Babesia bovis Rad51 ortholog influences switching of *ves* genes but is not essential for segmental gene conversion in antigenic variation

Erin A. Mack[1#¶], Massimiliano S. Tagliamonte[2,3#], Yu-Ping Xiao[1], Samantha Quesada[1], and David R. Allred[1,2,3,4*]

[1]Department of Infectious Diseases and Immunology, University of Florida; [2]Department of Pathology, Immunology, and Laboratory Medicine, University of Florida; [3]Emerging Pathogens Institute, University of Florida; [4]Genetics Institute, University of Florida, Gainesville, Florida, United States of America

[*]Corresponding author
email: allredd@ufl.edu (DRA)

#These individuals contributed equally to this project.

¶Current address: Department of Biological Sciences, University of Idaho, Moscow, Idaho, United States of America.

short title: Rad51 and antigenic variation in *Babesia*

Abstract

The tick-borne apicomplexan parasite, *Babesia bovis*, a highly persistent bovine pathogen, expresses VESA1 proteins on the infected erythrocyte surface to mediate cytoadhesion. The cytoadhesion ligand, VESA1, which protects the parasite from splenic passage, is itself protected from a host immune response by rapid antigenic variation. *B. bovis* relies upon segmental gene conversion (SGC) as a major mechanism to vary VESA1 structure. Gene conversion has been considered a form of homologous recombination (HR), a process for which Rad51 proteins are considered pivotal components. This makes BbRad51 a choice target for development of inhibitors that could both interfere with parasite genome integrity and disrupt HR-dependent antigenic variation. Previously, we knocked out the Bb*rad51* gene from the *B. bovis* haploid genome, resulting in a phenotype of sensitivity to methylmethane sulfonate (MMS) and apparent loss of HR-dependent integration of exogenous DNA. In a further characterization of BbRad51, we demonstrate here a failure to upregulate the Bbrad51 gene in response to DNA damage. Moreover, we demonstrate that ΔBb*rad51* parasites are not more sensitive than wild-type to DNA damage induced by γ-irradiation, and repair their genome with similar kinetics. To assess the need for BbRad51 in SGC, RT-PCR was used to observe alterations to a highly variant region of *ves1*α transcripts over time. Mapping of these amplicons to the genome revealed a significant reduction of in situ transcriptional switching (isTS) among *ves* loci, but not cessation. By combining existing pipelines for analysis of the amplicons, we demonstrate that SGC continues unabated in ΔBb*rad51* parasites, albeit at an overall reduced rate, and a reduction in SGC tract lengths was observed. By contrast, no differences were observed in the lengths of homologous sequences at which recombination occurred. These results indicate that, whereas BbRad51 is not essential to babesial antigenic variation, it influences epigenetic control of *ves* loci, and its absence significantly reduces successful variation. These results necessitate a reconsideration of the likely enzymatic mechanism(s) underlying SGC and suggest the existence of additional targets for development of small molecule inhibitors.

Author summary

B. bovis establishes highly persistent infections in cattle, in part by using cytoadhesion

54 to avoid passage through the spleen. While protective, a host antibody response
55 targeting the cytoadhesion ligand is quickly rendered ineffective by antigenic variation.
56 In *B. bovis*, antigenic variation relies heavily upon segmental gene conversion (SGC),
57 presumed to be a form of homologous recombination (HR), to generate variants. As
58 Rad51 is generally considered essential to HR, we investigated its contribution to SGC.
59 While diminishing the parasite's capacity for HR-dependent integration of exogenous
60 DNA, the loss of BbRad51 did not affect the parasite's sensitivity to ionizing radiation,
61 overall genome stability, or competence for SGC. Instead, loss of BbRad51 diminished
62 the extent of in situ transcriptional switching (isTS) among *ves* gene loci, the
63 accumulation of SGC recombinants, and the mean lengths of SGC sequence tracts.
64 Given the overall reductions in VESA1 variability, compromise of the parasite's capacity
65 for in vivo persistence is predicted.

67 ## Introduction

68 The apicomplexan parasite, *Babesia bovis*, is a quick-change artist with the ability to
69 rapidly alter proteins it expresses on the infected erythrocyte surface. This ability is
70 needed because, during asexual reproduction *B. bovis* cytoadheres to the capillary and
71 post-capillary venous endothelium within its bovine mammalian host [1-3]. It is thought
72 that this behavior allows this microaerophilic parasite to avoid splenic clearance and to
73 complete asexual development under hypoxic conditions, analogous to the human
74 malarial parasite, *Plasmodium falciparum* [4]. Cytoadhesion is mediated by
75 heterodimeric variant erythrocyte surface antigen-1 (VESA1) proteins, which are
76 exported by the parasites to the erythrocyte where they integrate into the erythrocyte
77 membrane [5, 6]. Cytoadhesion is compromised by a host antibody response targeting
78 VESA1, preventing and potentially reversing cytoadhesion [6]. However, *B. bovis* has
79 evolved the ability to rapidly vary the structure and antigenicity of VESA1 polypeptides,
80 abrogating recognition by existing antibodies [7-10]. Antigenic variation in *B. bovis*
81 involves the *ves* multigene family encoding VESA1a and 1b polypeptides [9, 11, 12], and
82 possibly the *smorf* multigene family [13, 14]. *B. bovis* intraerythrocytic stages reproduce
83 asexually, with a haploid genome of only 8-8.5 Mbp [13]. Despite its small genome size,
84 approximately 135 genes comprise the *B. bovis ves* multigene family, amounting to
85 approximately 4.7% of all coding sequences [10, 13]. Transcription of *ves* genes is
86 monoparalogous, arising from a single *ves* locus at any one time (but typically involves
87 transcription of both a *ves*1α and *ves*1β gene from the same locus to encode both
88 subunits), whereas the remainder of the family remains transcriptionally inactive [15].
89 In situ transcriptional switching (isTS) from one *ves* locus to another over time has been
90 implicated in *B. bovis* antigenic variation [10, 15, 16], although segmental gene
91 conversion (SGC) is the only mechanism of variation critically demonstrated in this
92 parasite to date [12]. Progressive replacement of short sequence patches within the
93 actively transcribed *ves* genes by SGC, yields *ves* genes (and VESA1 polypeptides) that
94 are mosaics comprised of sequences from many *ves* loci [11, 12]. The short lengths of the
95 SGC conversion tracts, ability to acquire sequences from any chromosome, and
96 involvement of two similarly variant subunits may enable this gene family to provide
97 practically unlimited diversity in epitope structure [5, 7, 9, 11, 12].
98

99 Canonical gene conversion is a form of homologous recombination (HR)-mediated DNA
100 repair. In this process a damaged sequence is repaired by incorporating duplicated
101 homologous sequences from an undamaged allele or paralog to replace the damaged
102 sequences. At least three models have been proposed to explain this process (reviewed
103 in [17]). Common to all models is the assembly of a repair complex at the site of damage.
104 Activities of the repair complex include 5' to 3' resection of the broken ends of the
105 damaged molecule, and stabilization of the single-stranded 3' ends by assembly of RPA
106 ssDNA-binding proteins onto the strands (reviewed in [17, 18]). Rad52 may stabilize the
107 RPA and maintain spatial proximity of the broken ends [19]. The Rad52-RPA complexes
108 then are replaced by Rad51 protein, which forms helical filaments on the ssDNA
109 strands. The Rad51-ssDNA complexes, together with Rad54, mediate both a search for
110 homologous sequences elsewhere in the genome and strand invasion when such
111 sequences are found [20-22]. Once found, single-stranded invasion of the identified
112 sequence allows sequence acquisition by extension of the 3' end of the invading strand.
113 Depending upon the model, the inter-chromosomal entanglement involves either one or
114 two Holliday junction structures that are resolved to yield two independent
115 chromosomes again. Sequence acquisition may occur with or without crossover,
116 depending upon how the junction structures are resolved [23]. Like canonical gene
117 conversion, we hypothesized SGC to occur via HR, a possibility consistent with the
118 stretches of homologous sequence flanking SGC tracts that are shared between donor
119 and recipient. However, other factors call this into question. For example, the reasons
120 for consistently short conversion tracts are unknown, and crossover appears to be a rare
121 outcome in *B. bovis* [10]. These traits suggest that minimal end resection, acquisition of
122 only short tracts of differing sequence, and rapid resolution of intermolecular junction
123 structures all may define this process mechanistically. Among the existing models of
124 HR, those most consistent with these traits are synthesis-dependent strand annealing,
125 either with rapid disentanglement of the invading strand(s), or double-strand break
126 repair but with convergent branch migration and junction dissolution. By contrast,
127 resolution of distal Holliday junctions would be expected to result in frequent crossover
128 events [17]. Alternatively, a more exotic explanation may hold, such as the involvement
129 of a template-switching repair polymerase, but no direct evidence yet supports this
130 possibility.

131

132 Rad51 is considered essential to HR and the gene conversion process in eukaryotes.
133 Organisms with defective Rad51 consistently suffer reduced viability and enhanced
134 sensitivity to environmental insult [24-29]. For example, knockout of the *Rad51* gene in
135 mice resulted in embryonic lethality [30], whereas loss of the Tb*rad51* gene in the
136 kinetoplastid parasite, *Trypanosoma brucei*, yielded parasites that were compromised
137 in growth and hypersensitive to methyl methanesulfonate (MMS) [28]. Interestingly,
138 Tb*rad51* knockout parasites continued to undergo variation of their variant surface
139 glycoprotein (*VSG*) genes, both by isTS and gene conversion mechanisms, but the rate at
140 which variation occurred was slowed dramatically. This observation suggested a role for
141 TbRad51 in facilitating or regulating trypanosomal antigenic variation, but not an
142 essential role in catalysis. Moreover, TbRad51-independent mechanisms may act in
143 trypanosomal antigenic variation. Recently, we knocked out the Bb*rad51* gene of *B.*
144 *bovis*. Unlike higher eukaryotes, there was no apparent effect on parasite viability or
145 growth. However, parasites were made hypersensitive to MMS and failed to integrate
146 exogenous DNA, suggesting defects in HR [31]. Given the importance of SGC to *B. bovis*
147 antigenic variation and survival, and of Rad51 to gene conversion and HR, we
148 investigated further the interplay between DNA repair and antigenic variation. We
149 provide evidence that overall DNA repair remains highly robust in the absence of
150 BbRad51. SGC also continues, albeit at an overall reduced rate, concomitant with a
151 significant reduction in isTS. We hypothesize that these results reflect unique roles for
152 BbRad51 in antigenic variation, and suggest that alternative enzymes catalyze
153 recombination during SGC.

154

155 **Results**

156 ***Bbrad51 knockout parasites behave similarly to wild-type in response to***
157 ***DNA damage.*** Previously, we knocked out the Bb*rad51* gene, resulting in a parasite
158 phenotype of MMS-sensitivity and apparent loss of the ability to integrate exogenous
159 DNA by HR, but not significant difference in rates of growth [31]. Cells that suffer
160 environmental insult may up-regulate DNA repair proteins, including Rad51 [32]. To
161 test this in *B. bovis* we needed to create a reporter parasite line with which we could
162 easily quantify BbRad51 protein levels. This was done by integrating the linearized

163 plasmid, pBRNHLg, into the genome by homologous double-crossover, resulting in
164 Bb*rad51* coding sequences fused in-frame with NanoLuciferase and dual-HA tag
165 sequences. Immediately following the stop codon, the *E. coli glmS* ribozyme element, a
166 self-cleaving autoregulatory sequence controlling glucosamine-6-phosphate synthase
167 expression [33], was also incorporated. Upon double crossover integration, a *gfp-bsd*
168 selectable marker cassette also becomes incorporated downstream of the gene in this
169 approach [34]. This combination allowed us to conveniently measure levels of BbRad51
170 protein via the activity of the NanoLuciferase fusion partner, and to inducibly knock
171 down expression via the *glmS* ribozyme element by the addition of the D-glucosamine
172 cofactor [33], an approach used previously in *Plasmodium* [35]. As a negative control
173 for induction, samples of *B. bovis* CE11 parasites were transiently transfected with the
174 plasmid pHggb, in which NanoLuciferase-2xHA expression is driven by the *Babesia*
175 *divergens EF1*α-B promoter. This promoter was chosen because it would be unlikely to
176 respond to upregulation of DNA repair proteins. Structures of these plasmids are shown
177 in S1 Figure, and sequences are provided in S2_File and S3_File. Reporter parasites
178 were exposed to 250 μM MMS, a level that reproducibly inhibits growth by
179 approximately 10% after 48 hours in wild-type *B. bovis* CE11 parasites [31], presumably
180 sufficient to induce Bb*rad51* expression in response but low enough not to significantly
181 reduce viability. Under these conditions, no evidence of an alteration in BbRad51
182 expression was observed after either 3 hours (Figure 1A) or 16 hours of post-treatment
183 recovery (Figure 1B). Treatment with glcN resulted in significant knockdown of
184 NanoLuciferase levels, on the order of 80-95%. No upregulation of BbRad51 was
185 observed if knockdown of existing BbRad51 via the *glmS* element was performed. This
186 was done to allow more sensitive detection in the case of post-translational stabilization,
187 or to prevent any possible negative feedback (Figure 1). This result indicates that
188 BbRad51 levels are not regulated in response to repair of alkylated DNA following acute,
189 sublethal exposure to MMS.
190
191 A second measure of phenotype commonly used in studies of DNA repair is sensitivity to
192 γ-irradiation. Sensitivity of the knockout parasites to DNA damage was assessed by
193 exposure to γ-irradiation provided by a Cs^{137} source. Wild type CE11 parasites first were
194 titrated for sensitivity, over a range from 10 - 1230 gray (Gy). A dosage of 100 Gy

