

Drosophila bloom helicase maintains genome integrity by inhibiting recombination between divergent DNA sequences

Michael Kappeler, Elisabeth Kranz, Katrina Woolcock, Oleg Georgiev and Walter Schaffner*

Institut für Molekularbiologie der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received July 21, 2008; Revised October 8, 2008; Accepted October 12, 2008

ABSTRACT

DNA double strand breaks (DSB) can be repaired either via a sequence independent joining of DNA ends or via homologous recombination. We established a detection system in *Drosophila melanogaster* to investigate the impact of sequence constraints on the usage of the homology based DSB repair via single strand annealing (SSA), which leads to recombination between direct repeats with concomitant loss of one repeat copy. First of all, we find the SSA frequency to be inversely proportional to the spacer length between the repeats, for spacers up to 2.4 kb in length. We further show that SSA between divergent repeats (homeologous SSA) is suppressed in cell cultures and *in vivo* in a sensitive manner, recognizing sequence divergences smaller than 0.5%. Finally, we demonstrate that the suppression of homeologous SSA depends on the Bloom helicase (Blm), encoded by the *Drosophila* gene *mus309*. Suppression of homeologous recombination is a novel function of Blm in ensuring genomic integrity, not described to date in mammalian systems. Unexpectedly, distinct from its function in *Saccharomyces cerevisiae*, the mismatch repair factor Msh2 encoded by *spel1* does not suppress homeologous SSA in *Drosophila*.

INTRODUCTION

DNA double strand breaks (DSB) can be repaired in two fundamentally different ways, either via mechanisms involving homologous recombination, or without the use of extensive homology (1,2). Single strand annealing (SSA) is a mechanism based on homologous recombination. It can be used for repair if the DSB occurs between repeated sequences that have the same orientation

(direct repeats). The repeats recombine with each other, whereby one repeat and the sequence between the repeats is lost.

In *Escherichia coli* (3,4), *Drosophila melanogaster* (5,6) and mammalian cells (7,8), the efficiency of homologous recombination was shown to depend on the length of uninterrupted sequence identity, as well as on the percentage of sequence identity within the region of homology. It was suggested that the strict homology requirement serves to avoid ectopic (nonallelic) recombination between related but nonidentical sequences (5). Based on this line of reasoning it was proposed that diverged introns may help to maintain the integrity of duplicated genes. According to this hypothesis, fast-evolving introns interrupt the homology between otherwise conserved paralogs, thereby inhibiting ectopic homologous recombination between them (9).

Recombination between similar, nonidentical sequences is termed homeologous recombination, to distinguish from homologous recombination between perfectly matching sequences. In several species the mismatch repair (MMR) pathway was shown to be involved in the suppression of homeologous recombination. In *E. coli* the MMR components MutL, MutS and MutH suppress interspecies recombination (10). In *Saccharomyces cerevisiae* the MutS homologs Msh2 and Msh3 prevent homeologous recombination (11,12), and Msh2 and Msh6 were shown to suppress homeologous SSA (13). In murine cells mutant for *msh2*, the frequency of homeologous, intrachromosomal recombination and the gene targeting frequency with non-isogenic constructs is enhanced (14,15).

Besides the MMR components, RecQ helicases also contribute to genomic stability. In *E. coli*, RecQ prevents illegitimate, ectopic recombination between short stretches of identical sequences (16), and in *S. cerevisiae* the sole RecQ homolog, Sgs1, suppresses homeologous recombination, notably spontaneously occurring, chromosomal translocations between related genes (13,17–19). In *D. melanogaster* the RecQ helicase system is more

*To whom correspondence should be addressed. Tel: +41 44 635 31 50; Fax: +41 44 635 68 11; Email: Walter.Schaffner@molbio.uzh.ch

sophisticated, consisting of three paralogs (RecQ4, RecQ5 and Blm), similar to the situation in mammals with five RecQ members. *Drosophila* Blm was shown to partially rescue the sensitivity of *sgs1* mutant strains to a DNA alkylating agent (20). Loss of Blm causes sterility in *D. melanogaster*, the sperm show high frequencies of chromosome nondisjunction and chromosome loss (21). *Drosophila blm* mutants are also impaired in DNA synthesis during homologous repair (HR) (22). In a system setup that allows the detection of different modes of DNA repair, homologous repair from the homologous chromosome (HR-h) was found to be decreased in *blm* mutants, while SSA frequency increased (23). Furthermore, *blm* mutants are more prone to crossing over, deletions of flanking sequences and template disruption during DSB repair (DSBR) (23,24). Mice and humans mutant for *blm* exhibit enhanced sister chromatid exchange (SCE) (25) and murine cell lines deficient for *blm* exhibit chromosomal instability (CIN) (26).

A recent model summarizes the complex role of Blm in HR (27). It postulates that Blm displays both pro- and anti-recombination activities. On the one hand it removes Rad51 from the single stranded 3' overhang or disrupts the already formed D-loop, thereby inhibiting HR at an early stage (27). On the other hand Blm would also promote HR by unwinding the DNA double helix in front of the D-loop in order to allow for DNA synthesis (22); at a late stage it disrupts the D-loop during HR to facilitate DNA repair via synthesis dependent strand annealing (SDSA) (24), and it also dissolves double Holliday junctions (DHJ) during DSBR, thereby suppressing crossing over (28,29).

Here we have tested SSA frequencies in the fruitfly *D. melanogaster*, using a reporter system with two tandemly arranged genes and a rare-cutting restriction enzyme site in between. After intracellular cleavage, SSA (consisting of resection of DNA ends, annealing of remaining single strands and repair of gaps/protruding strands) results in the expression of a green fluorescent protein. We find that the shorter the spacer between the direct repeats and the more perfect their sequences are matching, the higher the SSA frequency. SSA is thus under tight control in *D. melanogaster* and highly sensitive to sequence divergence and distance between direct repeats. Besides the proposed functions of Bloom helicase in the repair of DSBs mentioned above, we found a novel role for Blm in the inhibition of homeologous recombination: it suppresses SSA between divergent sequences. This suggests an additional role for Blm in the maintenance of genomic integrity. In contrast to results in *S. cerevisiae*, *spell*, the *Drosophila* homolog of the mismatch repair gene *MSH2*, did not show any involvement in the suppression of homeologous SSA.

MATERIALS AND METHODS

Plasmid sequences

All plasmids that contain tester constructs were generated using standard cloning procedures, and corresponding sequences are available on request. For S2 cell culture

experiments, tester constructs were based on *pDrBB2* from Michele Calos' lab.

Tester constructs for *in vivo* experiments (including the *attB* site) were subsequently cloned into *pCasper-4*. I-SceI was expressed in cell cultures via *pAc-SceI*.

Calcium phosphate transfection of S2 cells

Transfection was performed in 12-well plates with Schneider's *Drosophila* Medium (GIBCO, Invitrogen Corporation, Carlsbad, CA, USA), containing 10% FBS (ICN Biomedicals, Irvine, CA, USA) and 1% PenStrep (GIBCO); 0.5 ml medium/well (1.5E6 cells/ml) was used; 1.2 µg tester construct plasmid and 1.2 µg I-SceI expressing plasmid were used per well. Plasmid DNA was mixed with 4× Ca mix and 2× Pi mix in this order, mixture was given to cells 4' after 2× Pi mix addition (4× Ca mix: 0.5 M CaCl₂, 0.1 M HEPES; 2× Pi mix: 0.05 M HEPES pH 7.05, 0.75 mM Na₂HPO₄, 0.75 mM NaH₂PO₄, 0.28 M NaCl). After 12 h 0.5 ml medium containing 500 µM CuSO₄ was added to each well. Flow cytometry analysis was done 48 h after copper addition. All transfections were performed in triplicate.

