Mechanisms underlying mutational outcomes of DNA double-strand break repair

Kamp, J.A.

Citation

CHAPTER 3

Helicase Q promotes homology-driven DNA double-strand break repair and prevents tandem duplications

J. A. Kamp, B. B. L. G. Lemmens, R. J. Romeijn, S. C. Changoer, R. van Schendel, M. Tijsterman

ABSTRACT

DNA double-strand breaks are a major threat to cellular survival and genetic integrity. In addition to high fidelity repair, three intrinsically mutagenic DNA break repair routes have been described, *i.e.* single strand annealing (SSA), polymerase theta-mediated end-joining (TMEJ) and residual ill-defined microhomology-mediated end-joining (MMEJ) activity. Here, we identify *C. elegans* Helicase Q (HELQ-1) as being essential for MMEJ as well as for SSA. We also find HELQ-1 to be crucial for the synthesis-dependent strand annealing (SDSA) mode of homologous recombination (HR). Loss of HELQ-1 leads to increased genome instability: patchwork insertions arise at deletion junctions due to abortive rounds of polymerase theta activity, and tandem duplications spontaneously accumulate in genomes of *helq-1* mutant animals as a result of TMEJ of abrogated HR intermediates. Our work thus implicates HELQ activity for all DSB repair modes guided by complementary base pairs and provides mechanistic insight into mutational signatures common in HR-defective cancers.

INTRODUCTION

DNA is subject to damage, which can lead to alteration of its nucleotide sequence with potentially detrimental outcomes for an organism. One of the most impactful damages is breakage of DNA, resulting in separation of chromosome parts. Traditionally, two mechanisms to repair double-strand breaks (DSBs) were described: non-homologous end-joining (NHEJ) and homologous recombination (HR)(Kanaar, Hoeijmakers et al. 1998), which have been intensely studied over the last decades. NHEJ is a conceptually simple mechanism: it directly joins broken chromosome ends together requiring the ligase LIG4 and the DNA-end binding scaffold Ku(Zhao, Rothenberg et al. 2020). While NHEJ can be error prone, because certain break ends require polishing to allow ligation, it is likely a *grosso modo* accurate DSB repair pathway(Bétermier, Bertrand et al. 2014). HR performs a more complex, multistep reaction that starts with resecting the broken DNA ends in a 5' to 3' direction, thus generating a crucial HR intermediate: a 3' protruding tail(San Filippo, Sung et al. 2008). This single-stranded DNA (ssDNA) end, aided by the RAD51 recombinase, can invade a homologous sequence which will serve as a donor for templated DNA synthesis at the 3' tip. Because the homologous sequence usually is the unbroken sister chromatid, the product of DNA extension is a sequence that is identical to the resected part of the other end of the break, providing the basis for error-free repair: upon dissociation of the extended end from its template, annealing of perfectly complimentary stretches of DNA can occur. Depending on the length of the extension and/or the degree of resection, additional gap filling or trimming needs to occur prior to ligation, which completes a mode of HR repair that is called synthesis dependent strand annealing (SDSA)(Pâques and Haber 1999). SDSA is thought to account for most HR repair in mitotic cells(Scully, Panday et al. 2019), whereas in meiotic cells DSB repair can also progress via the establishment of double-Holliday junctions that subsequently require additional biochemistry for resolution(Schwartz and Heyer 2011).

Over the years, it has become increasingly clear that NHEJ and HR cannot account for all DSB repair and other modes of repair that prevent chromosomal fragmentation exist(Ceccaldi, Rondinelli et al. 2016). At least two additional pathways have been defined: single strand annealing (SSA) and alternative end-joining (altEJ)(Bhargava, Onyango et al. 2016, Chang, Pannunzio et al. 2017). SSA is able to repair DSBs that are flanked by direct repeats (of at least about 50 bps) (Sugawara and Haber 1992, Bhargava, Onyango et al. 2016, Kelso, Lopezcolorado et al. 2019). In this mode of repair, end resection exposes the homologous sequences that can directly anneal and seal the break without having to pick-up complementary sequence from the sister chromatid, as is the case in SDSA. In a sense, SSA thus resembles SDSA, yet without

the synthesis step. The outcome of SSA is, however, very different from SDSA as the sequence in between the repeats as well as one repeat copy will be lost and SSA is thus intrinsically mutagenic. Apart from mechanistic communalities in SDSA and SSA (e.g. the need for end resection and strand annealing), there are also differences: because direct repeats that guide SSA will rarely be located precisely at the break ends, non-complementary 3' protruding DNA flaps that inevitably arise upon annealing need to be removed in SSA, a biochemical activity attributed to the structure-specific endonuclease ERCC1/XPF(Dehé and Gaillard 2017). Because of this necessity, ERCC1/XPF dependency is one frequently used criteria in defining SSA(Pontier and Tijsterman 2009), as well as a dependence on the ssDNA binding protein RAD52, which is able to stimulate base-pairing of complementary sequences(Symington 2002, Rothenberg, Grimme et al. 2008).

Until recently, altEJ has been the most ill-defined DSB repair pathway. Early work revealed an end-joining activity independent on canonical NHEJ factors(Boulton and Jackson 1996, Kabotyanski, Gomelsky et al. 1998, Ma, Kim et al. 2003). Because the cognate repair products were enriched for DNA junctions that appeared to have used small stretches of sequence homology that surrounded the break, the term microhomology-mediated end-joining (MMEJ) has also been used(McVey and Lee 2008). Recently, it was found that a large proportion of altEJ, yet not all, in many systems involves the action of polymerase theta(Chan, Yu et al. 2010, Roerink, van Schendel et al. 2014, Yousefzadeh, Wyatt et al. 2014, Mateos-Gomez, Gong et al. 2015, van Kregten, de Pater et al. 2016). The term TMEJ for polymerase Theta-Mediated End-Joining was introduced(Roerink, van Schendel et al. 2014) to discriminate between DSB repair that crucially depends on polymerase theta versus altEJ modes that does not, such as MMEJ in yeast - yeast does not encode polymerase theta-, which was shown to be reliant on usage of somewhat longer microhomologies (>5bp)(McVey and Lee 2008). However, also in species with a polymerase theta ortholog, additional altEJ exist: mammalian cells clearly have altEJ activity that involves longer microhomologies, which is independent on POLQ activity(Kelso, Lopezcolorado et al. 2019). At present, the underlying mechanism and the genetic requirements of this type of repair is unknown.

Here, we describe the identification of helicase Q (HELQ/HEL308/HELQ-1), a protein previously found to affect DNA repair in various biological systems(McCaffrey, St Johnston et al. 2006, Wei and Rong 2007, Muzzini, Plevani et al. 2008, Adelman, Lolo et al. 2013, Takata, Reh et al. 2013), to be essential for polymerase theta-independent MMEJ. We further show that *C. elegans* HELQ-1 is essential for other DSB repair pathways that require annealing of homologous nucleotide stretches, such as SSA and SDSA. We provide a unified model for

HELQ action that explains the redundancies observed in altEJ as well as how the inability of completing HR results in spontaneous accumulation of tandem duplications.

RESULTS

Polymerase theta-independent end-joining of G4-induced breaks

In previous work we have established that G4 motifs in the *C. elegans* genome provide a source for DNA breaks that depend on alternative end-joining for their repair(Koole, van Schendel et al. 2014, Lemmens, van Schendel et al. 2015). These motifs, which contain sequential stretches of guanines, can assemble into thermodynamically very stable secondary structures that have the potential to block ongoing DNA replication(Castillo Bosch, Segura-Bayona et al. 2014). In the absence of the G-quadruplex resolving helicase DOG-1/FANCJ, ssDNA gaps form immediately downstream of the replication fork impediment, which persist and are converted to DSBs in the next cell cycle(Lemmens, van Schendel et al. 2015) (Figure 1A). These replication-associated DSBs are predominantly, if not exclusively, repaired by polymerase theta-mediated end joining (TMEJ), resulting in deletions with a size reflective of the ssDNA gap across the replication impediment(Koole, van Schendel et al. 2014, Lemmens, van Schendel et al. 2015). Endogenous G4 motifs in *dog-1* deficient *C. elegans* provide a unique model substrate to study TMEJ *in vivo* because their mutagenic consequences occur at well-defined genomic positions, with a frequency that can be measured with a variety of molecular techniques, including PCR, transgenic and endogenous reporters, and whole genome sequencing(Cheung, Schertzer et al. 2002, Kruisselbrink, Guryev et al. 2008, Koole, van Schendel et al. 2014).

Interestingly, while the vast majority of G4-induced DNA breaks are repaired via TMEJ, we found a few cases that are not. In fact, in all genetic contexts in which TMEJ dominates the repair spectrum, such as for structural variations accumulating in *C. elegans* BRCA1 mutant animals(Kamp, van Schendel et al. 2020), we observed a small fraction of repair outcomes that are independent of polymerase theta action. As an example, Figure 1B presents all genome alterations we identified in *dog-1* and *dog-1 polq-1* mutant animals that were clonally grown for in total over 120 generations after which their genomes were sequenced. While TMEJ products are absent in *polq-1* mutant animals, we found a few larger deletions, which had pronounced microhomology at the deletion junction (Figure 1C). *C. elegans* TMEJ requires only a minute degree of microhomology - frequently not more than a single base pair(Roerink, van Schendel et al. 2014) and it is suggested that the polymerase action of POLQ-1 extends these microhomologies to promote DSB repair. We hypothesized that more elaborate sequence

