubMed
422 Central PMCID: PMCPMC4766357.
423 3. Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, Carrington M, et al. Identification
424 of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. Proceedings of
425 the National Academy of Sciences of the United States of America. 2011;108(9):3671-
426 6. Epub 2011/02/16. doi: 10.1073/pnas.1019423108. PubMed PMID: 21321215;
427 PubMed Central PMCID: PMCPmc3048101.
428 4. Peacock L, Bailey M, Carrington M, Gibson W. Meiosis and Haploid Gametes in the
429 Pathogen *Trypanosoma brucei*. Curr Biol. 2013. Epub 2014/01/07. doi:
430 10.1016/j.cub.2013.11.044. PubMed PMID: 24388851.

5. Cross GA, Kim HS, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Molecular and biochemical parasitology. 2014;195(1):59-73. Epub 2014/07/06. doi: 10.1016/j.molbiopara.2014.06.004. PubMed PMID: 24992042.
6. Kanmogne GD, Bailey M, Gibson WC. Wide variation in DNA content among isolates of Trypanosoma brucei ssp. Acta Trop. 1997;63(2-3):75-87. Epub 1997/02/01. doi: 10.1016/s0001-706x(96)00600-6. PubMed PMID: 9088421.
7. Melville SE, Leech V, Gerrard CS, Tait A, Blackwell JM. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. Molecular and biochemical parasitology. 1998;94(2):155-73. Epub 1998/09/25. doi: 10.1016/s0166-6851(98)00054-1. PubMed PMID: 9747967.
8. Melville SE, Leech V, Navarro M, Cross GAM. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* stock 427. Mol Biochem Parasitol. 2000;111:261-73.
9. Gibson W, Peacock L, Ferris V, Williams K, Bailey M. Analysis of a cross between green and red fluorescent trypanosomes. Biochem Soc Trans. 2006;34:557-9.
10. Hope M, MacLeod A, Leech V, Melville S, Sasse J, Tait A, et al. Analysis of ploidy (in megabase chromosomes) in Trypanosoma brucei after genetic exchange. Mol Biochem Parasitol. 1999;104(1):1-9. Epub 1999/12/10. doi: 10.1016/s0166-6851(99)00103-6. PubMed PMID: 10589977.
11. Gibson W, Garside L, Bailey M. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. Mol Biochem Parasitol. 1992;51(2):189-99. Epub 1992/04/01. doi: 10.1016/0166-6851(92)90069-v. PubMed PMID: 1349422.
12. Mulindwa J, Mercé C, Matovu E, Enyaru J, Clayton C. Transcriptomes of newly-isolated *Trypanosoma brucei rhodesiense* reveal hundreds of mRNAs that are co-regulated with stumpy-form markers. BMC genomics. 2015;16:1118. PubMed Central PMCID: PMC10.1186/s12864-015-2338-y, PMID: 26715446.
13. Radwanska M, Chamekh M, Vanhamme L, Claes F, Magez S, Magnus E, et al. The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. Am J Trop Med Hyg. 2002;67:684-90.
14. Kato CD, Alibu VP, Nanteza A, Mugasa CM, Matovu E. Population genetic structure and temporal stability among *Trypanosoma brucei rhodesiense* isolates in Uganda. Parasit Vectors. 2016;9:259. Epub 2016/05/05. doi: 10.1186/s13071-016-1542-1. PubMed PMID: 27142001; PubMed Central PMCID: PMCPMC4855840.
15. Dean S, Marchetti R, Kirk K, Matthews K. A surface transporter family conveys the trypanosome differentiation signal. Nature. 2009;459:213-7.
16. MacGregor P, Savill NJ, Hall D, Matthews KR. Transmission stages dominate trypanosome within-host dynamics during chronic infections. Cell Host Microbe. 2011;9(4):310-8. Epub 2011/04/20. doi: 10.1016/j.chom.2011.03.013. PubMed PMID: 21501830; PubMed Central PMCID: PMCPmc3094754.
17. Silvester E, Ivens A, Matthews KR. A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. PLoS neglected tropical diseases. 2018;12(10):e0006863. Epub 2018/10/12. doi: 10.1371/journal.pntd.0006863. PubMed PMID: 30307943; PubMed Central PMCID: PMCPMC6199001.
18. Naguleswaran A, Doiron N, Roditi I. RNA-Seq analysis validates the use of culture-derived Trypanosoma brucei and provides new markers for mammalian and insect life-cycle stages. BMC genomics. 2018;19(1):227. Epub 2018/04/03. doi: 10.1186/s12864-018-4600-6. PubMed PMID: 29606092; PubMed Central PMCID: PMCPMC5879877.
19. Mulindwa J, Leiss K, Ibberson D, Kamanyi Marucha K, Helbig C, Melo do Nascimento L, et al. Transcriptomes of *Trypanosoma brucei rhodesiense* from sleeping sickness patients, rodents and culture: Effects of strain, growth conditions and RNA preparation methods. PLoS neglected tropical diseases. 2018;12(2):e0006280. Epub 2018/02/24.

486 doi: 10.1371/journal.pntd.0006280. PubMed PMID: 29474390; PubMed Central
487 PMCID: PMCPMC5842037.
488 20. Mathieu-Daudé F, Stevens J, Welsh J, Tibayrenc M, McClelland M. Genetic diversity
489 and population structure of *Trypanosoma brucei*: clonality versus sexuality. Mol
490 Biochem Parasitol. 1995;72(1-2):89-101. Epub 1995/06/01. doi: 10.1016/0166-
491 6851(95)00083-d. PubMed PMID: 8538703.
492 21. Duffy CW, MacLean L, Sweeney L, Cooper A, Turner CM, Tait A, et al. Population
493 genetics of *Trypanosoma brucei rhodesiense*: clonality and diversity within and
494 between foci. PLoS Negl Trop Dis. 2013;7(11):e2526. Epub 2013/11/19. doi:
495 10.1371/journal.pntd.0002526. PubMed PMID: 24244771; PubMed Central PMCID:
496 PMCPMC3828156.
497 22. Alsford S, Horn D. Single-locus targeting constructs for reliable regulated RNAi and
498 transgene expression in *Trypanosoma brucei*. Molecular and biochemical parasitology.
499 2008;161:76-9.
500 23. Alibu VP, Storm L, Haile S, Clayton C, Horn D. A doubly inducible system for RNA
501 interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. Mol
502 Biochem Parasitol. 2004;139:75-82.
503 24. Balmer O, Beadell JS, Gibson W, Caccone A. Phylogeography and taxonomy of
504 *Trypanosoma brucei*. PLoS Negl Trop Dis. 2011;5(2):e961. Epub 2011/02/25. doi:
505 10.1371/journal.pntd.0000961. PubMed PMID: 21347445; PubMed Central PMCID:
506 PMCPMC3035665.
507 25. Szoor B, Ruberto I, Burchmore R, Matthews KR. A novel phosphatase cascade
508 regulates differentiation in *Trypanosoma brucei* via a glycosomal signaling pathway.
509 Genes Dev. 2010;24(12):1306-16. Epub 2010/06/17. doi: 10.1101/gad.570310.
510 PubMed PMID: 20551176; PubMed Central PMCID: PMC20551176.
511 26. Stortz JA, Serafim TD, Alsford S, Wilkes J, Fernandez-Cortes F, Hamilton G, et al.
512 Genome-wide and protein kinase-focused RNAi screens reveal conserved and novel
513 damage response pathways in *Trypanosoma brucei*. PLoS pathogens.
514 2017;13(7):e1006477. Epub 2017/07/26. doi: 10.1371/journal.ppat.1006477. PubMed
515 PMID: 28742144.
516 27. Archer SK, van Luu D, de Queiroz R, Brems S, Clayton CE. *Trypanosoma brucei*
517 PUF9 regulates mRNAs for proteins involved in replicative processes over the cell
518 cycle. PLoS Pathog. 2009;5:e1000565. doi: 10.1371/journal.ppat.1000565.
519 28. Gupta SK, Chikne V, Eliaz D, Tkacz ID, Naboishchikov I, Carmi S, et al. Two splicing
520 factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA
521 stability, and rRNA processing in *Trypanosoma brucei*. RNA biology. 2014;11(6). Epub
522 2014/06/13. PubMed Central PMCID: PMC24922194.
523 29. Benz C, Clayton C. The cyclin F box protein CFB2 is required for cytokinesis of
524 bloodstream-form *Trypanosoma brucei*. Molecular and biochemical parasitology.
525 2007;156:217-24.
526 30. Melo do Nascimento L, Egler F, Arnold K, Papavisiliou N, Clayton C, Erben E.
527 Functional insights from a surface antigen mRNA-bound proteome. eLife.
528 2021;10:e68136. doi: https://doi.org/10.7554/eLife.68136.
529 31. Gluenz E, Sharma R, Carrington M, Gull K. Functional characterization of cohesin
530 subunit SCC1 in *Trypanosoma brucei* and dissection of mutant phenotypes in two life
531 cycle stages. Molecular microbiology. 2008;69(3):666-80. Epub 2008/06/17. doi:
532 10.1111/j.1365-2958.2008.06320.x. PubMed PMID: 18554326; PubMed Central
533 PMCID: PMCPMC2610385.
534 32. Peacock L, Ferris V, Bailey M, Gibson W. Fly transmission and mating of
535 *Trypanosoma brucei brucei* strain 427. Molecular and biochemical parasitology.
536 2008;160(2):100-6. Epub 2008/06/06. doi: 10.1016/j.molbiopara.2008.04.009. PubMed
537 PMID: 18524395; PubMed Central PMCID: PMCPMC2442871.
538 33. Fadda A, Ryten M, Droll D, Rojas F, Färber V, Haanstra J, et al. Transcriptome-wide
539 analysis of mRNA decay reveals complex degradation kinetics and suggests a role for