Flow cytometry

Cells were picked and washed with TBS (25 mM Tris, 137 mM NaCl, 5 mM KCl, 0.7 mM CaCl₂, 0.5 mM MgCl₂, 0.6 mM Na₂HPO₄) and resuspended in 0.5 ml ice-cold BSS (0.14 M NaCl, 5.4 mM KCl, 0.3 mM Na₂HPO₄, 0.4 mM KH₂PO₄, 1 mM CaCl₂, 0.8 mM MgSO₄) containing 2% FBS. Analysis was performed at Cytomics FC 500 (Beckman Coulter, Inc., Fullerton, CA, USA) with excitation at 488 nm, green detection (500 nm < x < 525 nm) and red detection (600 nm < x < 615 nm). At least 20 000 cells were measured per sample. Green fluorescence, as a measure for SSA, was quantified in each sample by multiplying the average green fluorescence intensity from the EGFP positive cells (\bar{x}_{EGFP}) with the number of EGFP positive cells (n_{EGFP}) and was then normalized to the corresponding product of the transfection control. Controls lacking an I-SceI expressing plasmid are not shown in here, but we included them in several pilot experiments. Background fluorescence in those experiments was not zero, but low (values were always lower than 5% of the sample values; data not shown).

FACS

EGFP positive cells were sorted for PCR analysis at the Zentrallabor für Zellsortierung ETH/UNI, Gloriastrasse 35, 8092 Zürich.

Plasmid DNA isolation

Plasmids were isolated from EGFP positive S2 cells applying GFXTM Micro Plasmid Prep Kit (Amersham, GE Healthcare, Waukesha, WI, USA).

Single fly DNA isolation

Flies were individually squashed in buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 0.2% Triton X 100) containing 200 µg/ml proteinase K and incubated for 30 min at 37°C followed by 15 min at 95°C.

PCR analysis of SSA products

For PCR amplification of SSA products MtnB5'-f1 (CCA GGCTTGCACACGACGTG) and EGFP-r (ACGTCG CCGTCCAGCTCGACCAG) were used. Sequencing of these PCR products was done with MtnB5'-f2 (GCAA TTTTGCACCTCGTTTCG).

The contrast of the PCR product gel picture in Figure 3B was enhanced with Photoshop (+57).

Fly stocks

For *C31 attP/attB* integration we used $y^1 w^{1118}; M\{RFP.attP\}zh86-Fb$ stocks with *attP* landing site on the third chromosome. For analysis of *blm* mutants $w^{1118}; p\{v+70I-SceI\}; mus309^{D2}/TM2 y^+$ and $mus309^{D3}/TM6B$ were used. For analysis of the *msh2* homolog *spell* the following stocks were used $y^1 w^{1118}; Df(2L)b84h1 p\{CaGal\}/CyO$ and $y^1 w^{1118}; Df(2L)k08712-rv21 p\{v+70I-SceI\}/CyO$.

Generation of transgenic flies

attB containing pCasper plasmids were co-injected with *C31* integrase capped mRNA into the fly stock mentioned above to generate stable fly lines as previously described (30).

Insertion of the same pCasper plasmids via P-element transformation was performed according to standard protocols.

SSA detection in male germline

To measure SSA frequencies in the male germline, experiments were generally performed as shown for the genomic *mtnB* direct repeat (*MtnBgen-gen*). Parental strains (P) $y^1 w^{1118}; p\{v+70I-SceI\}$ and $y^1 w^{1118}; M\{MtnBgen-gen\}zh86-Fb$ were crossed. Eggs were collected during 24 h before I-SceI endonuclease was induced by a heat shock at 37°C for 1 h. A second heat shock was applied 24 h later. Eclosing male flies (F1) were individually crossed with 3 *y w* virgins. Among the offspring (F2) the ratio of EGFP expressing flies to RFP positive flies was calculated to determine SSA frequency.

RESULTS

SSA frequency is reduced by sequence divergence and by increased spacer length between direct repeats in *Drosophila* S2 cells

Recombination by SSA can take place if a DSB occurs between direct repeats. In order to determine parameters that influence the frequency of SSA, we developed an SSA detection system in which different parameters can be modified. The tester construct contains the 18-bp cleavage site for the rare cutting endonuclease I-SceI, which allows for the induction of a DSB, producing a staggered cut with a 4 bp 3' overhang. A first copy of the metallothionein B gene (*mtnB*), including the regulatory enhancer/promoter DNA, is located upstream of this cleavage site. A second *mtnB* copy, without enhancer and promoter but fused to the enhanced green fluorescent protein (EGFP), is downstream of the cleavage site. Upon SSA of these two *mtnB*

copies, expression of the MtnB-EGFP fusion protein is inducible via the *mtnB* enhancer/promoter (Figure 1A). We first analyzed SSA frequencies in a transient transfection assay in *Drosophila* S2 cells. Plasmids containing the tester constructs were cotransfected with a second plasmid expressing I-SceI under the actin promoter. Expression of MtnB-EGFP was induced by CuSO₄ addition to the culture medium, and positive cells were detected by flow cytometry.

We measured SSA frequency in a construct with two identical genomic copies (259 bp) of *mtnB* (*MtnBgen-gen*) and compared it with two further constructs: one harboring a point mutation in the intron of the first *mtnB* copy (*MtnBgen*-gen*), and the other containing an intronless first *mtnB* copy (*MtnBcDNA-gen*) (Figure 1B). SSA in these constructs can form intermediate heteroduplex structures, including a non-matching DNA region, which consists of only 1 nt in *MtnBgen*-gen*, whereas in *MtnBcDNA-gen* the non-matching region comprises 61 nt. We found that the single mismatch in the heteroduplex of *MtnBgen*-gen* reduces SSA frequency to 50% of that observed with *MtnBgen-gen*. SSA in *MtnBcDNA-gen*, with the non-matching 61 bp intron, is even reduced to 25% (Figure 2). SSA in S2 cells therefore seems to be strictly regulated, sensing a 1-bp pair mismatch within a 259 bp DNA stretch. This points to the existence of a control mechanism to suppress homeologous SSA in *D. melanogaster*. Besides the impact of sequence divergence on SSA frequency, we also wished to examine the role of the spacer length between two direct repeats since in human cell culture experiments, efficiency of SSA decreases with increasing spacer length (31). The SSA constructs mentioned above have spacer sequences of 191 bp and 13 bp upstream and downstream of the I-SceI cleavage site, respectively. We added additional spacer sequences of 1 kb and 2.2 kb to either side of the I-SceI site and analyzed SSA frequencies. Identical as well as divergent repeats showed a decrease in SSA with increasing spacer length (Figure 2).