homology in the flanks of a DNA break may bypass the requirement for polymerase theta to create a thermodynamically stable repair intermediate. For purpose of clarity, in the remainder of the text we will refer to this polymerase-theta independent, extended micro-homology-mediated, alternative end joining mode of repair as eMMEJ. To further investigate the parameters and genetic requirements of eMMEJ, we searched the *C. elegans* genome for loci in which endogenous G4 motifs are flanked by small stretches of sequence homology and selected four loci for in-depth analysis: Qua1277, Qua213 and Qua915, which have an increasing degree of sequence similarity surrounding the G4 motif of about 20-30 base pairs (bps) in size (Figure 1D), whereas Qua317 is devoid of substantially repeated sequences up and downstream of the motif, and hence serves as a control. Using a PCR-based approach, we found all four G4 motifs to induce deletions of the TMEJ type in *dog-1*-deficient animals (Figure 1E): one deletion junction being located immediately upstream of the G4, the other at varying positions 50 to 300 bps downstream (Figure 1F). However, while deletion formation at Qua317, which lacks extended micro-homologous sequences (eMH), proved to be almost completely dependent on functional POLQ, deletion formation at Qua1277, Qua213 and Qua915 was less affected: we detected repair products at all of these sites in *polq-1 dog-1* mutant animals (Figure 1E). The dependency of break repair on POLQ was found to be inversely correlated with the degree of homology present in the flanks of the motifs (Figure 1G). Upon close inspection, we noticed that individual animals (in separately grown populations) often showed deletions of identical size, suggesting a preferred repair outcome specific for each locus (Figure 1E). Indeed, sequence analysis revealed all TMEJ-independent deletion events used eMH flanking the G4 sites (Figure 2A). Interestingly, the frequency of deletion formation was comparable in POLQ-1 proficient and deficient genetic backgrounds when substantial eMH was present (Figure 1G). These data support the notion that came forward from the whole genome sequence analysis, *i.e.* that G4-induced DSBs can be repaired independent of POLQ, but only when the sequence context supports the usage of homologous sequences at either end of the break.
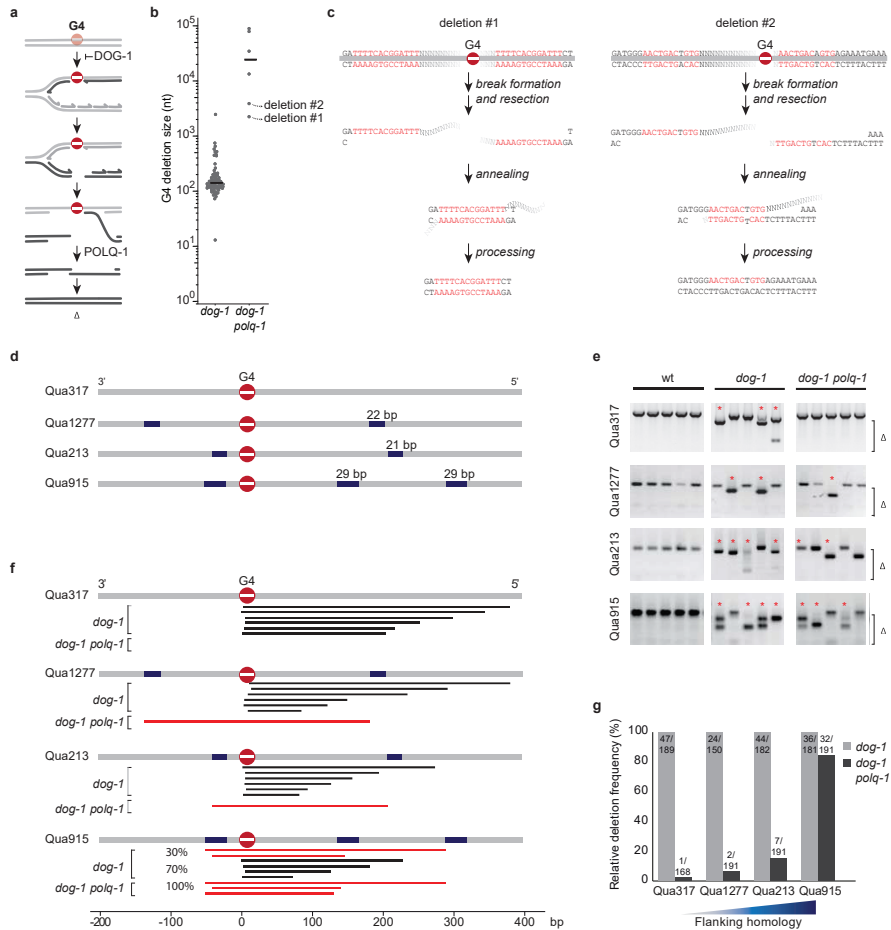
Figure 1. Polymerase theta-independent repair of G-quadruplex-induced DSBs is driven by flanking sequence homology a. Model explaining double-strand break (DSB) formation and DSB repair at replication-blocking G-quadruplex structures(Lemmens, van Schendel et al. 2015). b. Size distribution of deletions that accumulate in the genomes of *dog-1* and *dog-1 polq-1* mutant animals (*dog-1* data from (Koole, van Schendel et al. 2014)). c. Potential mechanism explaining the formation of the deletions marked in b. d. Schematic diagram of four endogenous G4 loci with different degrees of flanking homology. The most prominent repeated sequences are indicated in blue. e. Representative images of PCR-based analysis of the indicated G4-containing loci. Each lane contains the genomic DNA of three adult animals. Asterisks indicate stochastic deletions, which manifest as shorter than wild-type products and Δ indicates the size-range of the PCR-amplified deletion products. Deletion sequences are provided as a Source Data file. f. Graphic illustration of G4 deletions profiles at four endogenous G4 loci. For each locus typical G4 deletions in *dog-1* and *dog-1 polq-1* animals are depicted. Black bars represent homology-independent deletions; red bars represent homology-dependent events. g. Histogram depicting relative deletion frequencies at the indicated G4 loci as determined by the presence of deletion bands in the PCR-based assay on single worms of *dog-1* (grey bars) and *dog-1 polq-1* mutant animals (black bars). Depicted frequencies are relative to the deletion frequency in *dog-1* single mutants to allow comparison of loci expressing different stochastic G4 deletion rates. The number of analysed animals is depicted on top of the bars. Gel counts are provided as a Source Data file.
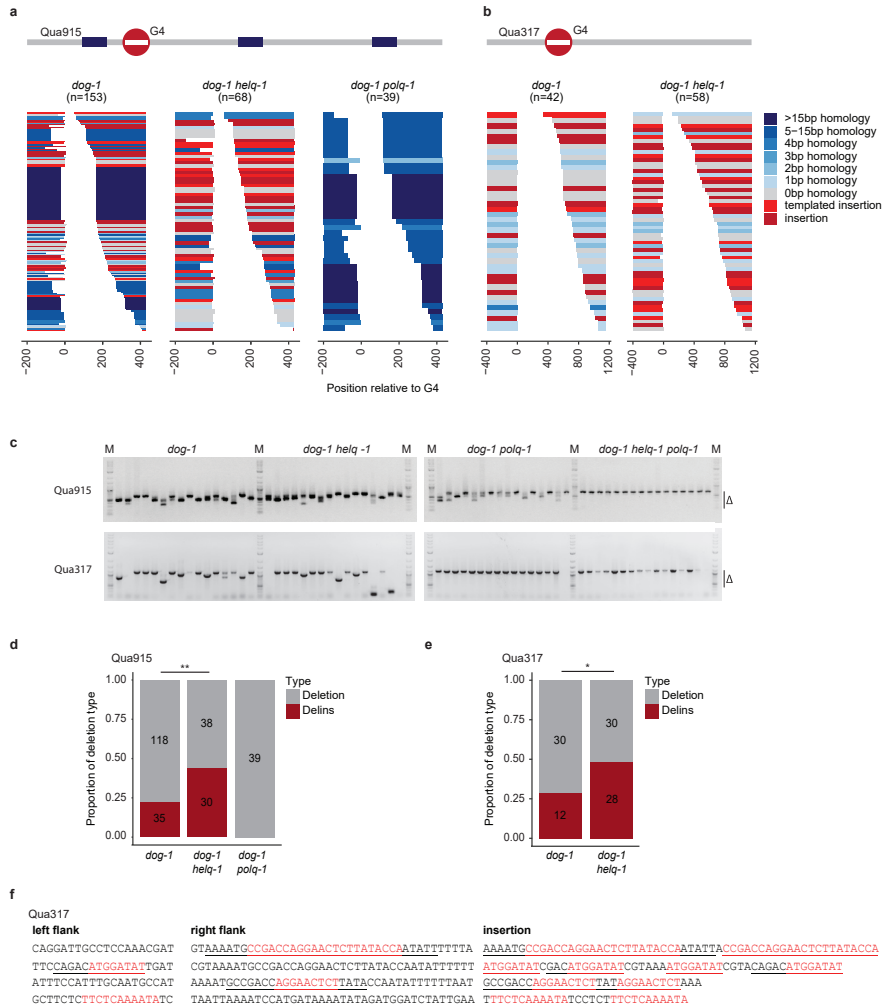
Helicase Q is essential for eMMEJ of G4-induced breaks

A logical explanation for stretches of sequence homology to bypass POLQ is to propose a specific role for POLQ in TMEJ: generating a sufficiently-sized stretch of base-pairing (a primer) starting with only a few bases of complementarity at either end of a break. In such a scenario, eMH already present in the immediate vicinity of a DSB makes such activity redundant (Supplementary Figure 1). However, this explanation does not take into account that POLQ, apart from a polymerase domain, contains a helicase domain that is also required for TMEJ, and which has been suggested to make DSBs accessible for the DNA synthesis step(Chan, Yu et al. 2010, Mateos-Gomez, Kent et al. 2017). We envisaged that another protein may substitute for this helicase function in eMMEJ, and focused our attention on a logical candidate: Helicase Q (HELQ-1), as this protein's helicase domain is highly similar to the helicase domain within polymerase theta(Marini and Wood 2002).

To address this hypothesis, we crossed a previously characterized mutant allele of *helq-1* (*helq-1(tm2134)* (Muzzini, Plevani et al. 2008)) into *dog-1* and *dog-1 polq-1* and analyzed G4-induced mutagenesis at the eMH containing locus Qua915 and the eMH lacking locus Qua317. We obtained mutation spectra in which we categorized the obtained deletions by the degree of MH at the junction, and whether or not they contained inserted nucleotides at the junction. In HELQ proficient *dog-1* animals the Qua915 spectrum is dominated by deletion junctions with eMH (Figure 2A), whereas a typical TMEJ spectrum is observed for Qua317 (Figure 2B): minute microhomology at deletion junctions and insertions that, when of sufficient size, can be reliably mapped to flanking sequences.

In support for a causal role of HELQ-1 in eMMEJ, we observed a complete absence of eMMEJ-driven cases at Qua915 in *helq-1 dog-1* animals (Figure 2A); the spectrum now resembles that of Qua317, arguing that these repair products completely rely on TMEJ. Indeed, no deletions were found at Qua915 in *helq-1 polq-1 dog-1* animals (Figure 2C). Conversely, the Qua915 spectrum in *polq-1 dog-1* mutant animals is entirely composed of deletions of the eMH type, arguing for a clearly distinct division of labor in altEJ. Together, these data demonstrate that DNA breaks originating from replication-obstructing G-quadruplexes are repaired via TMEJ or via HELQ-1 dependent eMMEJ.