195 resulted in approximately 80-90% lethality by 24 hours, but allowed some parasite
196 survival at 48 hours and beyond, whereas at 200 Gy no parasites were observed to
197 survive 48 hours post-irradiation (S2 Figure). Radiation sensitivities of three
198 independently-derived Bb*rad51* knockout lines (all on a CE11 genetic background)
199 therefore was compared with CE11 wild-type parasites over a range from 0-200 Gy.
200 There were no reproducible differences among any of the three knockout lines and CE11
201 wild type parasites in growth assays (Figure 2).
202
203 Because no differential survival phenotype was apparent following DNA damage by γ-
204 irradiation, we asked whether there were any detectable differences in overall rates of
205 DNA repair. To assess this, pulsed-field gel electrophoresis (PFGE) was used to monitor
206 the disintegration of chromosomes and their subsequent reassembly. Parasites were
207 subjected to 100 Gy γ-irradiation, allowed to recover for up to 24 hours, and then were
208 processed for PFGE analysis. In growth experiments, 100 Gy had resulted in killing of
209 80-90% of the parasites. Consistent with this level of killing, 100 Gy γ-irradiation
210 severely damaged *B. bovis* chromosomes, virtually eliminating full-length chromosomes
211 3 and 4, and greatly diminishing the proportion of intact chromosomes 1 and 2.
212 Remarkably, by 24 hours post-irradiation both wild-type and knockout parasites had
213 reassembled their genomes into full-length chromosomes of apparently normal size,
214 recovering approximately 50% of non-irradiated control values (Figure 3). Given the 10-
215 20% viability of parasites receiving this dosage (S2 Figure) it is likely that a large
216 proportion is not viable in the longer-term, but remains metabolically active long
217 enough to reassemble chromosomes. These data, when considered together,
218 demonstrate that the loss of BbRad51 has little, if any, effect on the extensive DNA
219 repair required to recover from such damage and suggest that BbRad51-dependent HR
220 plays little role in this type of repair.
221
222 ***Bbrad51 knockout did not prevent in situ transcriptional switching.***
223 Although overall DNA repair following damage from ionizing radiation was not
224 measurably impaired by the loss of BbRad51, sensitivity to MMS and loss of ability to
225 integrate selectable plasmids via long sequence tracts suggested that BbRad51 plays a
226 role in aspects of HR [31], and perhaps in repair of stalled replication forks [36].

227 Previously, it was demonstrated that SGC is a major mechanism of antigenic variation in
228 *B. bovis* [12]. Given the seemingly conflicting outcomes obtained with MMS and
229 ionizing radiation, we wished to determine whether BbRad51 plays any role in SGC. In
230 order to assess the nature of any changes occurring in transcribed *ves1α* genes of wild
231 type and Bb*rad51* knockout parasites, we adapted a previously published assay in which
232 the highly variant cysteine-lysine-rich domain (CKRD) region of *ves1α* transcripts
233 (Figure 4A) is amplified by RT-PCR, and the amplicons undergo deep sequencing [15].
234 Three immediate subclones of CE11 wild-type and three independent knockout clonal
235 lines were studied; their origin is described in [31]. Total RNAs were collected from each
236 clonal line at one month and five months post-cloning. These two timepoints were used
237 to observe for increases in the numbers of unique recombinants over time, and for some
238 analyses were pooled to minimize the loss of unique variants from the population over
239 time. The forward primer, *ves*UniF2, was selected because it represents a sequence
240 almost universally conserved among *ves*1α genes, and in combination with the highly-
241 conserved primer PD1R was anticipated to generate amplicons of approximately 340-
242 460 bp. By constructing bar-coded paired-end libraries from the amplicons and
243 generating 250 bp reads, reads could be merged with high-confidence overlaps of 55 -
244 185 bp. Merged, full-length sequences were obtained from 69.8 - 84.5% of amplicon
245 reads. Following the removal of ambiguous and low-quality reads, adaptor and primer
246 sequences, and sequences found likely to be PCR chimeras, individual libraries ranged
247 from a minimum of 853,688 to a maximum of 1,738,649 merged reads. Mean merged
248 read lengths ranged from 326.2 ± 60.3 to 334.7 ± 52.9 bp. To determine their probable
249 loci of origin, reads were mapped (using non-global settings) onto the *B. bovis* C9.1 line
250 genomic sequence (available at Wellcome Trust Sanger Institute;
251 ftp://ftp.sanger.ac.uk/pub/pathogens/Babesia/). The C9.1 line genome was used
252 because we do not currently have a high quality genome for the CE11 line. However, as
253 these are closely-related clonal sibling lines [12] this allowed us to easily identify the
254 probable locus of origin for nearly all reads. In all six knockout and wild-type lines, the
255 earlier time-point *ves1α* transcripts mapped predominantly, sometimes almost solely, to
256 a single locus (Figure 4B), consistent with prior observations of monoparalogous *ves*
257 gene transcription in the C9.1 clonal line [15]. At the latter time-point, lines ko1^{H5} and
258 CE11^{B8} continued to transcribe almost solely from the original locus, and all lines still

259 transcribed most heavily from the original locus. Minor but significant subpopulations
260 were detected in all lines that had switched to transcription from alternative loci.
261 Without immune pressure there is no obvious selection for parasites expressing specific
262 VESA1a isoforms to predominate. Regardless, detectable transcription occurred from
263 more alternative loci in wild type CE11 subclones B8, C2, and C5 (ranging from 28.7 ±
264 4.0 at 1 month to 33.3 ± 1.5 loci at 5 months) than were observed for the three knockout
265 lines, which ranged from 22.0 ± 3.5 at 1 month to 23.3 ± 1.5 at 5 months (Figure 4C).
266 While transcribing most heavily from a single locus, the CE11 C5 subclone (as a
267 population) also transcribed significantly from several alternative loci at the early time-
268 point, but by the 5-month time-point transcription levels had been reduced from all but
269 the single, major locus. Interestingly, the same alternative loci seemed to dominate as
270 sites to which switching occurred (S3 Figure), suggesting a hierarchy in locus
271 transcription. However, nothing can be inferred from these data regarding an order in
272 switching like that documented for *P. falciparum var* genes [37].
273
274 ***Bbrad51 knockout failed to prevent segmental gene conversion.*** The ideal
275 situation for accurate identification and characterization of SGC tracts would be to
276 directly map amplicon sequences against a reference genome. Although the CE11 and
277 C9.1 lines are closely related sibling clonal progeny of the MO7 clonal line, they have
278 quite different histories [12]. Given the nature and rapidity of SGC many loci would have
279 been extensively modified. The ideal would be to map to a high quality genome from
280 each subclone, so that unique variants in each line would be known, but this was not
281 feasible. As a more feasible alternative, we chose to observe for recombination among
282 loci represented by the *ves1α* transcript amplicons, considering only sequences unique
283 to a given line, on a line by line basis. By taking this approach, we could assess for
284 recombination among sequences that were definitively present in each line at the time
285 the experiment was performed. Using an RT-PCR strategy employed previously to
286 characterize variation in *ves1α* transcripts [15], we identified all unique transcribed
287 sequences by cluster analysis, then identified representative reads for each cluster.
288 Among those, we then identified sequences for which there was very strong statistical
289 support for them being the result of a true recombination event between two other
290 unique sequences, based upon a consensus of 4 out of 7 statistical analyses (see Methods

291 for details). For comparative analysis, we included only those recombinant sequences
292 found in a single clonal line, on the assumption that any sequences found in more than
293 one line was present prior to the act of parasite cloning and not a result of post-cloning
294 recombination. Information on all identified SGC tracts are provided in S1_File.zip. We
295 propose that most of the unique sequences were in fact recombinant, as a plot of all
296 unique sequences against all statistically-supported recombinant sequences (both
297 normalized per million reads) a regression of $R^2 = 0.9228$ was obtained. Thus, the use of
298 a statistical consensus approach was highly conservative and likely underestimates the
299 true number of recombinant sequences (S4 Figure), but allows for rigorous comparative
300 results.
301
302 Variants with strong statistical support as true recombinant SGC tracts were observed
303 among the transcripts of all six clonal lines, but clear distinctions are seen between
304 knockout and wild type. The mean lengths of conversion tracts differed between groups,
305 decreasing from means of 109.01 ± 39.39 in wild type parasites to 91.77 ± 40.78 in
306 knockouts ($p < 0.001$; Figure 5). The distributions of SGC tract lengths, plotted as
307 cumulative proportions of all SGC tracts from that population, resulted in wild type and
308 knockout medians of 107.0 and 81.0 (S5 Figure; $p < 0.001$). Interestingly, the difference
309 between the two populations arose primarily at tract lengths <150 bp. The two
310 populations were not distinct above that length, suggesting the possibility that more
311 than one mechanism gave rise to the SGC tracts. The number of unique SGC tracts
312 arising per active *ves* locus was not significantly different, ranging from 0.35- 1.20 SGC
313 tracts per locus (S6 Figure; $p = 0.22$ among all group comparisons). Although there was
314 a rise in the frequency of recombinants per locus in CE11 wt parasites over time this was
315 not statistically significant ($p = 0.18$). However, given the differences in the numbers of
316 transcriptionally active *ves* loci in wild type and knockout parasites, this led to
317 significantly larger total numbers of unique SGC tracts per million reads among
318 members of the wild type population (Figure 6; at 5 months, $p < 0.02$). Thus, while the
319 frequency of SGC alterations that may be observed at any given transcriptionally active
320 *ves* locus is approximately constant, the numbers of *ves* loci that are activated is
321 significantly higher in the presence of BbRad51. The number of unique SGC tracts per
322 million reads may be considered a surrogate measure of overall levels of variability in

323 antigen structure presented by the population of parasites. Taken together, these data
324 demonstrate clearly that neither SGC nor apparent isTS to alternative loci is abrogated
325 by knockout of the Bb*rad51* gene, although statistically significant quantitative effects
326 on the lengths of SGC tracts and on the frequency and extent of *ves* locus switching were
327 observed.
328
329 ***Involvement of homologous sequences flanking SGC breakpoints.*** We
330 wished to assess whether there is anything shared at SGC breakpoint sites, or unique
331 where SGC tracts differed from the active locus of *ves* transcription (LAT). To do this,
332 alignments were made of each pair of sequences identified as having given rise to a
333 unique recombinant, and the recombinant sequence itself. Regions of homology
334 between all three sequences were then identified manually that represented a region in
335 which transition occurred from one parent locus providing the sequence to the other (S7
336 Figure).. No significant differences were observed among wild type or Bb*rad51* knockout
337 genotype parasites in the lengths of homology patches (Figure 7). While this is clearly
338 not an exhaustive analysis of all SGC tracts, the results of this subsampling indicate that
339 there is no apparent difference between wild type and Bb*rad51* knockout parasites with
340 regard to the lengths of the homologous patches possibly involved in recombination.
341 Importantly, for SGC to occur patches of homology between donor and recipient may
342 not be required, as patches of as little as 2 bp were observed. In a few instances no
343 transition region was present, and in still others there was a brief patch in which no
344 match existed between the three sequences at the site of transition, suggesting a
345 sometimes chaotic process (S7 Figure).
346

347 ## Discussion

348 SGC is a major mechanism of antigenic variation in *B. bovis*, and to date the only one
349 that has been demonstrated critically [12]. This phenomenon, in which short DNA
350 patches are duplicated from a donor to a recipient gene, typically occurs without
351 modification of the donor and at least superficially resembles HR-mediated DNA repair.
352 DNA repair in apicomplexan and other protozoal parasites overall is not well
353 understood [38, 39]. It is even difficult to predict from the parasites' proteomes which
354 repair pathways may be active. Orthologs are recognizable for only a fraction of the

