A MUTAGENIC REPAIR PATHWAY DEPENDENT ON POLYMERASE θ PROTECTS THE *C. ELEGANS* GENOME AGAINST LARGE CHROMOSOMAL DELETIONS

3

Sophie F. Roerink*, Robin van Schendel*, Marcel Tijsterman
*equally contributing authors

Leiden University Medical Center,
Department of Toxicogenetics

ABSTRACT

DNA lesions that block replication fork progression are drivers of cancer-associated genome alterations, but the error prone DNA repair mechanisms acting on collapsed replication are incompletely understood, and there contribution to genome evolution largely unexplored. Here, by whole genome sequencing of mutation accumulation line, we identify a distinct class of deletions that spontaneously accumulate in *C. elegans* strains lacking translesion synthesis (TLS) polymerases. Emerging DNA double-strand breaks are repaired via an error-prone mechanism in which the outermost nucleotide of one end serves to prime DNA synthesis on the other end. This pathway critically depends on the A-family polymerase θ, which protects the genome against gross chromosomal rearrangements. By comparing the genomes of isolates of *C. elegans* from different geographical regions, we found that in fact most spontaneously evolving structural variations match the signature of polymerase Theta-mediated end joining (TMEJ), arguing that this pathway is an important source of genetic diversification.

INTRODUCTION

Identifying the mechanisms that fuel genome change is crucial for understanding evolution and carcinogenesis. Spontaneous mutagenesis is caused predominantly by misinsertions or slippage events of replicative polymerases that are missed by their proofreading domains, and not corrected by mismatch repair (Lynch *et al.* 2008). Less frequently, but with a potentially much more detrimental effect, mutations can arise when DNA damages obstruct progression of DNA replication; and stalled replication forks eventually collapse, resulting in highly mutagenic double stranded breaks (DSBs). While error-free homologous repair, where the sister chromatid is used as a template, restores the original sequence, infrequent error-prone end joining processes may give rise to spontaneous deletions and tumor-promoting translocations (Mitelman *et al.* 2007).

To circumvent fork collapse at DNA damage, cells employ various alternative polymerases that are capable of incorporating nucleotides across DNA lesions, and are hence called translesion synthesis (TLS) polymerases. TLS acts on a wide variety of DNA lesions that can result from endogenous as well as exogenous genotoxic sources: DNA lesions that result from UV-light exposure, for instance, are efficiently bypassed by the well-conserved TLS polymerase η, inactivation of which in humans leads to the variant form of the skin cancer predisposition syndrome Xeroderma Pigmentosum (Johnson *et al.* 1999; Masutani *et al.* 1999). Abundant *in vitro* studies demonstrate the involvement of TLS proteins Pol η and Pol κ in bypass of lesions that are produced by endogenous reactive compounds, arguing that these polymerases are also essential for protection of the genome under unchallenged conditions (Haracska *et al.* 2000; Kusumoto *et al.* 2002; Fischhaber *et al.* 2002).

Although error-prone while replicating, and thus potentially causing misinsertions, TLS polymerases are thought to protect cells against the more mutagenic effects of replication fork collapse (Lynch *et al.* 2008; Knobel and Marti 2011). Here, we investigate the contribution of TLS polymerases on the maintentance of genome stability and the mechanisms acting on stalled DNA replication, by characterizing *C. elegans* strains that are defective for the Y-family polymerases η and κ.

RESULTS AND DISCUSSION

TLS polymerases protect genomes against spontaneous deletions

In previous work, we established the role of the *C. elegans* homologs of TLS polymerases η and κ in protection against various a wide range of exogenous DNA damaging agents (Mitelman *et al.* 2007; Roerink *et al.* 2012). Anecdotal evidence in these studies, where we observed readily recognizable mutant phenotypes during normal culturing of *polh-1*

and *polk-1* double mutant animals, prompted us to suspect a prominent role for these Y-family of TLS polymerases in the prevention of spontaneous mutations (Figure S1). To assess mutagenesis in an unbiased way, we generated MA lines for single and double mutants in TLS for 60 generations, thus allowing spontaneous mutations to accumulate, and then sequenced their genomes (Figure 1A, Table S1 and Supplemental data file). Mutation accumulation lines of a wildtype strain (Bristol N2) and of the mismatch repair deficient strain *msh-6* - for which an ~100-fold higher mutation frequency has been reported (Johnson *et al.* 1999; Masutani *et al.* 1999; Tijsterman *et al.* 2002; Denver *et al.* 2005) - were sequenced as references.

Although Pol η and κ have reduced accuracy while replicating from undamaged as well as damaged DNA templates (Matsuda *et al.* 2000; Haracska *et al.* 2000; Ohashi *et al.* 2000; Kusumoto *et al.* 2002; Fischhaber *et al.* 2002), we found that they hardly, if at all,

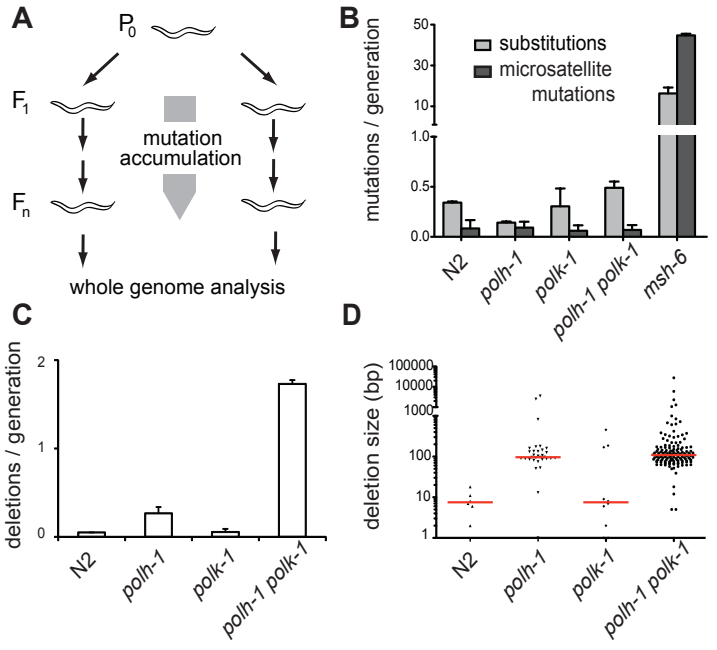


FIGURE 1. Spontaneous mutagenesis in TLS deficient strains. A. Generation of mutation accumulation (MA) lines. For each genotype multiple populations were started by cloning out single worms from a single hermaphrodite P0. Cultures were propagated by transferring animals to new plates each generation. At generation Fn, a single animal was grown to a full population of which genomic DNA was isolated and subjected to whole genome sequencing on an Illumina HiSeq. **B.** Substitution and microsatellite mutation rates for the indicated genotypes. Mutation rates are expressed as the number of mutations per generation divided by the number of nucleotides analysed. **C.** Rates of structural variations for the indicated genotypes. (D) Size distribution of deletions in the different mutant backgrounds. The median sizes are indicated in red.

contribute to base substitution processes under normal growth conditions (Figure 1B). Erroneous replication of microsatellites is another dominant basis of mutagenesis under wildtype growth conditions(Denver *et al.* 2009) for which activity of TLS polymerases has been implicated (Burr *et al.* 2006; Hile *et al.* 2011). However, our whole genome sequencing data show no elevation in the microsatellite mutation rate in *polh-1polk-1* worms compared to wildtype controls (Figure 1B), arguing that other types of genome alteration, or epigenetic changes (Sarkies *et al.* 2010) must underlie the phenotypic variations observed in TLS deficient cultures.

To detect any other structural variations, we employed Pindel software, developed to identify deletions and/or insertions in whole-genome sequencing data (Ye *et al.* 2009). Strikingly, a unique class of deletions emerged in *polh-1* and *polh-1 polk-1* mutants, which were not associated with repetitive loci or any other obvious genomic trait, and occurred at seemingly random locations throughout the genome (Figures 1C and S2). The vast majority of deletions ranged between 50 and 200 bp in size, with just a few exceptions being larger or smaller (Figure 1D). The median size, of 107 bp, was similar for *polh-1* and *polh-1polk-1*-derived deletions (Figure 1D). Control wildtype and *msh-6* samples did not display any mutations from this class. Deletions occurred in *polh-1* single mutants with a rate of ~0.4 per animal generation, which translates to an average of ~0.03 deletion per genome per cell division. *Polk-1* single mutants hardly suffered from deletions; however, *polh-1polk-1* double mutants had 5-fold increased rates of deletion induction as compared to *polh-1* single mutant animals, implying that *C. elegans* Pol η and Pol κ function redundantly on a subset of endogenous lesions.

These data also argue that Pol κ and Pol η prevent the induction of spontaneous deletions in a largely error-free manner: their joined action prevents ~2 deletions per animal generation without significantly affecting the overall substitution rate. In fact, the number of substitutions (~0.3 per generation) in TLS-proficient strains compares to only 15 percent of the number of deletions that are being induced in the absence of Pol η and Pol κ indicating that the vast majority of endogenous replication blocking lesions that use TLS to avoid deletion formation, are not leading to mutations (Figure 1B).

Footprints of error-prone DSB repair

To further investigate the origin of the high number of deletions in *polh-1polk-1* deficient strains, we looked for manifestations of genomic instability in germ cells of these animals.