3

Figure 2. HELQ is essential for eMMEJ of G4-induced DSBs and a facilitator of TMEJ

a. Spectra of deletions occurring at the Qua915 locus, which is schematically depicted above, for the indicated genotype. Single deletion events are piled and sorted from top to bottom by deletion end-point relative to the G4 motif set at 0. Deletion events are colour-coded according to the following mutational classification: grey for simple deletions without apparent MH at the junction, blue for simple deletions with MH at the junction, wherein the saturation level of the blue colour increases with increasing amount of homology identified. Deletions containing insertions are in red: bright red for insertions that can be reliably mapped to the flank of the deletion, dark red for insertions of which the origin could not be determined with certainty. Deletion sequences are provided as a Source Data file. b. Spectra of deletions occurring at the Qua317 locus, which is schematically depicted above, for the indicated genotype. Colour coding is identical to a). Deletion sequences are provided as a Source Data file. c. Representative images of PCR-based analysis of the indicated G4-containing loci for the indicated genotypes. Each well contains DNA of 10 animals. 2-Log DNA ladders (indicated by 'M') are used as markers for size reference. Uncropped gel pictures are provided as a Source Data file. d. Proportion of deletion types at Qua915. Difference in ratio between

deletions without and with insertion were tested using Chi square test (** P<0.01). e. Proportion of deletion types at Qua317⁻. Difference in ratio between deletions without and with insertion were tested using Chi square test (* P<0.05). f. Examples of deletions with a complex configuration of templated insertions. For each case, the insertion as well as the left and right flank of the corresponding deletion is depicted. Inserted nucleotide stretches that are identical to the flank of the deletion junction are underscored in both the insertion and the cognate flank. Nucleotide stretches present multiple times within one insertion are depicted in red.
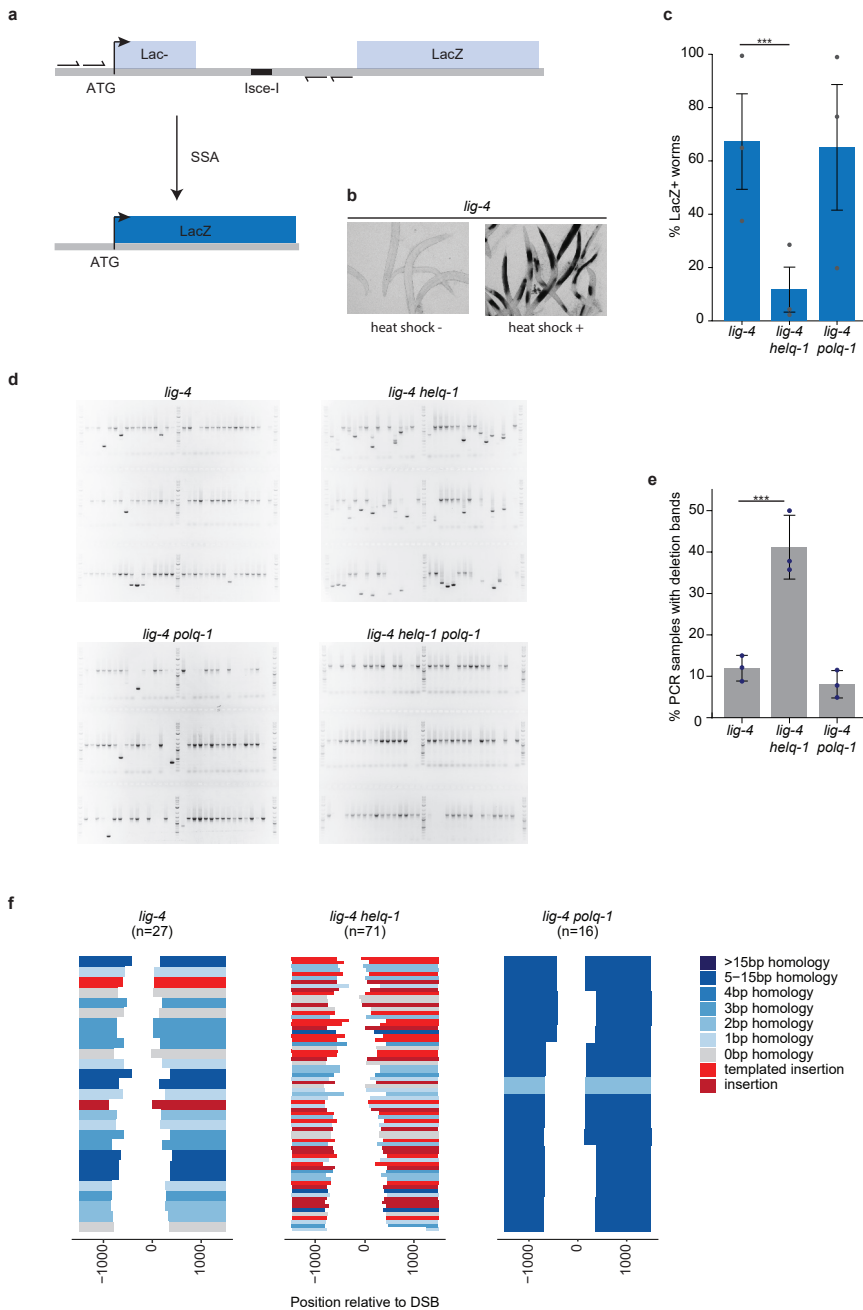
Helicase Q is a facilitator of TMEJ activity

Close inspection of the Qua 915 deletion spectrum revealed a striking peculiarity: we found that the ratio of (templated) insertions over simple deletions is increased in *helq-1* mutant animals when deletions are compared to those derived from HELQ-1 proficient animals (44% versus 23%, respectively, Figure 2D). One potential explanation is that breaks that are processed to use eMMEJ in HELQ-1 proficient genetic backgrounds have a higher tendency to result in deletions that also contain insertions when HELQ-1 is absent. An alternative explanation is that HELQ-1 is facilitating efficient, non-interrupted TMEJ, and in its absence one or more successive round of abortive TMEJ leads to increased levels of templated insertions (see Supplementary Figure 1 for a schematic illustration). To discriminate between these explanations we compared the spectrum of deletions at Qua317, a locus not flanked by obvious eMH, and which is entirely dependent on POLQ-1 for its DSB repair (Figure 2C). The frequency of insertions at this locus in HELQ-1 proficient animals is 28,6%, similar to the 28% that was previously described for *C. elegans* G4 sites genome-wide(Koole, van Schendel et al. 2014). However, this fraction is considerably higher (48%) in animals that lack HELQ-1 (Figure 2E). Moreover, also the molecular configuration of the templated insertions derived from *helq-1* mutant animals are different: whereas in other genotypes TMEJ products typically contain templated insertions with only one DNA stretch that maps to break-flanking sequences, we found 14% of TMEJ cases in *helq-1* mutants to have a "patchwork" appearance, containing multiple stretches of identical sequences (Figure 2F). We consider these cases products of iterative rounds of primer-template switching during TMEJ, and their increased appearance in *helq-1* mutant animals may hint towards an obstacle to TMEJ progression that can be relieved by HELQ-1 activity. HELQ-1 is thus not required for TMEJ but does affect the fidelity of this repair route.

HELQ-1 is essential for *C. elegans* single strand annealing

The eMMEJ defect in *helq*-1 mutant animals as well as the more subtle TMEJ phenotype fits well with a possible role for HELQ in making ssDNA ends available for repair-stimulating base-pairing interaction, *e.g.* by stripping potentially interfering ssDNA binding proteins.

A similar biochemical activity is proposed to facilitate yet another mode of DSB repair, *i.e.* single-strand annealing (SSA), which has been described as a pathway that repairs DSBs by making use of large stretches of near identical sequences (>50bp) up- and downstream of the break site. To assess whether HELQ-1 acts in SSA we used a reporter assay that we previously developed and validated for *C. elegans*. In this transgenic reporter, schematically illustrated in Figure 3A, a recognition site of the I-SceI endonuclease is flanked by two large (~250 bps) homologous sequences, which are part of a LacZ gene. Using these large stretches of homology to repair the I-SceI-induced DSB will lead to restoration of an otherwise interrupted LacZ open reading frame (Figure 3A). LacZ expression, quantified by staining animals for B-galactosidase activity, thus serves as a proxy for SSA activity. To optimize the utility of this reporter system, we performed all experiments in a NHEJ (*lig-4*) deficient genetic background, in which SSA is more prominently detectable (but surprisingly not reliant on the structure-specific endonuclease XPF (Supplementary Figure 2A)): ~70% of *lig*-4 deficient reporter animals display patches of blue cells (Figure 3B, 3C). We next assayed *helq-1 lig-4* animals and found a profound decreased number of LacZ-positive animals, indicative of impaired SSA (Figure 3C). It should be noted that a quantification based on the number of animals that display LacZ positive cells severely underestimates the SSA defect in *helq-1 lig-4* animals: while *lig-4* control animals display large segments of LacZ positive cells within individual animals, *helq-1 lig-4* animals typically had only one or two blue cells (Supplementary Figure 3). Identical staining patterns and intensities were found for *lig-4* and *lig-4 polq-1* animals (Figure 3C), demonstrating that POLQ-1 (in contrast to HELQ-1) does not act in SSA, nor suppresses it. From these data we conclude that HELQ-1 is critical for SSA repair of targeted DNA breaks, in addition to being essential for eMMEJ of replication-associated DNA breaks.

Figure 3. SSA deficiency in *C. elegans* lacking HELQ-1

a. Schematic representation of the single-strand annealing (SSA) reporter(Pontier and Tijsterman 2009). A DSB can be introduced at the I-SceI recognition site by the I-SceI endonuclease. DSB repair by

annealing of ~250 bp region of sequence identity placed up and downstream of the I-SceI recognition site leads to a functioning LacZ open reading frame. Small arrows depict primers used in the PCR analysis. b. Representative pictures of *lig-4* mutant animals carrying the reporter transgenes that were heat shocked or mock-treated to induce I-SceI expression, followed by staining for B-galactosidase expression. c. Histogram depicting the percentage of LacZ positive worms for the indicated genotype. Experiments were performed in triplicate (*** P<0.01; Chi square test). Each dot represents the average percentage of each replicate. Error bars represent SEM. Staining quantifications are provided as a Source Data file. d. Representative images of PCR-based analysis of the reporter locus at the I-SceI site for the indicated genotypes. Each well contains DNA of 1 animal. Wild-type bands are expected because many cells in the animal are insensitive to heat-shock driven I-SceI expression. 2-Log DNA ladders are used as markers for size reference. e. Histogram depicting the percentage of PCR samples containing a deletion. Each dot represents the percentage of deletions found in 96 samples. Experiments were performed in triplicate (*** P<0.01; Chi square test). Error bars represent SEM. Deletion counts are provided as a Source Data file. f. Spectra of deletions occurring at the I-SceI site of the SSA reporter for the indicated genotype. Single deletion events are piled and sorted from top to bottom by deletion end-point relative to the G4 motif set at 0. Deletion events are colour-coded according to the following mutational classification: grey for simple deletions without apparent MH at the junction, blue for simple deletions with MH at the junction, wherein the saturation level of the blue colour increases with increasing amount of homology identified. Deletions containing insertions are in red: bright red for insertions that can be reliably mapped to the flank of the deletion, dark red for insertions of which the origin could not be determined with certainty. Deletion sequences are provided as a Source Data file.

SSA is dominant over TMEJ

We next considered the fate of DNA breaks that cannot be repaired by SSA in *helq-1* mutant animals: are these now a substrate for TMEJ? To address this question, we PCR-amplified the reporter locus two days after DSB induction at the I-SceI recognition site in order to inspect the DNA for repair outcomes other than SSA products. While some deletion products that upon sequencing displayed all hallmarks of TMEJ were found in *lig-4* control animals, their incidence profoundly increased in *lig-4 helq-1* mutant animals (Figure 3D, 3E). We confirmed that these deletions were brought about by TMEJ activity as no smaller than wild-type bands were seen in *lig-4 helq-1 polq-1* animals (Figure 3D). We conclude from this outcome that in cases where profound homology is present, SSA is the preferred pathway for DSB repair and that TMEJ will act as a back-up if SSA is compromised. The notion that SSA substrates can be channeled into TMEJ suggests that POLQ can act on DSBs that are resected and likely covered by ssDNA binding proteins, as suggested previously(Mateos-Gomez, Kent et al. 2017).

In addition, and in agreement with our observations at G-quadruplex-induced DSBs, also at endonuclease-induced DSBs we find evidence for i) HELQ being essential for eMMEJ, and ii) having a facilitating role in TMEJ. On the first note, we observed residual deletion mutagenesis in *lig-4 polq-1* double mutant animals but not in *lig-4 polq-1 helq-1* triple mutant animals (Figure 3D); and also here we find HELQ-dependent repair products to be characterized by eMH at the junctions (Figure 3F). On the second note, we once more observed a markedly higher percentage of (templated) insertions in the HELQ deficient genetic background: nearly

half of all deletions (34 out of 71) contain an insertion (Figure 3F, Supplementary Figure 3). Of these, 19 can be classified as templated insertions, including three which have a complex, repetitive configurations (Supplementary Figure 3) similar to the patchwork insertions observed at G4 sites (Figure 2e).

