Universiteit Leiden
The Netherlands

# Roadblocks & bypasses : protection of genome stability by translesion DNA synthesis in C. elegans

Bostelen, I. van

**Citation**

Bostelen, I. van. (2019, December 3). *Roadblocks & bypasses : protection of genome stability by translesion DNA synthesis in C. elegans*. Retrieved from https://hdl.handle.net/1887/81315

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/81315 |

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

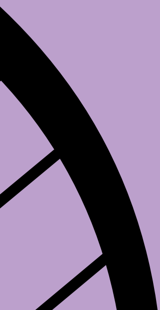
The handle http://hdl.handle.net/1887/81315 holds various files of this Leiden University dissertation.

**Author**: Bostelen, I. van
**Title**: Roadblocks & bypasses : protection of genome stability by translesion DNA synthesis in C. elegans
**Issue Date**: 2019-12-03

# 3

# Suppression of genome instability by Y-family TLS polymerase REV-1 in *C. elegans*

Ivo van Bostelen, Robin van Schendel, Sophie Roerink & Marcel Tijsterman



## Abstract

Translesion synthesis (TLS) polymerases promote the replication of damaged DNA through their ability to catalyze nucleotide addition to growing DNA chains at sites where DNA lesions block the replicative polymerases. By this action, TLS polymerases allow cells to complete the duplication of damage-containing genomes and thus prevent checkpoint activation, genome instability and cell death. The price to pay is mutation induction because TLS polymerases have, in comparison to replicative polymerases, reduced fidelity and are thus generally considered to be pro-mutagenic. In this study we have used *C. elegans* to determine the contribution of the Y-family polymerase REV-1 on long-term stability of an animal genome. Surprisingly, we found REV-1 to both stimulate and suppress spontaneous mutagenesis during unperturbed propagation. By stimulating bypass REV-1 prevents the persistence of ssDNA gaps that are converted to small deletions by alternative end joining of ensuing double-strand breaks. Thus, opposite of what is the current dogma, the action of REV-1 during unperturbed growth is predominantly anti-mutagenic: it prevents the accumulation of deletions at the cost of less detrimental SNVs. In addition, we found the level of spontaneous lesions that depend on REV-1 action to be surprisingly low: only 1 lesion in ~100 rounds of genome replication, which equates to 1 block per ~$10^{10}$ bases, requires REV-1 action. Our findings augment the concept that ensuring replication progression outweighs near-perfect conservation of genetic information in animal cells.

## Introduction

Although mutagenesis is a prerequisite for evolution, spontaneous mutations are also life threatening as they are at the basis of inborn diseases and age-related pathologies like cancer. To suppress these detrimental effects several mechanisms have evolved to prevent the occurrence of mutations. For example, during the copying of DNA, the combined action of exonuclease activity of the replicative polymerases, which removes erroneously incorporated nucleotides during synthesis, and the mismatch repair (MMR) pathway, which repairs mismatched bases after the fact, provides an estimated 10.000-fold increase in copying accuracy [1,2]. Apart from replication errors, so-called spontaneous mutations can result from replicating damaged DNA caused by endogenous processes in the cell. For example, oxidative metabolites can react with DNA, damage bases, and in that way hamper replication [3,4]. Efficient and unperturbed DNA synthesis is essential for survival since stalling of replication can lead to collapse of the replication fork followed by formation of highly toxic and mutagenic DNA double stranded breaks (DSBs) that may result in genomic rearrangements or cell death. A number of pathways have evolved to remove potential replication blocks and repair the DNA, including base excision repair (BER) and nucleotide excision repair (NER), where newly synthesized DNA replaces the damaged DNA at the site of the potential replication block [4,5]. Additionally, damage tolerance pathways have evolved to allow for continuation of the cell cycle in the presence of DNA lesions. A well-studied mechanism to tolerate DNA damage is translesion synthesis (TLS). While replicative polymerases stall at damaged bases, specialized TLS polymerases are able to synthesize DNA opposite of these blocks. Lesions can be bypassed directly when the replicative polymerase is temporarily switched with a TLS polymerase at the replication fork during S-phase, or single strand DNA gaps at the site of lesions remain and bypass and gap filling occurs after S-phase [6].