We observed a mild but statistically significant increase in the number of foci of the DSB marker RAD-51 in proliferating germ cells of *polh-1polk-1* mutant animals (Figure S3A-B). Elevated levels of DSBs are also suggested by the spontaneous emergence of dominant *him* mutants in *polh-1polk-1* mutant populations (Figure S1). This phenotype, which is defined by dominant inheritance of an increased number of males (XO) in

predominantly hermaphroditic (XX) populations, has previously been found upon exposure to γ-irradiation and in mutants with enhanced telomere shortening, where it proved to result from X/autosome translocations (Herman *et al.* 1982; Meier *et al.* 2009). Despite these manifestations of enhanced replication stress in *polh-1polk-1* mutants, the levels of DSBs were insufficient to activate the two DNA damage checkpoints that operate in the *C. elegans* germline: cell cycle arrest and apoptosis (Gartner *et al.* 2000). We found neither a reduction in germ cell proliferation nor an increase of apoptotic bodies in *polh-1polk-1* mutant germlines, suggesting that TLS compromised germ cells proliferate in the presence of elevated levels of DSBs, with genomic deletions as a consequence (Figure S3C-E).

To obtain mechanistic insight on the biology of deletion formation, we performed a detailed analysis on the sequence context of 141 *polh-1polk-1*-derived deletions (Supplemental data file). While the majority had simple deletion junctions (without inserts), about 25 percent of the footprints showed insertions of short sequence stretches (Figure 2A). Cases with inserts sufficiently long to faithfully trace their origin revealed that the inserted stretch, or part of it, is identical to sequences flanking the deletion (Figure 2B-C). This finding strongly suggests that DNA close to the break site was used as a template for *de novo* synthesis before both DNA ends were joined. The close proximity of insertions to their template also suggests that the extendable end of the DSB is not subjected to extensive trimming.

A DSB repair mechanism involving DNA synthesis is also suggested by the notion of a 'priming' nucleotide in more than 80 percent of all deletions: 83 of the 102 deletions without insert contain at the junction at least one nucleotide that could have originated from either flank; in 51 cases this is restricted to a single nucleotide. To systematically assess the significance of this observation, we constructed deletion junction heat maps, which reflect the level of (micro)homology between 5' and 3' junctions (Figure 2D-F). We scored the degree of sequence identity in an 8 nt window, encompassing the 4 outermost nucleotides of the flanking sequence and the 4 nucleotides of the adjacent, but deleted, sequence. Indeed, compared to a randomly distributed simulated set, we found a very high similarity score for the nucleotide at the -1 position of the deletions and the +1 position of the opposing flanks ($p=7.3x10^{-17}$). Importantly, this profound degree of microhomology is restricted to only a single nucleotide, arguing that the underlying mechanism is distinct from previously described repair pathways involving more extended microhomology (Payen *et al.* 2008; McVey and Lee 2008; Hastings *et al.* 2009).

We investigated whether the deletion specifics would reveal the nature of the spontaneous damage underlying fork stalling and break formation in TLS compromised animals. We hypothesized that if the nascent strand, blocked at the site of base damage, defines one end of a DSBs, then the -1 position of the corresponding junction will represent the nucleotide complimentary to the damaged base: it is the first base not to be incorporated. To test this hypothesis, we plotted the base distribution for each

position of the junction and indeed found it not to be random at the -1 position, but rather being dominated by cytosines (Figure 2G). This result strongly suggests that spontaneous base damage primarily at guanines requires Pol κ and Pol η to avoid DSB induction, which may point towards N2-dG and/or 8-oxo-dG adducted sites as a primary source of spontaneous mutagenesis, as bypass activities of Pol κ and Pol η have been reported for these lesions (Haracska *et al.* 2000; Avkin *et al.* 2004; Choi *et al.* 2006). This conclusion is further supported by the notion that the degree of spontaneous deletion induction in these mutant backgrounds, follows their sensitivity towards the guanine alkylator MMS, where Pol η mutant animals are much more sensitive than Pol κ, but not as sensitive as double knockout worms (Roerink *et al.* 2012).

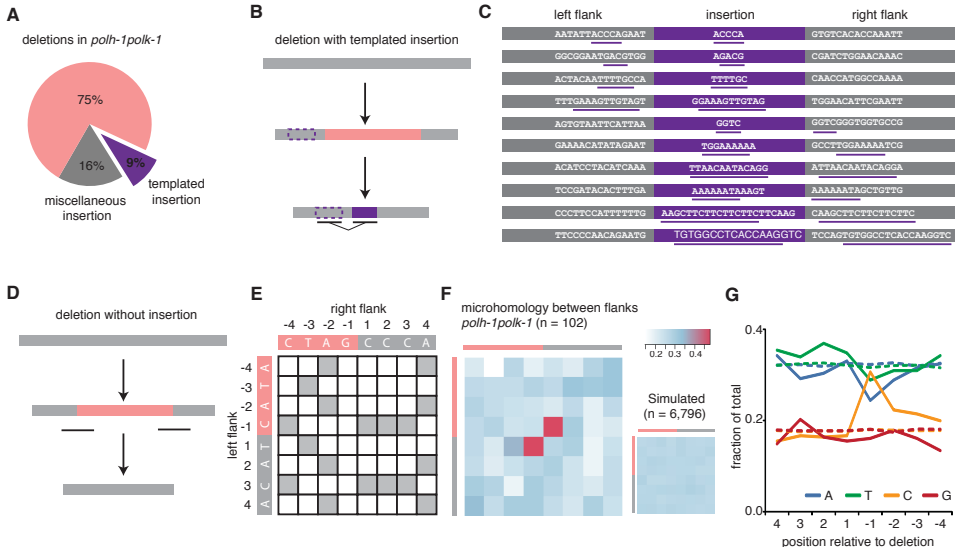


FIGURE 2. Deletion footprints in TLS mutants indicate a priming-based end joining mechanism. A. Distribution of deletion footprints in *polh-1polk-1* mutants. (B) Schematic illustration of a deletion associated with a templated insertion. Deleted sequence in pink; newly inserted sequence in purple and its template boxed; non-altered DNA in grey. **C.** Sequence context of deletions with templated insertions derived from *polh-1polk-1* animals. Matching sequences are underlined. **D.** Schematic illustration of a deletion not accompanied by insertions. Deleted sequence in pink; non-altered DNA in grey. The eight nucleotide window -capturing neighboring flanking and deleted sequences- that is used for the generation of the heat maps is underlined. **E.** The strategy to score junction homology: for each simple deletion, matching bases between the 5′ and 3′ junction were scored 1, non-matching bases were scored 0, thus creating one map per deletion. **F.** A heat map representing the sum of all individual deletion maps derived from *polh-1polk-1* animals. (n=102). A heat map for a simulated set of deletions (n=6796) with random distribution is displayed on the right. **G.** Base composition at the 5′ and 3′ junctions. The flanking sequences have positive numbers, the deleted sequence have negative; -1 being the first nucleotide within the deletion. Dotted lines indicate the relative abundance of a particular base for a simulated set of deletions (n=6796).

Polymerase θ mediates deletion formation

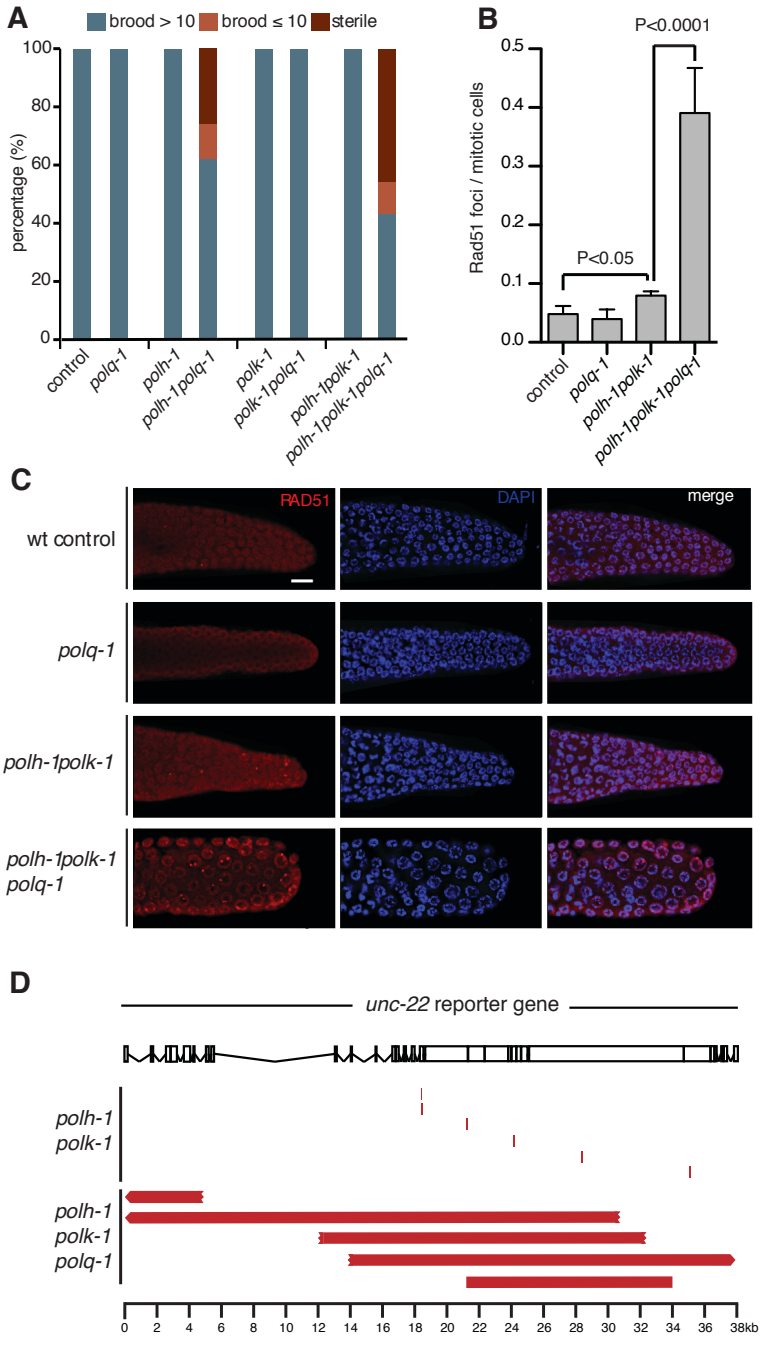
The footprints of the deletions that are suppressed by TLS polymerases fit best with a model in which one end of a DSB, induced at replication-blocking dG bases, is extended using the other end as a template, with just a single base-paired nucleotide as a primer (explaining both single nucleotide homology and templated insertions). In this model, templated inserts can be explained as the result of iterative rounds of annealing and extension (Figure S4).