Helicase Q is essential for synthesis-dependent strand-annealing

Having identified an essential role for HELQ-1 in two mutagenic DSB repair pathways, *i.e.* eMMEJ and SSA, which both involve annealing of complementary bases, we next wished to investigate whether HELQ-1 acts in error-free repair. In particular, it seems logical to assay the sub-pathway of homologous recombination that also requires an annealing step: SDSA. In a late step of SDSA, two single-stranded overhangs, that are complementary in sequence resulting from extending one end of the break using the sister chromatid as a template, anneal to finalize repair without the loss of DNA sequence (see Figure 4A for a visual representation). To test a potential involvement of HELQ-1 in SDSA, we made use of a previously described DSB-repair reporter that can read out HR in somatic cells of individual animals(Johnson, Lemmens et al. 2013). This reporter, schematically illustrated in Figure 4B, consists of a GFP expression cassette, but having the GFP open reading frame (ORF) disrupted by an I-SceI recognition site. Downstream of this cassette, a DNA segment has been engineered that has the disrupted part of the GFP ORF as well as up and downstream sequences, but by itself does not encode for a functional GFP protein. This downstream DNA segment can be used as a template for repair of a DSB introduced by the I-SceI endonuclease within the corrupted GFP, eventually leading to restoration of the GFP ORF. Reporter animals thus express GFP in cells in which HR repair took place(Johnson, Lemmens et al. 2013, Kamp, van Schendel et al. 2020). We previously established that this reporter monitors BRCA1/BRC-1 dependent SDSA(Kamp, van Schendel et al. 2020) (double holiday junction-mediated HR in worms is independent of BRC-1(Adamo, Montemauri et al. 2008, Garcia-Muse, Galindo-Diaz et al. 2019)). Figure 4C displays a severe reduction in GFP ORF correction that results from HELQ-1 loss. Similar to *brc-1* deficiency, loss of *helq-1* causes a significant reduction in GFP open reading frame correction, indicative of a severe defect in SDSA. Despite a comparable reduction in SDSA for *brc-1* and *helq-1* mutants, epistatic analysis argues also for non-overlapping roles of BRCA1/BARD1 and HELQ in DSB repair: GFP correction is further reduced in animals double deficient for HELQ-1 and BARD1/BRD-1 (Supplementary Figure 2B), the obligatory binding partner of BRCA1/BRC-1(Boulton, Martin et al. 2004) . In support for this notion, we found that *helq-1 brd-1* double mutant animals are also more hypersensitive to DSB-inducing

radiation than either of the single mutants, which by themselves are more sensitive to radiation than wild-type animals (Supplementary Figure 2C).

To further validate HELQ-1 action in DSB repair that includes a DNA synthesis and annealing steps (as is SDSA),we used an independent method(Arribere, Bell et al. 2014) in which templated repair of CRISPR/Cas9-induced breaks can be assessed by measuring gene-correction using single stranded oligonucleotides (ssODNs). In this approach, schematically illustrated in Figure 4D, animals are injected with plasmids that drive Cas9 expression and an sgRNA that targets the endogenous *dpy-10* locus, together with a short (101 nucleotides) ssODN containing a specific *dpy-10* mutation. This base-pair mutation when introduced into the genome leads to a morphological change that is distinguishable from *dpy-10* loss of function-inducing sequence resulting from mutagenic end-joining(Paix, Schmidt et al. 2016, Paix, Folkmann et al. 2017).

Per injected genotype 24-40 animals were isolated that based on their morphological phenotype had an altered *dpy-10* locus. Subsequent sequencing analysis revealed that in repair proficient controls ~84% of the animals were altered via ssODN-mediated gene correction. In contrast, in *helq-1* mutant animals only 17% of the repair proved ssODN-mediated (Figure 4E), hence strengthening a fundamental role for HELQ-1 in promoting SDSA. In line with the notion that TMEJ is responsible for mutagenic end-joining in the *C. elegans* germline(van Schendel, Roerink et al. 2015), we exclusively found ssODN-mediated events in *polq-1* mutant animals. No altered alleles were obtained in *helq-1 polq-1* double mutant animals, thus excluding other mutagenic repair pathways acting on these CRISPR-induced DSBs.

Helicase Q prevents tandem duplications

We next investigated whether HELQ-1 loss affects genomic integrity also during non-challenged growth, given that animal growth and development appears unaffected in *helq-1* mutants(Ward, Muzzini et al. 2010). Previously, we found that *brc-1* and *brd-1* mutant animals, which despite having impaired HR also grow and develop similar to wild-type animals, accumulate structural variants in their genomes(Kamp, van Schendel et al. 2020). This increased level of mutagenesis argues that SDSA plays a role in the maintenance of genetic integrity, likely through error-free repair of DSBs that may be of meiotic origin or spontaneously arise during proliferation. To monitor mutagenesis at a genome-wide scale and in an unbiased fashion we established independent mutation accumulation lines: we separately grew populations of *helq-1* deficient animals for 50 generations after which we performed whole-genome sequencing (WGS). For this analysis we used a previously described null allele of *helq-1* as well as a newly-
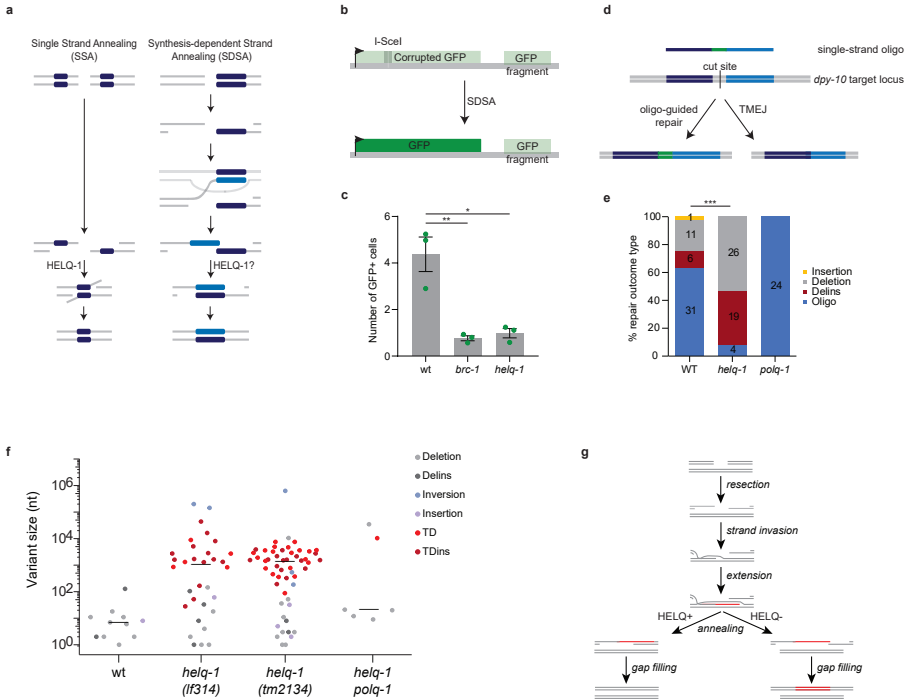
Figure 4 HELQ-1 acts in synthesis-dependent strand annealing and suppresses tandem duplications

a. Models for SSA and SDSA repair of DSBs. Homologous stretches of DNA are depicted in dark blue. b. Schematic representation of the SDSA reporter(Johnson, Lemmens et al. 2013). A DSB can be introduced by the I-SceI endonuclease at its recognition site positioned within a corrupted GFP expression cassette. SDSA repair using a GFP segment cloned downstream of the expression cassette will restore the GFP ORF leading to GFP expressing animals. c. Histograms depicting the average number of GFP expressing cells per worm for the indicated genotype. Experiments are performed in triplicate (* P<0.05, **P<0.01; two-sided t-tests). P-value WT vs. *brc-1*: 0.0085. P-value WT vs. *helq-1*: 0.0115. Each dot represents the average number of each replicate. Error bars represent SEM. Quantifications are provided as a Source Data file. d. Schematic representation of the CRISPR/Cas-9-induced gene correction assay. CRISPR-induced DSB at the *dpy-10* locus can be repaired via end joining or using a injected single-stranded oligo as a template. The sequence of the oligo differs from the target locus at the centre (in green). Mutation sequences are provided as a Source Data file. e. Histogram depicting the different categories of repair outcomes. Insertions are depicted in yellow, simple deletions in grey, deletions containing insertion in red, outcomes of repair using the oligo are depicted in blue. Difference in ratio between ssODN-guided repair and indels were tested using Chi square test (*** P<0.001). f. Size distribution of structural variations that accumulate in the genomes of the indicated genotype (wildtype data derived from(van Schendel, Roerink et al. 2015)). Mutation sequences are provided as a Source Data file. g. Mechanistic model for the aetiology of tandem duplications (TDs) in *helq-1* mutant genetic backgrounds.

made knockout generated by CRISPR. Strikingly, for both mutant strains we found mainly one specific type of genome alterations accumulating, i.e. Tandem Duplications (TDs) (Figure 4F), corroborating recent data describing mutational signatures in *C. elegans* DNA repair mutants(Volkova, Meier et al. 2020, Meier, Volkova et al. 2021). These TDs, which are rarely found in wild type nematodes(Kamp, van Schendel et al. 2020, Volkova, Meier et al. 2020),

range in size from ~ 100bp to 10 kb, with a median of ~ 1kb. Importantly, the DNA junctions that arise between the two direct repeats show all hallmarks of TMEJ: overrepresentation of micro-homology and the occasional presence of templated insertions (Supplementary Figure 4). This outcome fits a model in which TMEJ acts on failed SDSA intermediates: in the absence of HELQ-1, extended DNA ends (using a sister homolog as a template) will fail to anneal to the complementary sequence of the other break-end. Direct end joining of both ends now results in a duplication of the sequence that was copied during the synthesis step of SDSA (Figure 4G). Indeed, TD formation in *helq-1* deficient conditions proved to be dependent on functional POLQ-1 (Figure 4F). We find the combined loss of *helq-1* and *polq-1* to only mildly affect population growth, which may not be surprising given the low rate in which TDs accumulate in *helq-1* mutants under non-stressed conditions (*i.e.* 0.1 - 0.2 per animal generation). Increasing the genomic damage load by exposing animals to ionizing radiation indeed reveals redundancy for these proteins in the cellular response to DNA damage (Supplementary Figure 2D).

DISCUSSION

Here, by using validated transgenic DSB-reporter animals and mutational footprint analysis at G-quadruplex and CRISPR-induced DSBs, we show in *C. elegans* that HELQ-1 is necessary for all DSB-repair mechanisms that are guided by annealing of extensive stretches of complementary bases at break ends.