In eukaryotes TLS is mediated by Y-family polymerases Polη, Polκ, Polι and Rev1 and the B-family polymerase Polζ composed of catalytic subunit Rev3 and regulatory subunit Rev7. These TLS polymerases lack proofreading activity and have wide catalytic centers to allow for replication across damaged bases and DNA synthesis from misaligned primer termini. These characteristics cause TLS polymerases to have lower fidelity then replicative polymerases, making them inherently error prone [7,8]. Whereas some types of lesions require only a specific Y-family polymerase, other types require the sequential action of a two or more TLS polymerases [9-11]. Two modes for TLS have been proposed. The first acts on lesions that have low impact on the DNA helix structure and are still ‘readable’. For these lesions the default TLS polymerase Polη is suggested, mostly resulting in error-free bypass. The second mode of action is required for lesions that are harder to read and strongly distort the DNA helix. In that case the concerted efforts of one or more Y-family polymerases in combination with the Rev1

dependent activity of Polζ are required. This second mechanism is thought to be much more mutagenic [6,12].

REV1 has a similar structure as the other members of the Y-family, Polη, Polκ and Polι, but its catalytic activity is limited to the incorporation of deoxycytidine (dC) residues [7]. *In vitro*, REV1 was shown incorporate dCs across undamaged or damaged guanines but also opposite adenines, uracil and abasic sites [13-15]. *In vivo*, REV1 plays a role in bypass of lesions that are caused by lipid peroxidation [16,17]. Various studies have demonstrated that REV1 plays non-catalytic roles via interactions with other proteins. The BRCT domain of REV1 interacts with PCNA and is involved in the bypass of UV-C induced lesions in mouse ES cells [18-20]. The C-terminal part of mammalian REV1 contains ubiquitin binding motifs (UBMs) that can interact with ubiquitinated proteins like PCNA-K164Ub, a region able to interact with other Y-family polymerases, and a motif that interacts with the REV7 subunit of B-family TLS polymerase Polζ [7,21-24] (Suppl. fig. 1). REV1 could act as a master regulator of TLS, instead of acting as a true polymerase: via its interactions with PCNA, Y-family polymerases and Polζ, REV1 may provide a 'molecular scaffold' that is central to TLS [6,12]. REV1 might also have functions outside of canonical TLS since research in avian DT40 cells has shown a role for this polymerase in maintenance of epigenetic stability at G-quadruplex structures, possibly by facilitating replication through these hard-to-replicate secondary structures [25,26]. Finally, a yet unexplained role for REV1 (together with Polζ) in homologous recombination (HR) break repair has been described [27,28].

The action of TLS polymerases results in base substitutions, which may appear a detrimental process for cells, but an occasional base substitution may outweigh the deleterious consequences of complete replication fork blocks. Unrepaired breaks can result from collapsed replication and lead to cell cycle arrest and cell death [29,30]. Although the molecular details of translesion synthesis become more and more understood it remains unclear how TLS action affects genome maintenance or influences spontaneous mutagenesis either positively or negatively, on a genome-wide scale. The model system *C. elegans* is well suited to address this question through whole genome sequencing (WGS) revealing the diverse types of mutations, such as base substitutions, insertions/deletions (indels) or large genomic rearrangements that accumulate over generations. Here, we report that the TLS polymerase REV-1 safeguards replicative potential and genomic stability that are threatened by spontaneous DNA lesions.