We next wished to identify the polymerase responsible for inserting new nucleotides at breaks that result from TLS deficiency. One candidate is *C. elegans* Polymerase θ. Pol θ is a polymerase from the A-family, previously implicated in repair of interstrand crosslinks in various models (Shima *et al.* 2004; Muzzini *et al.* 2008; Yousefzadeh and Wood 2012), and repair of transposition-induced DSBs in Drosophila. Pol θ mutant mice display spontaneous genomic instability and increased radiosensitivity. The molecular function of this protein in these phenotypes is, however, largely unknown (Shima *et al.* 2004). In an independent study, we recently identified Pol θ in prevention of genomic instability at endogenous sequences that are able to fold into potentially replication blocking G-quadruplex structures (Koole et al, article submitted).

To test a possible role for Pol θ in deletion formation at spontaneous damage, we generated animals defective for *polh-1polk-1* and the *C. elegans* Pol θ homolog *polq-1*. Strikingly, these animals are severely compromised in normal growth: while *polq-1* and *polh-1polk-1* animals had nearly wildtype growth characteristics, *polh-1polk-1polq-1* triple mutant animals had very much reduced fertility, albeit in a stochastic fashion, ranging from complete sterility to brood sizes of 30 percent of wildtype levels (Figure 3A). Associated with these fertility defects, we observed a profound increase in the number of RAD-51 foci in the proliferative zone of the germline as well as activation of the DNA damage checkpoint (Figure 3B,C), suggesting increased DNA end-resection and DSB signaling, (Figure S3E). From this we conclude that when damage cannot be bypassed Pol θ action safeguards animal fertility by preventing undesired HR-related processing of replication-associated breaks, which would trigger checkpoint activation and prohibit proliferation.

To study the role of Pol θ in deletion formation on a molecular level, we assessed mutagenesis using an endogenous *unc-22* reporter gene (Figure 3D). We isolated spontaneous *unc-22* mutants from *polh-1polk-1* and *polh-1polk-1polq-1* populations

FIGURE 3. Pol θ mediates end joining of breaks in Pol η and Pol κ deficient animals. A. Fecundity of single, double and triple knockout mutants of Polymerase θ and TLS Polymerases η and κ. **B.** Quantification and (**C**) representative pictures of RAD-51 immunostainings on germlines of the indicated genotype. Scale bar, 10 μm **D.** Schematic representation of the *unc-22* reporter gene and spontaneous deletions (in red) isolated from either *polh-1polk-1* or *polh-1polk-1polq-1* mutant animals. Three out of five deletions extended beyond the borders of the *unc-22* locus.


A
brood > 10
brood ≤ 10
sterile
percentage (%)
100
80
60
40
20
0
control
polq-1
polh-1
polh-1polq-1
polk-1
polk-1polq-1
polh-1polk-1
polh-1polk-1polq-1
B
P<0.0001
P<0.05
Rad51 foci / mitotic cells
0.5
0.4
0.3
0.2
0.1
0.0
control
polq-1
polh-1polk-1
polh-1polk-1polq-1
C
RAD51
DAPI
merge
wt control
polq-1
polh-1polk-1
polh-1polk-1
polq-1
D
unc-22 reporter gene
polh-1
polk-1
polh-1
polk-1
polq-1
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38kb

and determined their molecular nature using PCR and Sanger sequencing. In perfect agreement with our whole-genome sequencing data, all *unc-22* mutations derived from *polh-1polk-1* animals were 50-200 bp deletions characterized by single nucleotide homology and templated insertions (Figure 3D, Table S2). In sharp contrast, *unc-22* mutants derived from *polh-1polk-1polq-1* triple mutants, while being induced at comparable rates, were of a completely different size category. Here, deletions were typically larger than 5 kb, with some spanning over 30 kb of genomic sequence, thus amplifying the deleterious impact of replication stalling lesions more than 100-fold (Figure 3D, Tables S2 and S3).

We conclude that a Pol θ-mediated end joining mechanism (TMEJ) is responsible for the small-sized deletions induced at replication fork stalling endogenous lesions. In its absence, large stretches of DNA surrounding DSBs are resected, resulting in abundant RAD-51 filament formation, mitotic checkpoint activation and a Pol θ-independent repair process accompanied by excessive loss of DNA. Reduced viability of the triple mutant may be the consequence of both processes: unscheduled checkpoint activation and loss of large genomic regions encoding essential genes.

TMEJ in wildtype *C. elegans* strains

The notion that we have uncovered TMEJ acting at spontaneous damage but only in TLS mutants, raises the question: do cells rely on TMEJ under TLS proficient conditions, or in other words, how relevant is TMEJ for animal fitness? We hypothesized that the action of an error-prone repair mechanism with such a clear and distinct signature, *i.e.* a distinct size class, single nucleotide homology and templated insertions, may leave its fingerprint in evolving genomes. For this reason, we compared the genomes of different natural isolates of *C. elegans*, to identify structural variations and defined their characteristics (Figure 4). The majority of deletions is of small size - 60 percent being smaller than 10 bp - while the number of deletions decreases with increasing size in an exponential manner. However, we found deletions in the size range 50-200 bp much more abundantly present than expected from this exponentially declining trend (Figure 4B), suggesting that the typical TMEJ-mediated deletions seen in TLS-deficient circumstances also contribute to mutation induction in wildtype natural strains (Figure 4B). Indeed, deletions in this size range bear the TMEJ signature including templated insertions and a strong overrepresentation (over 80 percent) of having at least one nucleotide homology (Figure 4C).Unexpectedly, we observed also templated insertions (2%) in the small size range of deletions, and found this class to also be dominated by ≥1 nucleotide homology at the junction (Figure 4C-D), which may argue that TMEJ is a very prominent source of mutagenesis in *C. elegans* evolution.

To further test the involvement of TMEJ in spontaneous mutation induction under non-challenged growth we used a forward mutagenesis assay that is based on the uncoordinated movement of animals carrying a dominant mutation (e1500) in the

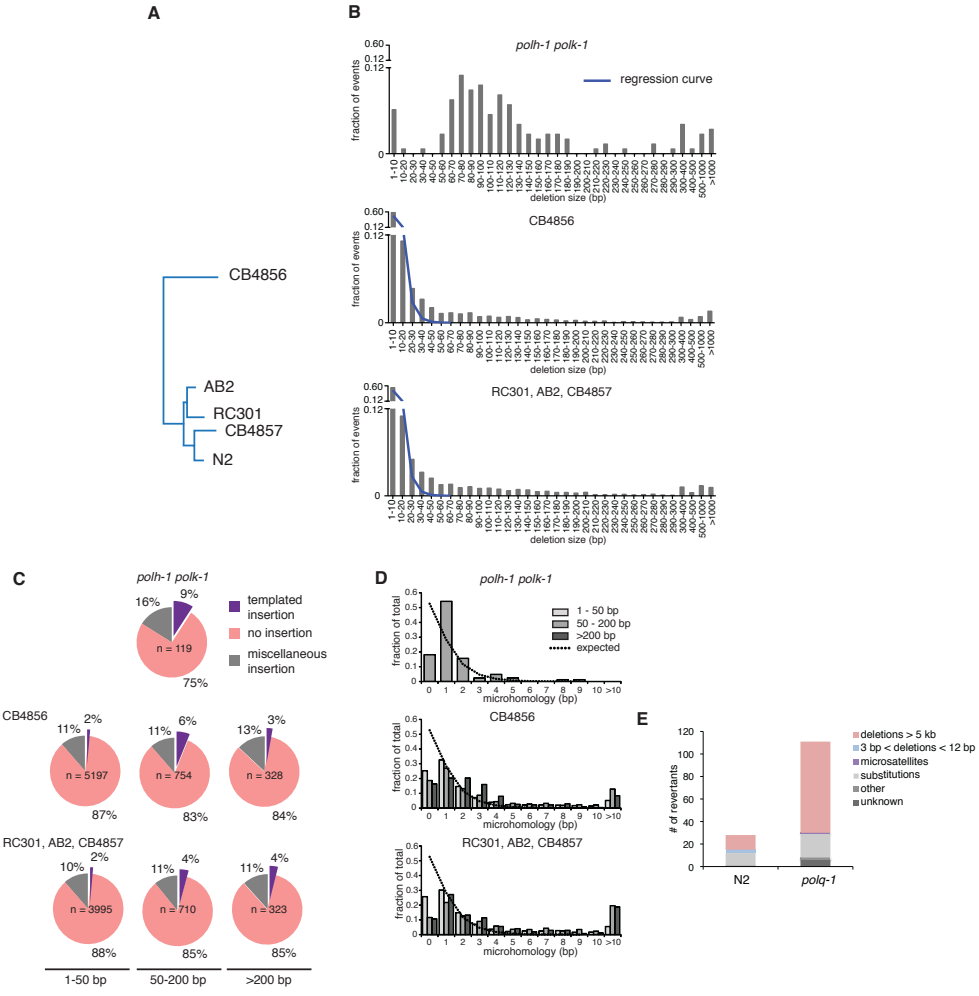


FIGURE 4. Signature of Pol θ - mediated end joining in natural isolates of *C. elegans*.
A. Phylogenetic tree diagram of the different isolates of *C. elegans* used in this study. **B.** Size distribution of deletions of evolutionary distinct *C. elegans* species compared to size distribution of *polh-1polk-1* derived deletions. An exponential regression curves describes the size distribution of deletions in both natural isolates up to 20 bp; deletions up to 300 bp are overrepresented. **C.** Deletions in natural isolates, especially in size class 50-300 bp show templated insertions analogously to deletion footprints in *polh-1polk-1* animals. **D.** Microhomology for deletions in natural isolates as compared to deletions in *polh-1polk-1* animals. **E.** *unc-93* mutagenesis in *polq-1* worms and wildtype controls.