HELQ is an ATP-dependent 3'-5' DNA helicase of the HEL308 family that is present in archaea and eukaryotes(Marini and Wood 2002, Adelman and Boulton 2010). Loss of HELQ in mouse or human cells causes germ cell attrition and ICL sensitivity(Adelman, Lolo et al. 2013, Luebben, Kawabata et al. 2013, Takata, Reh et al. 2013). Because of persistent RAD51 foci in *C. elegans*(Ward, Muzzini et al. 2010) and mammalian cells(Adelman, Lolo et al. 2013) lacking HELQ, a post-synaptic role in HR was proposed. More recently it was found that HELQ acts redundantly with a pathway involving the proteins HROB, MCM8, and MCM9 – cells lacking both HROB and HELQ had severely impaired HR - leading to the suggestion that HELQ could act in parallel to these proteins in HR to promote bubble migration after D-loop establishment(Hustedt, Saito et al. 2019). However, our finding that HELQ also acts in SSA and MMEJ -pathways that do not entail processing of D-loops- points to a role in DSB repair that is shared between these pathways.

In this respect, it is noteworthy that the annealing step in SDSA is conceptually indistinguishable from that in SSA: perfect alignment of large stretches of complementary nucleotides that are present at the ends of a break in a ssDNA configuration. Early work already pointed

towards a role for HELQ in SSA in *Drosophila*(Johnson-Schlitz, Flores et al. 2007), although different reporter systems have led to seemingly incompatible data(Wei and Rong 2007), likely because of different experimental features such as the distance between, and the length of, the stretch of homologous sequence flanking a DSB. SSA is most well-studied in yeast which, however, does not encode HelQ, and where *RAD52* and its paralogue *RAD59* have shown to be required(Krogh and Symington 2004). While RAD52 also promotes SSA in mammalian cells, there is strong experimental support for the existence of additional, partly redundant annealing factors(Mendez-Dorantes, Bhargava et al. 2018, Kelso, Lopezcolorado et al. 2019). For instance, RAD52 inactivation in U2OS cells reduced SSA at CRISPR-induced breaks to approximately 50% of SSA in RAD52 proficient cells(Kelso, Lopezcolorado et al. 2019). *C. elegans* does not encode a RAD52 ortholog and the absolute requirement for HELQ-1 in *C. elegans* SSA further supports this notion of functional redundancy. However, it may well be that only HELQ promotes MMEJ. While we demonstrate this here for worms, recent work in human cells revealed that repair using stretches of homology of about 20bp did not require RAD52, nor polymerase theta(Kelso, Lopezcolorado et al. 2019), thus directly pointing towards other genetic requirements for pol theta independent altEJ.

Our data provoke the idea that HELQ supports a wide range of MH usage, in fact arguing that a strict separation between MMEJ and SSA is not sustainable as the biochemistry and genetic requirements may be identical. Yet the usage of this mode of repair is being dictated by the kinetics potentially as a function of MH length and/or strength of the base pairing interaction. In DSB contexts where annealing results in 3' flaps that need to be processed before gap filling can occur, an additional genetic requirement for a structure specific endonuclease is logical (hence the need for XPF in many experiments addressing SSA).

With respect to usage of HELQ dependent MMEJ/SSA over TMEJ, it is of interest to note that the lower limit of HELQ-dependent MMEJ (>5bp) is overlapping with TMEJ, and it was recently shown that somewhat larger microhomologies positively influence TMEJ efficacy, making the reaction less prone to abortive synthesis[51]. As previously proposed(Kelso, Lopezcolorado et al. 2019, Carvajal-Garcia, Cho et al. 2020, Kamp, van Schendel et al. 2020), the sole function of polymerase theta in alternative end-joining may thus be to convert minimal MH (1-5 bp) into larger stretches that have sufficient thermodynamic stability and priming potential for error-free DNA synthesis, *e.g.* by Pol delta aided by other helicases such as HELQ. In case such a "primer" can be generated by base pairing of small complementary stretches of sequence, the need for POLQ in altEJ becomes obsolete(Kelso, Lopezcolorado et al. 2019), as evident in our experiments as well as in mammalian cell culture work(Kelso, Lopezcolorado et al. 2019) (Figures 1-3).

The mechanism by which HELQ makes ssDNA available for base-pairing interactions is currently unknown, but primarily because of recent *in vitro* work we favor a function in removing potentially annealing-interfering ssDNA binding proteins such as RPA or RAD51. HELQ, purified from archaeal or human source, was found to interact with RPA and able to modulate the RPA-DNA binding including its displacement from ssDNA(Woodman, Brammer et al. 2011, Jenkins, Northall et al. 2021); *C. elegans* HELQ-1 was shown to promote the disassembly of RAD51 from DNA *in vitro*(Ward, Muzzini et al. 2010). Also, the helicase domain in POLQ, which is highly similar to HELQ, was capable of removing RPA from resected DSB ends *in vitro* to allow their annealing(Mateos-Gomez, Kent et al. 2017). Furthermore, the increased incidence of templated insertions (occasionally having more complex sequence configurations) in HELQ-1 deficient contexts suggests an obstacle to DNA synthesis leading to multiple abortive rounds of TMEJ. It is tempting to speculate that TMEJ can initiate DNA synthesis using minute MH available at the terminal ends of a DSB(Wyatt, Feng et al. 2016, Carvajal-Garcia, Cho et al. 2020) yet require helicase activity (within polymerase theta itself or borrowed from HELQ) for removing ssDNA binding proteins, as suggested(Mateos-Gomez, Kent et al. 2017). One way to explain how HELQ suppresses templated insertions is to envision two possible repair routes acting on DSB ends that result from abortive TMEJ. *i.e.* eMMEJ and (another round of) TMEJ. Aborted TMEJ can produce 3' tails that now contain stretches of eMH, because in TMEJ one 3' ssDNA end is extended using the other end as a template. Repair of such intermediates via HELQ-mediated eMMEJ will result in footprints in which the aborted reaction escapes detection. However, a new round of TMEJ that uses the outermost nucleotides of opposing break ends will then produce a templated insertion. According to this rationale, the frequency of templated insertions will increase in a HELQ deficient genetic background. Of particular interest in this respect, TMEJ outcomes in plants, which do not encode HELQ (while encoding RAD52 and POLQ), are far more complex than those observed in *C. elegans*, the majority having templated insertions, which very frequently have a patchwork configuration(Kleinboelting, Huep et al. 2015, van Kregten, de Pater et al. 2016).

Our data also allow us to address hierarchy in DSB-repair pathway usage. In a direct comparison, HELQ-mediated SSA was shown to dominate over TMEJ and loss of SSA in *helq-1* mutants coincided with more frequent TMEJ products at induced DSB. This refunneling halfway during SSA repair suggests that TMEJ can act on resected DSB ends, likely covered by ssDNA binding proteins, as already discussed above. It also suggests that TMEJ functions as a back-up for annealing, which is in line with previous findings that TMEJ can compensate for impaired HR processes(Chan, Yu et al. 2010, Ceccaldi, Liu et al. 2015, Mateos-Gomez, Gong et al. 2015, Carvajal-Garcia, Cho et al. 2020, Kamp, van Schendel et al. 2020). Indeed, a further

testimony for TMEJ's ability to act on HR substrate is the spontaneous accumulation of TDs in the genomes of HELQ-1 deficient animals when grown in unchallenged conditions – a phenotype previously linked to the HR deficiency in BRCA1 mutant cells(Kamp, van Schendel et al. 2020). Also in *C. elegans* BRCA1 deficiency leads to the accumulation of TDs, and we have previously suggested that impaired annealing of an extended strand during SDSA (as a consequence of impaired resection) underlies this phenotype(Chandramouly, Kwok et al. 2013). Similarly, we suggest that TDs arise in HELQ-1 deficient animals because of an inability to complete a downstream step in SDSA: the annealing of an extended break end after being released from its template within the D-loop. Instead, this extended end is (end-)joined to the other DSB end by TMEJ resulting in a TD, a signature motif in certain cancer genomes.

In conclusion, we demonstrate that HELQ-1 is needed for DSB repair that involves annealing of complementary nucleotides, making it an essential factor in SSA, MMEJ and SDSA. We find that altEJ in worms is the sum of two pathways, both requiring a HELQ-family helicase: HELQ in eMMEJ and POLQ in TMEJ. AltEJ may be the rescue for chromosomal breaks that cannot be repaired by the proficient and predominantly error-free pathways NHEJ and HR. While perhaps not being as critical for cell survival as NHEJ or HR, altEJ pathways are intrinsically mutagenic and may thus be the primary cause for DSB-induced genomic alterations(van Schendel, Roerink et al. 2015, Schimmel, van Schendel et al. 2019, Kamp, van Schendel et al. 2020).

METHODS

C. elegans genetics

Nematodes were cultured on standard NGM plates seeded with OP50 bacteria at 20 °C(Brenner 1974). The following alleles were used in this study: *brc-1(lf288)*, *dog-1(gk10)*, *helq-1(tm2134)*, *lig-4(ok716)*, *polq-1(tm2026)*. Using CRISPR/Cas9 mutagenesis a novel *helq-1* allele (*lf314*) was generated (guide RNA sequence 5'- GCAGATTTGCACCTTCGTAT-3') by injecting plasmids in germlines of N2 Bristol worms using standard *C. elegans* microinjection procedures. The injected plasmids encode Cas9 protein (pDD162 (Peft-3::Cas9, Addgene 47549)), the guide RNA (pMB70 (pJJR50)and phenotypic markers (mCherry (pCFJ90, pGH8 and pCFJ104) and rol (pRF4)) to identify transgenic F1 progeny.

Mutation accumulation assays

Mutation accumulation lines were generated by cloning out ten F1 animals from one hermaphrodite. Three nematodes were transferred to new plates each generation. MA lines

were maintained for 40-60 generations (Supplementary table 1). Single animals from the last generation were transferred to new NGM plates. When sufficient offspring was present on these plates, nematodes were washed off with water and incubated for 2 hours while shaking to remove bacteria from the intestine. Genomic DNA was isolated using a Blood and Tissue Culture Kit (Qiagen). DNA was sequenced on an Illumina HiSeq platform (2 × 150 bp paired-end reads). The median sequencing depth of the samples is presented in Supplementary table 1.

Bioinformatic analysis

Image analysis, base calling and error calibration were performed using standard Illumina software. BCL output from the HiSeqX and Novaseq6000 platform was converted using bcl2fastq tool. Raw reads were mapped to the *C. elegans* reference genome (Wormbase release 235) by BWA(Li and Durbin 2009) and SAMtools(Li, Handsaker et al. 2009). Pindel(Ye, Schulz et al. 2009) was used to call G4-induced deletions. Pindel, Manta(Chen, Schulz-Trieglaff et al. 2016) and GRIDSS(Cameron, Schroder et al. 2017) were used for calling structural variations. Variations were considered true if they were covered by both forward and reverse reads and supported by at least five reads. Events were only considered if they were uniquely present in one of the samples. All events were inspected by IGV(Robinson, Thorvaldsdottir et al. 2011) to ensure correctness of the call.