540 co-transcriptional degradation in determining mRNA levels. Molecular microbiology.
541 2014;94:307-26. doi: 10.1111/mmi.12764. PubMed Central PMCID: PMCdoi: .
542 34. Minia I, Merce C, Terrao M, Clayton C. Translation regulation and RNA granule
543 formation after heat shock of procyclic form *Trypanosoma brucei*: many heat-induced
544 mRNAs are increased during differentiation to mammalian-infective forms. PLoS
545 neglected tropical diseases. 2016;10:e0004982. PubMed Central PMCID:
546 PMC10.1371/journal.pntd.0004982.
547 35. Leiss K, Merce C, Muchunga E, Clayton C. TrypRNAseq - A easy to use pipeline for
548 *Trypanosoma* RNAseq data. Zenodo. 2016;http://doi.org/10.5281/zenodo.158920.
549 36. Leiss K, Clayton C. DESeqUI - Trypanosome RNAseq analysis made easy. Zenodo.
550 2016;doi:10.5281/zenodo.165132.
551 37. Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system
552 for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma*
553 *brucei*. Molecular and biochemical parasitology. 1999;99(1):89-101. Epub 1999/04/24.
554 doi: 10.1016/s0166-6851(99)00002-x. PubMed PMID: 10215027.
555

556

557 **Tables**

558
559 **Table 1**. Genes showing copy number increases after culture of MAK65 and MAK98 *T. brucei*. Diploid gene copy
560 numbers are shown. The haploid version is in Supplementary Table S2, sheet 1.
561

GeneID	Annotation	Class	MAK 65	MAK 98	MAK 65 cult-ure A	MAK 65 cult-ure B	MAK 65 cult-ure C	MAK 65 cult-ure D	MAK 98 cult-ure B	Aver age chan ge
11.11330	heat shock protein 70	Chaperone	18.3	20.3	23.6	25	26.6	25.1	26.9	+6.8
7.2820	histone H2A	Chromatin	46.4	43.2	51.7	54	56.5	54.5	55.6	+8.8
1.2430	histone H3	Chromatin	37.6	26.5	55	58.5	61.3	57.9	31.2	+17.4
5.4170	histone H4	Chromatin	42.3	63.4	49.4	49.6	51.5	50.6	70.5	+7.8
3.4290	paraflagellar rod protein PFR1	Cytoskeleton	9.7	10.5	12	12.9	13.3	13	13.1	+3
3.3770	paraflagellar rod protein PFC6	Cytoskeleton	3.9	4.2	5	5.2	5.7	5.2	5.6	+1.4
8.4970	paraflagellar rod protein PFR2	Cytoskeleton	7.5	7.9	8.9	9	9.4	8.9	9.3	+1.6
9.12550	glycosomal glycerol kinase GLK1	Glucose-glycerol	8.6	10.8	10.9	11.5	11.9	11.8	15.2	+3.2
8.1620	MSP-B, putative	Membrane protein	9.4	14.3	14	14.1	14.7	14.4	15.4	+4.2
11.15490	Tb-291 membrane protein	Membrane protein	6.3	6	9	9.5	10.6	9.4	9.1	+3.2
10.6510	mitochondrial HSP60	Mito.bio-genesis	5.5	5.9	6.9	7.4	7.8	7.4	7.8	+1.8
8.7260	kinetoplast-associated protein	Mito.DNA	4.3	3.1	6	6.6	7.3	6.1	4.5	+2
11.9570	hypothetical protein	Mito.pathway	2.8	5.2	6.2	6.6	6.7	6.4	6.6	+3.2
10.9720	RNA-editing protein REAP-1	Mito.RNA	3.2	6.7	5.2	5.5	5.5	5.7	8.8	+2.2
6.2230	TbRGG1	Mito.RNA	3	3.2	4.2	4.5	4.9	4.6	5	+1.6
10.2110	EF1-alpha	Translation	20.8	22.4	27	28.9	30.8	28.5	31.5	+8.2
10.4570	EF2	Translation	5.5	6.6	7.9	8.2	8.5	8.1	8.7	+2.6
1.4540	cyclin-like F-box protein CFB1	Ubiquitin	15.3	14.9	19.3	20.7	21.3	21	20.1	+5.2
1.4650	cyclin-like F-box protein CFB2	Ubiquitin	6.4	6.6	7.8	7.9	8.1	8.3	8	+1.6
5.1780	hypothetical protein, conserved	Unknown	3.6	5.7	4.6	4.7	5	4.9	7.9	+1.4
3.4740	glycerol-3-phosphate dehydrogenase-like	Unknown	3.5	3.8	4.9	5.1	5.8	5.4	5.7	+1.8
4.3580	Domain found in a transcription factor	Unknown	4.6	4.9	5.8	5.8	6.1	5.7	6.3	+1.2

562
563

564 **Table 2.**
565 Genes showing copy number decreases after culture of MAK65 and MAK98 *T. brucei*. Diploid gene copy
566 numbers are shown. The haploid version is in Supplementary Table S2, sheet 1.
567

GeneID	Annotation	Class	MAK 65	MAK 98	MAK 65 cult-ure A	MAK 65 cult-ure B	MAK 65 cult-ure C	MAK 65 cult-ure D	MA K98 cult-ure B	Aver-age chan ge
8.4700	AAT5 Arginine transporter	Amino acid transport	14.8	10.6	12.1	11.4	11.6	11.9	9.3	-2.6
11.13020	calmodulin	Calcium	10	9.3	8	8	8.3	8.8	7.6	-1.6
11.7740	dynein light chain	Cyto-skeleton	7.3	7.4	4.3	3.8	3.8	4.4	5.6	-3
9.13820	KMP-11	Cyto-skeleton	8.8	8.3	7.3	6.5	6.6	6.9	6.6	-2
11.8680	DNA polymerase kappa	DNA	11	12.4	9.9	9.5	9.4	9.7	10.4	-1.4
9.6090	PTP1-interacting protein PIP39	Protein phos-phatase	30.5	15.9	28.2	26.5	27	27.9	14.8	-2.8
10.8300	U1A small nuclear ribonucleoprotein RBP14A	RNA pro-cessing	4.2	4.5	2.7	2.6	2.5	2.6	3.4	-1.6
11.18680	dynein light chain LC8 DYNLL1	RNA synthesis	4.7	4.6	3.2	3.2	2.9	3.3	3.1	-1.6
2.1931	18S ribosomal RNA	rRNA	24.1	23.5	18.8	18.2	18.1	18	22.1	-5
3.3441	28S alpha ribosomal RNA	rRNA	28.6	30	22.5	21.1	20.9	20.9	25.7	-6.6
7.6885	28S beta ribosomal RNA	rRNA	28.6	32.1	25.2	24.2	23.7	23.9	28.5	-4.2
3.3431	5.8S ribosomal RNA	rRNA	42.3	38.8	25.4	22.8	22.5	24.9	28.2	-16.8
9.2840	Fam77 cell surface phylome protein	Unknown	13.9	12.7	12.4	10.9	10.7	11.7	9.5	-2.6
1.40	hypothetical protein	Unknown	8.3	11.3	4.9	4.1	4	4.7	8.1	-3.8
5.297b	hypothetical protein	Unknown	14	14.2	10.4	9.5	9.4	10	11.9	-3.8
10.130	hypothetical protein	Unknown	7.8	7.4	4.4	3.8	4	4.3	5.5	-3.4
5.293b	hypothetical protein	Unknown	10.6	11.1	7.6	7.2	6.9	7.4	9.2	-3

568
569
570

571 **Legends**

572 **Figure 1**

573 Growth of MAK65 and MAK98 strains in rodents.

574 A. Parasitaemias after infection of 3 mice with 1000 strain MAK65 parasites. "d" indicates
575 death and "k" indicates killing because of clear symptoms. The y axis scale indicates the
576 number by which the parasite concentrations were multiplied to obtain the scale. Thus
577 "100" indicates 10^8.