To test whether the green fluorescence is truly caused by intramolecular SSA, rather than intermolecular recombination, we cotransfected *Drosophila* S2 cells with a first plasmid, containing *mtnB* (including promoter), and a second, separate plasmid, containing the promoterless *mtnB-EGFP* fusion construct. As usual, the I-SceI endonuclease gene was provided on a different plasmid. This resulted in only a few green cells, which confirms that most of the EGFP expression observed in the regular SSA tester constructs is caused by intra-plasmid recombination events (Figure 3A). To further analyze whether these intra-plasmid recombinations result in the predicted SSA products, we transfected S2 cells with *MtnBgen-gen*, *MtnBgen-cDNA*, *MtnBcDNA-gen*, or *MtnBcDNA-cDNA* plasmid constructs. We sorted EGFP positive cells, isolated plasmid DNA from these cells and performed PCR with primers spanning the direct repeats. Bands of the expected size for the SSA product were obtained with plasmid DNA retrieved from fluorescent S2 cells (Figure 3B).

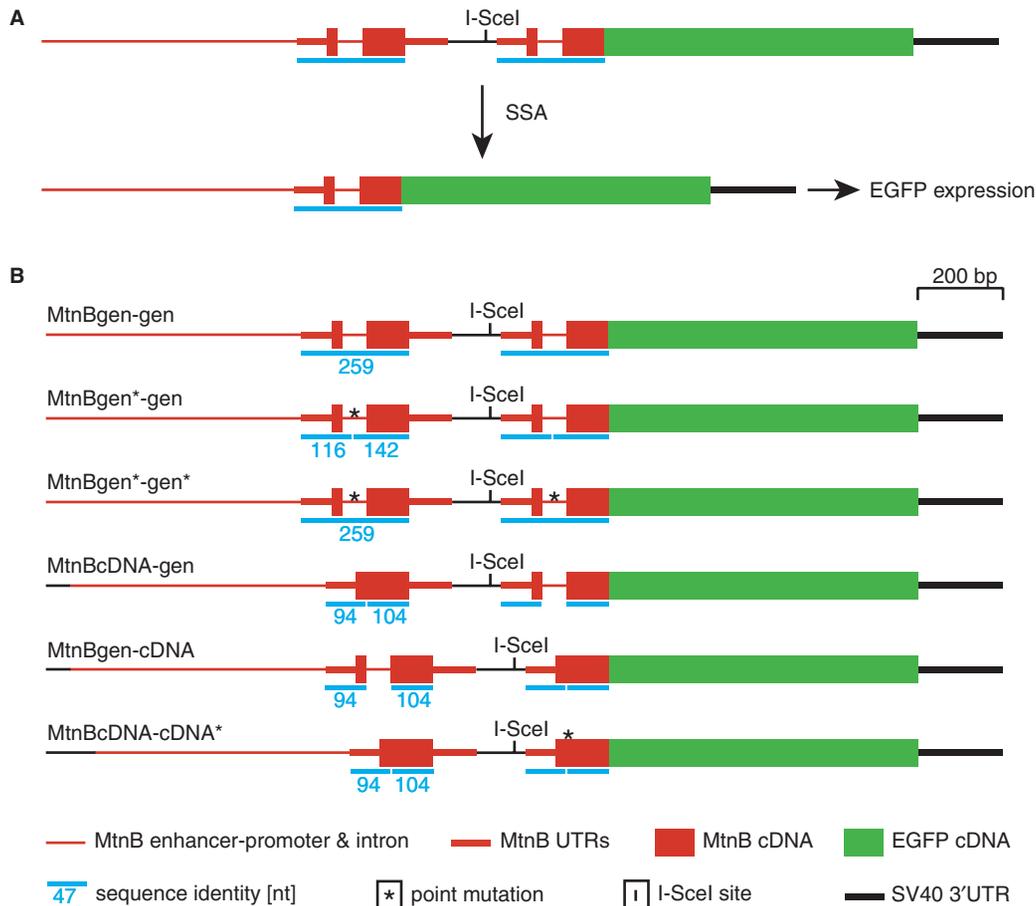


Figure 1. Schematic representation of DNA constructs for SSA detection. **(A)** One of the two direct repeats, the one upstream of the I-SceI endonuclease cleavage site, is under the control of the *mtnB* enhancer/promoter. The coding region of the downstream repeat, lacking regulatory sequences for transcription, is fused in-frame to the *EGFP* cDNA. Induction of transcription does not lead to expression of EGFP protein, unless the DSB, generated by the endonuclease I-SceI, is repaired via SSA of the flanking metallothionein copies. **(B)** All constructs contain an I-SceI endonuclease recognition site (I-SceI), which is flanked by genomic (gen) or cDNA copies of *mtnB*. The length of sequence identity between tandem duplicates is indicated by blue bars and numbers, which represent nucleotide numbers. The longest stretch of sequence identity is 259 nt in the construct containing two genomic *mtnB* copies (MtnBgen-gen). This identity is disturbed in several constructs either by a silent point mutation (MtnBgen*-gen, MtnBcDNA-cDNA*) or the absence of the intron (MtnBcDNA-gen, MtnBgen-cDNA).

Sequence divergence and spacer sequence also decrease frequency of SSA in transgenic flies

To find out if the sensitive SSA control mechanism is also present *in vivo*, we analyzed SSA in the male germline of *D. melanogaster*. We expressed I-SceI during embryonic and larval stages in male flies harboring the SSA tester construct (as outlined in Materials and Methods section). Offspring of these males were then screened for green flies. We first investigated whether the integration of the same construct into different genomic loci influences SSA frequency. For this we inserted the MtnBgen-gen transgene into the genome either randomly with the aid of a P-element transposase, or at a predefined *attP* 'landing site' on the third chromosome by the phage *C31* integrase (30,32). There was considerable variation in the percentage of green progeny of flies with insertions at different loci (Figure 4A). Therefore, to more accurately compare SSA frequencies of different tester transgenes, we chose to insert them at the same genomic locus via the *C31*

integrase system. We started out by testing if SSA frequency depends on the spacer length as in S2 cells and human cell lines. For this we compared MtnBgen-gen with and without the intervening 2.2 kb spacer DNA. Similar to the cell culture results, the spacer reduced the frequency by 91% (Figure 4B). We next investigated whether sequence divergence also reduces SSA frequency *in vivo* like in S2 cells. Indeed, SSA proved to be very sensitive to mismatches, since 1 nt difference in a 259 nt stretch reduced SSA frequency by 58% (Figure 4C). As a control to ensure that the point mutation itself does not influence the readout, we also tested MtnBgen*-gen*, where the same point mutation is present in both *mtnB* copies, thereby restoring a 259 nt repeat of identical sequence. The SSA frequency of MtnBgen*-gen* was as high as that of MtnBgen-gen (Figure 4D). We next analyzed the influence of a longer stretch of unpaired DNA, located within the duplication, on SSA. MtnBcDNA-gen showed a pronounced and highly significant reduction of SSA in comparison to MtnBgen*-gen.