355 proteins known to be important in higher eukaryotes, and many orthologs are simply
356 not present. For example, the proteins RPA, Rad51, Rad52, Rad54, and ATM are
357 considered key participants in DNA repair pathways, including HR and gene
358 conversion. Yet, in *B. bovis* orthologs may be identified only for RPA, Rad51, and Rad54,
359 and many similar examples of "missing" proteins hold [13, 31], suggesting the merging
360 of functions. In this study, we wished to understand the contributions of BbRad51 to
361 SGC because proteins of this family are considered essential to HR and gene conversion
362 in other systems, including other apicomplexans [29, 40]. Our prior identification of
363 BbRad51 as the true Rad51 ortholog was based upon several criteria, including sequence
364 and structural similarities to established Rad51 proteins, a greatly reduced or abrogated
365 ability to achieve HR-dependent integration of exogenous sequences in the absence of
366 BbRad51, and enhanced sensitivity to MMS in knockouts that could be complemented
367 by Bb*rad51* coding sequences [31].
368
369 In contrast with our prior work, we present evidence herein that the absence of BbRad51
370 does not influence the parasite's survival or extent and rate of general repair of dsDNA
371 breaks engendered by acute exposure to γ-irradiation, and that BbRad51 expression fails
372 to up-regulate in response to MMS-induced damage. The apparent insignificance of
373 BbRad51 to repair and survival of IR-induced DSBs may be attributable to the haploid
374 nature of the asexual stages studied here. A large proportion of IR-caused DSBs would
375 occur in unique regions of the genome, with no intact second copy of the damaged
376 sequence available to support true gene conversion, except briefly during mitosis. Thus,
377 in the absence of available sequence donors for HR, BbRad51 may be largely superfluous
378 to surviving heavy dosages of IR, yielding similar ionizing radiation survival outcomes in
379 wild-type and knockout parasites. *B. bovis* may depend instead upon error-prone end-
380 joining reactions for survival of such significant damage. Unlike most eukaryotes,
381 including *Toxoplasma* [41], *B. bovis* lacks genes for key players in canonical non-
382 homologous end-joining repair, such as Ku70/80 and DNA ligase 4 ([13]; this study).
383 This parasite instead may depend upon a synthesis-dependent microhomology-
384 mediated end-joining mechanism like that demonstrated in *P. falciparum* [42], but this
385 remains to be determined. In contrast with ionizing radiation, the absence of BbRad51
386 does render *B. bovis* sensitive to alkylation damage caused by acute exposure to MMS

387 [31]. In diploid organisms recovery from either type of insult is typically compromised
388 by loss of Rad51 [43]. However, in diploids such repair makes frequent use of the second
389 allele for repair through gene conversion. Unlike IR, MMS does not directly cause DSBs,
390 but rather alkylates adenosine and guanidine bases which must be removed and
391 replaced [36]. DSBs still may result when abasic sites or single-strand breaks created
392 during base excision repair of the methylated bases stall or cause the collapse of mitotic
393 replication forks [44]. In diploid cells HR is the major mechanism used in repairing
394 such DSBs during replication, and in at least one mechanism makes use of Rad51 [43,
395 45]. In TbRad51-intact *T. brucei* (a diploid parasite), the repair of DSBs created within
396 *VSG* bloodstream-stage expression sites resulted in a massive increase in the rate of
397 *VSG* switching through gene conversion, presumably due to the ready availability of
398 related sequences [46]. This also may be accomplished in haploid organisms, if
399 appropriate sequences have already been replicated on the opposite strand, but would
400 limit the timing of gene conversion only to S phase, prior to separation of sister
401 chromatids.
402
403 The distribution and structure of *ves* genes might suggest that they have evolved for
404 efficient application of the SGC mechanism. In most microorganisms that undergo
405 antigenic variation the variant multigene families involved typically are arranged in
406 large, subtelomeric clusters [47]. By contrast, *ves* genes are found in numerous small
407 clusters scattered throughout the genome, and often are interspersed with *smorf* genes
408 [10, 13]. Despite extremely high overall variability, *ves* genes possess periodic tracts of
409 highly conserved sequence, and *ves* genes on the same or different chromosomes may
410 provide targets for strand invasion via such conserved tracts [12]. A small-scale
411 chromatin conformation capture (3C) query of sequences in proximity to the transcribed
412 *ves* genes of the LAT suggested close proximity primarily to other *ves* genes [16]. In that
413 organization, the ability to find short patches of homologous sequence in which to
414 initiate strand invasion might occur via those conserved patches. However, when the
415 sequences flanking the breakpoint sites were compared, patches of homologous
416 sequence unrelated to the highly conserved sequence regions were observed at the SGC
417 tract breakpoints between the LAT and donor sequences. These stretches of homologous
418 sequence ranged from 2 bp to 35 bp in length (Figure 7), and in some sequences no

419 identifiable homologous sequences could be identified. These data suggest that, for SGC
420 to proceed only local microhomology is required, and perhaps no homology. While these
421 data do not specifically identify the process responsible they are consistent with SGC not
422 relying upon classical HR mechanisms. The gene family- and subfamily-specific
423 conserved tracts found in most *ves* genes instead likely serve other functions, either in
424 the chromatin or in the VESA1 polypeptides they encode. Interestingly, very similar
425 patterns were reported in *P. falciparum* of recombination between *var2CSA* genes
426 associated with adhesion of infected erythrocytes to chondroitin sulfate in placental
427 malaria [48].
428
429 When initiating this study we hypothesized that, as Rad51 proteins are considered
430 essential to HR and SGC is thought to be a form of HR, BbRad51 should be essential to
431 this process and its loss should abrogate SGC. Our results disprove that hypothesis. The
432 analysis of sequences surrounding SGC breakpoints supports error-free recombination
433 via homologous sequences in most, but not all, instances. However, unlike Rad51-
434 mediated HR where tracts of homologous sequence at least 8 bp long are needed for
435 synapse formation [49, 50], only local sequence microhomology of at least 2 bp, or
436 perhaps no homology, appears to be needed (Figure 7). It is possible that this is an
437 artifactual result arising from misidentification of recombining sequences during
438 analysis. Alternatively, this may be the actual biological result, due to mismatch repair
439 or deletion subsequent to sequence acquisition or replication. Whereas these results
440 reflect the mechanism(s) responsible, they do not identify the cause. This result is
441 consistent with loss of the ability to incorporate exogenous plasmid sequences into the
442 genome via HR in the absence of BbRad51 [31], whereas only the rate and product
443 lengths of SGC are affected. The reason for the apparent shortening of SGC tracts
444 observed in knockouts is not clear. One possibility is that it may reflect the loss of some
445 recombinants where longer unique tracts were involved, which might require BbRad51
446 for successful disengagement of the invading and donor strands, repair of breaks
447 created during the disengagement process, or stabilization of a longer crossover region
448 during exchange. Regardless of any secondary effects, BbRad51 is clearly not essential to
449 the SGC process.
450

451 The overall reduction in isTS observed among Bb*rad51* knockouts, based upon the
452 numbers of alternative loci to which transcripts mapped in each line, is particularly
453 intriguing. At the 5-month time point, in the knockout lines transcripts appear to have
454 arisen from a mean of only 23.3 *ves* loci, whereas in wild-type parasites a mean of 33.3
455 loci had been activated within the population. Although statistically significant (p=
456 0.014), the biological significance of this difference is not as clear. The reason is because
457 these data are derived from *ves1*α genes only. Also, the maximal number of *ves* loci that
458 are competent to be activated is not known and the ability to modify a single locus is
459 extensive. About half of the *ves* family is organized in divergent (head-to-head) pairs
460 (approximately 33 pairs) of *ves1*α/*ves1*β or *ves1*α/*ves1*α genes that flank quasi-
461 palindromic, bidirectional promoter regions [12, 16]. The remainder are present as
462 individual *ves* genes with potentially unidirectional promoters. Among the *ves* loci
463 putatively activated in this study, the ratio of loci with divergent/ unidirectional
464 promoters ranged from 1.07 - 1.60 (mean 1.28 ± 0.24) in CE11, and 1.25 - 2.00 (mean
465 1.61 ± 0.29) in knockouts. Although wild type parasites, on average, appeared to activate
466 a higher proportion of unidirectional promoters than did knockouts, this did not reach
467 statistical significance (p = 0.059; S1 Table). The functionality of several bidirectional
468 *ves* promoters has been demonstrated experimentally [16], but function has not yet been
469 similarly tested for promoters preceding individual *ves* genes. This result clearly
470 indicates that a significant proportion of *ves* loci have the potential to be activated,
471 including individual *ves* genes, consistent with a study on the transcriptomes of pairs of
472 virulent *B. bovis* lines and attenuated lines derived from them. In that study, virulent
473 lines transcribed from a significantly wider variety of *ves* loci that included both
474 divergent and non-divergent loci [51]. Interestingly, among attenuated parasites *ves*
475 transcription was upregulated only from loci that are not divergently-oriented. The
476 difference in *ves* transcriptional behavior observed here and in the attenuation study
477 suggests that in vivo attenuation of *B. bovis* is unlikely to be related to BbRad51
478 expression or function. This conclusion is supported by the unperturbed Bb*rad51*
479 transcription observed in attenuated parasites [51].
480
481 The basis for a connection between the SGC and isTS mechanisms is not clear. Indirect
482 evidence allows us to propose at least three feasible explanations, each with varying

483 levels of support. (i) First, of the various *ves* loci represented among the transcripts, a
484 subset of the sequences may reflect the complete replacement of the observed region
485 within the original locus of active *ves* transcription (LAT) by much longer conversion
486 tracts, rather than by isTS. In this case, such a long replacement sequence would cause
487 the read to map artifactually to the donor locus rather than to the locus from which the
488 full *ves* gene was actually being transcribed. Given the inability to integrate exogenous
489 DNAs into the genome of Bb*rad51* knockout parasites [31] it is anticipated that this
490 would be a rare event in knockouts. From our data, we cannot rule out this possibility.
491 (ii) The sites failing to activate in Bb*rad51* knockouts may be unusually sensitive to
492 recombination, with most such events leading to lethality. However, comparison of all
493 the putatively activated *ves* loci in wild-type and knockout parasites reveals that there is
494 essentially complete overlap in the *ves* gene clusters that can be activated (S3 Figure),
495 directly arguing against this explanation. Rather, the knockouts appear to achieve
496 comparable inter-locus switches, but with a lower frequency than wild-type. (iii) A
497 mechanistically distinct possibility is that Bb*rad51* knockouts have a diminished
498 capacity to activate *ves* loci epigenetically. As a part of DNA repair, chromatin first must
499 be remodeled to make it accessible to the repair machinery. This occurs in part by local
500 chromatin decondensation through histone acetylation [52-54]. Rad51 has been
501 proposed to assist in the assembly of the histone acetylation machinery during repair of
502 dsDNA breaks and stalled replication forks [55]. In the presence of BbRad51, double-
503 stranded breaks in silenced *ves* loci may be successfully acetylated and repaired, but
504 may not always be remodeled again for silencing. With transcription of a single *ves* locus
505 being the default state, a choice would have to be made between the existing LAT (the
506 single active locus of *ves* transcription [15]) and the newly repaired/acetylated *ves* locus.
507 Silencing of the existing LAT would lead to isTS and establishment of a new LAT. Thus,
508 BbRad51 may be epigenetically effecting isTS of *ves* genes as an unintended side-effect
509 of the DNA repair process. If true, then in Bb*rad51* knockouts the absence of BbRad51
510 would be anticipated to result in frequent failure to assemble the full repertoire of repair
511 machinery, including epigenetic modifiers. Accordingly, isTS would be a less common
512 event. While explanation (i) is consistent with well-established Rad51 protein functions,
513 possibility (iii) is neither implausible nor inconsistent with less well-characterized
514 functions, and even could provide a potential mechanism for the stochastic switch

515 events of isTS. If substantiated experimentally, this could provide a direct link between
516 DNA repair and antigenic variation via isTS as well as recombination. Our currently
517 available evidence does not distinguish these two possibilities, and this question
518 warrants further investigation.
519
520 The limitations of our study include the lack of genomic data for our culture
521 populations. High coverage whole genome sequencing, and good quality assemblies
522 might improve the assessment of SGC and recombination. Still, our results demonstrate
523 that BbRad51 is not necessary for survival of asexual *B. bovis* in vitro or for overall
524 genome stability in the absence of environmental insult. Moreover, this protein is
525 dispensable to SGC-based antigenic variation in *B. bovis*, although it influences the rates
526 of SGC antigenic variation and isTS. It is not clear whether its absence would be
527 similarly benign during in vivo infection, where there is strong selection by host immune
528 responses. Evolutionary retention of BbRad51 and its involvement in recovery from
529 alkylation damage indicates clearly that it plays some role(s) in parasite DNA repair,
530 including in asexual developmental stages. The clear implication of this work is that
531 some component besides BbRad51 provides for the recombination observed in SGC.
532 Whether this is from a more distantly related member of the RecA/RadA/Rad51
533 superfamily proteins encoded by the *B. bovis* genome, or another enzyme class
534 altogether, awaits experimental evidence.
535

536 ## Materials and Methods

537 ***Parasite culture, transfection, and selection.*** This project used the clonal *B.*
538 *bovis* CE11 parasite line as starting material [6]. In vitro parasite cultures were
539 maintained as microaerophilous stationary phase cultures under 90% N_2/ 5% O_2/ 5%
540 CO_2 (v/v), essentially as described [5, 56]. Cloning of parasites was conducted by two
541 sequential rounds of limiting dilution cloning as described previously [7]. Parasites were
542 transfected with DNAs purified from *E. coli* DH5α, using EndoFree Plasmid Maxi kits
543 (Qiagen; Valencia, CA). Both parasitized erythrocytes and DNAs were suspended in
544 cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO_4 pH 7.6, 25 mM HEPES
545 pH 7.6, 2 mM EGTA, 5 mM $MgCl_2$, pH 7.6) [57] prior to electroporation.
546 Electroporation was performed in 2 mm-gap cuvettes, using 5 pmol linearized DNA at