UNC-93 protein that affects muscle contraction (De Stasio *et al.* 1997; Greenwald and Horvitz 2003). This easily recognizable phenotype is suppressed by complete loss of function of *unc-93,* or by loss of one of several extragenic suppressor genes (*e.g. sup-9*, *sup-10*). We propagated 400 populations of wildtype and *polq-1* mutant animals out of which we isolated and moleculary characterized revertants animals that had normal movement. Strikingly, the total number of revertants was increased fourfold in *polq-1* mutants, suggesting that in the absence of Pol θ, TMEJ substrates are shuttled into more mutagenic pathways (Figure 4E, Tables S4 and S5). Indeed, about 75 percent of all *polq-1*-deficient revertants were attributed to large chromosomal deletions in *unc-93* or the suppressor genes *sup-9* and *sup-10* (Figure S5); while *sup-9* and *sup-10* occupy less genomic space than *unc-93*, deletions at these sites dominate the overall spectrum, most likely because they, in contrast to *unc-93*, are located in a genomic region that is devoid of essential genes, and are therefore selective targets for large chromosomal deletions (Figure S5).

The selective increase of large deletions in *polq-1* mutants may replace mutations that are otherwise the products of TMEJ. Indeed, one class of mutations, *i.e.* small deletions of a size ≥3 bp, was exclusively found in wildtype animals (3/28 versus 0/111 in *polq-1*), suggesting that this class is the result of TMEJ action. Similar events in intronic and moreover in intragenic regions flanking these genes will give rise to large chromosomal deletions in the absence of Pol θ, explaining the increased mutagenesis in these strains.

TMEJ footprints in evolving genomes

In this study, we have identified a role for polymerase θ in preventing mitotic crisis and loss of DNA sequence at sites of stalled replication. Our data provide a mechanistic model in which Pol θ acts in an error-prone DSB repair pathway to extend one end of a DSB using the other end as a template for new DNA synthesis. This results in a stable association of both ends in the form of a canonical DNA helix, which can be further processed. In this respect, TMEJ provides an alternative to homologous recombination repair in which the sister chromatid is used as a template to generate new DNA that can subsequently be used to anneal to the other broken end. We hypothesize that TMEJ provides the cells with the ability to repair replication-associated breaks in cases where the sister chromatid cannot be used as a template because that still contains the original replication-blocking lesion (Figure S6).

The observation that Pol θ suppresses mutagenesis in wildtype animals, together with the notion that the signature of TMEJ is apparent in mutation profiles under laboratory conditions and in the genomes of natural isolates of *C. elegans* argues for a prominent role of this error-prone pathway to protect genomes against large chromosomal rearrangements.

While mutagenic processes are drivers of evolution, they also fuel malignant

transformation of cells. It is a current challenge to recognize specific classes of mutations in cancer genomes and attribute these either to underlying sources of DNA damage or to error-prone repair mechanisms. Identifying mutational signatures typifying specific repair processes is pivotal to this ambition. Templated insertions and the use of minimal homology - two characteristics of TMEJ - have frequently been observed in higher order eukaryotes and in cancer tissues (Chen *et al.* 2010; Nik-Zainal *et al.* 2012; Carvalho *et al.* 2013), and have been ascribed to either classical nonhomologous end-joining or the molecularly ill-defined process of microhomology-mediated end-joining (Honma *et al.* 2007; Kloosterman *et al.* 2012). Here, we describe a mechanistic alternative for repair of DSBs induced at stalled forks, which leaves a distinct and well-defined footprint in evolving genomes.

EXPERIMENTAL PROCEDURES

C. elegans genetics

All strains were cultured according to standard methods(Brenner 1974). Wildtype N2 (Bristol) worms were used in all control experiments. Alleles used in this study are: *polh-1 (lf31)*; *polh-1 (ok3317)*; *polk-1 (lf29)*; *polq-1(tm2026); msh-6(pk2504)*; *bcIs39[P(lim-7) ced-1::GFP + lin-15(+)]); unc-93(e1500)*. All mutant strains were backcrossed six times before performing experiments.

Whole genome sequencing of MA lines

Mutation accumulation (MA) lines were generated by cloning out F1 animals from one hermaphrodite. Each generation about five worms were transferred to new plates. MA lines were maintained for 60 generations or until severe growth defects developed. Single animals were then cloned out and propagated to obtain full plates for DNA isolation. Worms were washed off with M9 and incubated for one hour at room temperature while shaking, to remove bacteria from the animal's intestine. After two washes, worm pellets were lysed for two hours at 65°C with SDS containing lysis buffer. Genomic DNA was purified by using a DNeasy kit (Qiagen). Paired end (PE) libraries for whole genome sequencing (HiSeq2000 Illumina) were constructed from genomic DNA according to manufacturers' protocols with some adaptations. Shortly, 5 μg DNA was sheared using a Covaris S220 ultrasonicator, followed by DNA end-repair, formation of 3'A overhangs using Klenow and ligation to Illumina PE adapters. Adapter-ligated products were purified on Qiaquick spin columns (Qiagen) and PCR-amplified using Phusion DNA polymerase and barcoded Illumina PE primers for 10 cycles. PCR products of the 300 - 400 bp size range were selected on a 2% ultrapure agarose gel and purified on Qiaquick spin columns. DNA quality was assessed and quantified using an Agilent

DNA 1000 assay. Four to five barcoded libraries were pooled in one lane for sequencing on a HiSeq.

Bio-informatic analysis

Image analysis, basecalling and error calibration was performed using standard Illumina software. For the analysis of the natural isolates paired-end whole genome sequence data was downloaded from the NCBI Short Read Archive (SRP011413) (Grishkevich *et al.* 2012), and sequence reads were mapped to the *C. elegans* reference genome (Wormbase release 225) by BWA. Samtools was used for SNP and indel calling, with BAQ calculation turned off. All non-unique SNPs and indels are considered to be pre-existing and were filtered out using custom Perl scripts. To identify microsatellite mutations and deletions we used Pindel, developed by Ye et al (Ye *et al.* 2009).
A more detailed description of the bio-informatic procedures is enclosed in the supplemental information.

Microscopy

To study RAD-51 foci formation, germlines were dissected, freeze cracked and subsequently washed with 1% Triton and methanol (-20°C). RAD-51 was visualized by using an anti-RAD-51 rabbit monoclonal antibody and an Alexa488-labelled goat-anti-rabbit secondary antibody (Molecular Probes Inc), combined with 10 μg/mL DAPI. Dissected worms and eggs were mounted using Vectashield. Apoptosis was monitored using a *lim-7* driven *CED-1*::GFP fusion, which visualises sheath cells surrounding apoptotic germ cells. All microscopy was performed with a Leica DM6000 microscope.

unc-22 mutagenesis assay

To identify spontaneous mutations in the *unc-22* muscle gene we started 50 populations by transferring a single animal to 9 cm plates seeded with OP50. In the case of the synthetically sick *polh-1polk-1polq-1* mutant, we started 200 populations with 5 worms per plate. Animals were washed off with 2 mM levamisole and transferred to 6-well plates to facilitate scoring of *unc-22* mutants, which are insensitive to the hypercontracting effects of the drug levamisole. Independent *unc-22* mutant animals were isolated. Genomic DNA was isolated from homozygous animals for subsequent PCR and sequence analysis.

unc-93 (e1500) mutagenesis assay

To generate a complete spectrum of spontaneous mutations we used a mutagenesis assay based on reversion of the socalled 'rubber band' phenotype, caused by a dominant

mutation in the muscle gene *unc-93* (De Stasio *et al.* 1997; Greenwald and Horvitz 2003). Reversion of the *unc-93(e1500)* phenotype is caused by homozygous loss of *unc-93* or one of the suppressor genes *sup-9*, *sup-10*, *sup-11* and *sup-18*. *polq-1(tm2026) unc-93(e1500)* and *unc-93(e1500)* animals were singled to 2 x 400 6 cm plates. These plates were grown till starvation and equal fractions (chunks of 2 x 2 cm) were then transferred to 9 cm plates. Before these plates were fully grown, they were inspected for wildtype moving animals. From each starting culture only one revertant animal was isolated to ensure independent events.