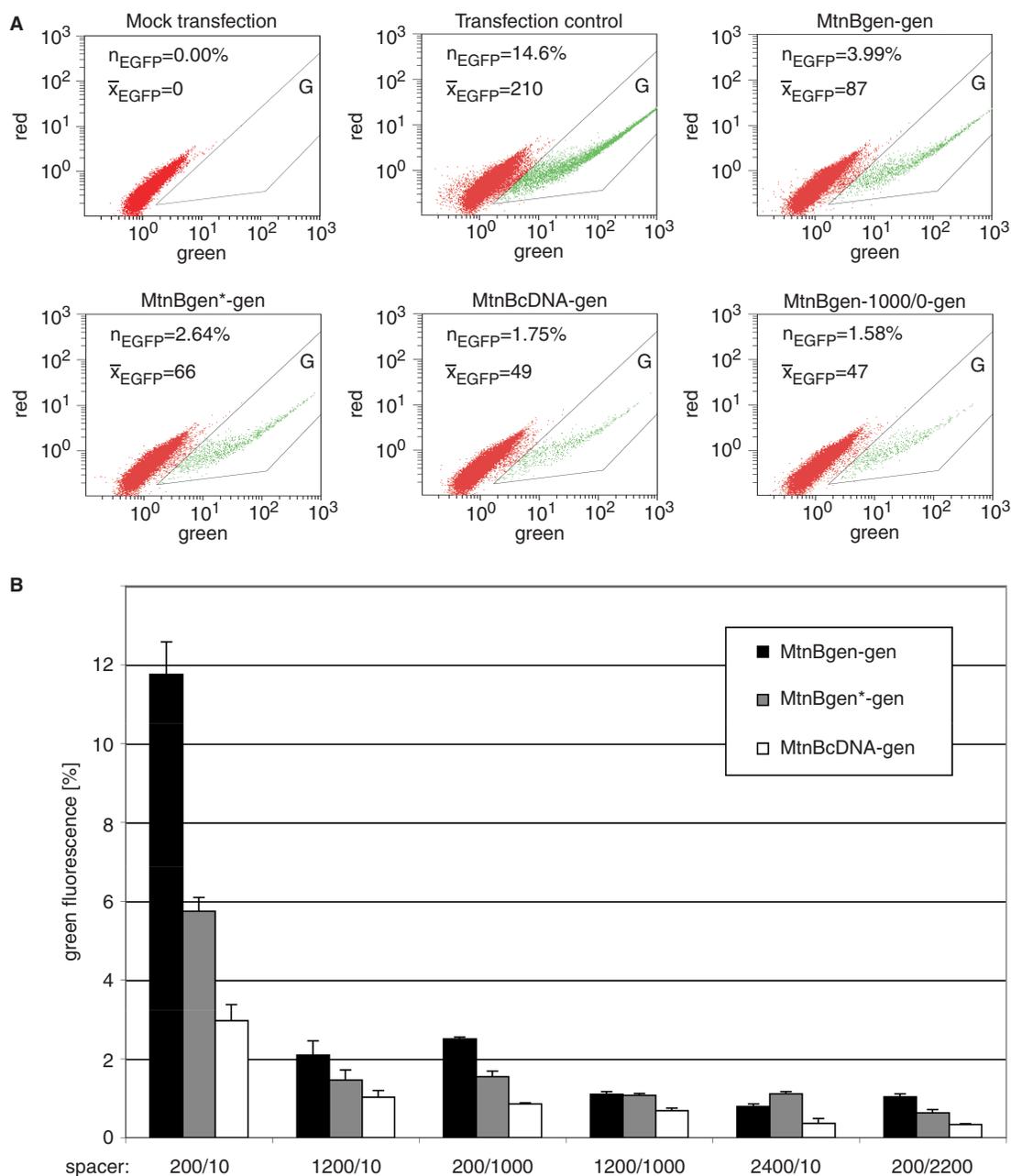


Figure 2. SSA recombination is impaired by mismatches and spacer DNA in transiently transfected S2 cells. S2 cells were transfected with two plasmids, one containing the SSA tester construct and the other encoding the endonuclease I-SceI under the actin promoter. SSA products resulted in *EGFP* cDNA driven by the copper inducible *mtnB* enhancer/promoter. 12 h after transfection EGFP expression was induced by addition of 250 μ M $CuSO_4$. (A) In order to reduce the background luminescence, green fluorescence was plotted against red. All cells with specific green fluorescence are shifted to the right. An area for EGFP positive cells was defined (G). For every sample the percentage of EGFP positive cells (η_{EGFP}) and the mean fluorescence intensity of these cells (\bar{x}_{EGFP}) were determined. All indicated constructs are depicted in Figure 1B except MtnBgen-1000/0-gen, which contains an additional 1000 bp spacer upstream of the I-SceI recognition site. (B) On the x-axis the length of spacer upstream/downstream of the I-SceI site are indicated. Green fluorescence as a measure for SSA was calculated as described in Materials and Methods section. Error bars represent standard deviations of three experiments.

As expected the same reduction was also seen when we reversed the order of the genomic and the cDNA copy of *mtnB* (MtnBgen-cDNA) (Figure 4E). The reduction is probably due to the fact that in a heteroduplex generated by MtnBcDNA-gen the 61 bp intron of the genomic copy bulges out as a large non-matching loop, whereas in MtnBgen*-gen the heteroduplex is distorted only by a

1-bp mismatch. However, an alternative explanation is the reduced length of identical sequence flanking the site of interruption (Figure S1). We therefore analyzed SSA of MtnBcDNA-cDNA*, where we had introduced a silent point mutation in the *mtnB* cDNA at the site where the intron is normally located. Interestingly, the SSA frequency in MtnBcDNA-cDNA* was significantly higher