547 1.25 kV, 25 μF, and 200 Ω, as described by Wang et al. [16], with plating at 1.25% packed
548 cell volume. After 24h recovery in culture, selection was initiated by addition of
549 blasticidin-s hydrochloride (TOKU-E; Bellingham, WA) to a final concentration of 15 μg
550 ml^{-1} (32.7 μM) [58]. Every three days medium was removed, and replaced with a 2.5%
551 packed cell volume of uninfected erythrocytes in medium plus blasticidin-s. Once viable
552 parasites emerged, usually after approximately two weeks, they were maintained under
553 drug selection and were immediately cloned.
554
555 ***Validation of Bbrad51 knock-out.*** Validation of Bb*rad51* gene knock-out was
556 performed by diagnostic PCR, Southern blotting, RT-PCR, and sequencing of the
557 Bb*rad51* locus. These data are presented in [31].
558
559 ***Phenotypic analyses.***
560 *Parasite growth assays*. Parasite growth was assayed by counting Giemsa-stained
561 smears, with samples collected at 0, 24, and 48h growth (approximately 0, 3 and 6 cell
562 cycles [59]). Alternatively, in some experiments a DNA-based SYBR Green I method was
563 performed, essentially as described [60, 61], on parasites grown in bovine erythrocytes
564 depleted of leukocytes [62]. For experiments involving γ-irradiation, parasites were
565 exposed to a calibrated [Cs^{137}] source (Gammacell GC-10 gamma irradiator), on ice.
566 Control cells were maintained on ice for the duration of the treatment time. Samples
567 were immediately diluted into fresh medium containing 10% packed cell volume
568 uninfected erythrocytes, and placed into culture. For experiments involving MMS
569 exposure, parasites were exposed to MMS (diluted in complete medium) for 90 minutes,
570 followed by washout as described [31].
571
572 *Creation of a Bbrad51-NanoLuciferase-tagged parasite line*. The Bb*rad51* locus from
573 nucleotide 1,858,332- 1,860,393 of the *B. bovis* C9.1 line genome, already fused with 2x
574 HA tag and a glmS element in plasmid pBbRad51HAglmS, was opened by inverse PCR
575 with primers DA350 and DA351 [63]. NanoLuciferase coding sequences were amplified
576 from plasmid p2xHAglmS-gfp-bsd (pHggb) with primers DA348 and DA349, and
577 inserted into the opened plasmid with NEBuilder reagents (New England Biolabs;
578 Berverley, MA) to create plasmid pBbRad51NLHAglmS (pBRNLHg; S2_File). This

579 plasmid was linearized and transfected into *B. bovis* CE11 parasites [16]. Transfectants
580 were selected for resistance to blasticidin-S. Recovered transformants had experienced
581 fusion of the cassette to the 3' end of Bb*rad51* coding sequences by double crossover
582 homologous recombination, as described [31].
583
584 *Chromosome reassembly*. Parasites grown in leukocyte-depleted erythrocytes as
585 described above. Cultures, at 2.5% parasitized erythrocytes, were given 100 Gy exposure
586 to [^{137}Cs] on ice to fragment chromosomes. Irradiated cells were placed back into culture
587 to recover for designated times, then were processed for pulsed-field gel analysis [64].
588 Plugs were embedded into 1% SeaKem Gold agarose in 0.5x TBE buffer, and
589 electrophoresed for 23.5 hours at 180V, with a 50-165 second ramped switch time [65].
590 Gels were stained with SYBR Gold Nucleic Acid Stain (Invitrogen) for DNA
591 visualization, and photographed. Integrated pixel intensities were plotted for each
592 chromosome and the "smear" of DNA below chromosome 1 using ImageJ v. 1.52 "Gels"
593 and "Measure" algorithms. Corresponding blank gel regions were used for background
594 correction.
595
596 *SGC-mediated antigenic variation assay*. This experiment was performed with three
597 biological replicates per genotype, comprised of one clone each from three independent
598 Bbrad51 knockout lines (CE11Δrad51^{ko1} H5, CE11Δrad51^{ko2} E8, and CE11Δrad51^{ko3} A5;
599 referred to as ko1^{H5}, ko2^{E8}, and ko3^{A5}), and three subclones of wild type CE11 line
600 parasites (CE11 B8, CE11 C2, and CE11 C5). Bb*rad51* knockout and wild type parasites
601 were cloned by limiting dilution [7]. RNAs were isolated one and five months after
602 parasite cloning, using Ribozol (Amresco). RNAs were treated two times with
603 TurboDNase (Ambion) supplemented with 1 mM $MnCl_2$ [66], followed by inactivation
604 with DNase Inactivation Reagent (Ambion). M-MuLV reverse transcriptase (New
605 England Biolabs) and oligo-d(T) primers were used to make cDNAs. A hypervariable
606 segment containing most of the CKRD domain of *ves*1α transcripts was amplified by RT-
607 PCR, using "universal" primers vesUniF2 (TGGCACAGGTACTCAGTG) and PD1R
608 (TACAANAACACTTGCAGCA) as described [15].
609
610 *Sequencing and recombination analysis*. Four independent amplifications of each

611 cDNA were pooled in stoichiometrically equal amounts to maximize detection of rare
612 variants and minimize single-sample PCR artifacts during sequencing. Paired-end
613 amplicon libraries were generated with NEBNext reagents (New England Biolabs) by
614 the University of Florida NextGen Sequencing Core Laboratory, incorporating Illumina
615 TruSeq index sequences. Libraries, spiked with 8% PhiX genomic library as internal
616 control, were sequenced on the Illumina MiSeq platform, using the Illumina Pipeline
617 1.8. Fastq reads were analyzed using Qiime2 pipeline [67]. Quality and adapter
618 trimming were performed using CutAdapt [68, 69]. Further de-noising and amplicon
619 sequence variant (ASV)-calling were performed using DADA2 [70], truncating the reads
620 at 230 nt and allowing a maximum of 5 expected errors per read. In order to remove
621 potential contaminating sequences, ASVs were aligned with the bovine genome, using
622 BLAST [71]. In order to identify recombinant sequences, ASVs were first aligned using
623 CLC Main Workbench, version 6.9.2. Recombination analyses were then performed on
624 the alignments with RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiSscan, and
625 3Seq, as implemented in RDP4 [72]. Only recombinant events identified by a minimum
626 of 4 out of seven tests (at $p \leq 0.05$) were considered statistically supported and included
627 in downstream analyses. Note that no genome is currently available for the CE11 line.
628 Therefore, the C9.1 line genome was used, as the C9.1 and CE11 lines are sibling clonal
629 lines derived from the MO7 clonal line [6, 7, 12]. Recombinant results, including
630 extraction of SGC tracts, were summarized using custom scripts written with R version
631 3.6.3 [73] through the RStudio shell. Mean lengths of SGC tracts were compared by one-
632 way ANOVA, whereas tract length distributions were compared by the Mann-Whitney
633 Rank Sum test, without expectation of a normal distribution of variance, using
634 SigmaPlot version 11.0 (Systat Software, Inc.; San Jose, CA).
635
636 **Data availability.** Raw sequence reads are publicly available for download from
637 NCBI, as BioProject #PRJNA357248 (accession numbers SRR5110992-SRR5111003).
638 All recombinant transfection constructs are available upon request.
639

640 ## Acknowledgments

641 This work was supported by funds from National Institutes of Health grants R01
642 AI0055864 and T32 AI0007110, USDA Animal Health project FLA-VME-005011, and

U.F. College of Veterinary Medicine. The authors thank Allison Vansickle for assistance with animal care and handling, and Kevin Brown, Linda Bloom, Salvador Castaneda Barba, Eva Top, and Olivia Kosterlitz for helpful comments.

Author contributions

Conceptualization: DRA, EAM
Data curation: DRA
Experimental design: DRA, EAM, Y-PX
Formal analysis: DRA, EAM, MST
Funding acquisition: DRA
Investigation: DRA, EAM, SQ, Y-PX
Methodology: DRA, EAM, MST, Y-PX
Project administration: DRA
Resources: DRA
Supervision: DRA
Writing- original draft: DRA, EAM
Writing- review and editing: DRA, EAM, MST

661 ## Figure Captions

662 **Figure 1. BbRad51 levels are not upregulated in response to DNA damage.** *B.*
663 *bovis* CE11 parasites were given acute exposure to 250 μM MMS to induce DNA
664 alkylation, then placed back into culture to allow upregulation of DNA repair proteins.
665 Cells were harvested and assayed for expression levels of NanoLuciferase. 1-4 samples
666 are transiently transfected parasites expressing NanoLuciferase from episomal plasmid
667 pHggb; 5-8 samples are expressing BbRad51-NanoLuciferase-2xHA fusion protein from
668 the endogenous Bb*rad51* locus and under control of the Bb*rad51* promoter. Because
669 expression was at quite different levels depending upon the source, values have been
670 normalized as a proportion of the signal produced by the untreated control in each set
671 (samples 1). "1" and "5", mock-treated, grown in complete medium (CM); "2" and "6",
672 MMS-treated, CM; "3" and "7", mock-treated, grown in CM containing 400 μM glcN;
673 "4" and "8", MMS-treated, 400 μM glcN. Parasites were allowed to recover from MMS
674 treatment for 3 hours **(A)** or for 16h **(B)**. Asterisks indicate statistically significance
675 difference from "1" control samples of each series ($p < 0.001$). No other differences were
676 statistically significant.
677

678 **Figure 2. Survival and growth of Bb*rad51* knockouts following γ-irradiation**
679 **is not different from wild-type. A.** Parasites were exposed to varying dosages of γ-
680 irradiation, placed back into culture, sampled at 24 and 48h post-irradiation, and
681 percent parasitized erythrocytes determined microscopically. In this individual
682 experiment a significant difference was observed at 48h in the non-irradiated control
683 samples only. This was not a reproducible difference. This assay shows that there is no
684 increase in percent parasitized erythrocytes over time after 200 Gy exposure. **B.**
685 Comparison of CE11 wild-type with knockout clonal lines $ko2^{E8}$ and $ko3^{A5}$, determined
686 by a Sybr Green DNA-detection assay [31]. Sybr Green data are plotted as the
687 proportion of the signal exhibited by the non-irradiated control samples (± 1 s.d.). No
688 reproducible significant differences were observed among the parasite lines at these
689 dosages. Note that the $ko1^{H5}$ line was derived approximately one year prior to $ko2^{E8}$ and
690 $ko3^{A5}$ lines and was assessed differently, by microscopic observation of percent
691 parasitized erythrocytes. In each case, this experiment was repeated three times with
692 four replicates per sample. No reproducible differences were observed between *B. bovis*

693 CE11 wild-type and any of the knockout lines.

694

695 **Figure 3. Chromosomes of γ-irradiated parasites reassembled with the same**
696 **kinetics.** *B. bovis* CE11 wild type and ko1^{H5} clonal line were exposed to 100 Gy γ-
697 irradiation, then allowed allowed to recover for 0, 8, 16, or 24h. **A.** Reassembly of
698 chromosomes was assessed by PFGE (CE11 wild type is shown as an example). **B.** Plot of
699 relative signal intensities of *B. bovis* CE11 wild type and ko1^{H5} DNAs within individual
700 chromosomes, and in the "smear" below chromosome 1. The plot includes the means (±
701 1 s.d.) of four independent experiments. Statistical analyses using Student's T-test
702 revealed no significant differences between wild type and knockout parasites at any time
703 point. NC, non-irradiated control.