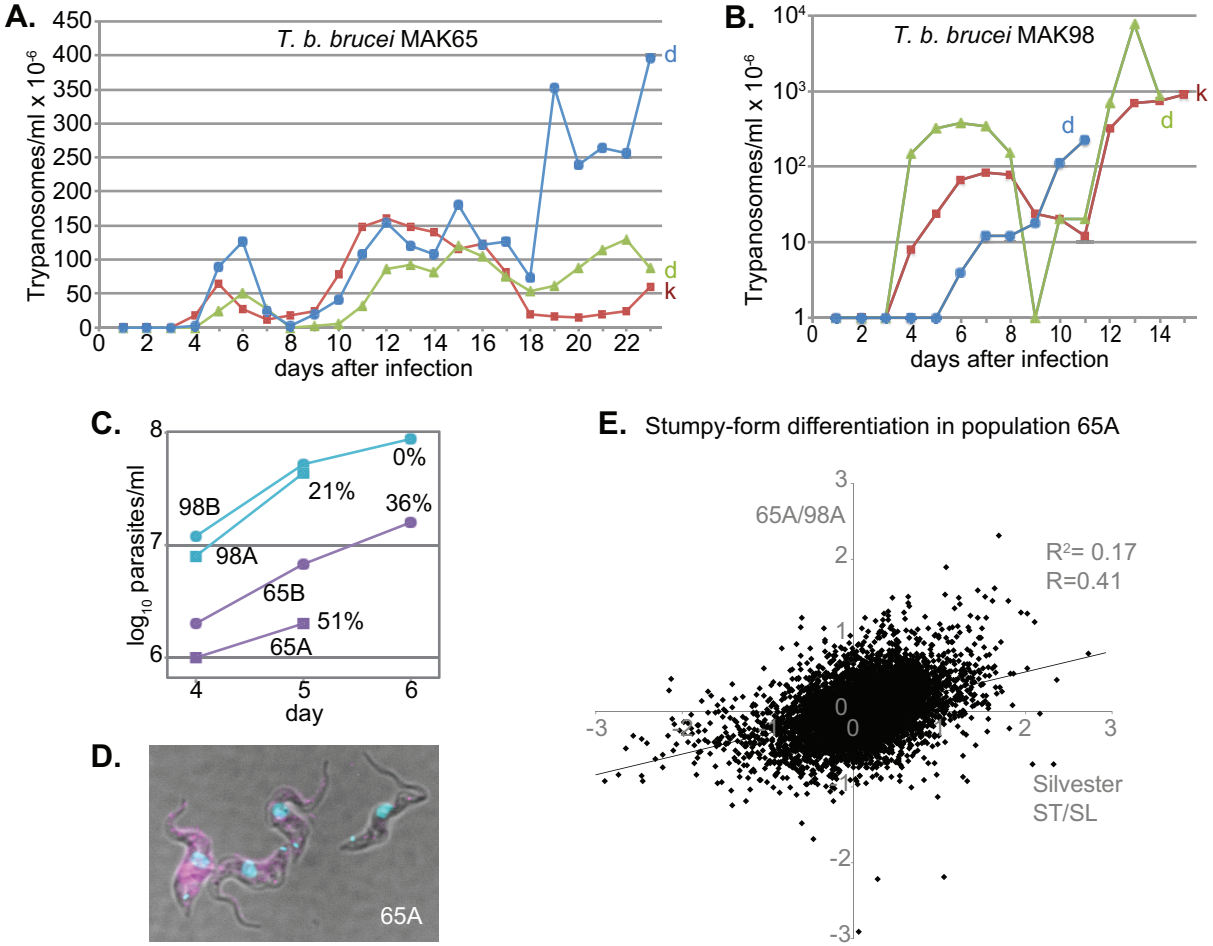
578 B. Parasitaemias after infection of 3 mice with 1000 strain MAK98 parasites.

579 C. Growth of both strains in rats before harvest for RNA purification. The percentage of
580 PAD1 positive cells at the time of harvest is indicated.

581 D. Example of PAD1 staining for sample 65A.

582 E. The transcriptomes of population 65A and 98A parasites (panel C) were compared and
583 the log2 ratio (65A/98A, Supplementary Table S1, sheet 1) is on the y-axis. The x-axis shows
584 published stumpy/long slender results for trypanosomes grown in mice [17].