Large chromosomal deletions in *unc-93*, *sup-9* and *sup-10* were identified by PCR amplification of exonic regions and two regions 5 kb upstream and downstream of the respective genes. Smaller genetic changes and substitutions were first classified into events in either the *unc-93* gene or in one of the suppressor genes by their ability to complement a known *unc-93* deletion allele. All *unc-93* exons were sequenced in revertant animals that failed to complement *unc-93*, whereas all exons of *sup-9* and *sup-10* were sequenced in revertants that complemented *unc-93*. *sup-11* or *sup-18* could not be subjected to molecular analysis due to lack of sequence data. Revertants that complemented *unc-93* but had not detectable mutation of *sup-9* or *sup-10*, were classified as 'unkown'.

ACKNOWLEDGEMENTS

We thank the *C. elegans* Knockout Consortium, Shohei Mitani and the *C. elegans* Genetics Center for providing strains. We thank the Leiden Genome Technology Center for support in whole genome sequencing. We thank Wouter Koole and Jane van Heteren for critical reading of the manuscript and Bennie Lemmens and Harry Vrieling for discussions. M. T. is supported by grants from the European Research Council (203379, DSBrepair) and ZonMW/NGI-Horizon.

REFERENCES

Avkin S., Goldsmith M., Velasco-Miguel S., Geacintov N., Friedberg E. C., Livneh Z., 2004 Quantitative analysis of translesion DNA synthesis across a benzo[a]pyrene-guanine adduct in mammalian cells: the role of DNA polymerase kappa. J Biol Chem **279**: 53298–53305.

Brenner S., 1974 The genetics of Caenorhabditis elegans. Genetics **77**: 71–94.

Burr K. L.-A., Velasco-Miguel S., Duvvuri V. S., McDaniel L. D., Friedberg E. C., Dubrova Y. E., 2006 Elevated mutation rates in the germline of Polκ mutant male mice. DNA Repair **5**: 860–862.

Carvalho C. M. B., Pehlivan D., Ramocki M. B., Fang P., Alleva B., Franco L. M., Belmont J. W., Hastings P. J., Lupski J. R., 2013 Replicative mechanisms for CNV formation are error prone. Nat Genet.

Chen J.-M., Cooper D. N., Férec C., Kehrer-Sawatzki H., Patrinos G. P., 2010 Genomic rearrangements in inherited disease and cancer. Semin. Cancer Biol. **20**: 222–233.

Choi J.-Y., Angel K. C., Guengerich F. P., 2006 Translesion synthesis across bulky N2-alkyl guanine DNA adducts by human DNA polymerase kappa. J Biol Chem **281**: 21062–21072.

De Stasio E., Lephoto C., Azuma L., Holst C., Stanislaus D., Uttam J., 1997 Characterization of revertants of unc-93(e1500) in Caenorhabditis elegans induced by N-ethyl-N-nitrosourea. Genetics **147**: 597–608.

Denver D. R., Dolan P. C., Wilhelm L. J., Sung W., Lucas-Lledó J. I., Howe D. K., Lewis S. C., Okamoto K., Thomas W. K., Lynch M., Baer C. F., 2009 A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc Natl Acad Sci USA **106**: 16310–16314.

Denver D. R., Feinberg S., Estes S., Thomas W. K., Lynch M., 2005 Mutation rates, spectra and hotspots in mismatch repair-deficient Caenorhabditis elegans. Genetics **170**: 107–113.

Fischhaber P. L., Gerlach V. L., Feaver W. J., Hatahet Z., Wallace S. S., Friedberg E. C., 2002 Human DNA polymerase kappa bypasses and extends beyond thymine glycols during translesion synthesis in vitro, preferentially incorporating correct nucleotides. J Biol Chem **277**: 37604–37611.

Gartner A., Milstein S., Ahmed S., Hodgkin J., Hengartner M. O., 2000 A conserved checkpoint pathway mediates DNA damage--induced apoptosis and cell cycle arrest in C. elegans. Mol Cell **5**: 435–443.

Greenwald I. S., Horvitz H. R., 2003 unc-93(e1500): A behavioral mutant of Caenorhabditis elegans that defines a gene with a wild-type null phenotype. Genetics **96**: 147–164.

Grishkevich V., Ben-Elazar S., Hashimshony T., Schott D. H., Hunter C. P., Yanai I., 2012 A genomic bias for genotype-environment interactions in C. elegans. Mol. Syst. Biol. **8**: 587.

Haracska L., Yu S. L., Johnson R. E., Prakash L., Prakash S., 2000 Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. Nat Genet **25**: 458–461.

Hastings P. J., Ira G., Lupski J. R., 2009 A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet **5**: e1000327.

Herman R. K., Kari C. K., Hartman P. S., 1982 Dominant X-chromosome nondisjunction mutants of Caenorhabditis elegans. Genetics **102**: 379–400.

Hile S. E., Wang X., Lee M. Y. W. T., Eckert K. A., 2011 Beyond translesion synthesis: polymerase {kappa} fidelity as a potential determinant of microsatellite stability. Nucleic Acids Research.

Honma M., Sakuraba M., Koizumi T., Takashima Y., Sakamoto H., Hayashi M., 2007 Non-homologous end-joining for repairing I-SceI-induced DNA double strand breaks in human cells. DNA Repair **6**: 781–788.

Johnson R. E., Prakash S., Prakash L., 1999 Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. Science **283**: 1001–1004.

Kloosterman W. P., Tavakoli-Yaraki M., van Roosmalen M. J., van Binsbergen E., Renkens I., Duran K., Ballarati L., Vergult S., Giardino D., Hansson K., Ruivenkamp C. A. L., Jager M., van Haeringen A., Ippel E. F., Haaf T., Passarge E., Hochstenbach R., Menten B., Larizza L., Guryev V., Poot M., Cuppen E., 2012 Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. CellReports **1**: 648–655.

Knobel P. A., Marti T. M., 2011 Translesion DNA synthesis in the context of cancer research. Cancer Cell Int. **11**: 39.

Kusumoto R., Masutani C., Iwai S., Hanaoka F., 2002 Translesion synthesis by human DNA polymerase eta across thymine glycol lesions. Biochemistry **41**: 6090–6099.

Lynch M., Sung W., Morris K., Coffey N., Landry C. R., Dopman E. B., Dickinson W. J., Okamoto K., Kulkarni S., Hartl D. L., Thomas W. K., 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. Proceedings of the National Academy of Sciences **105**: 9272–9277.

Masutani C., Araki M., Yamada A., Kusumoto R., Nogimori T., Maekawa T., Iwai S., Hanaoka F., 1999 Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. EMBO J **18**: 3491–3501.

Matsuda T., Bebenek K., Masutani C., Hanaoka F., Kunkel T. A., 2000 Low fidelity DNA synthesis by human DNA polymerase-eta. Nature **404**: 1011–1013.

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

Meier B., Barber L. J., Liu Y., Shtessel L., Boulton S. J., Gartner A., Ahmed S., 2009 The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in Caenorhabditis elegans. EMBO J **28**: 3549–3563.

Mitelman F., Johansson B., Mertens F., 2007 The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer **7**: 233–245.

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

Nik-Zainal S., Alexandrov L. B., Wedge D. C., Van Loo P., Greenman C. D., Raine K., Jones D., Hinton J., Marshall J., Stebbings L. A., Menzies A., Martin S., Leung K., Chen L., Leroy C., Ramakrishna M., Rance R., Lau K. W., Mudie L. J., Varela I., McBride D. J., Bignell G. R., Cooke S. L., Shlien A., Gamble J., Whitmore I., Maddison M., Tarpey P. S., Davies H. R., Papaemmanuil E., Stephens P. J., McLaren S., Butler A. P., Teague J. W., Jönsson G., Garber J. E., Silver D., Miron P., Fatima A., Boyault S., Langerød A., Tutt A., Martens J. W. M., Aparicio S. A. J. R., Borg Å., Salomon A. V., Thomas G., Børresen-Dale A.-L., Richardson A. L., Neuberger M. S., Futreal P. A., Campbell P. J., Stratton M. R., the Breast Cancer Working Group of the International Cancer Genome Consortium, 2012 Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell **149**: 979–993.

Ohashi E., Bebenek K., Matsuda T., Feaver W. J., Gerlach V. L., Friedberg E. C., Ohmori H., Kunkel T. A., 2000 Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. J Biol Chem **275**: 39678–39684.

Payen C., Koszul R., Dujon B., Fischer G., 2008 Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. PLoS Genet **4**: e1000175.

Roerink S. F., Koole W., Stapel L. C., Romeijn R. J., Tijsterman M., 2012 A Broad Requirement for TLS Polymerases η and κ, and Interacting Sumoylation and Nuclear Pore Proteins, in Lesion Bypass during C. elegans Embryogenesis. PLoS Genet **8**: e1002800.

Sarkies P., Reams C., Simpson L. J., Sale J. E., 2010 Epigenetic instability due to defective replication of structured DNA. Mol Cell **40**: 703–713.

Shima N., Munroe R. J., Schimenti J. C., 2004 The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. Mol Cell Biol **24**: 10381–10389.

Tijsterman M., Pothof J., Plasterk R. H. A., 2002 Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient Caenorhabditis elegans. Genetics **161**: 651–660.

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

Yousefzadeh M. J., Wood R. D., 2012 DNA polymerase POLQ and cellular defense against DNA damage. DNA Repair.

SUPPLEMENTARY INFORMATION

Bio-informatic analysis

Image analysis, basecalling and error calibration was performed using standard Illumina software. Alignments to the annotated sequence of *C. elegans* available at WormBase WS225 were performed by BWA. Samtools was used for SNP and indel calling, with BAQ calculation turned off. All non-unique SNPs and indels are considered to be pre-existing and were filtered out using custom Perl scripts. We considered a SNP to be real if at least 80% of the called bases were non-wildtype for SNPs that are covered ≥4 times. As a second filter at least one of MA lines of the same genotype should be called as wildtype: having a coverage ≥4 of which ≥80% was of wildtype nature according to pileup generated with mpileup. Sanger sequencing of predicted SNPs validated these criteria.