704

705 **Figure 4. Variation arose over time in the sequences of *ves*1α transcripts.**
706 RT-PCR amplification of the highly variable CKRD domain of *ves*1α transcripts and
707 deep sequencing of amplicons was used to assess variation arising over time. Three
708 freshly cloned lines of *B. bovis* CE11 wild type (lines CE11^{B8}, CE11^{C2}, and CE11^{C5}), and
709 one clonal line from each of three independent CE11Δ*rad51* knockout lines were
710 observed 1 and 5 months post-cloning. **A.** Typical structure of *ves*1α genes. Positions of
711 primers vesUniF2 and PD1R used in PCR reactions are shown, delineating the region
712 amplified (modified from [15]). **B.** All merged amplicon sequences were mapped to the
713 *B. bovis* C9.1 genome, using CLC Genomics Workbench. Mapping tracks are shown for
714 each sample. The X axis represents the 8 Mbp *B. bovis* C9.1 genome as a single linear
715 sequence, the Y axis represents the proportion of sequences aligning to a given *ves* locus
716 within the genome, and the alternating blue and white bands represent data from 1
717 month (blue) and 5 month (white) time-points for each parasite line. The area of each
718 peak is proportional to the numbers of reads mapping to that locus relative to the total.
719 Transcripts from many activated loci are present in too low abundance to be observable
720 in this plot. These data confirm earlier results suggesting that *ves* transcription is
721 monoparalogous [15], but also show that isTS appears to occur over time. **C.** Numbers of
722 transcriptionally active *ves* loci detected at each time point (data is pooled for knockouts
723 and for CE11 samples). Slotted crossbars indicate means, solid crossbars indicate
724 medians, and box boundaries represent 25% and 75% confidence intervals. Statistical

725 significance of differences among samples, based upon one-way ANOVA, are indicated.
726
727 **Figure 5. SGC tracts in knockout parasite populations were shorter than in**
728 **wild-type.** The lengths of SGC tracts identified in amplicons with full statistical
729 support were compared for each of the six clonal lines under study. Data from 1 mo. and
730 5 mo. samples were pooled for each sample. A statistically significant difference ($p <$
731 0.001) was observed between wild-type and knockout parasites SGC tracts.
732
733 **Figure 6. The numbers of unique SGC tracts created over time by knockouts**
734 **is reduced.** The numbers of unique recombinant SGC tracts with full statistical
735 support were plotted, normalized per million merged reads. This value provides a
736 surrogate measure of the total variation presented by a parasite population. Slotted
737 crossbars indicate means, solid crossbars indicate medians, and the box boundaries
738 represent 25% and 75% confidence intervals. Asterisk: a statistically significant
739 difference of $p = 0.035$ was observed between wt and knockout parasites when
740 comparing the total populations of unique recombinants (Student's t-test). wt, pooled
741 CE11 wild-type samples; ko, pooled knockout parasite samples.
742
743 **Figure 7. Lengths of homology patches at SGC breakpoints did not differ.**
744 Patches of homologous sequence were observed at the breakpoints between the
745 sequences of transcripts representing loci that had served as sequence donors and those
746 of the recipient loci during SGC modifications of actively transcribed ves genes. These
747 tracts of homology were identified through alignments of donor and recipient sequences
748 (S7 Figure), and plotted for a minimum of 15 recombinant sequences from each sample.
749 Slotted crossbars indicate means, solid crossbars indicate medians, and the box
750 boundaries represent 25% and 75% confidence intervals. No statistically significant
751 differences were observed between any of the samples ($p= 0.133$; Kruskal-Wallis one
752 way analysis of variance on ranks).
753

755 ## Supporting Information Captions

756 **S1 Figure. Structures of plasmids pBbRad51-NanoLuciferase-2xHA-glmS**
757 **and p2xHA-glmS-gfp-bsd plasmids.** Complete sequences of these plasmids are
758 provided in S2_File.fa and S3_File.fa, in fasta format. Prior to transfection, pBbRad51-
759 NanoLuciferase-2xHA-glmS was linearized at the NotI site to facilitate double crossover
760 integration as described in Results.
761

762 **S2 Figure. Titration of γ-irradiation dosage on *B. bovis* CE11 (wild type).**
763 Parasites were exposed to 0- 1000 Gy irradiation from a calibrated Cs^{137} source, then
764 placed back into culture and the percent parasitized erythrocytes determined from
765 Giemsa-stained smears made at 0, 24, and 48h growth. For experiments requiring
766 survival of a proportion of the parasites the 100 Gy dosage was chosen.
767

768 **S3 Figure. Locations of *ves* loci from which transcripts and SGC tracts**
769 **arose.** The genomic locations of *ves* loci to which amplicons mapped (plotting the full
770 8 Mbp genome as a single linear element), and from which they were presumed to be
771 transcribed, is plotted relative to the numbers of parasite lines transcribing from that
772 locus. No *ves* gene clusters within the genome were transcribed by CE11 parasites that
773 could not also be transcribed by knockout parasites, and vice versa. For this plot, data
774 were pooled from both time-points of all three wt or all three knockout lines (i.e.,
775 maximally 6 possible per *ves* locus).
776

777 **S4 Figure. Statistically supported recombinants *vs*. total unique amplicons.**
778 A regression plot was made of the frequencies of merged amplicons with full statistical
779 support for identification as true recombinants against the total numbers of unique
780 amplicons observed for each population, with each type of value normalized per million
781 reads (R^2= 0.9228).
782

783 **S5 Figure. Cumulative distribution of unique SGC tract lengths in wild-type**
784 **and knockout parasites.** The length distribution of unique SGC tracts was plotted as
785 the cumulative proportion of total SGC tracts against tract lengths. Samples were pooled
786 for all six samples for wild-type and knockout parasites. These pooled data are from the

787 same CE11 wild-type and knockout lines shown individually in Figure 4B. The vertical
788 index lines indicates the median values of each population. Differences among the
789 distributions in each sample type were determined by the Mann-Whitney Rank Sum test
790 (p = <0.001).
791
792 **S6 Figure. Unique SGC tracts per active *ves* locus.** The numbers of unique SGC
793 tracts with full statistical support per transcriptionally active *ves* locus is shown for *B.*
794 *bovis* CE11 wild-type and knockout populations at 1 and 5 months growth post-cloning.
795 The data were pooled from all three lines of each population type. Slotted crossbars
796 indicate means, solid crossbars indicate medians, and the box boundaries represent 25%
797 and 75% confidence intervals. Values did not vary significantly among samples (p=
798 0.986, based upon one-way ANOVA).
799
800 **S7 Figure. Examples of homology patches observed in recombinant**
801 **amplicons.** In each case, the top sequence is the minor parent, the bottom the major
802 parent, and the middle sequence is the recombinant sequence. The gray boxes identify
803 the regions of homology within which recombination presumably occurred. Illustrated
804 are patches of (A) 2 bp, (B) 9 bp, and (C) a region wherein no identifiable patch of
805 homology may be found.
806
807 **S1 Table. Organizational nature of the *ves* loci to which reads mapped.** Loci
808 to which amplicons mapped are listed for all 12 samples (i.e., 1-month and 5-month
809 RNAs from all three wild-type and knockout lines), along with mean lengths and the
810 nature of each locus. Both divergent, bidirectional loci and unidirectional loci were
811 active in transcription at some level in all samples.
812
813 **S1_File.zip.** This .zip file contains individual files of SGC tracts identified for all six
814 clonal lines, in .csv format.
815
816 **S2_File.fa.** This file contains the full sequence of the pBbRad51-NanoLuciferase-
817 2xHA-glmS plasmid (pBRNLHg), in fasta format.
818

819 **S3_File.fa.** This file contains the full sequence of the p2xHA-glmS-gfp-bsd plasmid
820 (pHggb), in fasta format.

822 # References

823

824 1. Wright IG. An electron microscopic study of intravascular agglutination in the
825 cerebral cortex due to *Babesia argentina* infection. Int J Parasitol. 1972;2:209-15.

826 2. Wright IG. Ultrastructural changes in *Babesia argentina*-infected erythrocytes in
827 kidney capillaries. J Parasitol. 1973;59:735-6.

828 3. Wright IG, Goodger BV, McKenna RV, Mahoney DF. Acute *Babesia bovis*
829 infection: A study of the vascular lesions in kidney and lung. Z Parasitenkd. 1979;60:19-
830 27.

831 4. Allred DR, Al-Khedery B. Antigenic variation and cytoadhesion in *Babesia bovis*
832 and *Plasmodium falciparum*: different logics achieve the same goal. Mol Biochem
833 Parasitol. 2004;134:27-35.

834 5. Allred DR, Hines SA, Ahrens KP. Isolate-specific parasite antigens of the *Babesia*
835 *bovis*-infected erythrocyte surface. Mol Biochem Parasitol. 1993;60:121-32.

836 6. O'Connor RM, Allred DR. Selection of *Babesia bovis*-infected erythrocytes for
837 adhesion to endothelial cells co-selects for altered variant erythrocyte surface antigen
838 isoforms. J Immunol. 2000;164:2037-45.

839 7. Allred DR, Cinque RM, Lane TJ, Ahrens KP. Antigenic variation of parasite-
840 derived antigens on the surface of *Babesia bovis*-infected erythrocytes. Infect Immun.
841 1994;62:91-8.

842 8. O'Connor RM, Lane TJ, Stroup SE, Allred DR. Characterization of a variant
843 erythrocyte surface antigen (VESA1) expressed by *Babesia bovis* during antigenic
844 variation. Mol Biochem Parasitol. 1997;89:259-70.

845 9. Xiao Y-P, Al-Khedery B, Allred DR. The *Babesia bovis* VESA1 virulence factor
846 subunit 1b is encoded by the 1β branch of the *ves* multigene family. Mol Biochem
847 Parasitol. 2010;171:81-8.

848 10. Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, et al. The
849 evolutionary dynamics of variant antigen genes in *Babesia* reveal a history of genomic
850 innovation underlying host-parasite interaction. Nucleic Acids Res. 2014;42:7113-31.

851 11. Allred DR, Carlton JM-R, Satcher RL, Long JA, Brown WC, Patterson PE, et al.
852 The *ves* multigene family of *B. bovis* encodes components of rapid antigenic variation at
853 the infected erythrocyte surface. Mol Cell. 2000;5:153-62.

854 12. Al-Khedery B, Allred DR. Antigenic variation in *Babesia bovis* occurs through
855 segmental gene conversion of the *ves* multigene family, within a bidirectional site of
856 transcription. Mol Microbiol. 2006;59:402-14.

857 13. Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, et al.
858 Genome sequence of *Babesia bovis* and comparative analysis of Apicomplexan
859 hemoprotozoa. PLoS Pathog. 2007;3:e148.

860 14. Ferreri LM, Brayton KA, Sondgeroth KS, Lau AOT, Suarez CE, McElwain TF.
861 Expression and strain variation of the novel "*small open reading frame*" (*smorf*)
862 multigene family in *Babesia bovis*. Int J Parasitol. 2012;42:131-8.

863 15. Żupańska AK, Drummond PB, Swetnam DM, Al-Khedery B, Allred DR. Universal
864 primers suitable to assess population dynamics reveal apparent mutually exclusive
865 transcription of the *Babesia bovis ves*1α gene. Mol Biochem Parasitol. 2009;166:47-53.

866 16. Wang X, Bouchut A, Xiao Y-P, Al-Khedery B, Allred DR. Characterization of the
867 unusual bidirectional *ves* promoters driving VESA1 expression and associated with
868 antigenic variation in *Babesia bovis*. Eukaryotic Cell. 2012;11:260-9.

869 17. Symington LS, Rothestein R, Lisby M. Mechanisms and regulation of mitotic
870 recombination in *Saccharomyces cerevisiae*. Genetics. 2014;198:795-835.

871 18. Wyman C, Kanaar R. DNA double-strand break repair: All's well that ends well.
872 Annu Rev Genet. 2006;40:363-83.

873 19. Sugiyama T, Kantake N. Dynamic regulatory interactions of Rad51, Rad52, and
874 Replicaton Protein-A in recombination intermediates. J Mol Biol. 2009;390:45-55.

875 20. Petukhova G, Stratton S, Sung P. Catalysis of homologous DNA pairing by yeast
876 Rad51 and Rad54 proteins. Nature. 1998;393:91-4.

877 21. Inbar O, Kupiec M. Homology search and choice of homologous partner during
878 mitotic recombination. Mol Cell Biol. 1999;19(6):4134-42.

879 22. Bugreev DV, Mazina OM, Mazin AV. Rad54 protein promotes branch migration
880 of Holliday junctions. Nature. 2006;442:590-3.

881 23. Mortimer RK, Fogel S. Genetical interference and gene conversion. In: Grell RF,
882 editor. Mechanisms in Recombination. New York: Plenum Publishing Corporation;
883 1974. p. 263-75.

884 24. Anderson DG, McKay LL. Isolation of a recombination-deficient mutant of
885 *Streptococcus lactis* ML3. J Bacteriol. 1983;155(2):930-2.

886 25. Morita T, Yoshimura Y, Yamamoto A, Murata K, Mori M, Yamamoto H, et al. A
887 mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae RAD51*
888 genes. Proc Natl Acad Sci (USA). 1993;90(14):6577-80.

889 26. Siede W, Friedl AA, Dianova I, Eckardt-Schupp F, Friedberg EC. The
890 *Saccharomyces cerevisiae* Ku autoantigen homologue affects radiosensitivity only in the
891 absence of homologous recombination. Genetics. 1996;142(1):91-102.

892 27. Essers J, Hendriks RW, Swagemakers SM, Troelstra C, de Wit J, Bootsma D, et
893 al. Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous
894 recombination. Cell. 1997;89(2):195-204.

895 28. McCulloch R, Barry JD. A role for RAD51 and homologous recombination in
896 *Trypanosoma brucei* antigenic variation. Genes Dev. 1999;13:2875-88.

897 29. Roy N, Bhattacharyya S, Chakrabarty S, Laskar S, Babu SM, Bhattacharyya MK.
898 Dominant negative mutant of *Plasmodium* Rad51 causes reduced parasite burden in
899 host by abrogating DNA double strand break repair. Mol Microbiol. 2014;94(2):353-66.

900 30. Tsuzuki T, Fukii Y, Sakumi K, Tominaga Y, Nakao K, Sekiguchi M, et al. Targeted
901 disruption of the *Rad51* gene leads to lethality in embryonic mice. Proc Natl Acad Sci
902 (USA). 1996;93:6236-40.

903 31. Mack EA, Xiao Y-P, Allred DR. Knockout of *Babesia bovis rad51* ortholog and its
904 complementation by expression from the BbACc3 artificial chromosome platform. PLoS
905 One. 2019;14(8):e0215882. doi: https://doi.org/10.1371/journal.pone.0215882.