585



586
587
588

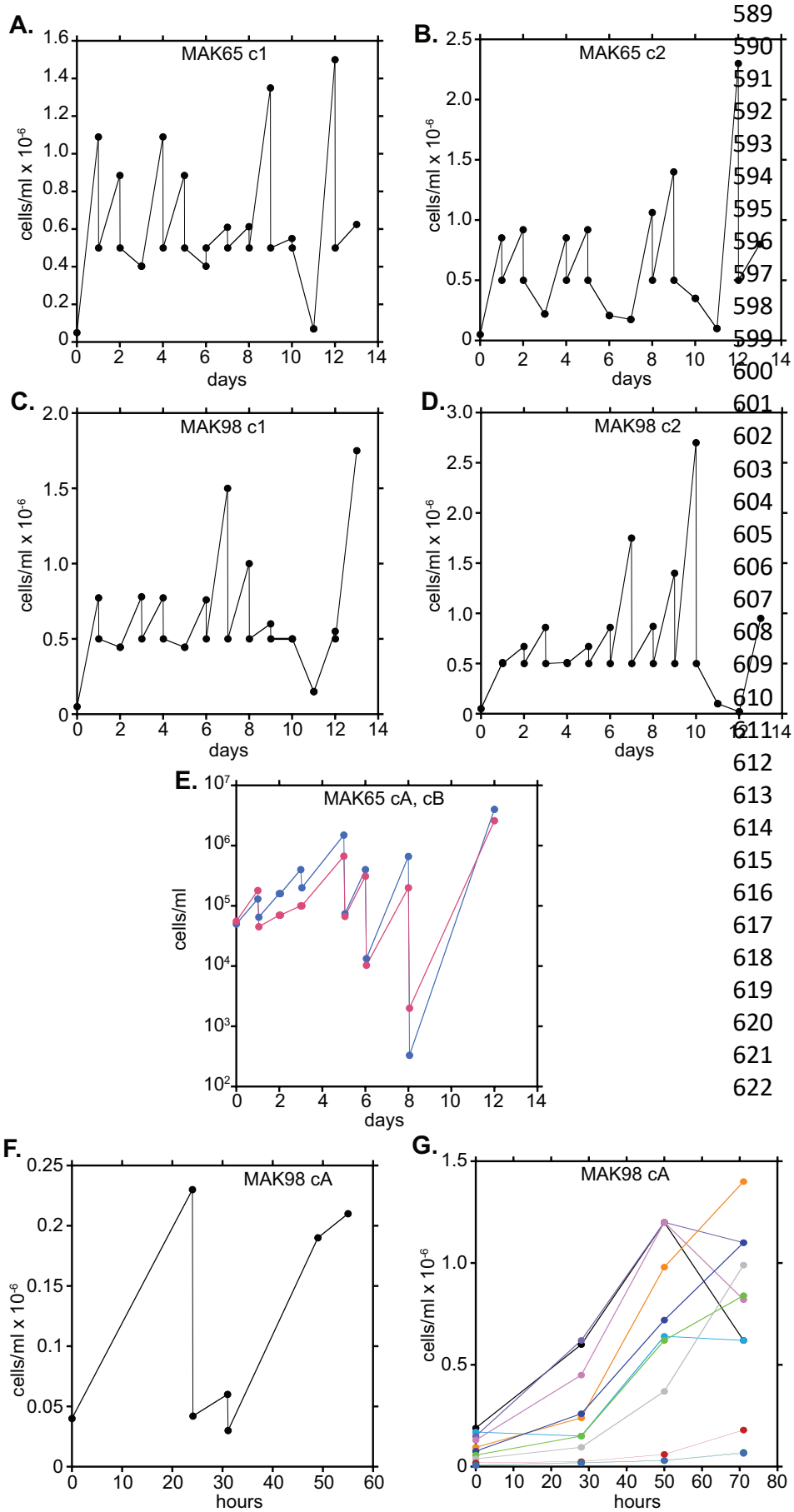


Figure 2. Parasite densities in cultures from blood of mice infected with MAK65 or MAK98. A, B: cultures from MAK65; C,D: cultures from MAK98; E: MAK65 cultures that were continued for DNA preparation (cultures A and B); F: MAK98 culture used both for "original strain" DNA preparation and for prolonged culture (cultures MAK98 A and B). G: MAK98 cultures initiated at different densities in order to determine the maximal cell density.

Figure 3.

Copy number distributions for different trypanosome strains and cultures.

After genome sequencing all reads were allowed to align 20 times. For a set of unique genes, containing one representative each for repeated genes, reads per million reads were calculated. Copy numbers for each gene were than calculated based on the assumption that most genes are present once in every haploid genome (or 2 per diploid genome). The numbers of genes with copy numbers between 1 and 4 are plotted here for the strains and cultures indicated.

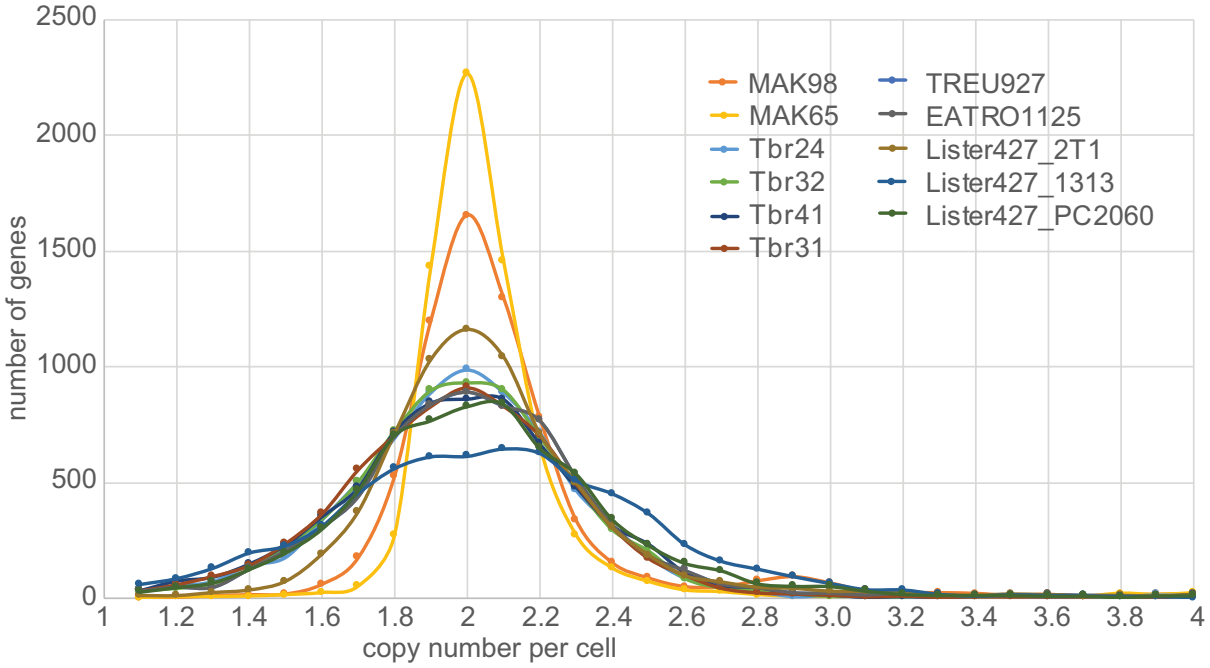


Figure 4.

Gene copy numbers of older lines plotted across the genome.

Copy numbers were determined as in Figure 3. The approximate chromosome boundaries are indicated on the plot; columns exceeding 8 copies are runcated at the top of each panel. Red bars indicate departures from diploidy. The strain or culture is indicated on each panel. A: TREU927 (genome strain); B: Cultured EATRO1125 bloodstream forms expressing the tet repressor; C: Lister 427 2T1 bloodstream forms expressing the tet repressor and T7 polymerase (427_2T1) [37]; D: Lister 427 bloodstream forms expressing the tet repressor (427_1313) [23]; E: Lister 427 procyclic forms (427_PC2060) expressing the tet repressor; F, G: *T. b. rhodesiense* from humans, passaged 3 times in rodents [12].

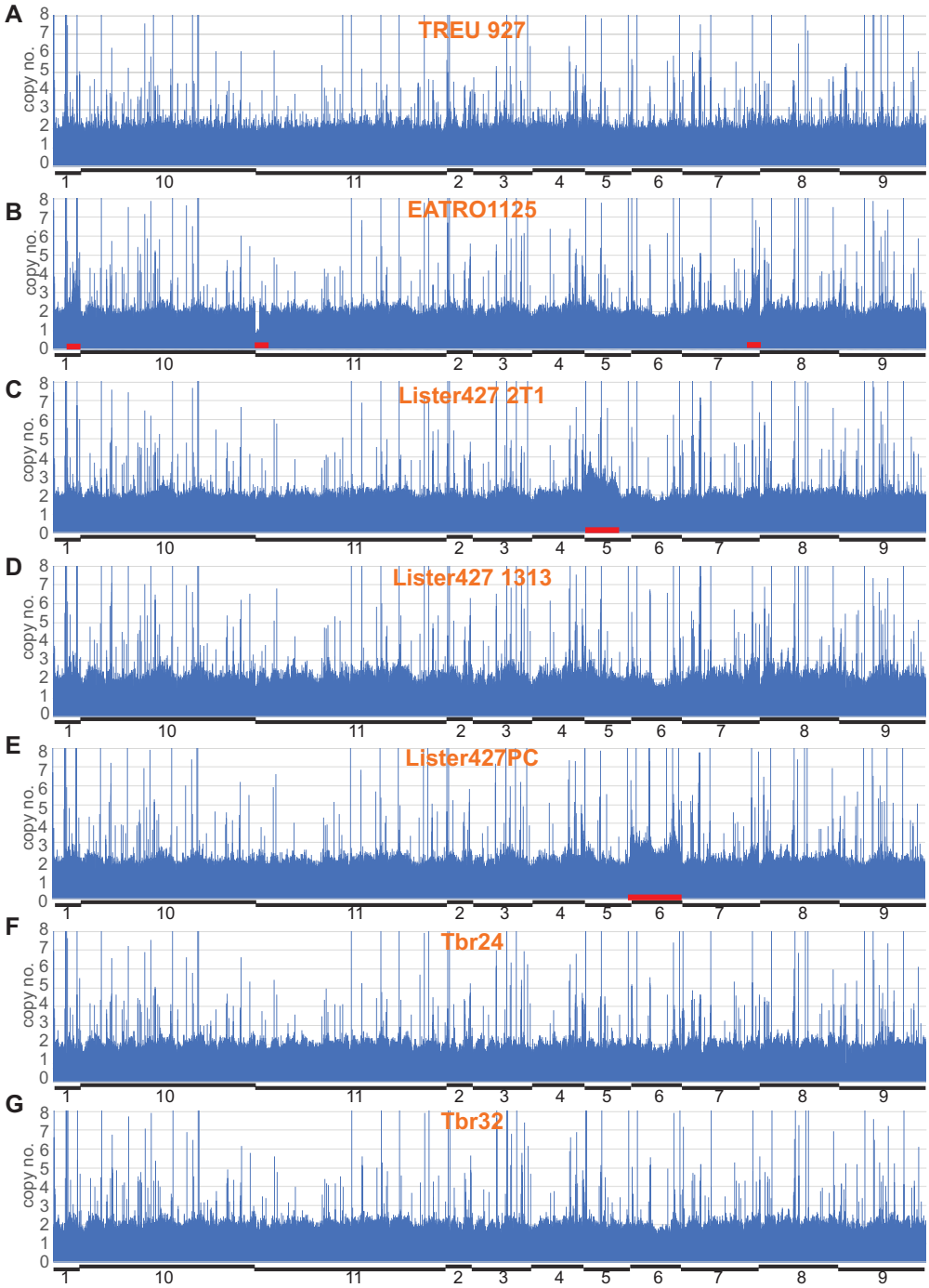


Figure 5.