PCR-based assays to identify deletions

Deletion formation at endogenous G4 DNA loci and at the SSA reporter was assayed using a PCR-based approach. Genomic DNA was isolated either from single worms or pools of worms and subjected to nested rounds of PCRs (see supplementary table 2 for primer sequences). L4 stage animals were lysed in 15 µl lysis buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM $MgCl_2.6H_2O$, 0.45 % NP-40 (IGEPAL CA 630), 0.45 % Tween-20) and 1 µl lysate was transferred into 15 µl external PCR mix. After the external PCR, ~ 0.1 µl external PCR mix was transferred to internal PCR mix using a 384 pin replicator (Genetix X5050) for the internal PCR. Deletion products were discriminated based on size by gel-electrophoresis using a 2-log DNA ladder (New England Biolabs N3200S). Repair footprints were analysed using Sanger sequencing. Uncropped blots are provided as a Source Data file.

Sanger sequence analysis

A custom Sanger Sequence-analyser was written (available upon request) to determine sequence alterations at the endogenous G4 loci, SSA reporter and CRISPR target site. Each Sanger

sequence was filtered prior to comparison with the reference sequence on the following criteria: a stretch of $\geq$ 40nt was present where each base had an error probability of < 0.05 surrounding the break site site. All other nucleotides were masked. The filtered high-quality sequence was then compared to the reference sequence and classified as insertion, deletion, wild-type or delins (deletion with insertion).

Ionising radiation sensitivity assay

L4 stage nematodes were exposed to ionising irradiation or mock-treated. Per experimental condition, three-seeded NGM plates containing three nematodes were prepared. The irradiated nematodes were removed from the plate after 48 hours of egg-laying. The number of hatched and unhatched progeny was quantified 24 hours after removal.

Reporter assays

SSA was read out using a LacZ reporter(Pontier and Tijsterman 2009) and SDSA was read out using a GFP reporter(Johnson, Lemmens et al. 2013) using methods described previously. *C.elegans* populations carrying a reporter were synchronised by incubating the worms in a 3:2 mixture of hypochlorite (Acros Organics) to 4 M NaOH until they were dissolved and only eggs remained. The eggs were washed with M9 buffer (22 mM KH_2PO_4, 42 mM Na_2HPO_4, 86 mM NaCl, 1 mM $MgSO_4$) to remove the residual bleach. Eggs were allowed to hatch overnight in M9 buffer. After hatching, larvae were plated on OP-50 seeded NGM plates and incubated at 34°C for 2 h (heat shock). One day after heat shock, I-SceI-mCherry expression was verified using a Leica DM6000 microscope. For SDSA experiments, nematodes were scored for GFP-positive intestinal nuclei using a Leica DM6000 microscope 72 hours after heat shock. For SSA experiments, single worms were fixated and stained in 5% X-gal or lysed for PCR analysis two days after heat shock. One hour prior LacZ staining, young adults were heat-shocked at 34°C for 120 minutes to induce SSA reporter expression.

CRISPR/Cas9 experiment

One day prior to injection, L4 animals were incubated at 15 degrees overnight. Gonads of young adults were injected with a mix containing 20 ng/µl Cas9 (pDD162 (Peft-3::Cas9, Addgene 47549)), 20 ng/µl mCherry marker (pCFJ90 and pCFJ104), 20 ng/µl pBluescript and 20 ng/µl *dpy-10* sgRNA coding plasmid (pMB70 (u6::sgRNA) and 20 ng/µl ssODN described previously(Arribere, Bell et al. 2014). Three days after injecting P0 young adults, F1 progeny were phenotypically screened and phenotypically altered animals (dpy, rol or mCherry

phenotype) were transferred to fresh plates. F2 animals were lysed for PCR analysis (primers: FW: 5'-caacgaactattcgcgtcag-3', RV: 5'-gtggtggctcacgaacttg-3').

Data availability

The raw sequence data generated in this study have been deposited in the NCBI SRA database under accession code PRJNA718436 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA718436. The N2 wild-type and *dog-1* sequence data were published previously and can be found at NCBI SRA (accession codes PRJNA260487 https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA260487 and PRJNA225882 https://www.ncbi.nlm.nih.gov/bioproject/225882). We used the C. elegans reference genome (Wormbase release 235, https://wormbase.org//about/ wormbase_release_WS235#10--10) in this study.

Author contributions

JK and MT conceived and designed the study. JK, BL, RR and SC performed the experiments. RvS performed the bioinformatical analyses. All authors interpreted the experimental results. JK and MT wrote the manuscript.

Competing interests

The authors declare no competing interests.

REFERENCES

Adamo, A., P. Montemauri, N. Silva, J. D. Ward, S. J. Boulton and A. La Volpe (2008). "BRC-1 acts in the inter-sister pathway of meiotic double-strand break repair." EMBO Rep **9**(3): 287-292.

Adelman, C. A. and S. J. Boulton (2010). "Metabolism of postsynaptic recombination intermediates." FEBS Lett **584**(17): 3709-3716.

Adelman, C. A., R. L. Lolo, N. J. Birkbak, O. Murina, K. Matsuzaki, Z. Horejsi, K. Parmar, V. Borel, J. M. Skehel, G. Stamp, A. D'Andrea, A. A. Sartori, C. Swanton and S. J. Boulton (2013). "HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis." Nature **502**(7471): 381-384.

Arribere, J. A., R. T. Bell, B. X. Fu, K. L. Artiles, P. S. Hartman and A. Z. Fire (2014). "Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans." Genetics **198**(3): 837-846.

Bétermier, M., P. Bertrand and B. S. Lopez (2014). "Is non-homologous end-joining really an inherently error-prone process?" PLoS Genet **10**(1): e1004086.

Bhargava, R., D. O. Onyango and J. M. Stark (2016). "Regulation of Single-Strand Annealing and its Role in Genome Maintenance." Trends Genet **32**(9): 566-575.

Boulton, S. J. and S. P. Jackson (1996). "Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance." Nucleic Acids Res **24**(23): 4639-4648.

Boulton, S. J., J. S. Martin, J. Polanowska, D. E. Hill, A. Gartner and M. Vidal (2004). "BRCA1/BARD1 orthologs required for DNA repair in Caenorhabditis elegans." Curr Biol **14**(1): 33-39.

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

Cameron, D. L., J. Schroder, J. S. Penington, H. Do, R. Molania, A. Dobrovic, T. P. Speed and A. T. Papenfuss (2017). "GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly." Genome Res **27**(12): 2050-2060.

Carvajal-Garcia, J., J. E. Cho, P. Carvajal-Garcia, W. Feng, R. D. Wood, J. Sekelsky, G. P. Gupta, S. A. Roberts and D. A. Ramsden (2020). "Mechanistic basis for microhomology identification and genome scarring by polymerase theta." Proc Natl Acad Sci U S A.

Carvajal-Garcia, J., J. E. Cho, P. Carvajal-Garcia, W. Feng, R. D. Wood, J. Sekelsky, G. P. Gupta, S. A. Roberts and D. A. Ramsden (2020). "Mechanistic basis for microhomology identification and genome scarring by polymerase theta." Proc Natl Acad Sci U S A **117**(15): 8476-8485.

Castillo Bosch, P., S. Segura-Bayona, W. Koole, J. T. van Heteren, J. M. Dewar, M. Tijsterman and P. Knipscheer (2014). "FANCJ promotes DNA synthesis through G-quadruplex structures." Embo j **33**(21): 2521-2533.

Ceccaldi, R., J. C. Liu, R. Amunugama, I. Hajdu, B. Primack, M. I. Petalcorin, K. W. O'Connor, P. A. Konstantinopoulos, S. J. Elledge, S. J. Boulton, T. Yusufzai and A. D. D'Andrea (2015). "Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair." Nature **518**(7538): 258-262.

Ceccaldi, R., B. Rondinelli and A. D. D'Andrea (2016). "Repair Pathway Choices and Consequences at the Double-Strand Break." Trends Cell Biol **26**(1): 52-64.

Chan, S. H., A. M. Yu and M. McVey (2010). "Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila." PLoS Genet **6**(7): e1001005.

Chandramouly, G., A. Kwok, B. Huang, N. A. Willis, A. Xie and R. Scully (2013). "BRCA1 and CtIP suppress long-tract gene conversion between sister chromatids." Nat Commun **4**: 2404.

Chang, H. H. Y., N. R. Pannunzio, N. Adachi and M. R. Lieber (2017). "Non-homologous DNA end joining and alternative pathways to double-strand break repair." Nat Rev Mol Cell Biol **18**(8): 495-506.

Chen, X., O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Kallberg, A. J. Cox, S. Kruglyak and C. T. Saunders (2016). "Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications." Bioinformatics **32**(8): 1220-1222.

Cheung, I., M. Schertzer, A. Rose and P. M. Lansdorp (2002). "Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA." Nat Genet **31**(4): 405-409.

Dehé, P. M. and P. H. L. Gaillard (2017). "Control of structure-specific endonucleases to maintain genome stability." Nat Rev Mol Cell Biol **18**(5): 315-330.

Garcia-Muse, T., U. Galindo-Diaz, M. Garcia-Rubio, J. S. Martin, J. Polanowska, N. O'Reilly, A. Aguilera and S. J. Boulton (2019). "A Meiotic Checkpoint Alters Repair Partner Bias to Permit Inter-sister Repair of Persistent DSBs." Cell Rep **26**(3): 775-787.e775.

Hustedt, N., Y. Saito, M. Zimmermann, A. Álvarez-Quilón, D. Setiaputra, S. Adam, A. McEwan, J. Y. Yuan, M. Olivieri, Y. Zhao, M. T. Kanemaki, A. Jurisicova and D. Durocher (2019). "Control of homologous recombination by the HROB-MCM8-MCM9 pathway." Genes Dev **33**(19-20): 1397-1415.

Jenkins, T., S. J. Northall, D. Ptchelkine, R. Lever, A. Cubbon, H. Betts, V. Taresco, C. D. O. Cooper, P. J. McHugh, P. Soultanas and E. L. Bolt (2021). "The HelQ human DNA repair helicase utilizes a PWI-like domain for DNA loading through interaction with RPA, triggering DNA unwinding by the HelQ helicase core." NAR Cancer **3**(1): zcaa043.

Johnson-Schlitz, D. M., C. Flores and W. R. Engels (2007). "Multiple-pathway analysis of double-strand break repair mutations in Drosophila." PLoS Genet **3**(4): e50.

Johnson, N. M., B. B. Lemmens and M. Tijsterman (2013). "A role for the malignant brain tumour (MBT) domain protein LIN-61 in DNA double-strand break repair by homologous recombination." PLoS Genet **9**(3): e1003339.

Kabotyanski, E. B., L. Gomelsky, J. O. Han, T. D. Stamato and D. B. Roth (1998). "Double-strand break repair in Ku86- and XRCC4-deficient cells." Nucleic Acids Res **26**(23): 5333-5342.

Kamp, J. A., R. van Schendel, I. W. Dilweg and M. Tijsterman (2020). "BRCA1-associated structural variations are a consequence of polymerase theta-mediated end-joining." Nat Commun **11**(1): 3615.

Kanaar, R., J. H. Hoeijmakers and D. C. van Gent (1998). "Molecular mechanisms of DNA double strand break repair." Trends Cell Biol **8**(12): 483-489.

Kelso, A. A., F. W. Lopezcolorado, R. Bhargava and J. M. Stark (2019). "Distinct roles of RAD52 and POLQ in chromosomal break repair and replication stress response." PLoS Genet **15**(8): e1008319.