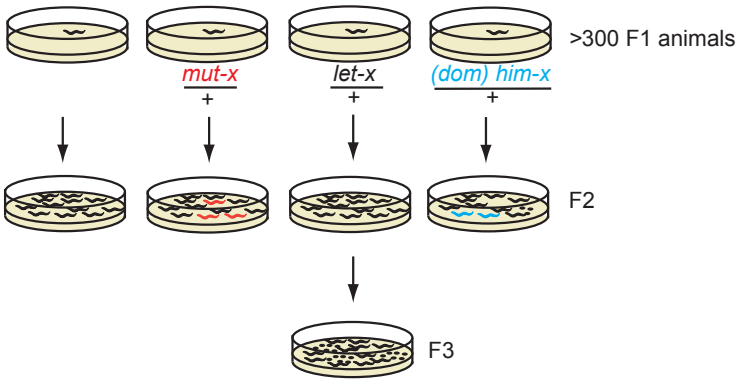
To identify microsatellite mutations and deletions we used Pindel, developed by Ye et al (Ye *et al.* 2009). Standard settings were used, except for changing the maximum allowed mismatch rate (-u) from 0.02 to 0.01. In addition, the sequencing error rate (-e) was adjusted to 0.0001. Pindel results were filtered using custom Java software that selects structural variations that are unique for one sample or the variation was supported by more than eight reads in one sample and eight times less often seen in any other sample. The latter criterium allowed us to also detect variations in homopolymers, which frequently contain sequencing errors. A deletion or deletion with template insertion of ≥20 nt was selected if the local average coverage was below five, while the directly surrounding area was covered >20 fold. Deletions and deletions with template insertions of <20 nt were included when both Pindel and samtools reported the variation and samtools reported it to be homozygous (≥80% non-wildtype). A representative set of deletions was validated by Sanger sequencing. The complete sequencing data have been submitted to the NCBI Short Read Archive (SRA) with accession ID SRP020555.

The simulated set was created in two steps. First, the N2 reference genome was edited by creating random deletions with a size distribution similar to *polh-1;polk-1* throughout the genome using a custom Perl script. Second, the software tool wgsim was used to generate pair-end reads with a 100x coverage based on the modified genome. Subsequently, the reads were analyzed using the same pipeline as used for the MA lines.

For the analysis of the natural isolates paired-end whole genome sequence data was downloaded from the NCBI Short Read Archive (SRP011413), and sequence reads were mapped to the *C. elegans* reference genome (Wormbase release 225). The average base coverage was 176x, 164x, 166x and 75x for AB2, CB4857, RC301 and CB4856, respectively. Pindel was used to detect structural variations (SVs) in the natural isolates as compared to the N2 reference genome. We included only SVs that had at least 10 reads supporting the SV and no reads supporting the reference genome. Next, we grouped

all SVs from AB2, CB4857, RC301 together as the phylogenetic tree indicate that they are closely related (as compared to CB4856. As a second criterion, we collected only those SVs that were N2-like in CB4856, but were flagged as a SV in the AB2, CB4857 or RC301 group (and vice versa). We found 6,279 and 5,028 deletions of at least 1bp for CB4856 and the combined group that consists of CB4857, AB2 and RC301, respectively. The regression curves displayed in figure 4 were created by iterating over the natural isolate datasets.

A



B

Genotype	# analyzed plates	Mutants found
N2	340	0
polh-1(ok3317)	340	0
polh-1(lf31)	340	0
polk-1(lf29)	340	0
polh-1(ok3317);polk-1(lf29)	740	*dpy(3); ste(1); let(15); him(6)*
polh-1(lf31);polk-1(lf29)	340	*dpy(3); let(5); him(3)*
msh-6	300	20 visible mutants

Figure S1. Occurrence of spontaneous visible mutants in TLS defective strains.
A. Experimental set-up to determine spontaneous mutagenesis: the F1 brood of non-mutant segregating hermaphrodites (P0) were singled to establish individual populations. These were inspected for mendelian segregation of abnormal phenotypes indicating the occurrence of a recessive mutations in the gametes of the P0. Mutants affecting body morphology (e.g. dumpy/dpy) or movement (i.e. uncoordinated/unc) can be scored in the F2 progeny. Mutations in essential genes (i.e. lethal/let) give rise to islands of dead eggs when populations are allowed to clear the food supply. Elevated numbers of males in the F2 progeny indicate a high incidence of males (him) phenotype, arguing for a dominant him mutation in the F1. **B.** Quantification of visible mutant phenotypes. The data for msh-6 mutants have been published previously (Tijsterman *et al.* 2002).

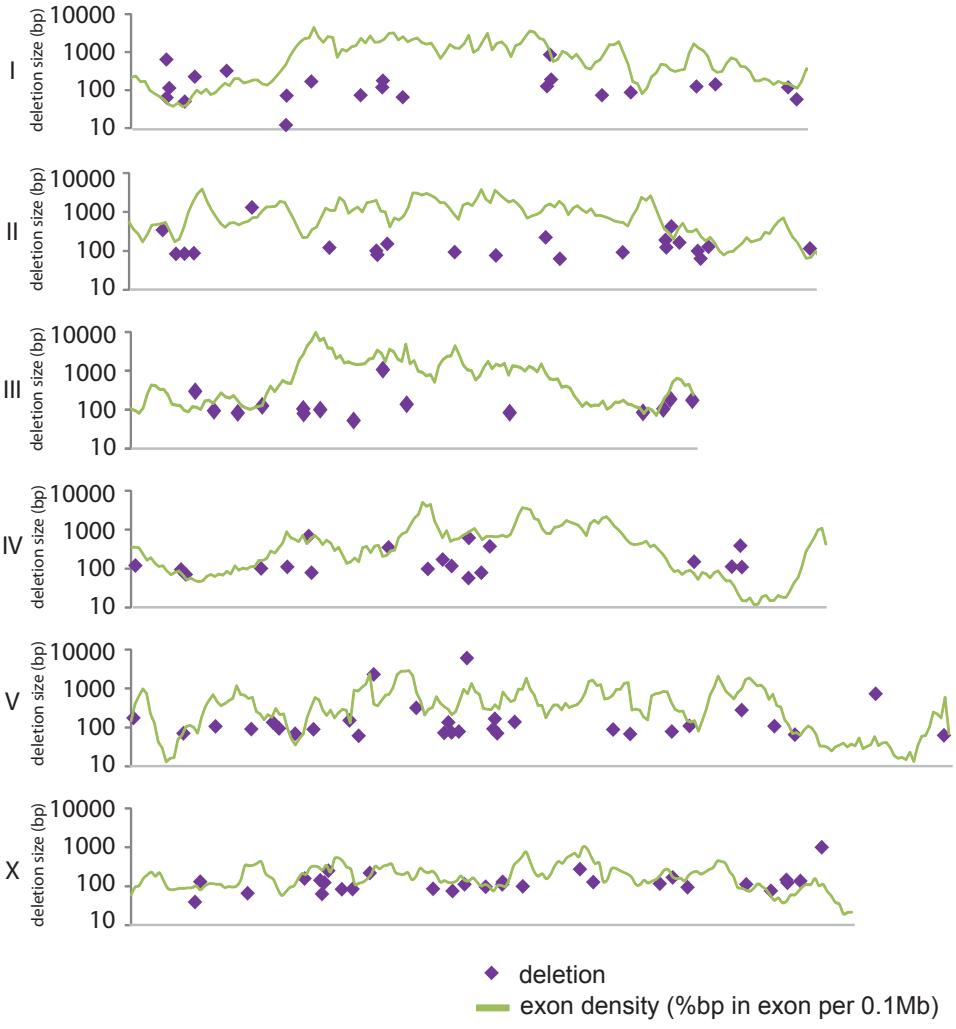


Figure S2. Genomic distribution of deletions in *polh-1polk-1* mutant animals.
Individual deletions (purple) were plotted onto a physical map of the *C. elegans* genome. The y-axisshows the size of the deletion on a logarithmic scale. The exon density is displayed in green (y-axis not shown). The length of the graph shows the size of the indicated chromosomes relative to each other.