Gene copy numbers for original and cultured MAK65 and MAK98

A. Copy numbers for original and cultured MAK65, as in Figure 3.

B. Copy numbers for original and cultured MAK98, as in Figure 3.

C - G: Copy numbers plotted across the genome, as in Figure 4. The courses of the cultures are shown in supplementary Figure 2.

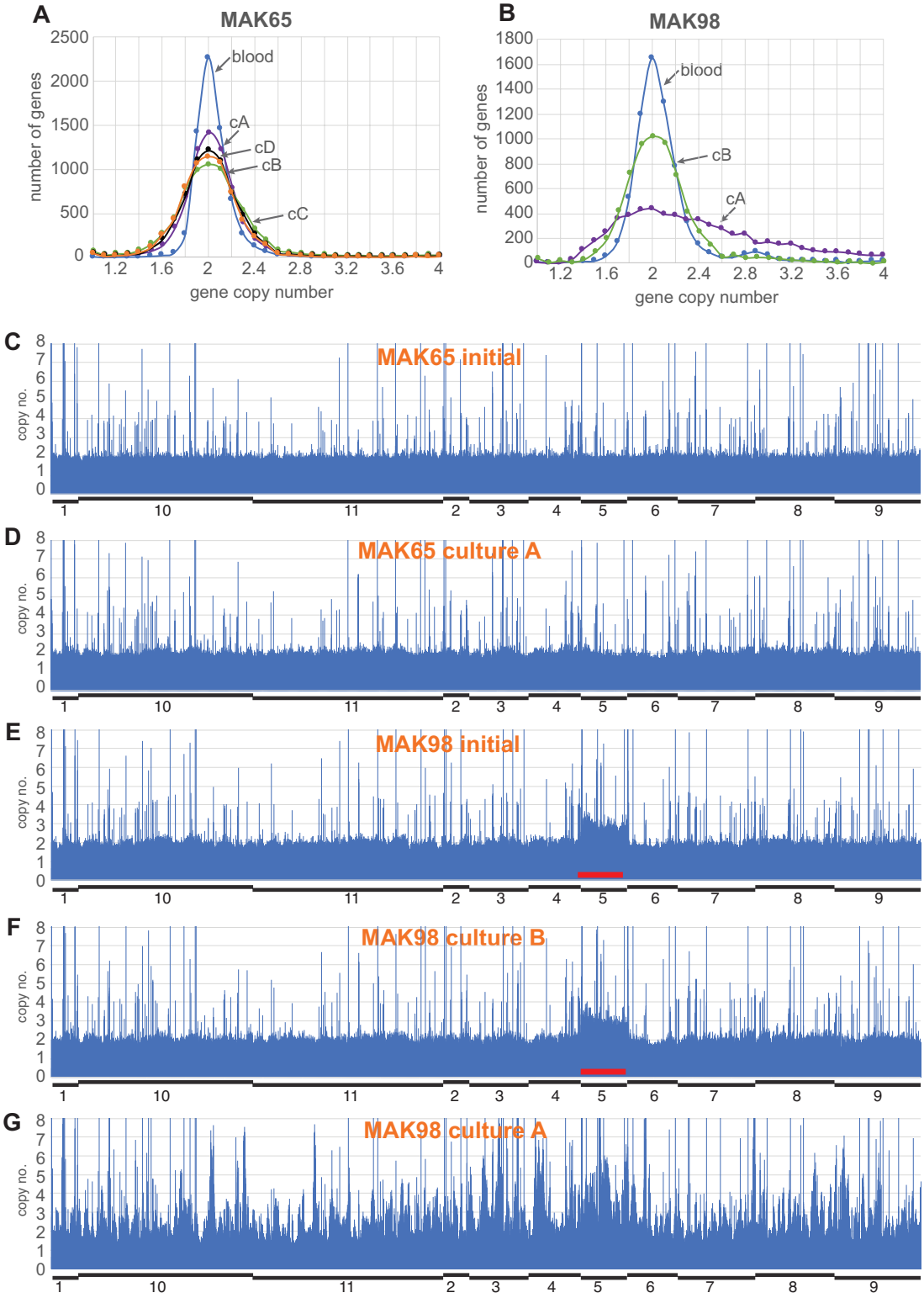
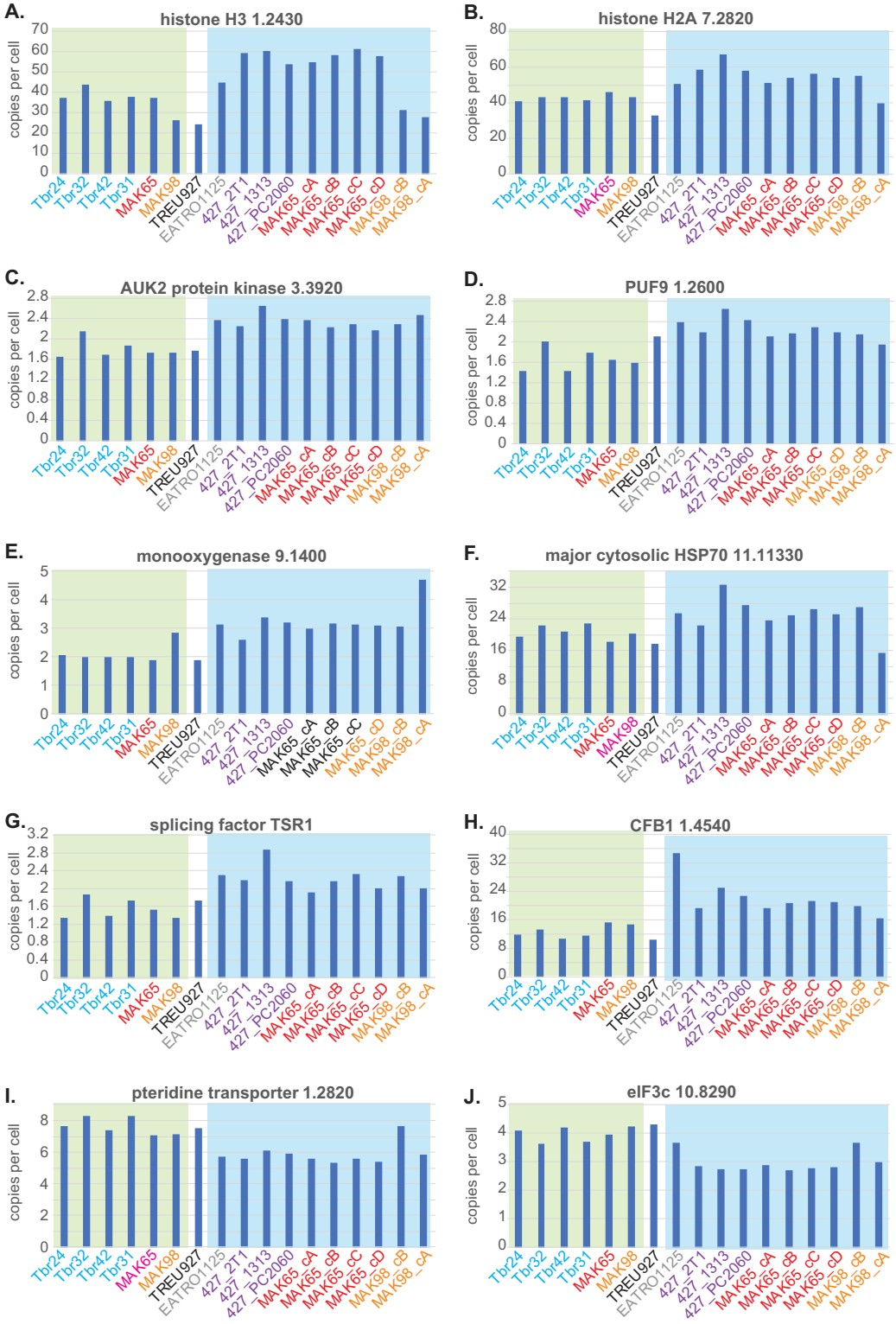


Figure 6.