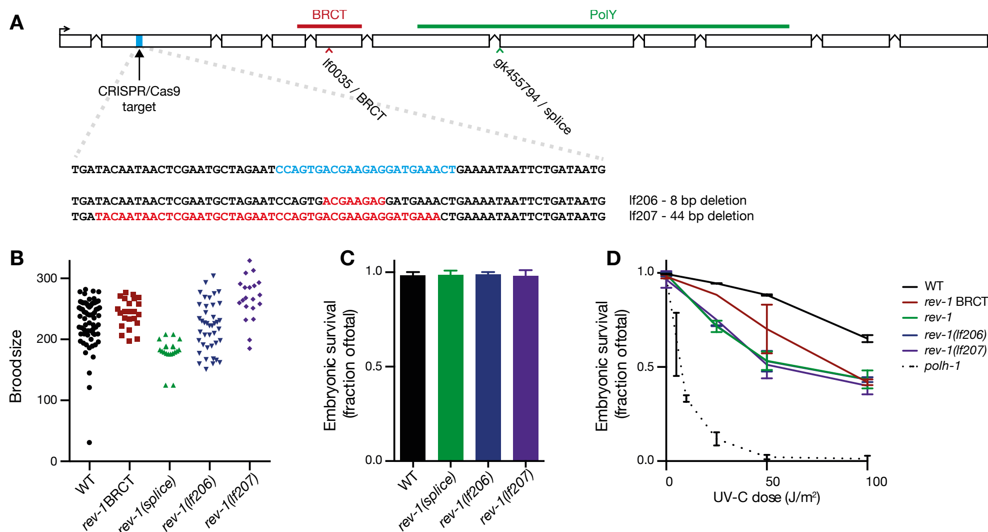



***Figure 1. Generation and characterization of rev-1 alleles.*** *A) Schematic representation of the rev-1 ORF. Exons are to scale, introns are not. In blue the CRISPR/Cas9 sgRNA target is indicated with the two generated knock-out alleles below. The red bar indicates the BRCT domain with the location of the point mutation (lf0035/BRCT) below. The green bar indicates the polymerase domain with the splice site mutation (gk455794) below. B) Brood size quantification of the different rev-1 mutant alleles. Each data point represents the total brood of one animal. C) Quantification of embryonic lethality of the same broods as in B. D) Embryonic survival in response to UV-C exposure of the indicated genotypes.*

# Results

## Generation and characterization of the *rev-1* alleles

To study the role of REV-1 in maintenance of genome stability we obtained several loss-of-function alleles. The allele most used in this study is *rev-1(gk455794)*, generated in the million mutation project [31], has a point mutation in the acceptor splice site of exon 7 that generates a potential truncated product lacking the polymerase domain and C-terminal part. Although it is likely that this mutant is a null allele because it lacks the whole Y-family polymerase domain and other C-terminal parts, we also generated two mutants targeting exon 2 of the rev-1 gene via the CRISPR/Cas-9 method [32-38]. We independently isolated two small genomic deletions at the CRISPR targeted site that both lead to frame shifts running into early stops (Fig. 1A). We consider the alleles *rev-1(gk455794)*, *rev-1(lf206)* and *rev-1(lf207)* to be null alleles and will refer to *rev-1(gk455794)* simply as *rev-1* from now on. Using a different technology [39], we also isolated a mutation causing a G283>D amino acid substitution in the highly conserved G283 residue of the BRCT domain: the *C. elegans* G283 residue aligns to G193 in yeast and

G76 in mice Rev1 (Suppl. fig. 1). The widely studied G193>R, G193>D and G193>V mutants in yeast, and the G76>R mutant in mice were shown to abolish the functionality of the BRCT domain [21,40,41]. We refer to this allele as *rev-1BRCT*.