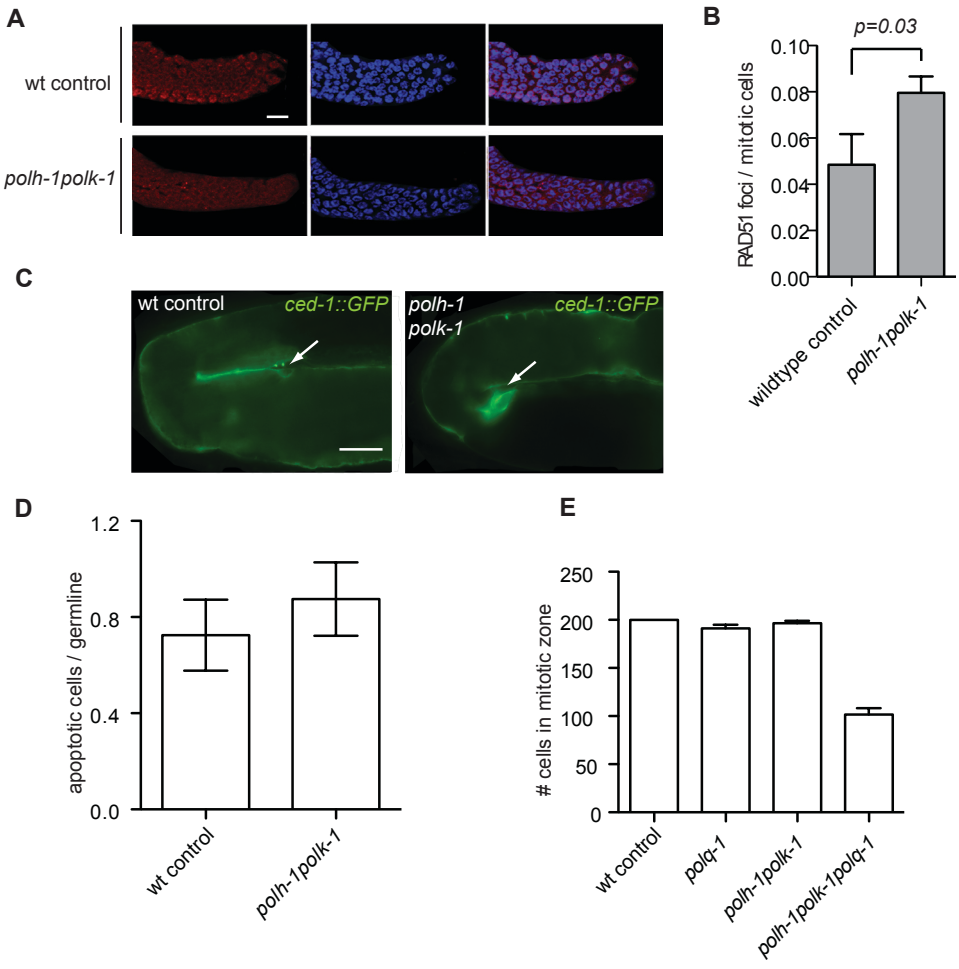


Figure S3. Analysis of checkpoint activation in the *C. elegans* germline.
A. Representative images and **B.** quantification of RAD-51 foci for the indicated genotype in nuclei present in the proliferative compartment of the *C. elegans* reproductive system. DAPI stainings in blue, RAD-51 in red. Scale bar, 10 μm **C.** Representative images of the bend of the gonad arm of animals transgenic for the apoptotic marker ced1::GFP; cells in the process of apoptotic engulfment are indicated with arrows. Scale bar, 10 μm **D.** Quantification of apoptotic cells in *polh-1polk-1* mutant animals and wildtype controls. **E.** Quantification of the number of nuclei in the mitotic region of the germline. A reduction in the number of cells in this region is an established outcome of checkpoint activation.

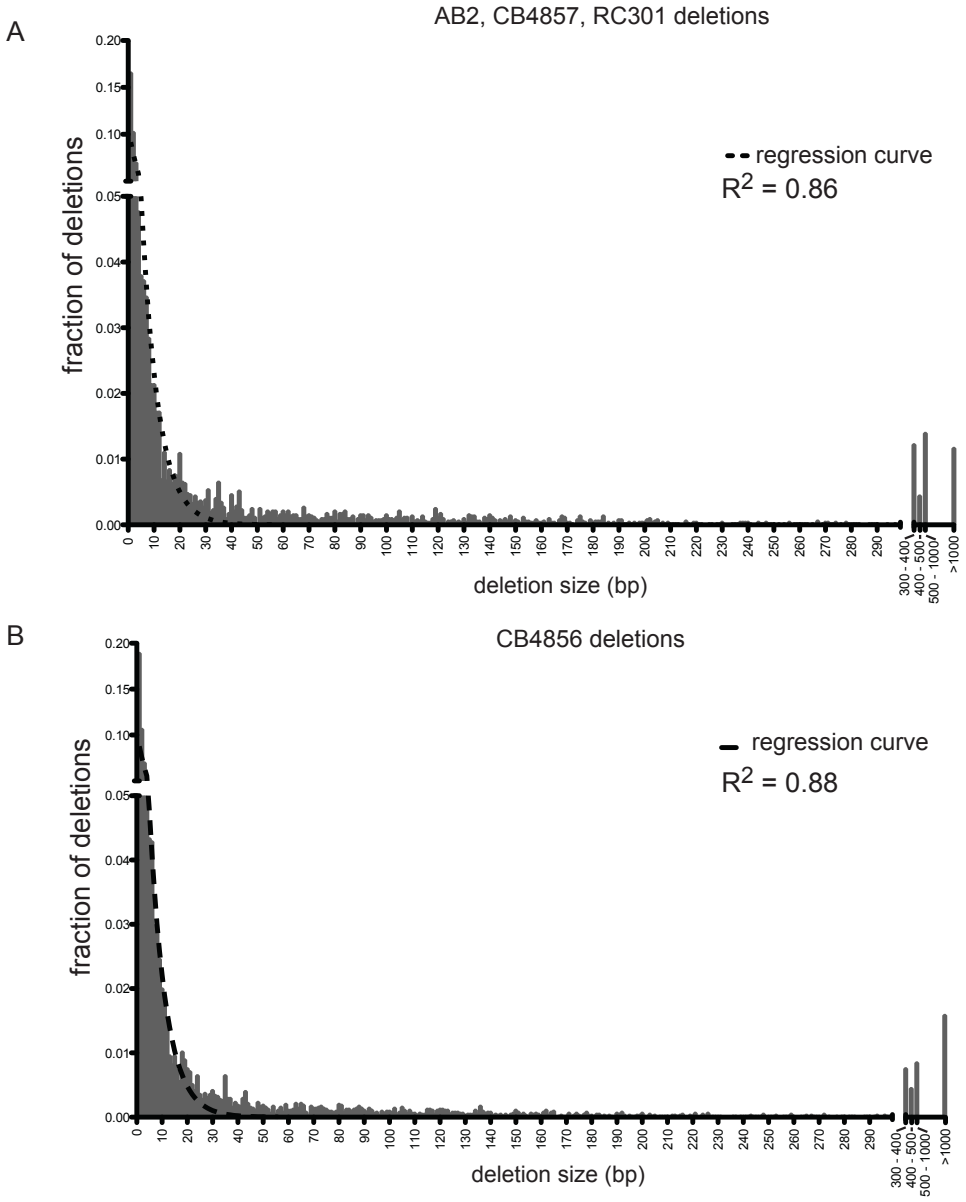
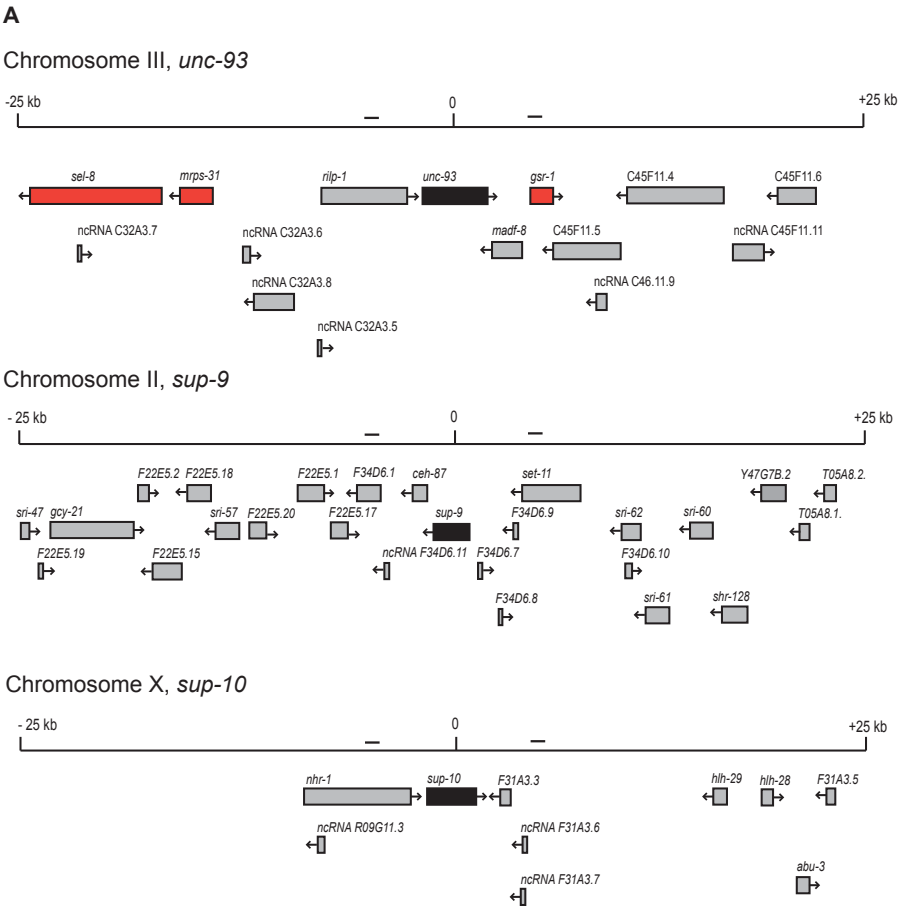


Figure S4. Histogram of size distribution of deletions in various *C. elegans* natural isolates that were analyzed.

Regression analysis showed that an exponential fit for deletion sizes up to 20bp approaches the actual distribution best. **A.** the grouped distribution for AB2, CB4857 and RC301. **B.** as in **A.** but now for CB4856.

A



B

	unc-93	*sup-9*	*sup-10*
wildtype	0	2	11
polq-1	6	20	55

Figure S5. Selective occurence of large chromosomal deletions in regions that are devoid of essential genes in the *unc-93* mutagenesis assay.
A. Schematic representation of 50 kb regions surrounding the *unc-93*, *sup-9* and *sup-10* genes. Known essential genes are depicted in red. While *unc-93* is flanked by two essential genes, no essential genes are known in the 50 kb intervals around *sup-9* and *sup-10*. To estimate deletion sizes, amplification of PCR products at -5kb and +5kb positions has been tested. **B.** Number of deletions larger than 5kb in *unc-93*, *sup-9* and *sup-10*.