Copy numbers of selected genes that appear to be influenced by culture. The gene products and numbers are shown above each graph. Columns with green underlay are for non-culture-selected parasites, and those with blue underlay were culture-selected. Colours indicate the origin. The four T. b. rhodesiense isolates (Tbr) are from independent patients but have been assigned the same colour because they were all taken from a single clinic. The same colours are used in Supplementary table 2.



Supplements

Supplementary text 1: Preliminary characterization

1. *T. brucei* isolates from cattle grown in mice

We inoculated and successfully grew 3 recently isolated *Trypanosoma brucei brucei* strains from cattle (Table 1) and made fresh stabilates.

Table 1

Isolate ID	Simplified name	Source Village/Parish	Source District	Date of Isolation
Tb065BAPC	MAK65	Bunya	Apac	25/5/2016
Tb236BAPC		Bunya	Apac	25/5/2016
Tb098AAPC	MAK98	Apuru	Apac	30/7/2017

For simplicity Tb065BAPC and Tb098AAPC are designated MAK65 and MAK98 in the paper, emphasizing their original characterization at Makerere University.

2. PCR characterization

Internal transcribed spacer (ITS)

We confirmed *Trypanozoon* status in all isolates by carrying out internal transcribed space (ITS) PCR, which yields a band size of approximately 480bp. We confirmed that the parasites were *T. brucei brucei,* and therefore not infective for humans, by PCR for the *SRA* gene [1]. Human-infective *T. brucei rhodesiense* yield a product of 284bp. In both cases *T. b. rhodesiense* LW042 [2] was used as a positive control.

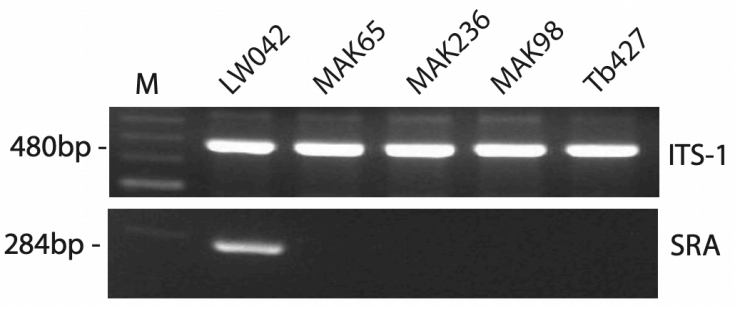


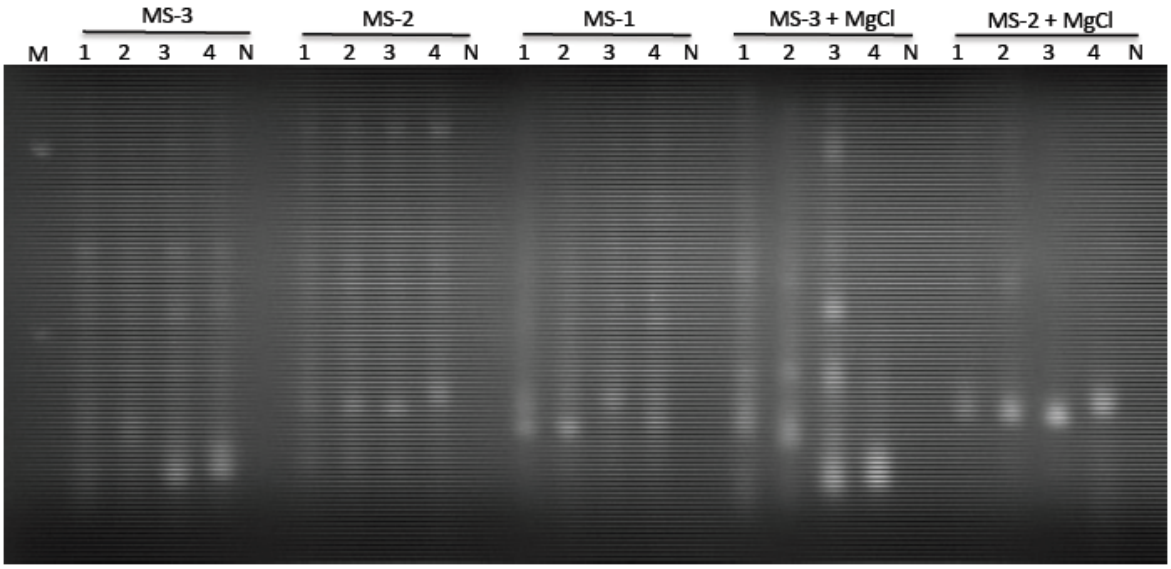
Figure 1: PCR confirmation that MAK65 and MAK98 are *T. brucei brucei*: positive PCR for ITS-1 and negative result for *SRA*.

Microsatellites

Genotyping of the *T. brucei* strains was carried out using microsatellites designated for *T. brucei* (Table 1) as described in [3].

The three recently isolated strains were again compared with *T. b. rhodesiense* human isolate LW042 (Figure 2). Tb236B and MAK65 were similar and could be distinguished from both MAK98 and LW042. This was probably because Tb236B and MAK65 were isolated from cattle in the same village/parish.

696 Figure 2: Microsatellite analysis.



697

698 Microsatellite analysis of 1-Tb236B, 2-Tb065B, 3-LW042, 4-Tb098A using the primers
699 shown in Table 2.

700 **Table 2**

Locus (code)	Primer (outer)	Sequence (5'-3')	Primer (nested)	Sequence (5'-3')
Ch2/PLC (MS-1)	PLC-G2 PLC-H4	ttaagtggacgacgaaataacaaca ttcaaacaccgtccccctcaataat	2/PLC-G 2/PLC-H3	caacgacgttggaagagtgtgaac ccactgacctttcatttgatcgctttc
Ch4/M12C12 (MS-2)	M12C12-C M12C12-B	aaacctcatccagtcgcactgg taccctcatcaagtggtcg	M12C12-A M12C12-D	tggacacacagaagcctaccg agtgtggtggtgcgtgcaaacttgg
Ch5/JS2 (MS-3)	JS2-C JS2-D	agtaatgggaatgagcgtcaccag gatcttcgcttacacaagcggtac	JS2-AFAM JS2-B	gattggcgcaacaactttcacatacg ctttcttccttggccattgttttactat