Kleinboelting, N., G. Huep, I. Appelhagen, P. Viehoever, Y. Li and B. Weisshaar (2015). "The Structural Features of Thousands of T-DNA Insertion Sites Are Consistent with a Double-Strand Break Repair-Based Insertion Mechanism." Mol Plant **8**(11): 1651-1664.

Koole, W., R. van Schendel, A. E. Karambelas, J. T. van Heteren, K. L. Okihara and M. Tijsterman (2014). "A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites." Nat Commun **5**: 3216.

Krogh, B. O. and L. S. Symington (2004). "Recombination proteins in yeast." Annu Rev Genet **38**: 233-271.

Kruisselbrink, E., V. Guryev, K. Brouwer, D. B. Pontier, E. Cuppen and M. Tijsterman (2008). "Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans." Curr Biol **18**(12): 900-905.

Lemmens, B., R. van Schendel and M. Tijsterman (2015). "Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers." Nat Commun **6**: 8909.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Luebben, S. W., T. Kawabata, M. K. Akre, W. L. Lee, C. S. Johnson, M. G. O'Sullivan and N. Shima (2013). "Helq acts in parallel to Fancc to suppress replication-associated genome instability." Nucleic Acids Res **41**(22): 10283-10297.

Ma, J. L., E. M. Kim, J. E. Haber and S. E. Lee (2003). "Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences." Mol Cell Biol **23**(23): 8820-8828.

Marini, F. and R. D. Wood (2002). "A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308." J Biol Chem **277**(10): 8716-8723.

Mateos-Gomez, P. A., F. Gong, N. Nair, K. M. Miller, E. Lazzerini-Denchi and A. Sfeir (2015). "Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination." Nature **518**(7538): 254-257.

Mateos-Gomez, P. A., T. Kent, S. K. Deng, S. McDevitt, E. Kashkina, T. M. Hoang, R. T. Pomerantz and A. Sfeir (2017). "The helicase domain of Poltheta counteracts RPA to promote alt-NHEJ." Nat Struct Mol Biol **24**(12): 1116-1123.

McCaffrey, R., D. St Johnston and A. González-Reyes (2006). "Drosophila mus301/spindle-C encodes a helicase with an essential role in double-strand DNA break repair and meiotic progression." Genetics **174**(3): 1273-1285.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Meier, B., N. V. Volkova, Y. Hong, S. Bertolini, V. González-Huici, T. Petrova, S. Boulton, P. J. Campbell, M. Gerstung and A. Gartner (2021). "Protection of the C. elegans germ cell genome depends on diverse DNA repair pathways during normal proliferation." PLoS One **16**(4): e0250291.

Mendez-Dorantes, C., R. Bhargava and J. M. Stark (2018). "Repeat-mediated deletions can be induced by a chromosomal break far from a repeat, but multiple pathways suppress such rearrangements." Genes Dev **32**(7-8): 524-536.

Muzzini, D. M., P. Plevani, S. J. Boulton, G. Cassata and F. Marini (2008). "Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways." DNA Repair (Amst) **7**(6): 941-950.

Paix, A., A. Folkmann, D. H. Goldman, H. Kulaga, M. J. Grzelak, D. Rasoloson, S. Paidemarry, R. Green, R. R. Reed and G. Seydoux (2017). "Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks." **114**(50): E10745-E10754.

Paix, A., H. Schmidt and G. Seydoux (2016). "Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs." Nucleic Acids Res **44**(15): e128.

Pâques, F. and J. E. Haber (1999). "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **63**(2): 349-404.

Pontier, D. B. and M. Tijsterman (2009). "A robust network of double-strand break repair pathways governs genome integrity during C. elegans development." Curr Biol **19**(16): 1384-1388.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Roerink, S. F., R. van Schendel and M. Tijsterman (2014). "Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans." Genome Res **24**(6): 954-962.

Rothenberg, E., J. M. Grimme, M. Spies and T. Ha (2008). "Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes." Proc Natl Acad Sci U S A **105**(51): 20274-20279.

San Filippo, J., P. Sung and H. Klein (2008). "Mechanism of eukaryotic homologous recombination." Annu Rev Biochem **77**: 229-257.

Schimmel, J., R. van Schendel, J. T. den Dunnen and M. Tijsterman (2019). "Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining." Trends Genet **35**(9): 632-644.

Schwartz, E. K. and W. D. Heyer (2011). "Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes." Chromosoma **120**(2): 109-127.

Scully, R., A. Panday, R. Elango and N. A. Willis (2019). "DNA double-strand break repair-pathway choice in somatic mammalian cells." Nat Rev Mol Cell Biol **20**(11): 698-714.

Sugawara, N. and J. E. Haber (1992). "Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation." Mol Cell Biol **12**(2): 563-575.

Symington, L. S. (2002). "Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair." Microbiol Mol Biol Rev **66**(4): 630-670, table of contents.

Takata, K., S. Reh, J. Tomida, M. D. Person and R. D. Wood (2013). "Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs." Nat Commun **4**: 2338.

van Kregten, M., S. de Pater, R. Romeijn, R. van Schendel, P. J. Hooykaas and M. Tijsterman (2016). "T-DNA integration in plants results from polymerase-θ-mediated DNA repair." Nat Plants **2**(11): 16164.

van Schendel, R., S. F. Roerink, V. Portegijs, S. van den Heuvel and M. Tijsterman (2015). "Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis." Nat Commun **6**: 7394.

Volkova, N. V., B. Meier, V. González-Huici, S. Bertolini, S. Gonzalez, H. Vöhringer, F. Abascal, I. Martincorena, P. J. Campbell, A. Gartner and M. Gerstung (2020). "Mutational signatures are jointly shaped by DNA damage and repair." Nat Commun **11**(1): 2169.

Ward, J. D., D. M. Muzzini, M. I. Petalcorin, E. Martinez-Perez, J. S. Martin, P. Plevani, G. Cassata, F. Marini and S. J. Boulton (2010). "Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair." Mol Cell **37**(2): 259-272.

3

Wei, D. S. and Y. S. Rong (2007). "A genetic screen for DNA double-strand break repair mutations in Drosophila." <u>Genetics</u> **177**(1): 63-77.

Woodman, I. L., K. Brammer and E. L. Bolt (2011). "Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA)." <u>DNA Repair (Amst)</u> **10**(3): 306-313.

Wyatt, D. W., W. Feng, M. P. Conlin, M. J. Yousefzadeh, S. A. Roberts, P. Mieczkowski, R. D. Wood, G. P. Gupta and D. A. Ramsden (2016). "Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks." <u>Mol Cell</u> **63**(4): 662-673.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler and Z. Ning (2009). "Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads." <u>Bioinformatics</u> **25**(21): 2865-2871.

Yousefzadeh, M. J., D. W. Wyatt, K. Takata, Y. Mu, S. C. Hensley, J. Tomida, G. O. Bylund, S. Doublié, E. Johansson, D. A. Ramsden, K. M. McBride and R. D. Wood (2014). "Mechanism of suppression of chromosomal instability by DNA polymerase POLQ." <u>PLoS Genet</u> **10**(10): e1004654.

Zhao, B., E. Rothenberg, D. A. Ramsden and M. R. Lieber (2020). "The molecular basis and disease relevance of non-homologous DNA end joining." <u>Nat Rev Mol Cell Biol</u> **21**(12): 765-781.

SUPPLEMENTARY INFORMATION
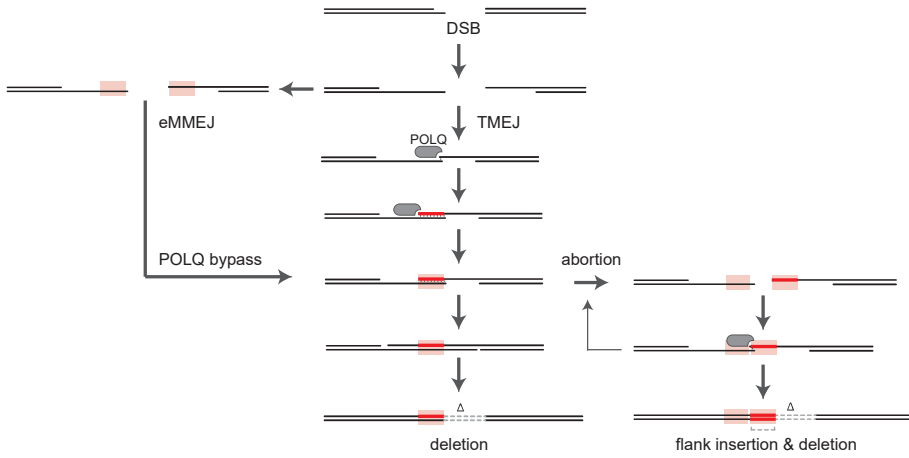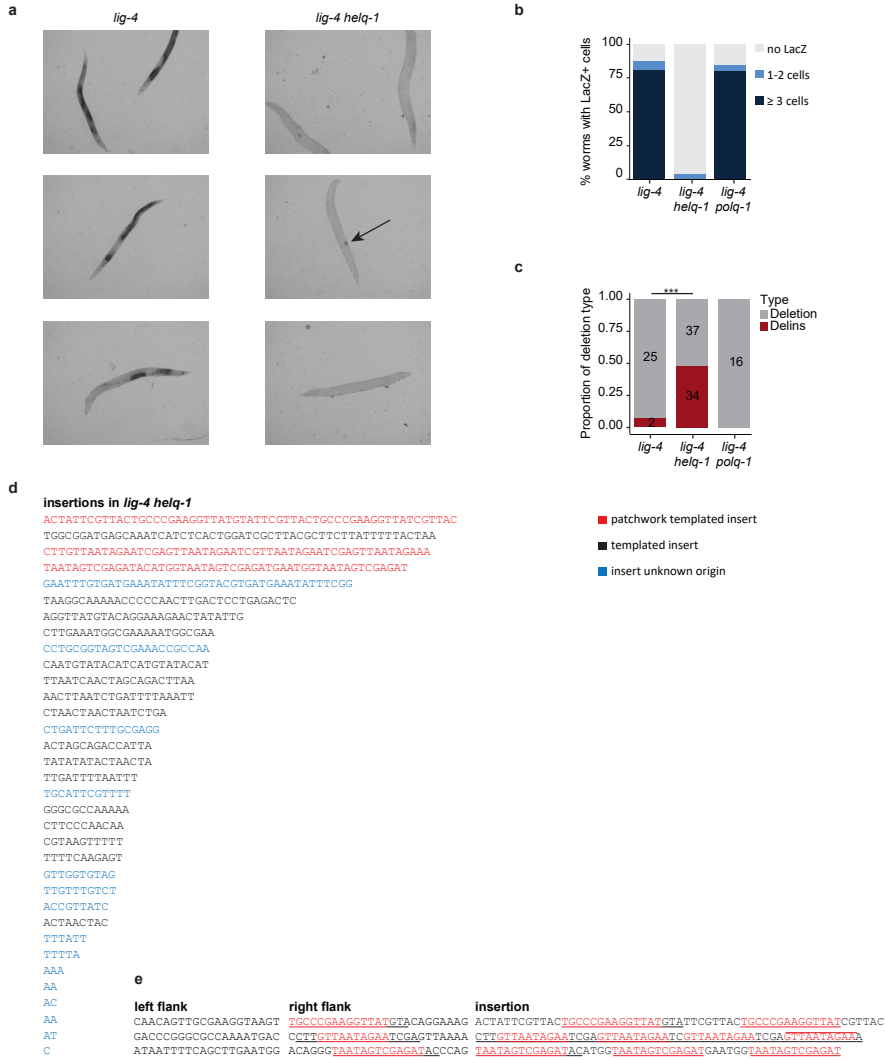