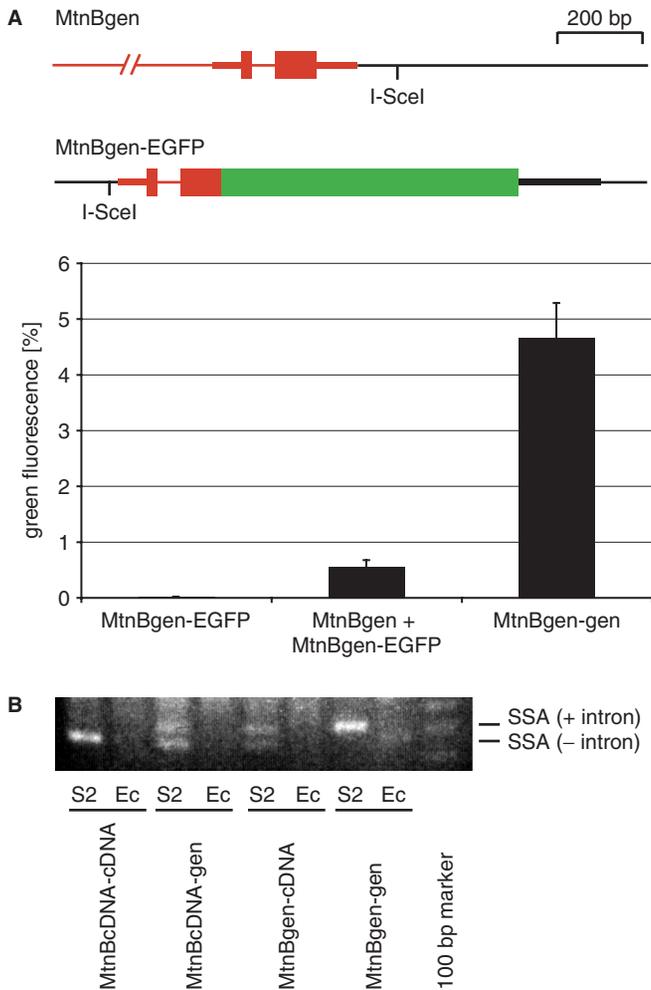


Figure 3. Verification of SSA in transiently transfected S2 cells. **(A)** Constructs for transfection are depicted: enhancer/promoter-containing *mtnB* copy (MtnBgen) and enhancer/promoter-less *mtnB-EGFP* fusion copy (MtnBgen-EGFP). For the positive control construct for intramolecular SSA (MtnBgen-gen), see Figure 1B. Constructs indicated in the bar diagram were always cotransfected with a plasmid expressing the endonuclease I-SceI. Green fluorescence as a measure for SSA was calculated as described in Materials and Methods section. Error bars represent standard deviations of three experiments. **(B)** Isolation of SSA products from S2 cells. EGFP positive cells were first sorted, and plasmid DNA was isolated. PCR to detect SSA products was performed. The indicated tester constructs are depicted in Figure 1B. PCR products of plasmid DNA from EGFP positive S2 cells after DSB induction (S2) and from *E. coli* without induction of DSB (Ec) are shown. Expected PCR bands for SSA products are 431 bp and 370 bp (with and without the intron). The three marker bands shown indicate the position of 300 bp, 400 bp and 500 bp.

than in the intron-containing construct MtnBcDNA-gen. Thus, not only the length of the identical sequence but also the size of the unmatched DNA sequence itself influences the efficiency of SSA. (Figure 4F). We also verified that repair occurred via SSA, rather than via unequal crossing over, by exploiting the fact that unlike SSA, DSB repair via homologous repair depends on Rad51 (33,34). In a *rad51* mutant background, EGFP positive flies emerged with similar frequency as in controls, indicating that most, if not all repair occurred via SSA (data not shown).

We also did an experiment to see if homologous repair from an ectopic donor sequence is suppressed by mismatches in the donor sequence. To this end, the transgene arrangement was chosen such that conversion to a functional GFP reporter required homologous recombination, rather than SSA (Figure S7). However, in spite of testing many offspring, recombination frequencies were too low for significant results. An identical donor sequence yielded two recombination events among 1650 flies, 1 nt mismatch gave four events in 2141 flies, while the deletion of the *mtnB* intron gave one event in 3142 flies. Successful recombination events were verified by PCR (data not shown).

Drosophila Bloom helicase (Blm) suppresses homeologous SSA *in vivo*

In budding yeast the RecQ helicase Sgs1 and the mismatch repair factors Msh2 and Msh6 were shown to be involved in the suppression of homeologous SSA (13). As mentioned, *Drosophila mus309*, encoding Blm, is an ortholog of yeast *sgs1*, while *spell* is an ortholog of *msh2*. We were interested to know if in *D. melanogaster* these orthologs also have a similar role in SSA. For this purpose we crossed heterozygous mutants and assayed DSB repair via SSA in the offspring's germline. Using this approach, offspring consisted of homozygous mutants as well as heterozygous and wildtype controls.

For the analysis of Blm, we used compound heterozygotes with two mutant alleles, *mus309^{D2}* and *mus309^{D3}* (Figure S2), since both alleles behave as genetic nulls (21). PCR of the SSA product with subsequent sequencing was performed in all EGFP positive flies to subtract NHEJ events leading to rare random in-frame ligated *mtnB-EGFP* fusion products. The values below are based on corrected numbers. No significant change in SSA frequency between *blm* mutant flies and the controls could be seen with 100% sequence identity (MtnBgen-gen). A small increase of SSA frequency was seen in mutant flies, when we analyzed repair of a construct with a 1-bp mismatch (MtnBgen*-gen). With the more divergent construct MtnBcDNA-gen, a strong increase of SSA frequency was detected in mutants (Figure 5A).

We then wondered if Bloom helicase is also responsible for the suppression of recombination between even more divergent sequences like *mtnB* and *mtnD* cDNA, derived from the most closely related paralogs of the four metallothionein genes in *D. melanogaster*. These cDNAs are 129 nt long and exhibit only 81% sequence identity. Analysis of the MtnB-MtnD construct (Figure 6A and Figure S3) revealed that also in this case, suppression of homeologous SSA is relaxed in Blm deficient flies, resulting in 4.5-fold elevated SSA frequencies (Figure 6B). Sequencing analysis of the repair products revealed an interesting fact. Neither mutant, nor control flies showed any evidence for discontinuous mismatch repair of the heteroduplex DNA (540 potential mismatches in total), since all the MtnB-MtnD SSA repair products contained pure *mtnB* type sequence upstream and *mtnD* type sequence downstream of the recombination site; in other words, discontinuous repair patterns were never seen