906 32. Cohen Y, Dardalhon M, Averbeck D. Homologous recombination is essential for
907 RAD51 up-regulation in *Saccharomyces cerevisiae* following DNA crosslinking damage.
908 Nucleic Acids Res. 2002;30(5):1224-32.

909 33. Collins JA, Irnov I, Baker S, Winkler WC. Mechanism of mRNA destabilization by
910 the *glmS* ribozyme. Genes Dev. 2007;21:3356-68.

911 34. Suarez CE, McElwain TF. Stable expression of a GFP-BSD fusion protein in
912 *Babesia bovis* merozoites. Int J Parasitol. 2008;39:289-97.

913 35. Prommana P, Uthaipibull C, Wongsombat C, Kamchonwongpaisan S, Yuthavong
914 Y, Knuepfer E, et al. Inducible knockdown of *Plasmodium* gene expression using the
915 *glmS* ribozyme. PLoS One. 2013;8(8):e73783.

916 36. Lundin C, North M, Erixon K, Walters K, Jenssen D, Goldman AS, et al. Methyl
917 methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo
918 DNA double-strand breaks. Nucleic Acids Res. 2005;33(12):3799-811.

919 37. Horrocks P, Pinches R, Christodoulou Z, Kyes SA, Newbold CI. Variable *var*
920 transition rates underlie antigenic variation in malaria. Proc Natl Acad Sci (USA).
921 2004;101:11129-34.

922 38. Kelso AA, Waldvogel SM, Luthman AJ, Sehorn MG. Homologous recombination
923 in protozoan parasites and recombinase inhibitors. Front Microbiol. 2017;8:1716.

924 39. Bhattacharyya MK, Norris DE, Kumar N. Molecular players of homologous
925 recombination in protozoan parasites: implication for generating antigenic variation.
926 Infect Genet Evol. 2004;4:91-8.

927 40. Achanta SS, Varunan SM, Bhattacharyya S, Bhattacharyya MK. Characterization
928 of Rad51 from apicomplexan parasite *Toxoplasma gondii*: an implication for inefficient
929 gene targeting. PLoS One. 2012;7(7):e41925.

930 41. Fox BA, Falla A, Rommereim LM, Tomita T, Gigley JP, Mercier C, et al. Type II
931 *Toxoplasma gondii* KU80 knockout strains enable functional analysis of genes required
932 for cyst development and latent infection. Eukaryotic Cell. 2011;10:1193-206.

933 42. Kirkman LA, Lawrence EA, Deitsch KW. Malaria parasites utilize both
934 homologous recombination and alternative end joining pathways to maintain genome
935 integrity. Nucleic Acids Res. 2014.

936 43. Krogh BO, Symington LS. Recombination proteins in yeast. Annu Rev Genet.
937 2004;38:233-71.

938 44. Ensminger M, Iloff L, Ebel C, Nikolova T, Kaina B, Löbrich M. DNA breaks and
939 chromosomal aberrations arise when replication meets base excision repair. J Cell Biol.
940 2014;206(1):29-43.

941 45. Yeeles JTP, Poli J, Marians KJ, Pasero P. Rescuing stalled or damaged replication
942 forks. Cold Spring Harb Perspect Biol. 2013;5:a012815.

943 46. Boothroyd CE, Dreesen O, Leonova T, Ly KI, Figueiredo LM, Cross GAM, et al. A
944 yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma*
945 *brucei*. Nature. 2009;459:278-81. doi: doi:10.1038/nature07982.

946 47. Deitsch KW, Moxon ER, Wellems TE. Shared themes of antigenic variation and
947 virulence in bacterial, protozoal, and fungal infections. Microbiol Mol Biol Rev.
948 1997;61:281-93.

949 48. Talundzic E, Shah S, Fawole O, Owino S, Moore JM, Peterson DS. Sequence
950 polymorphism, segmental recombination and toggling amino acid residues within the
951 DBL3X domain of the VAR2CSA placental malaria antigen. PLoS One.
952 2012;7(2):e31565. doi: doi:10.1371/journal.pone.0031565.

953 49. Hsieh P, Camerini-Otero CS, Camerini-Otero RD. The synapsis event in the
954 homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of
955 DNA. Proc Natl Acad Sci (USA). 1992;89:6492-6.

956 50. Lee JY, Terakawa T, Zi Z, Steinfeld JB, Redding S, Kwon Y, et al. Base triplet
957 stepping by the Rad51/RecA family of recombinases. Science. 2015;349(6251):977-81.

958 51. Pedroni MJ, Sondgeroth KS, Gallego-Lopez GM, Echaide I, Lau AOT.
959 Comparative transcriptome analysis of geographically distinct virulent and attenuated
960 *Babesia bovis* strains reveals similar gene expression changes through attenuation.
961 BMC Genomics. 2013;14:763.1-13.

962 52. Sivanand S, Rhoades S, Jiang Q, Lee JV, Benci J, Zhang J, et al. Nuclear acetyl-
963 CoA production by ACLY promotes homologous recombination. Mol Cell. 2017;67:252-
964 65.

965 53. Gong F, Miller KM. Mammalian DNA repair: HATs and HDACs make their mark
966 through histone acetylation. Mutat Res. 2013;750:23-30.

967 54. Price BD, D'Andrea AD. Chromatin remodeling at DNA double-strand breaks.
968 Cell. 2013;152:1344-54.

969 55. Rowlands H, Dhavarasa P, Cheng A, Yankulov K. Forks on the run: can the
970 stalling of DNA replication promote epigenetic changes? Frontiers in Genetics.
971 2017;8:86.

972 56. Levy MG, Ristic M. *Babesia bovis*: continuous cultivation in a microaerophilous
973 stationary phase culture. Science. 1980;207:1218-20.

974 57. van den Hoff MJB, Moorman AFM, Lamers WH. Electroporation in 'intracellular'
975 buffer increases cell survival. Nucleic Acids Res. 1992;20:2902.

976 58. Mack E. Unexpected effects of Rad51 deletion on viability, chromosome repair,
977 and antigenic variation in *Babesia bovis* [Ph.D. Dissertation]: University of Florida;
978 2014.

979 59. Franssen FFJ, Gaffar FR, Yatsuda AP, de Vries E. Characterisation of erythrocyte
980 invasion by *Babesia bovis* merozoites efficiently released from their host cell after high-
981 voltage pulsing. Microbes Inf. 2003;5:365-72.

982 60. Smilkstein M, Sriwilaijaroen N, Kelly JX, Wilairat P, Riscoe M. Simple and
983 inexpensive fluorescence-baseed technique for high-throughput antimalarial drug
984 screening. Antimicrob Agents Chemother. 2004;48(5):1803-6.

985 61. Bennett TN, Paguio M, Gligorijevic B, Seudieu C, Kosar AD, Davidson E, et al.
986 Novel, rapid, and inexpensive cell-based quantification of antimalarial drug efficacy.
987 Antimicrob Agents Chemother. 2004;48(5):1807-10.

988 62. Ambrosio RE, Potgieter FT, Nel N. A column purification procedure for the
989 removal of leucocytes from parasite-infected bovine blood. Onderstepoort J Vet Res.
990 1986;53:179-80.

991 63. Triglia T, Peterson MG, Kemp DJ. A procedure for in vitro amplification of DNA
992 segments that lie outside the boundaries of known sequences. Nucleic Acids Res.
993 1988;16:8186.

994 64. Ray BK, Bailey CW, Jensen JB, Carson CA. Chromosomes of *Babesia bovis* and
995 *Babesia bigemina*. Mol Biochem Parasitol. 1992;52:123-6.

996 65. Alleman AR, Kamper SM, Viseshakul N, Barbet AF. Analysis of the *Anaplasma*
997 *marginale* genome by pulsed-field electrophoresis. J Gen Microbiol. 1993;139:2439-44.

998 66. Bauer P, Rolfs A, Regitz-Zagrosek V, Hildebrandt A, Fleck E. Use of manganese
999 in RT-PCR eliminates PCR artifacts resulting from DNase I digestion. Biotechniques.
1000 1997;22:1128-32.

1001 67. Bolyen E, Riedeout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.
1002 Reproducible, interactive, scalable and extensible microbiome data science using QIIME
1003 2. Nat Biotechnol. 2019;37:852-7.

1004 68. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al.
1005 Quality-filtering vastly improves diversity estimates from Illumina amplicon
1006 sequencing. Nature Methods. 2013;10(1):57-9. doi: DOI:10.1038/NMETH.2276.

1007 69. Martin M. Cutadapt removes adapter sequences from high-throughput
1008 sequencing reads. EMBnetJournal. 2011;17(1):1-3.

1009 70. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.
1010 DADA2: high-resolution sample inference from Illumina amplicon data. Nature
1011 Methods. 2016;13(7):581-3. doi: DOI:10.1038/NMETH.3869.

1012 71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
1013 search tool. J Mol Biol. 1990;215:403-10.

1014 72. Martin DP, Murrell B, golden M, Khoosal A, Muhire B. RDP4: detection and
1015 analysis of recombination patterns in virus genomes. Virus Evolution. 2015;1(1):vev003.
1016 doi: doi: 10.1093/ve/vev003.

1017 73. R_Core_Team. R: A language and environment for statistical computing Vienna,
1018 Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.r-
1019 project.org/.
1020