Supplementary Figure 1
Model illustrating i) how eMMEJ can bypass the requirement for POLQ in altEJ; ii) how abortive TMEJ results in templated insertions within deletion junctions.
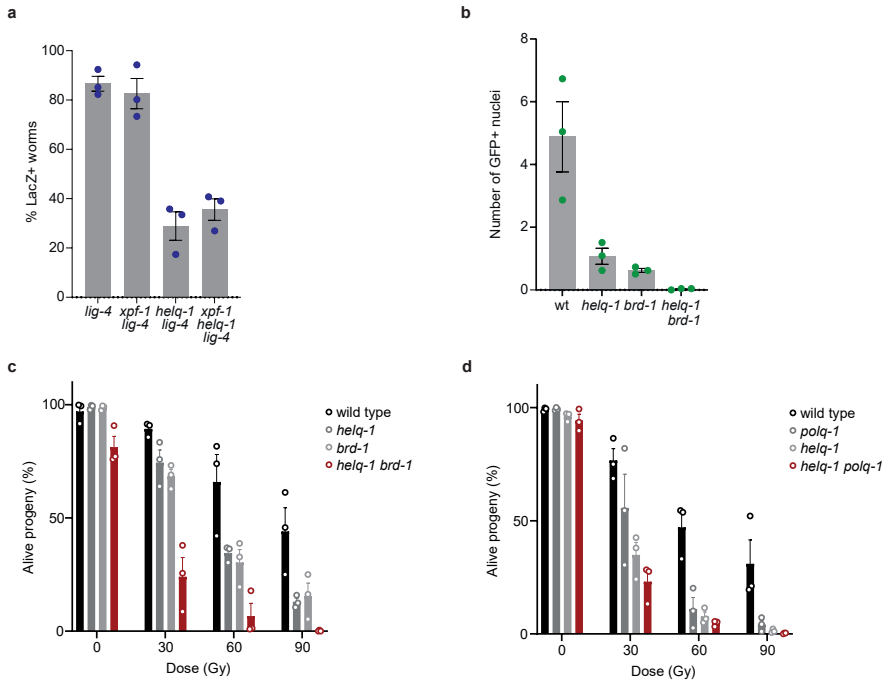
**Supplementary figure 2 Effect of deficiency in POLQ-1 and HELQ-1 on sensitivity to ionizing radiation.**

a. Histogram depicting the percentage of LacZ positive worms for the indicated genotype. Experiments were performed in triplicate. Each dot represents the average percentage of each replicate. Error bars represent SEM. b. Histograms depicting the average number of GFP expressing nuclei per worm for the indicated genotype. Experiments are performed in triplicate (n=45 per replicate). Each dot represents the average number of each replicate. Error bars represent SEM. c. Wild type (black), *helq-1* (grey), *brd-1* (light grey) and *helq-1 brd-1* (red) L4 larvae were exposed to 0, 30, 60 or 90 Gy IR (n = 3). Three days after irradiation, the numbers of alive offspring versus unviable eggs were quantified. The average survival in each experiment is represented in dots. Error bars represent SEM. d. Wild type (black), *polq-1* (grey), *helq-1* (light grey) and *helq-1 polq-1* (red) L4 larvae were exposed to 0, 30, 60 or 90 Gy IR (n = 3). Three days after irradiation, the numbers of alive offspring versus unviable eggs were quantified. The average survival in each experiment is represented in dots. Error bars represent SEM.

Supplementary figure 3 SSA deficiency and increased insertions in *C. elegans* lacking HELQ-1

a. Representative pictures of *lig-4* and *lig-4 helq-1* mutant animals carrying the reporter transgenes that were heat shocked to induce Isce-I expression, followed by staining for B-galactosidase expression. Arrow indicates staining scored as 1 LacZ+ cell. b. Histogram depicting the percentage of worms without LacZ+ cells, with one or two LacZ+ cells or more than two blue cells for the indicated genotype. Worms of three independent experiments were scored. c. Proportion of deletion types at the SSA reporter. Difference in ratio between deletions without and with insertion were tested between *lig-4* and *lig-4 helq-1* using Chi square test (*** P<0.001). d. Sequences of DNA insertions identified within deletions at the SSA reporter in *lig-4 helq-1* animals (sorted to size). Patchwork insertions are represented in red, templated insertions in black, and insertions of which the origin could not be determined are represented in blue. e. Examples of deletions with patchwork insertions. For each case, the insertion and flanks of the corresponding deletion are depicted. Inserted nucleotide stretches that are identical to the flank of the deletion junction are underscored in both the insertion and flank. Nucleotide stretches present multiple times within one insertion are depicted in red.
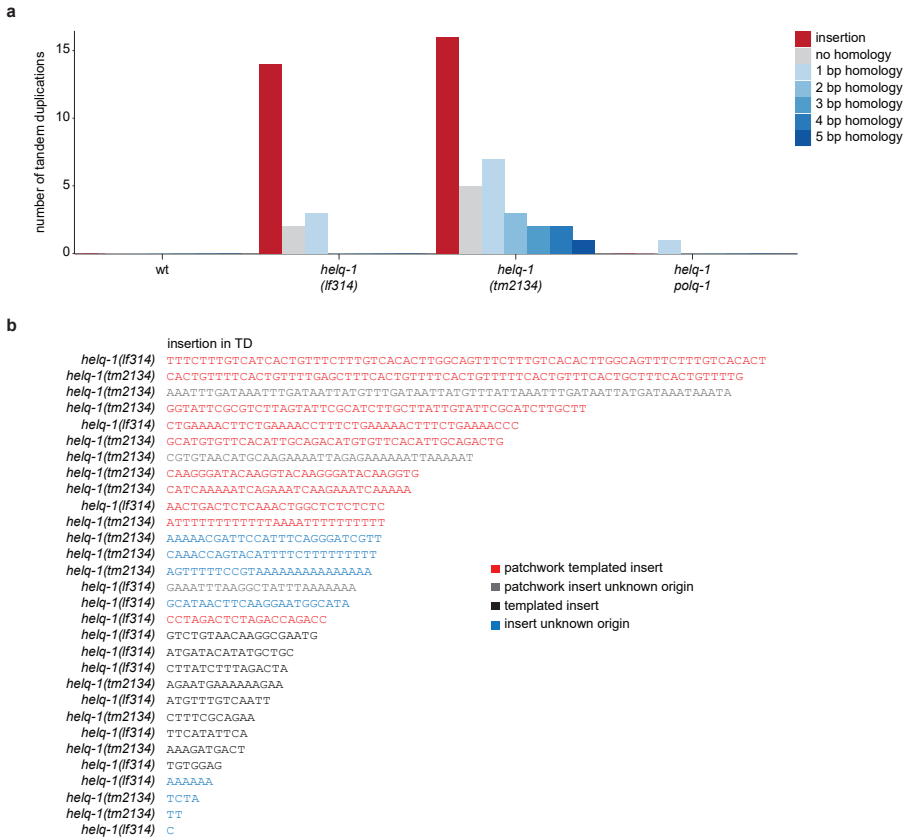
Supplementary figure 4 Mutational footprints of identified tandem duplications
a. Histogram depicting the different mutational footprints of tandem duplications that accumulated in genomes of indicated genotypes. Tandem duplications are colour-coded according to the following mutational classification: grey for simple deletions without apparent MH at the junction, blue for simple deletions with MH at the junction, wherein the saturation level of the blue colour increases with increasing amount of homology identified. Deletions containing insertions are depicted in red b. Insertions identified in tandem duplications that accumulated in genomes of *helq-1(lf314)* and *helq-1(tm2134)* animals. Insertions are sorted according to size. Patchwork templated insertions are represented in red, patchwork insertions of unknown origin in grey, simple templated insertions in black and insertions of which the origin could not be determined are represented in blue.

File	Genotype	Generations	Median Coverage
N2_2_60	Wildtype	60	34.38165
N2_3_60	Wildtype	60	57.39455
N2_4_60	Wildtype	60	18.4895
N2_50_60	Wildtype	60	26.0963
FX2134-D50	*helq-1(tm2134)*	50	36.62475
FX2134-G50	*helq-1(tm2134)*	50	43.48685
FX2134-I50	*helq-1(tm2134)*	50	39.7627
NL4716_50gen	*dog-1(pk2247)*	50	26.68345
NL4721_50gen	*dog-1(pk2247)*	50	9.99105
NL4722_50gen	*dog-1(pk2247)*	50	24.8927
XF1116_11_42	*dog-1(gk10) polq-1(tm2026)*	42	30.2128
XF1116_3_40	*dog-1(gk10) polq-1(tm2026)*	40	23.7144
XF1116_7_40	*dog-1(gk10) polq-1(tm2026)*	40	19.1549
XF1582-0	*helq-1(tm2134) polq-1(tm2026)*	0	37.10295
XF1582-B50	*helq-1(tm2134) polq-1(tm2026)*	50	43.79555
XF1582-E50	*helq-1(tm2134) polq-1(tm2026)*	50	43.77495
XF1582-G50	*helq-1(tm2134) polq-1(tm2026)*	50	36.13365
XF1700-0	*helq-1(lf314)*	0	29.9088
XF1700-50E	*helq-1(lf314)*	50	32.5171
XF1700-50H	*helq-1(lf314)*	50	34.87865
XF1700-50I	*helq-1(lf314)*	50	39.75315

Supplementary table 1 Mutation accumulation experiment sample characteristics Overview of the number of generations per sample and median sequencing depth per sample.

Locus	Primer function	Sequence
Qua213	External PCR forward	ctcagccaaggctacaaac
Qua213	External PCR reverse	ccggcaattacacatttgcc
Qua213	Internal PCR forward	gatacgtgtacatgaatagtc
Qua213	Internal PCR reverse	caaaactgtcgcctgacctc
Qua317	External PCR forward	cattgtgggaaaaatccgacg
Qua317	External PCR reverse	tttgccatcaaggttccagac
Qua317	Internal PCR forward	ctttagccaattttgagcac
Qua317	Internal PCR reverse	ggatttcacagcgtcaagag
Qua915	External PCR forward	ccgttttataagcccaaaccg
Qua915	External PCR reverse	agtttccagtgaaaaatcgcg
Qua915	Internal PCR forward	tgcgcggttttgtgtgtgg
Qua915	Internal PCR reverse	agtttccagtgaaaaatcgcg
Qua1277	External PCR forward	cctgacaaacgcctactctc
Qua1277	External PCR reverse	ggggagaagccgcatccaa
Qua1277	Internal PCR forward	gaatccctttaatttggcaatag
Qua1277	Internal PCR reverse	cacatggagacggagagaaac
SSA reporter	External PCR forward	cagacggaaaatgtatctgggg
SSA reporter	External PCR reverse	agttcatccatgccatgtg
SSA reporter	Internal PCR forward	caaactataatcatctcactgg
SSA reporter	Internal PCR reverse	ggacagggccatcgccaattgg

Supplementary table 2 Primer sequences

3