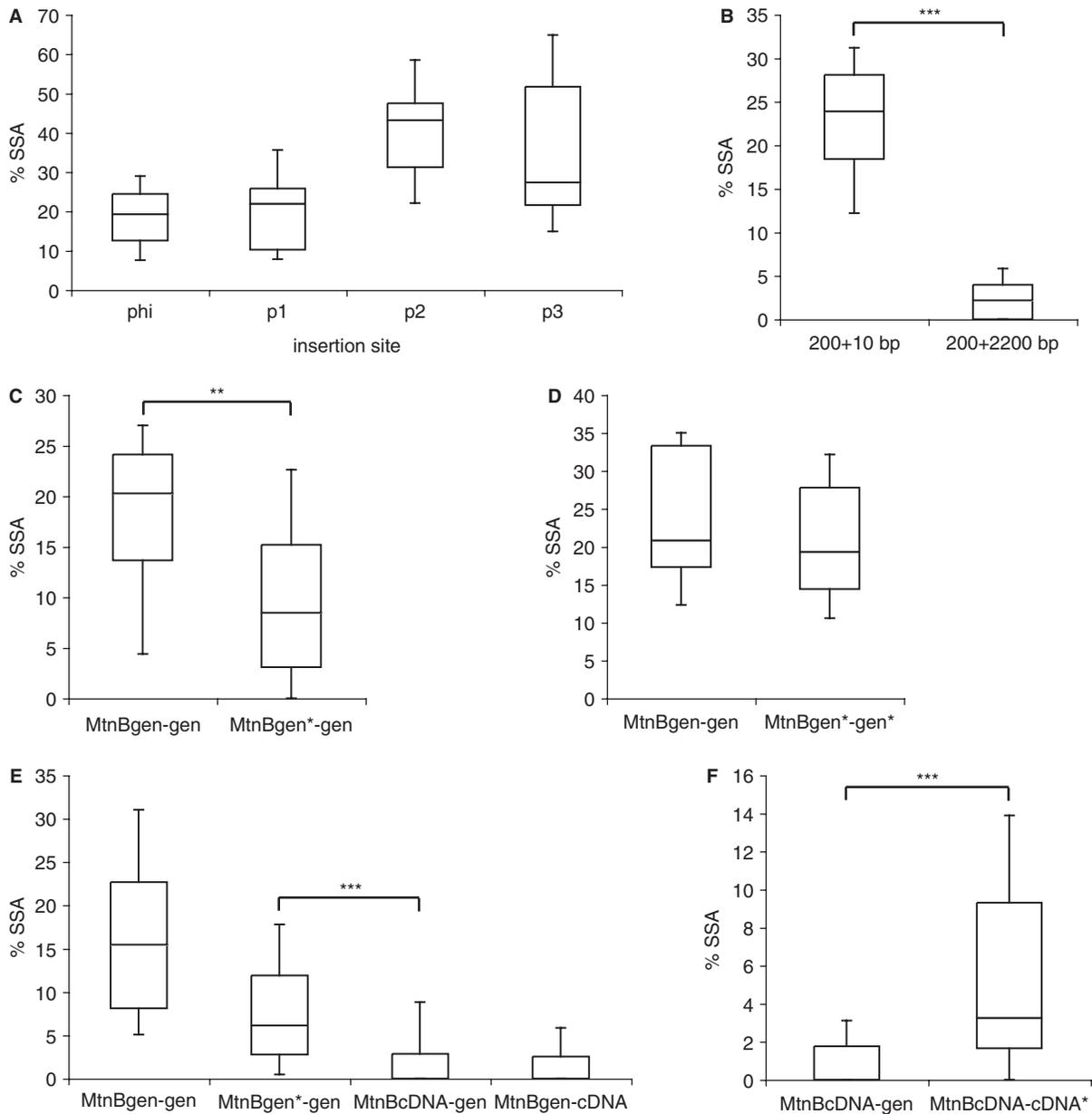


Figure 4. *In vivo* detection of SSA in the male germline. SSA frequency in flies was determined as described in Material and Methods section. Graphs show box plots for SSA frequencies with the middle line marking the median frequency of all tubes, boxes represent 25th and 75th percentiles and whiskers 10th and 90th percentiles. The two-tailed Mann Whitney U-Test was performed to assess statistical significance. $**P < 0.01$; $***P < 0.001$. All constructs are depicted in Figure 1B. (A) SSA frequencies of MtnBgen-gen tester transgenes inserted at four different loci in the genome. The test lines were generated by targeted integration using the *attP/attB* system (phi) or random insertion using the traditional P-element system (p1, p2, p3). (B) SSA frequency of the MtnBgen-gen tester construct (with direct repeats spaced by 200 nt) compared to MtnBgen-gen containing an additional 2.2 kb spacer sequence between the repeats, downstream of the I-SceI recognition site. (C–F) SSA frequencies of the tester constructs indicated.

(Figure 6C). In control PCRs with DNA from flies with uncleaved DNA, we always detected only the longer product corresponding to the uncleaved DNA, thus PCR artefacts can be excluded.

To investigate the influence of the *msh2* ortholog *spell* on homeologous SSA, we analyzed compound heterozygous flies, harboring two different deletions, which overlap at the *spell* locus (35). Both deletions are homozygous lethal. Analysis of the four constructs, MtnBgen-gen, MtnBgen*-gen, MtnBcDNA-gen (Figure S4) and

MtnB-MtnD (Figure S5), did not reveal a change in SSA frequency between *spell* deficient flies and heterozygous controls (Figure 5B). Because we already induce DSB formation in embryos, we performed a control experiment with *spell* deficient parents to exclude any contribution of *spell* gene product to the embryo by heterozygous parents (Figure S6). If *spell* suppresses SSA between divergent sequences, the effect should be most evident in the divergent MtnB-MtnD construct. However, analysis of this construct in flies with *spell* deficient parents showed that

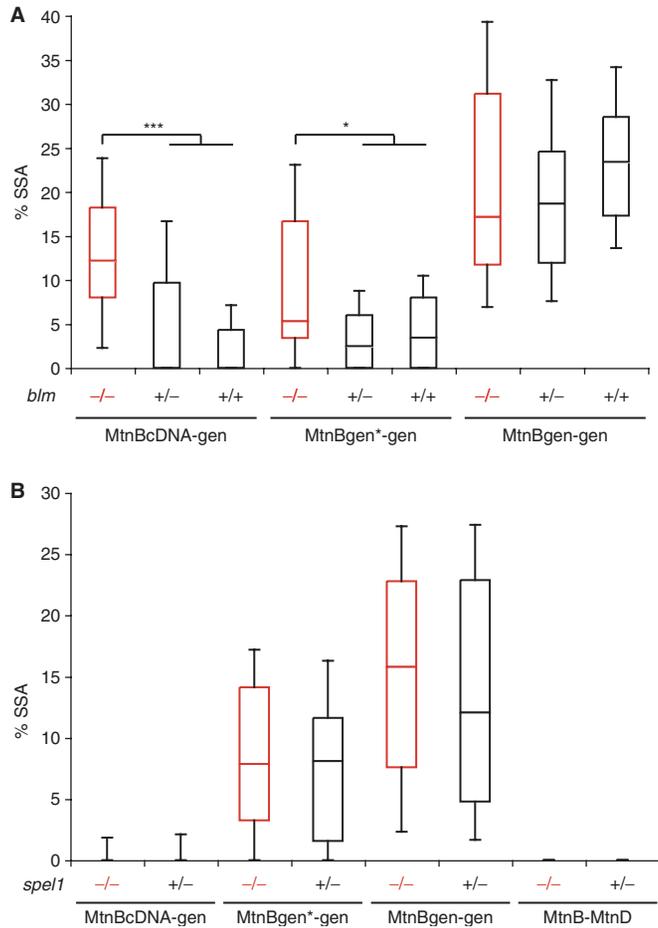


Figure 5. Homeologous SSA in male germline of *blm* (*mus309*) mutants and *spell* mutants. Box plots for SSA frequencies are depicted with the middle line marking the median of all tubes, boxes represent 25th and 75th percentiles and whiskers 10th and 90th percentiles. **(A)** Analysis of *blm* mutants. The rare cutting endonuclease I-SceI was expressed under control of the heat shock promoter in *blm* compound heterozygotes (*mus309^{D3}/mus309^{D2}*), in heterozygotes (*mus309^{D3}/+*) and wt controls; they are labeled in the graph with *-/-*, *+/-* and *+/+*, respectively. All constructs (MtnBgen-gen, MtnBgen*-gen and MtnBcDNA-gen) are depicted in Figure 1B. SSA PCR products of all EGFP positive flies were sequenced to exclude random in-frame NHEJ events, graphs represent corrected data. The two-tailed Mann Whitney U-Test was performed to test for statistical significance between flies deficient (red) and proficient (black) for Blm. * $P < 0.05$; *** $P < 0.001$. In this experiment, tester constructs were inserted with the help of P-elements at different genomic loci on the second chromosome. SSA frequencies can therefore only be compared within the same tester construct. **(B)** Analysis of SSA in flies mutant for the *msh2* homolog *spell*. SSA was analyzed in *spell* compound heterozygotes (*-/-*) and heterozygous controls (*+/-*), which were both derived from a cross of *spell* heterozygous parents. The tester constructs MtnBgen-gen, MtnBgen*-gen and MtnBcDNA-gen (Figure 1B) are integrated at the *attP* landing site on the third chromosome, and MtnB-MtnD (Figure 6A) was P-element inserted on the X-chromosome (line px-3).

SSA frequency (2 events/850 flies) was as low as in wildtype flies (5 events/1950 flies) (Figure 6B).

DISCUSSION

Here we describe some critical sequence constraints for recombination via single strand annealing (SSA) in

D. melanogaster. SSA frequency was highest with identical tandem repeats and a minimal length of spacer DNA between them. A single nucleotide difference between 259 nt long direct repeats is recognized and reduces SSA frequency by 50%. We identified *blm*, coding for the Bloom helicase of the RecQ helicase family, as a gene responsible for the suppression of SSA between divergent repeats (termed homeologous SSA), thereby preventing potentially deleterious recombination events such as chromosomal translocations between paralogous genes. In *S. cerevisiae*, the sole RecQ helicase Sgs1 is involved in inhibiting SSA between divergent sequences (13). In contrast to results in yeast, the gene product of the *D. melanogaster* Msh2 ortholog *spell* did not show any suppression of SSA between divergent sequences.