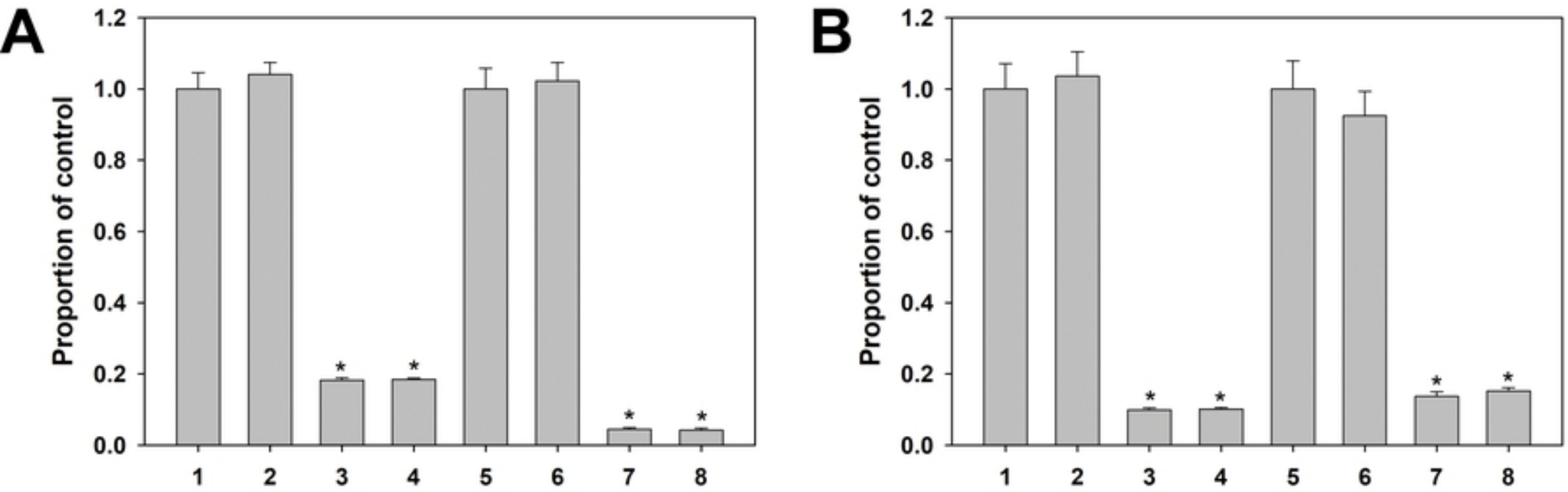


Figure 1

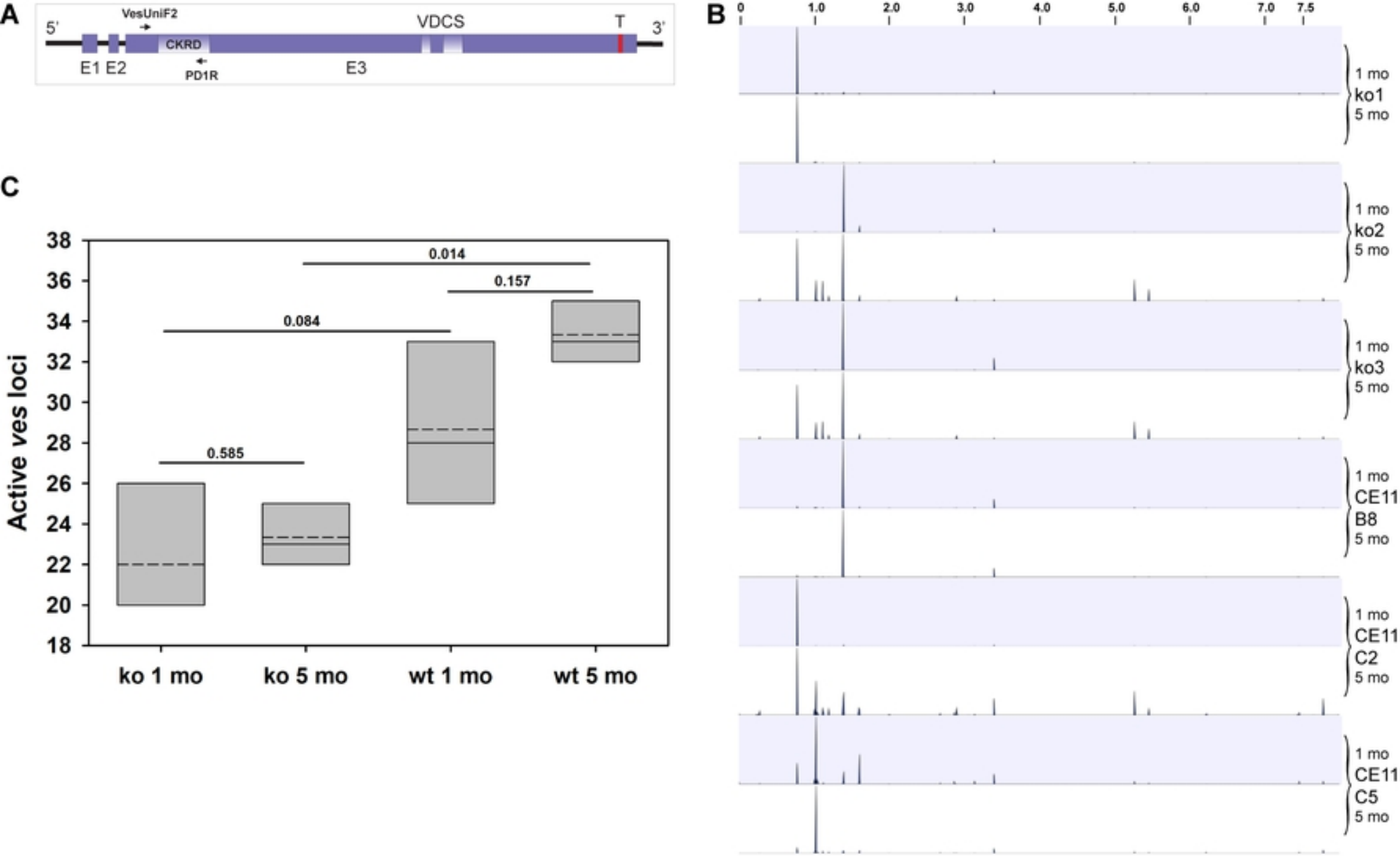

A
5'
VesUniF2
CKRD
VDCS
T
3'
E1 E2
PD1R
E3
B
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
7.5
1 mo
ko1
5 mo
1 mo
ko2
5 mo
1 mo
ko3
5 mo
1 mo
CE11
B8
5 mo
1 mo
CE11
C2
5 mo
1 mo
CE11
C5
5 mo
C
Active ves loci
38
36
34
32
30
28
26
24
22
20
18
0.014
0.157
0.084
0.585
ko 1 mo
ko 5 mo
wt 1 mo
wt 5 mo

Figure 4

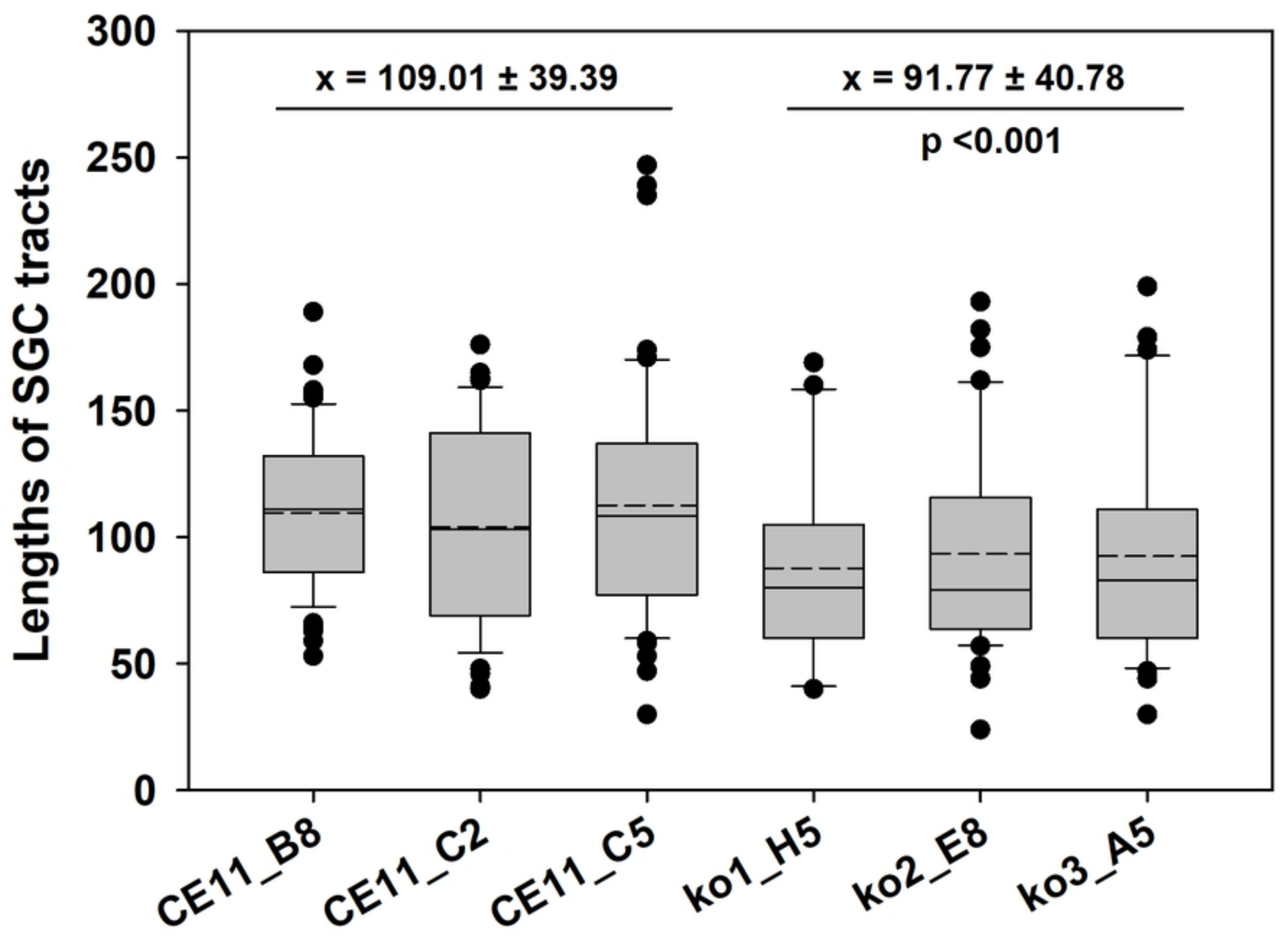

300
250
200
150
100
50
0
Lengths of SGC tracts
x = 109.01 ± 39.39
x = 91.77 ± 40.78
p <0.001
CE11_B8
CE11_C2
CE11_C5
ko1_H5
ko2_E8
ko3_A5

Figure 5

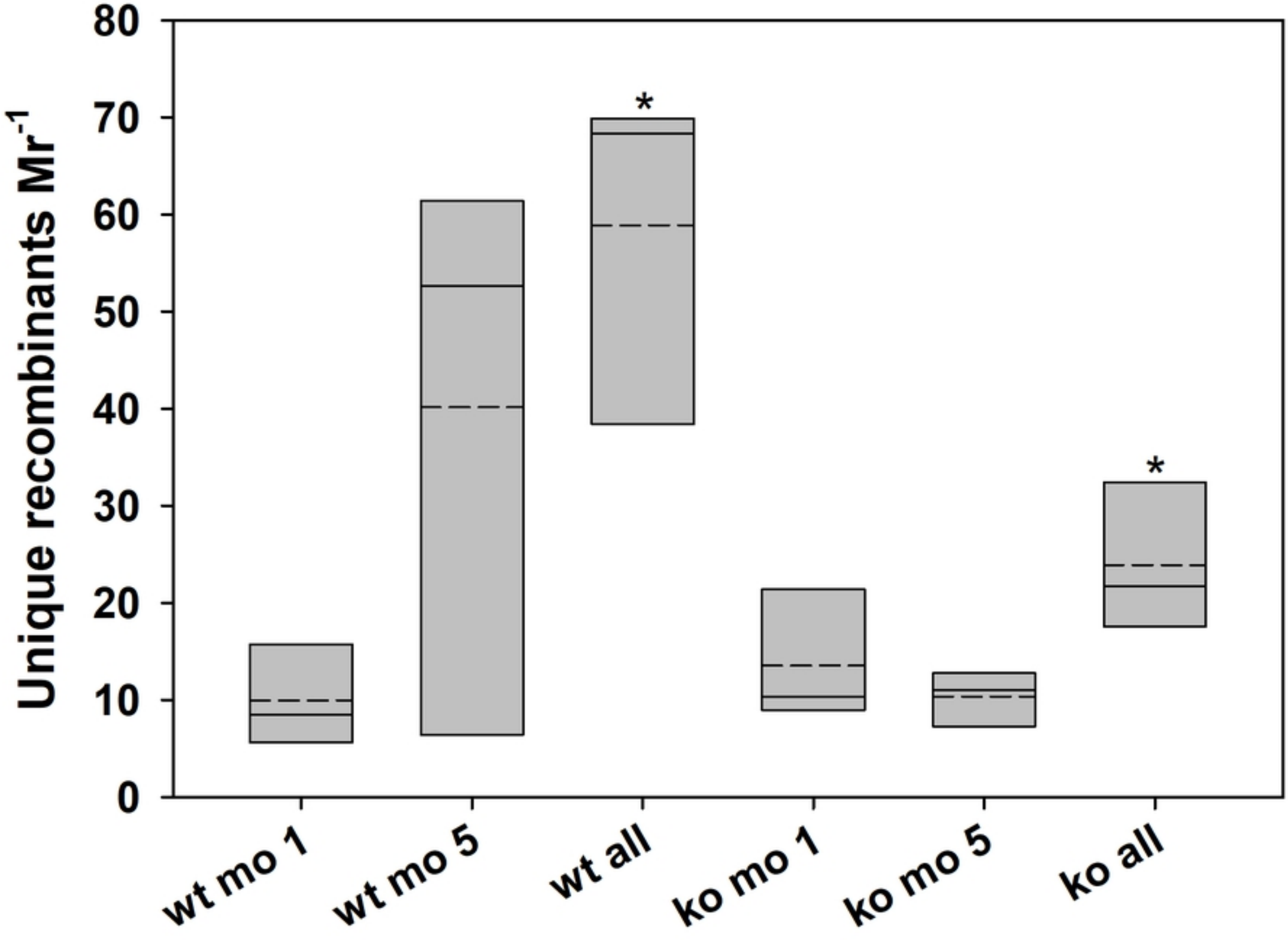

Unique recombinants Mr-1
80
70
60
50
40
30
20
10
0
*
*
wt mo 1
wt mo 5
wt all
ko mo 1
ko mo 5
ko all