701

702

703

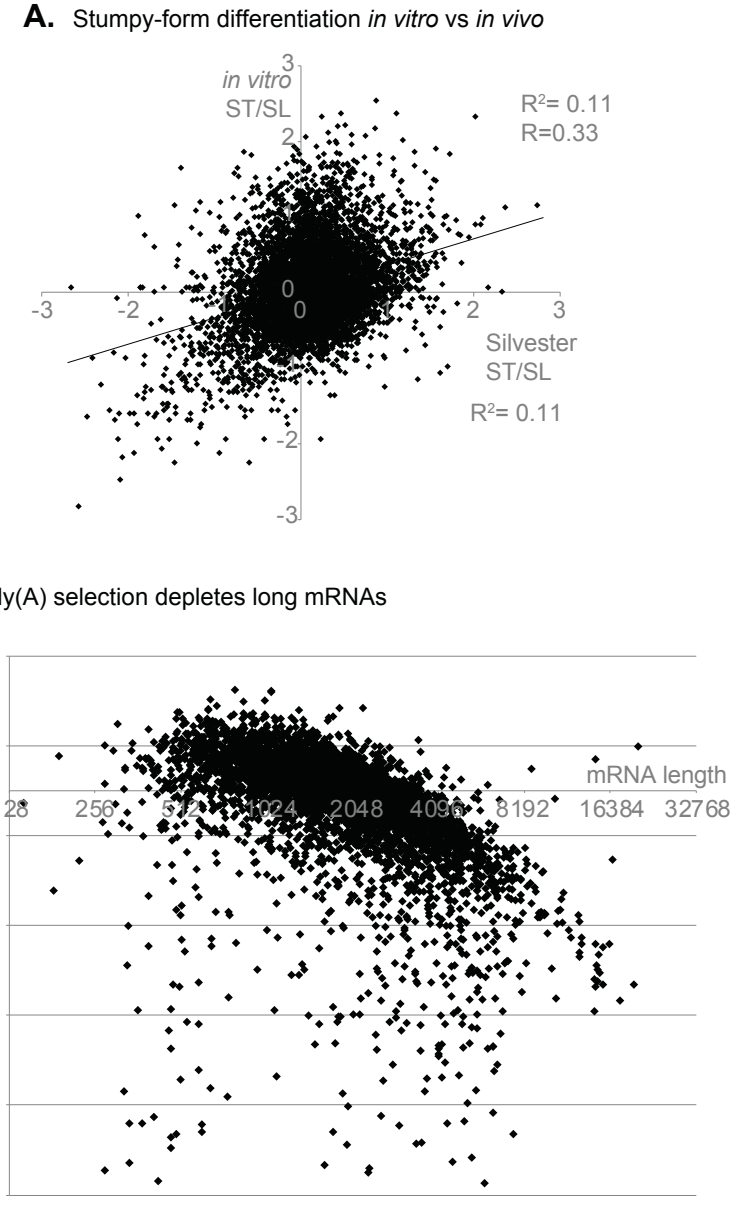
704 1. Radwanska M, Chamekh M, Vanhamme L, Claes F, Magez S, Magnus E, et al. The
705 serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma*
706 *brucei rhodesiense*. Am J Trop Med Hyg. 2002;67:684-90.
707 2. Mulindwa J, Mercé C, Matovu E, Enyaru J, Clayton C. Transcriptomes of newly-
708 isolated *Trypanosoma brucei rhodesiense* reveal hundreds of mRNAs that are co-regulated
709 with stumpy-form markers. BMC genomics. 2015;16:1118. PubMed Central PMCID:
710 PMC10.1186/s12864-015-2338-y, PMID: 26715446.
711 3. Kato CD, Alibu VP, Nanteza A, Mugasa CM, Matovu E. Population genetic structure
712 and temporal stability among *Trypanosoma brucei rhodesiense* isolates in Uganda. Parasit
713 Vectors. 2016;9:259. Epub 2016/05/05. doi: 10.1186/s13071-016-1542-1. PubMed PMID:
714 27142001; PubMed Central PMCID: PMCPMC4855840.

715

716

717

718 **S1 Figure**
719 Transcriptome comparisons.
720 A. Comparison of published results for *in vitro* [18] and *in vivo* [17]. In each case the $\log_2$
721 ratio of stumpy-form to long-slender form EATRO1125 is shown.
722 **B.** Total RNA from MAK98 trypanosomes (sample A) was either selected on oligo d(T) to give
723 poly(A)+ RNA, or treated with RNase H and oligonucleotides complementary to the rRNA in
724 order to give ribo-minus RNA. The $\log_2$ ratio of poly(A)+ to ribo-minus was is on the y-axis
725 and the annotated mRNA length (log scale) on the x-axis. Results are in Supplementary table
726 S1, sheets 2 and 5.



727
728
729
730

S2 Figure

Cultures of MAK65 and MAK98 used for genome sequencing.

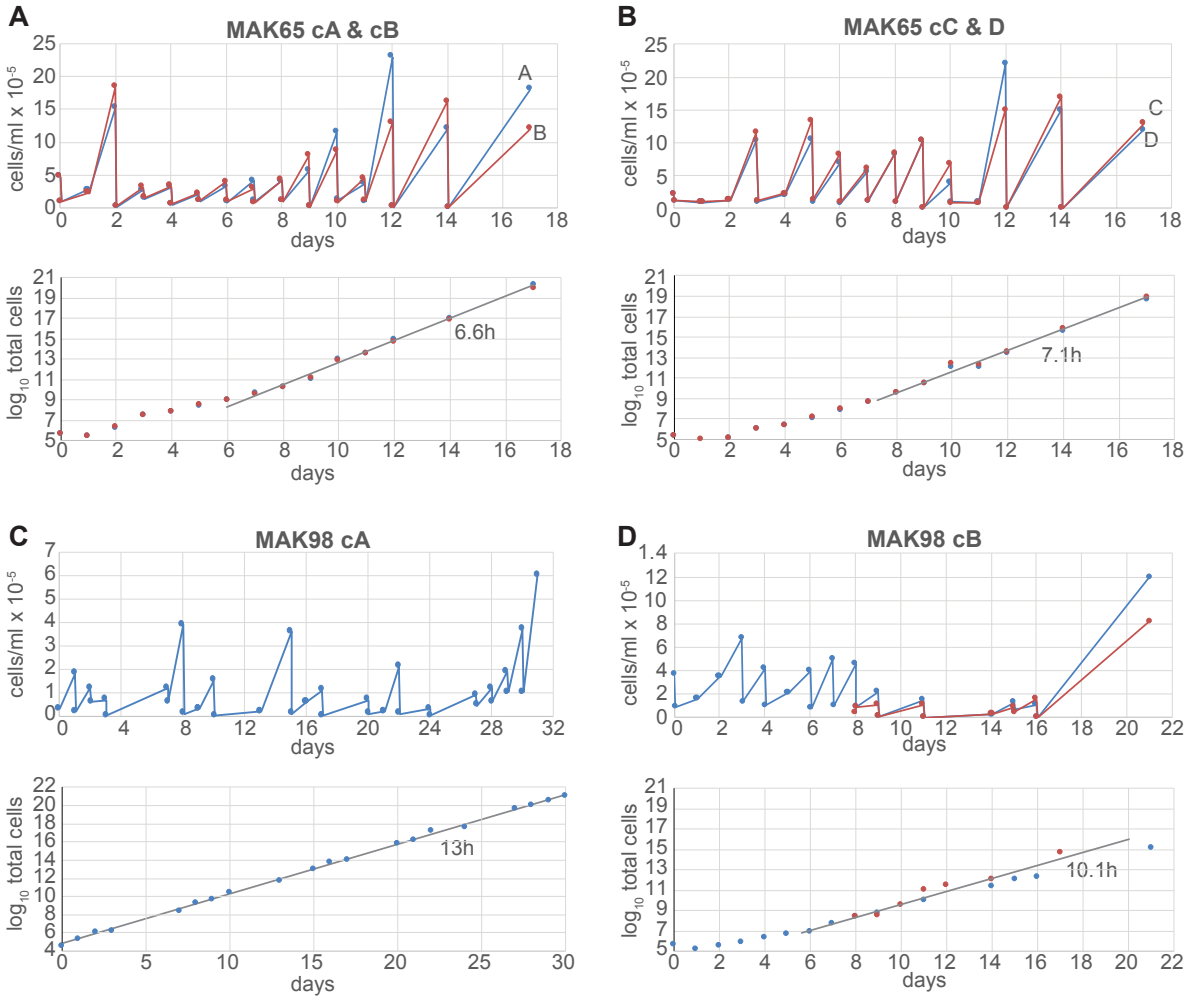
In each case the upper panel shows parasite densities on a linear scale and the lower panel shows cumulative parasite numbers on a log scale. The lines on the log scale graphs indicate the part used for division time calculations. In each case DNA was harvested at the end of the culture period, and stabilates were made.

A. MAK65 cultures A and B (cA and cB) were continued from those shown in Figure 2E to give a total culture time of 30 days.

B. MAK65 cultures C and D (cC and cD) were freshly initiated from blood stabilates.

C. Mak98 culture A (cA) was continued directly from Figure 2F.

D. Mak98 culture B (cB) was continued from a frozen stabilate from Figure 2F. For genome analyses results from the two final cultures were >99% identical so the counts were pooled.