Most obviously, SSA can reduce any fortuitously duplicated DNA sequence back to one copy. Recent studies in both humans (36,37) and drosophilids (38) have revealed an unexpected genomic variability within one and the same species, notably duplications of DNA segments, both coding and noncoding. In most cases these are thought to be stochastic events without any benefit for the carrier, but such duplications undoubtedly also allow for the generation of gene families whose members have overlapping but distinct, important functions. Thus an organism faces the task of removing useless, if not detrimental DNA duplications yet preserving gene family members, which are particularly abundant in higher eukaryotes including insects and mammals. A typical example is the *Drosophila* metallothionein family with the closely related, chromosomally linked members *mtnB* and *mtnD* that must have arisen by duplication. Fortunately, the tendency of introns to diverge faster than coding sequences mounts an additional barrier against such undesired gene losses (9); in agreement with this notion, the introns are most divergent between *mtnB* and *D*. Compared to identical repeats, we indeed find SSA between *mtnB* and *mtnD* transgenes to be strongly reduced, and also show that the homolog of human Bloom helicase (but not the MSH2 homolog spellchecker), is an important discriminatory component. The guardian function of Bloom helicase in preventing ectopic SSA recombination appears particularly important in higher organisms, including *Drosophila*, which harbor many gene families.

In addition to these general considerations, a specific substrate for naturally occurring SSA could be the hybrid element insertion (HEI) process in *Drosophila*; the corresponding P-elements located on sister chromatids recombine in concert with a nearby locus on the homologous chromosome, resulting in a deletion and a duplication product. The generated duplication is prone to undergo SSA because the P-element is still located between the duplicates (39). In this process, SSA would preserve the original genome sequence.

An alternative way of looking at SSA is in the context of the synthesis dependent strand annealing (SDSA) mechanism, which was demonstrated to be a major pathway for DSB repair in *D. melanogaster* (34). In the SDSA model, a DSB is followed by 5' resection of the DNA ends, as it is the case in SSA. The resulting 3' overhangs may

two or more DSBs occur simultaneously in similar sequences. Recombination between these sequences would lead to chromosomal translocations, which would most likely be deleterious (40).

So far no evidence existed for the involvement of Blm in the control of sequence constraints during homologous recombination, despite an established role of Blm in safeguarding the genome during DSB repair. A model published recently summarizes the complex involvement of Blm in homologous repair (HR) (27). It includes the removal of Rad51 from the single stranded 3' overhang, dissociation of the D-loop, unwinding of the DNA double helix in front of the D-loop and dissolution of double Holliday junctions (DHJ) (27). This essential role in HR makes it difficult to investigate a possible additional function of Blm in the suppression of HR between divergent sequences. Different from the situation in HR, such an analysis is possible in SSA, where Blm is not essential.

We show here with several substrate constructs that Blm suppresses SSA between divergent sequences but not between identical ones. As discussed before, this might be important during SDSA in regions containing repeat sequences. In addition one might speculate that Blm also suppresses strand invasion at ectopic templates during HR, which would be in line with the yeast data, where Sgs1 inhibits chromosomal translocations between similar but divergent genes (17). However, analyzing homologous repair with an ectopic donor sequence, we could not detect significant differences in repair frequencies between constructs that contained an identical donor sequence and others that contained donor sequences harboring mismatches since repair events occurred very rarely.

In *Drosophila blm* mutants, a general shift from HR towards SSA was reported, but only if HR used the homologous chromosome as a template (HR-h) and not if the sister chromatid was used (HR-s) (23,33). In our experiments the SSA tester constructs were always in a heterozygous state, hence HR-h was not possible. As observed by others under comparable conditions, we did not see a general shift towards SSA in *blm* mutants if the repeats were identical. We find however that the SSA frequency was only elevated in *blm* mutants when heteroduplexes could be generated between divergent repeats. Data from *S. cerevisiae* suggest that heteroduplexes are unwound rather than degraded (13). We therefore propose that the Blm helicase unwinds heteroduplexes during DSB repair in *D. melanogaster*.

Although we saw a pronounced, robust effect in *blm* mutants, it is likely that we still underestimate the impact of Blm on the suppression of homeologous recombination. Firstly, *S. cerevisiae* has only one RecQ helicase, Sgs1, which suppresses recombination between divergent sequences, whereas *D. melanogaster* has three RecQ helicases: RecQ4, RecQ5 and Blm. Redundancy between these helicases in the suppression of homeologous recombination appears possible. Secondly, the parents of the investigated flies were heterozygous for *blm* since Blm deficient females are sterile. Thus a parental contribution of Blm mRNA or protein cannot be excluded. Indeed we have observed that maternal contribution of DSB repair

factors can improve DNA repair efficiency in embryos (unpublished results).

The mismatch repair (MMR) system was found to suppress recombination between divergent sequences in various species, such as *E. coli* (10), budding yeast (11–13) and the mouse (14,15). The *E. coli* MMR factor MutS as well as the MutS homologs Msh2 in yeast and mice suppress homeologous recombination. Thus it was quite unexpected that *spell* mutant flies, encoding the *Drosophila* Msh2 homolog, did not show increased SSA frequency between divergent sequences. In these experiments residual *spell* activity can be excluded because the *spell* gene is deleted. Furthermore, a control experiment where both parents were homozygous mutants yielded the same result. This suggests that in contrast to the situation in other model organisms, MMR is not involved in the suppression of homeologous recombination in *D. melanogaster*. However, we cannot exclude that another component of the MMR machinery in *D. melanogaster*, Msh6, contributes to the suppression of homeologous recombination.

Recapitulating the results presented in this work, one can state that in *D. melanogaster* Blm revealed a novel function in suppressing DSB repair between divergent sequences. In contrast to the situation in mammals and yeast where Msh2 also suppresses homeologous recombination, in *D. melanogaster*, the Msh2 ortholog, Spell, was not suppressing recombination between divergent sequences. Based on these data we propose that Bloom helicase, by increasing recombination fidelity, helps to inhibit chromosomal rearrangements, and thus might specifically prevent the loss of related genes that originated by duplication. Our studies have shed light on the way different species handle the difficult task of maintaining genome integrity while repairing damage to their DNA.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Amy C. Groth and Michele P. Calos for the plasmid *pDrBB2*, Johannes Bischof and Konrad Basler for *attP* flies, Carlos Flores and William Engels for fly strains mutant for *mus309* and deficient for *spellchecker1*. We further thank Eva Niederer for fluorescence activated cell sorting and George Hausmann and Albert Pastink for valuable discussions.

FUNDING

'Controlled Gene Integration' project of the European Union (LSHG-CT-2003-503303 to W.S.); the Schweizerischer Nationalfonds; the Kanton Zürich. Funding for open access charge: Kanton Zürich.

Conflict of interest statement. None declared.