Figure 6

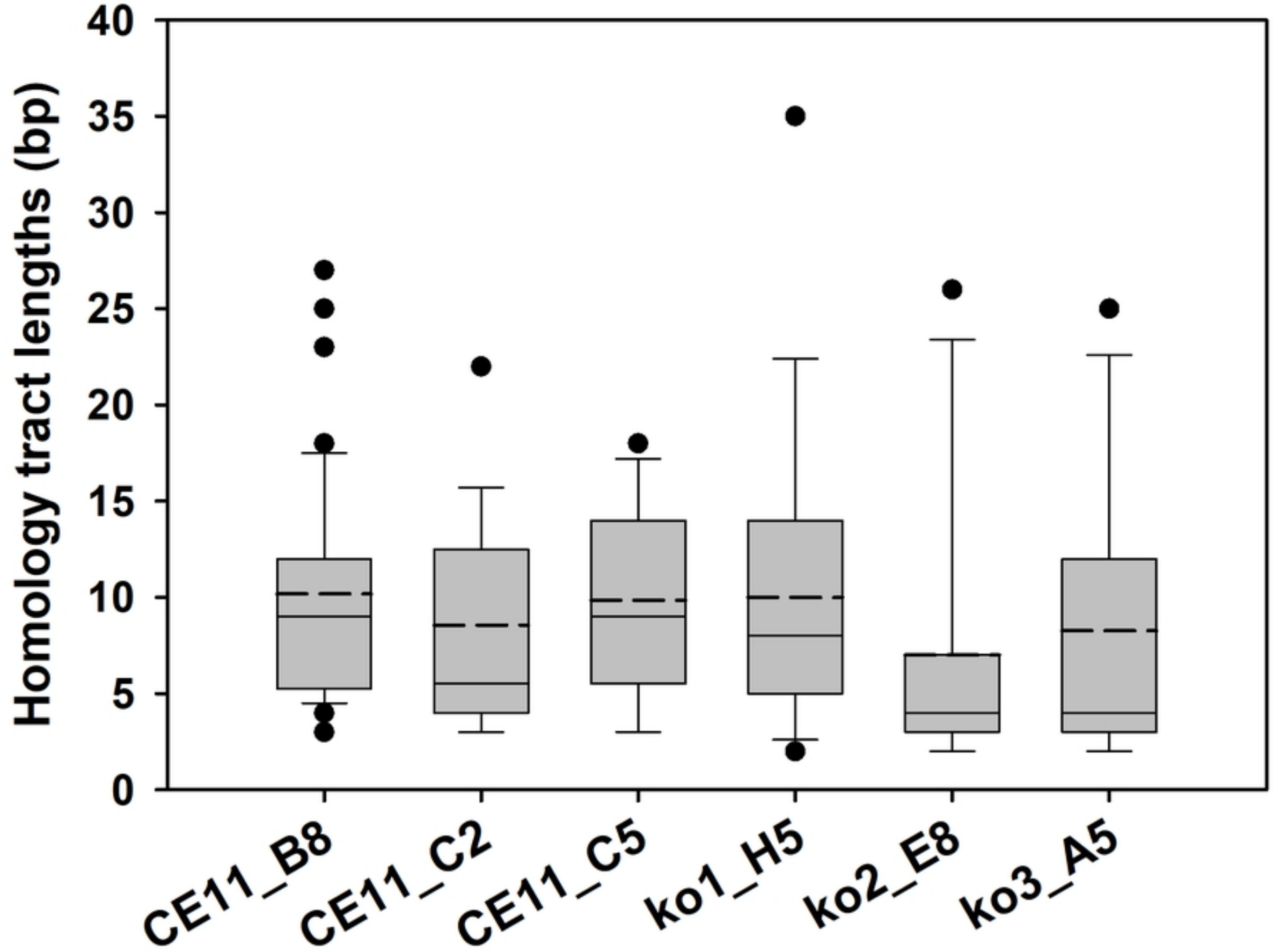
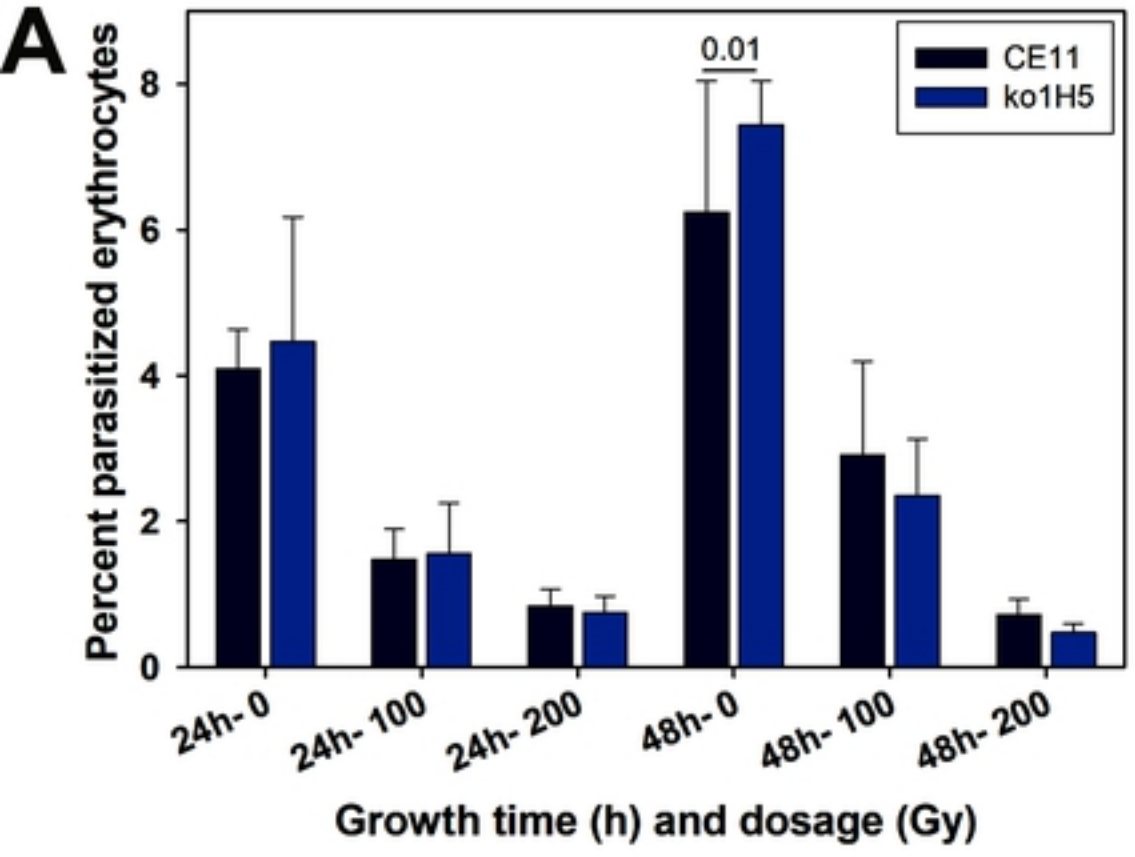

Homology tract lengths (bp)
40
35
30
25
20
15
10
5
0
CE11_B8
CE11_C2
CE11_C5
ko1_H5
ko2_E8
ko3_A5

Figure 7


A
Percent parasitized erythrocytes
0
2
4
6
8
0.01
CE11
ko1H5
24h- 0
24h- 100
24h- 200
48h- 0
48h- 100
48h- 200
Growth time (h) and dosage (Gy)

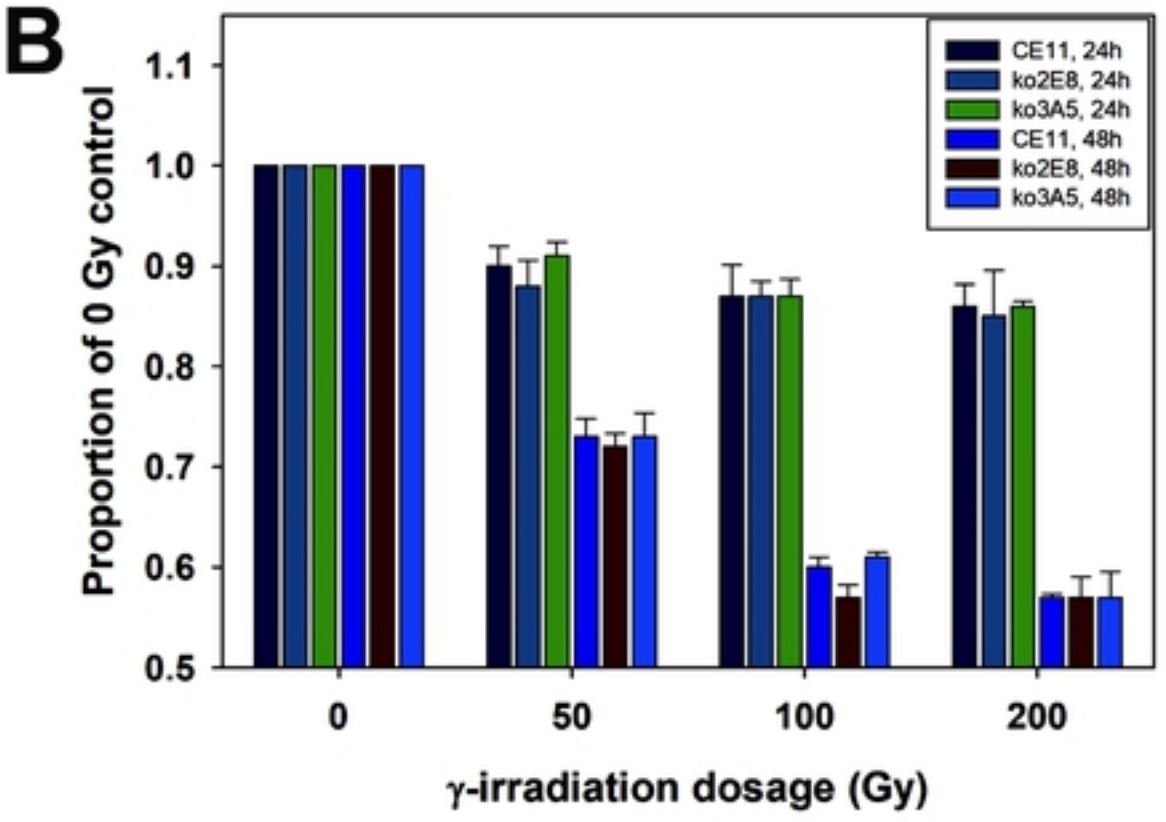

B
Proportion of 0 Gy control
0.5
0.6
0.7
0.8
0.9
1.0
1.1
CE11, 24h
ko2E8, 24h
ko3A5, 24h
CE11, 48h
ko2E8, 48h
ko3A5, 48h
0
50
100
200
γ-irradiation dosage (Gy)

Figure 2

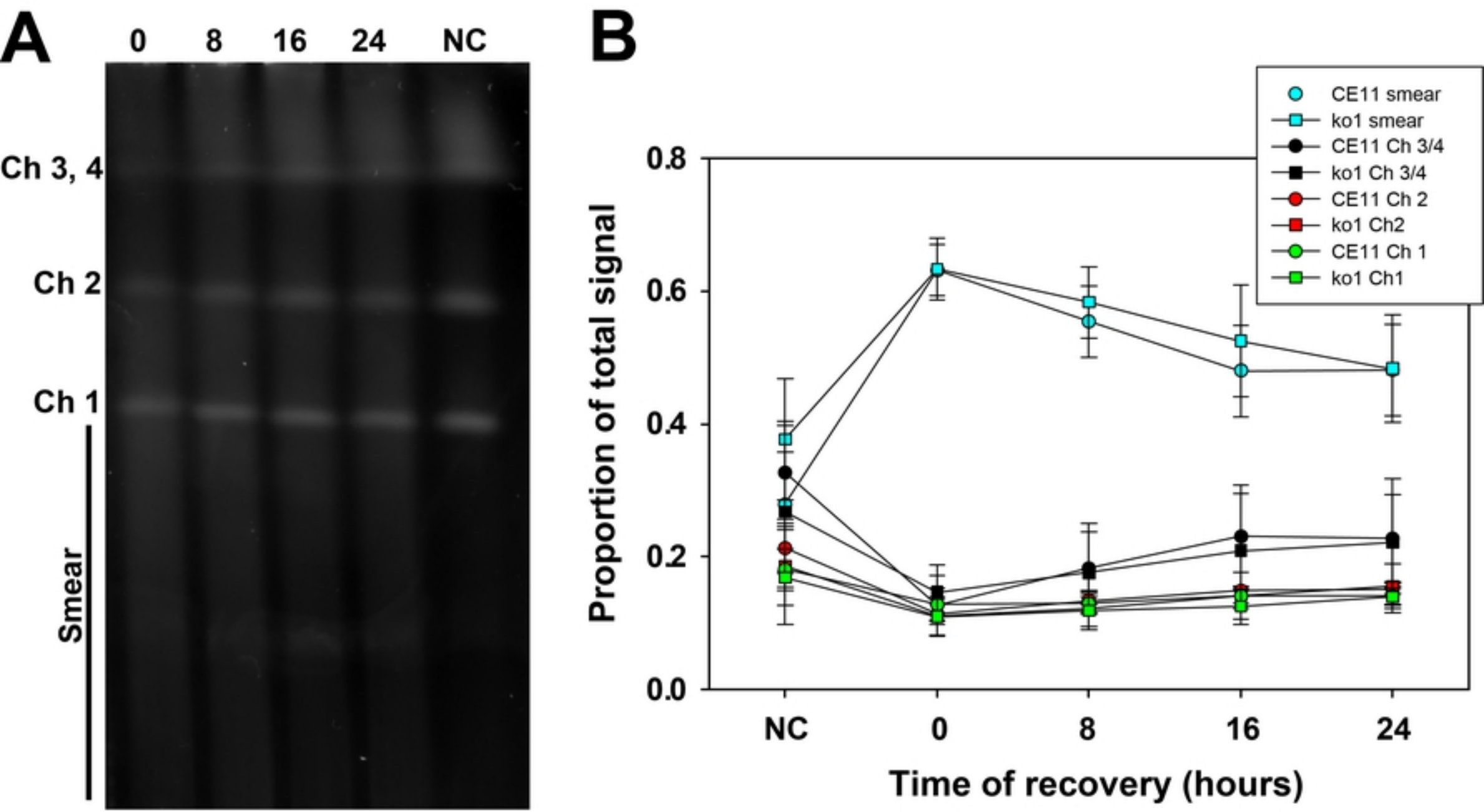

A
0
8
16
24
NC
Ch 3, 4
Ch 2
Ch 1
Smear
B
CE11 smear
ko1 smear
CE11 Ch 3/4
ko1 Ch 3/4
CE11 Ch 2
ko1 Ch2
CE11 Ch 1
ko1 Ch1
Proportion of total signal
0.0
0.2
0.4
0.6
0.8
NC
0
8
16
24
Time of recovery (hours)

Figure 3