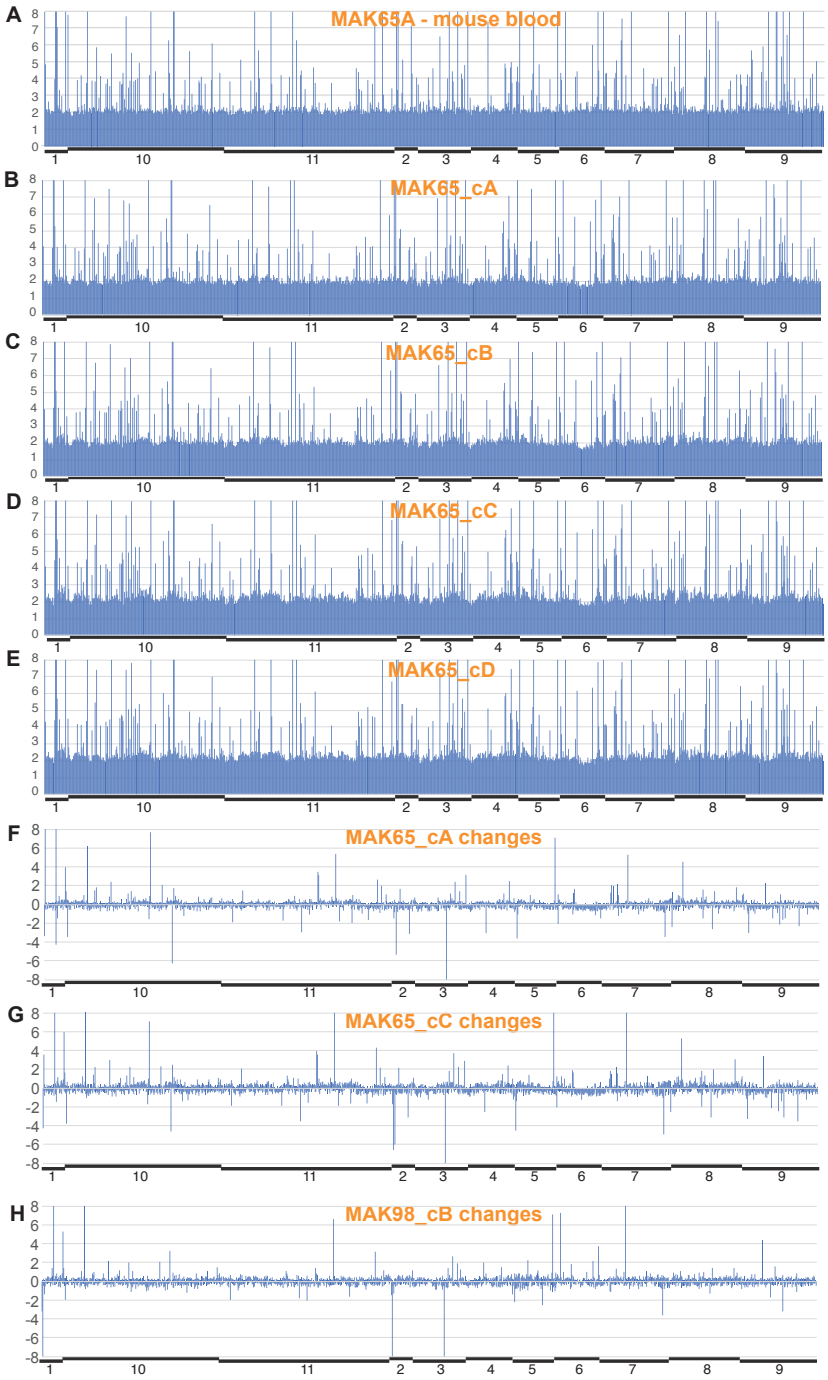


S3 Figure

Gene copy number distributions for replicate cultures of MAK65, and copy-number differences.

A-E. Copy numbers were determined as in Figure 3. The approximate chromosome boundaries are indicated on the plot; columns exceeding 8 copies are runcated at the top of each panel. Red bars indicate departures from diploidy. The strain or culture is indicated on each panel.

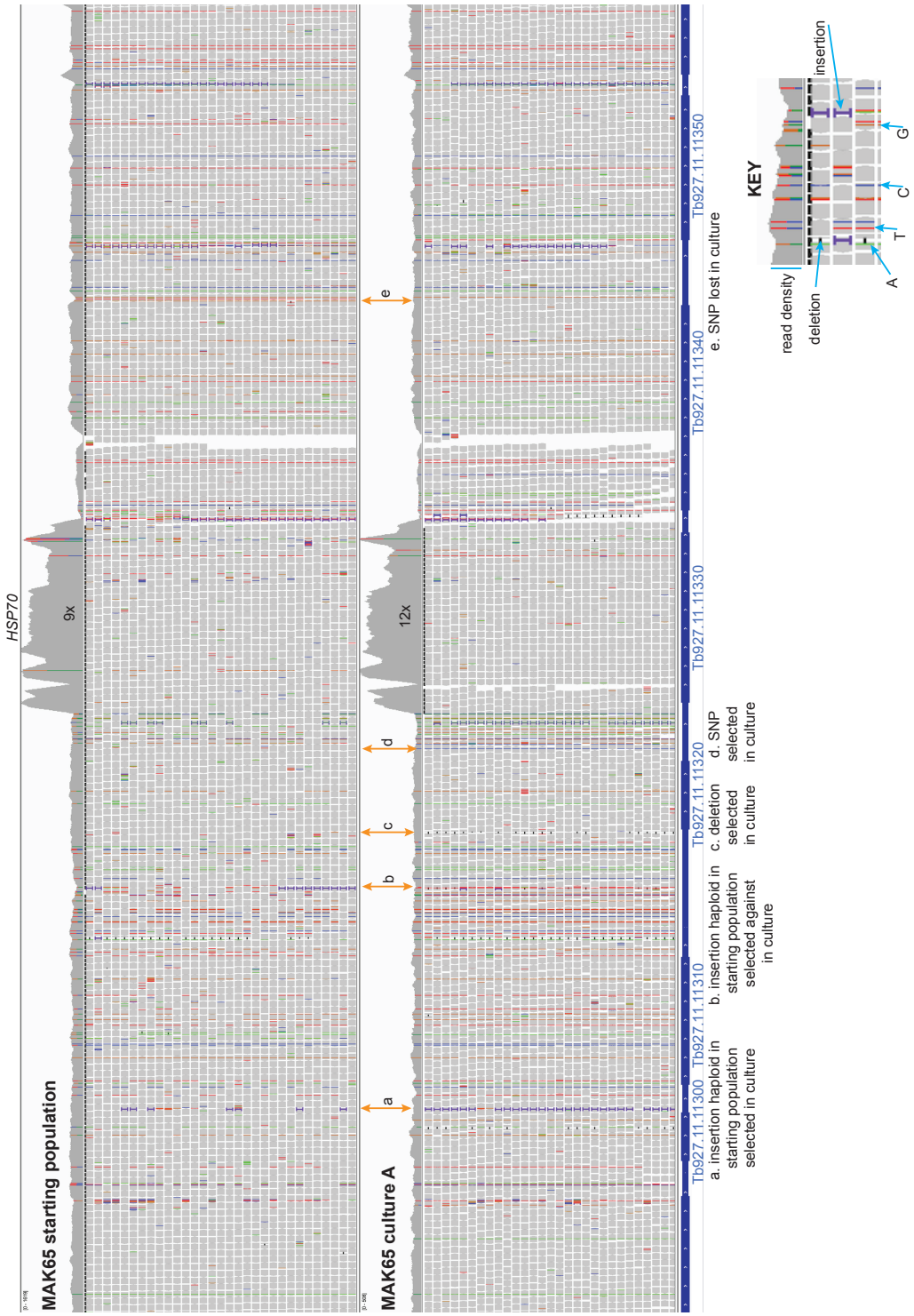
F-H: The plots are the same except that for each culture shown, the copy number in the starting population was subtracted.



S4 Figure

A detailed comparison for one MAK65 culture

All reads for MAK65 starting population, and culture A, were allowed to align once to the TREU927 genome. The resulting mapped reads were visualized using the Integrated genome viewer (Broad Institute). The region surrounding the gene encoding the major cytosolic HSP70 is shown. A key is below the alignment and few differences between the genomes are highlighted. The relative copy number for *HSP70* can be seen by comparing its read density with that over the surrounding single-copy regions.



768 **S1 Table**
769 Transcriptomes of MAK65 and MAK98 trypanosomes grown in rats. For details see the top
770 sheet of the table.

771 **S2 Table**
772 Gene copy numbers for MAK65 and MAK98 trypanosomes grown in rats and in culture. For
773 details see the top sheet of the table.

774