REFERENCES

- Shrivastav, M., De Haro, L.P. and Nickoloff, J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res.*, **18**, 134–147.
- Pastink, A., Eeken, J.C. and Lohman, P.H. (2001) Genomic integrity and the repair of double-strand DNA breaks. *Mutat. Res.*, **480**, 37–50.
- Shen, P. and Huang, H.V. (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics*, **112**, 441–457.
- Gonda, D.K. and Radding, C.M. (1983) By searching processively RecA protein pairs DNA molecules that share a limited stretch of homology. *Cell*, **34**, 647–654.
- Nassif, N. and Engels, W. (1993) DNA homology requirements for mitotic gap repair in *Drosophila*. *Proc. Natl Acad. Sci. USA*, **90**, 1262–1266.
- Coveny, A.M., Dray, T. and Gloor, G.B. (2002) The effect of heterologous insertions on gene conversion in mitotically dividing cells in *Drosophila melanogaster*. *Genetics*, **161**, 249–258.
- Waldman, A.S. and Liskay, R.M. (1987) Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc. Natl Acad. Sci. USA*, **84**, 5340–5344.
- Waldman, A.S. and Liskay, R.M. (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.*, **8**, 5350–5357.
- Matsuo, K., Clay, O., Kunzler, P., Georgiev, O., Urbanek, P. and Schaffner, W. (1994) Short introns interrupting the Oct-2 POU domain may prevent recombination between POU family genes without interfering with potential POU domain ‘shuffling’ in evolution. *Biol. Chem. Hoppe-Seyler*, **375**, 675–683.
- Raysiguier, C., Thaler, D.S. and Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature*, **342**, 396–401.
- Datta, A., Adjiri, A., New, L., Crouse, G.F. and Jinks Robertson, S. (1996) Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 1085–1093.
- Bailis, A.M. and Rothstein, R. (1990) A defect in mismatch repair in *Saccharomyces cerevisiae* stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics*, **126**, 535–547.
- Sugawara, N., Goldfarb, T., Studamire, B., Alani, E. and Haber, J.E. (2004) Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. *Proc. Natl Acad. Sci. USA*, **101**, 9315–9320.
- Elliott, B. and Jasin, M. (2001) Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol. Cell. Biol.*, **21**, 2671–2682.
- de Wind, N., Dekker, M., Berns, A., Radman, M. and te Riele, H. (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell*, **82**, 321–330.
- Hanada, K., Ukita, T., Kohno, Y., Saito, K., Kato, J. and Ikeda, H. (1997) RecQ DNA helicase is a suppressor of illegitimate recombination in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **94**, 3860–3865.
- Schmidt, K.H., Wu, J. and Kolodner, R.D. (2006) Control of translocations between highly diverged genes by Sgs1, the *Saccharomyces cerevisiae* homolog of the Bloom’s syndrome protein. *Mol. Cell. Biol.*, **26**, 5406–5420.
- Myung, K., Datta, A., Chen, C. and Kolodner, R.D. (2001) SGS1, the *Saccharomyces cerevisiae* homologue of BLM and WRN, suppresses genome instability and homeologous recombination. *Nat. Genet.*, **27**, 113–116.
- Spell, R.M. and Jinks-Robertson, S. (2004) Examination of the roles of Sgs1 and Srs2 helicases in the enforcement of recombination fidelity in *Saccharomyces cerevisiae*. *Genetics*, **168**, 1855–1865.
- Kusano, K., Berres, M.E. and Engels, W.R. (1999) Evolution of the RECQ family of helicases: a *drosophila* homolog, Dmblm, is similar to the human bloom syndrome gene. *Genetics*, **151**, 1027–1039.
- Kusano, K., Johnson-Schlitz, D.M. and Engels, W.R. (2001) Sterility of *Drosophila* with mutations in the Bloom syndrome gene—complementation by Ku70. *Science*, **291**, 2600–2602.
- Adams, M.D., McVey, M. and Sekelsky, J.J. (2003) *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science*, **299**, 265–267.
- Johnson-Schlitz, D. and Engels, W.R. (2006) Template disruptions and failure of double Holliday junction dissolution during double-strand break repair in *Drosophila* BLM mutants. *Proc. Natl Acad. Sci. USA*, **103**, 16840–16845.
- McVey, M., Larocque, J.R., Adams, M.D. and Sekelsky, J.J. (2004) Formation of deletions during double-strand break repair in *Drosophila* DmBlm mutants occurs after strand invasion. *Proc. Natl Acad. Sci. USA*, **101**, 15694–15699.
- Chaganti, R.S., Schonberg, S. and German, J. (1974) A manifold increase in sister chromatid exchanges in Bloom’s syndrome lymphocytes. *Proc. Natl Acad. Sci. USA*, **71**, 4508–4512.
- Chester, N., Babbe, H., Pinkas, J., Manning, C. and Leder, P. (2006) Mutation of the murine Bloom’s syndrome gene produces global genome destabilization. *Mol. Cell. Biol.*, **26**, 6713–6726.
- Bugreev, D.V., Yu, X., Egelman, E.H. and Mazin, A.V. (2007) Novel pro- and anti-recombination activities of the Bloom’s syndrome helicase. *Genes Dev.*, **21**, 3085–3094.
- Wu, L., Bachrati, C.Z., Ou, J., Xu, C., Yin, J., Chang, M., Wang, W., Li, L., Brown, G.W. and Hickson, I.D. (2006) BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates. *Proc. Natl Acad. Sci. USA*, **103**, 4068–4073.
- Raynard, S., Bussen, W. and Sung, P. (2006) A double Holliday junction dissolvasome comprising BLM, topoisomerase IIIalpha, and BLAP75. *J. Biol. Chem.*, **281**, 13861–13864.
- Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA*, **104**, 3312–3317.
- Schildkraut, E., Miller, C.A. and Nickoloff, J.A. (2005) Gene conversion and deletion frequencies during double-strand break repair in human cells are controlled by the distance between direct repeats. *Nucleic Acids Res.*, **33**, 1574–1580.
- Groth, A.C., Fish, M., Nusse, R. and Calos, M.P. (2004) Construction of Transgenic *Drosophila* by Using the Site-Specific Integrase From Phage {phi}C31. *Genetics*, **166**, 1775–1782.
- Wei, D.S. and Rong, Y.S. (2007) A genetic screen for DNA double-strand break repair mutations in *Drosophila*. *Genetics*, **177**, 63–77.
- McVey, M., Adams, M., Staeva-Vieira, E. and Sekelsky, J.J. (2004) Evidence for multiple cycles of strand invasion during repair of double-strand gaps in *Drosophila*. *Genetics*, **167**, 699–705.
- Flores, C. and Engels, W. (1999) Microsatellite instability in *Drosophila* spellchecker1 (MutS homolog) mutants. *Proc. Natl Acad. Sci. USA*, **96**, 2964–2969.
- Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L. *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*, **318**, 420–426.
- Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
- Emerson, J.J., Cardoso-Moreira, M., Borevitz, J.O. and Long, M. (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, **320**, 1629–1631.
- Preston, C.R., Sved, J.A. and Engels, W.R. (1996) Flanking duplications and deletions associated with P-induced male recombination in *Drosophila*. *Genetics*, **144**, 1623–1638.
- Preston, C.R., Engels, W. and Flores, C. (2002) Efficient repair of DNA breaks in *Drosophila*: evidence for single-strand annealing and competition with other repair pathways. *Genetics*, **161**, 711–720.