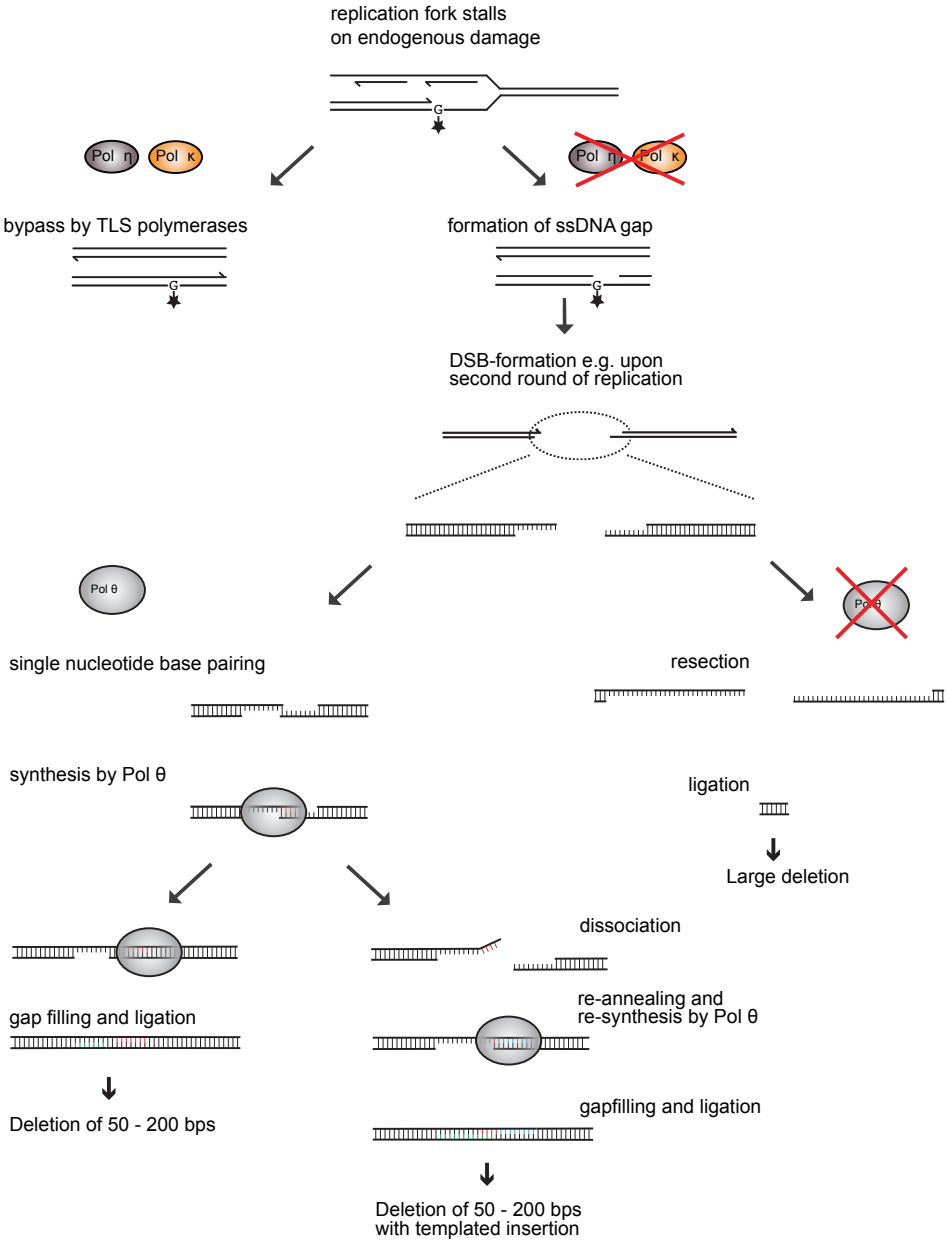


Figure S6. Proposed model for TMEJ of breaks induced at spontaneous replication fork barriers. Endogenous damage - spontaneously, or with increased frequency in the absence of functional TLS - causes replication fork blocks, leading to double stranded breaks. The broken ends can be joined by Pol θ-mediated end-joining (TMEJ) either in a continuous process, giving rise to deletions with a single nucleotide homology (left) or in subsequent steps, giving rise to deletions with templated insertions (right).

Table S1. Whole genome sequencing statistics.

genotype	sample	# generations	# reads	average coverage	# bp >= 4x covered
N2	N2	60	45,258,326	28x	100,140,732
	N4	60	23,693,826	16x	99,675,920
polh-1(lf31)	H7	60	46,203,688	39x	100,229,062
	H8	60	44,982,616	37x	100,238,324
polk-1(lf29)	K1	60	41,517,548	21x	99,970,233
	K4	60	39,275,458	30x	100,235,635
	K9	60	40,037,564	24x	100,120,773
polh-1(lf31);polk-1(lf29)	D4	32	46,284,780	21x	99,911,564
	D13	25	38,712,292	29x	100,224,845
	D14	25	59,163,976	27x	100,202,641
msh-6 (pk2504)	M13	10	48,338,722	19x	99,236,278
	M15	10	44,129,942	12x	99,799,729

Table S2. *unc-22* deletions in *polh-1polk-1* and *polh-1polk-1polq-1*.

	size	left flank	deletion left	deletion right	right flank	insertion
polh-1polk-1						
A	83 bp	GTACCTACTCA	CGTCCAAATG	TTATCGAAAA	GAACGTGTGC	-
B	74 bp	AATCCAGAAGT	CGATGACACC	CTTGGTTAGT	TATTTTTTGG	-
C	153 bp	ACAAGGCTGGG	CCTGGACAAC	TAAAGGCTGG	AGCCACTGTT	-
						AATCTGACTATCAAAGGAAATC
D	119 bp	GACTATCAAGG	CTGGTCAATC	TGATAACCCA	GAATACCAAT	TCAAGAATCTGACTATCAAAG
E	93 bp	CTTGCAAAGGA	TCCATTTGGA	CACGTGACAA	CGGTGGATCA	-
F	71 bp	TGTGAAGCCTT	ACGGAACTGA	ACCACCAGTT	GTTACTTGGC	-
G		not identified				
polh-1polk-1 polq-1						
A	>4.7 kb					
B	>30.5 kb					
C	19 - 20.6 kb					
D	12660 bp	AAATGAGCACA	CTATTCTGTG	GAACAGGAGC	ATTTGGAGTT	
E	> 23.7 kb					
F		not identified				

Table S3. Frequency of *unc-22* mutations in *polq-1*, *polh-1polk-1* and *polh-1polk-1polq-1*.

Strain		total # plates scored	# plates containing one or more twitchers	estimated mutation rate
N2	wildtype	40	0	0.00E+00
XF152	*polq-1*	40	0	0.00E+00
XF507	*polh-1polk-1*	46	7	8.00E-06
XF840	*polh-1polq-1polk-1*	39	6	8.00E-06

Table S4. Sequence analysis of reversion mutants for *unc-93(e1500)*.

wildtype			
unc-93			
deletions > 5kb	0		
substitutions	6	cagttt(g>a)tctggc; C>Y gacacg(t>a)cacagt; V>D tgtctg(g>c)aatact; G>A aaatat(c>t)gatttt; R>L ggaatc(g>a)cggctt; T>A tgttag(g>t)taatgg; splice	
other	1	gaatat(tcga>deleted)aaaactt	3bp > deletion > 12 bp
sup-9			
deletions > 5kb	2		
substitutions	1	ccattg(g>a)gactta; G>stop	
other	2	ccaata(gtga>deleted)cgtcat tctgta(ccgggtgggga>deleted)ggtctg	3bp > deletion > 12 bp 3bp > deletion > 12 bp
sup-10			
deletions > 5kb	11		
substitutions	3	cagttc(t>a)cttgta; L>H tggaat(a>g)tggtcgg; M>V* agccag(g>t)tttgta;; splice site mutation	
unknown	2		

*also tctttt(t>c)caacca in intron 150 bp upstream

Table S5. Sequence analysis of reversion mutants for *polq-1; unc-93(e1500)*.

polq-1			
unc-93			
deletions > 5kb	6		
substitutions	12	tgcgga(c>a)aagtcg; Q>K cgttga(c>a)gattttc; T>K gatctc(g>a)gatctg; G>R ttccat(c>t)atttat; S>L tttcta(c>a)ctcatg; T>N tttcat(g>t)attgta; M>I ggggag(c>a)caaatg; A>D aagtcg(t>a)cggaaa; V>D tccttt(c>t)gagaca; R>stop tctata(c>a)attgtc; Y>stop aatata(t>a)ttgctg; Y>stop tgttag(g>a)taatgg; splice site mutation	
other	2	acgtca(ca>deleted)gttgaa ttttac(t>deleted)ttttag	other microsatellite
sup-9			
deletions > 5kb	20		
substitutions	7	tcttcg(g>a)gctcac; G>E gggtac(c>a)agtgga; Q>K gtggag(c>a)atttta; A>E ccattg(g>a)gactta; G>stop aggcta(c>a)ggtcat; Y>stop tccctg(c>t)aaactc; Q>stop caagta(c>a)aacatg; Y>stop	
sup-10			
deletions > 5kb	55		
substitutions	1	atgtta(a>t)tataag; N>I	
other	1	gtgatg(a>deleted)catcaa	hairpin
unknown	7		