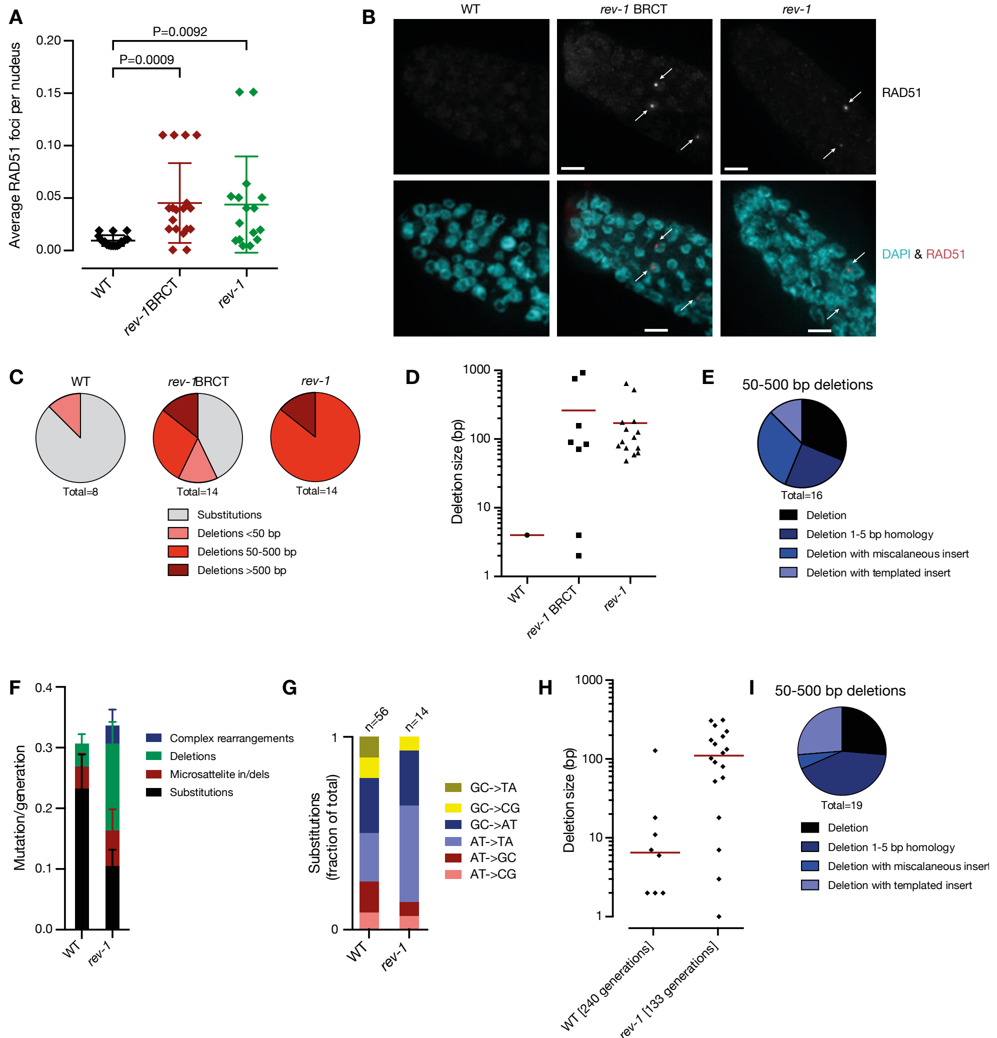
None of the *rev-1* mutants showed brood sizes that were substantially different from wild type (WT) controls (Fig. 1B). Also, *rev-1* knock-out mutants showed similar levels of very marginal embryonic lethality as WT controls (Fig. 1C). Both these observations show that REV-1 is not essential for proliferation under unchallenged conditions. The hypersensitivity to UV-C-induced DNA lesions observed in REV1 deficient yeast and MES cells suggests a conserved role for REV1 in TLS of photoproducts [18,42,43]. To test whether this is true for *C. elegans* REV-1 we exposed young adult worms to different doses of UV-C light and determined the embryonic lethality of their brood. We established that *rev 1* mutants have reduced embryonic survival as compared to WT controls in response to UV-C exposure but not to the extent as we showed for mutants defective in Polη, a TLS polymerase previously implicated to be essential in protection against UV induced damage [44,45] (Fig. 1D). A similar degree of sensitivity is observed in all three different *rev-1* knockout mutants; the *rev-1BRCT* hypomorph shows an intermediate phenotype, which is consistent with a partial loss of function.

### ***rev-1* and *rev-1BRCT* mutants show increased levels of spontaneous DSBs**

When TLS is impaired replication forks can collapse at sites of damaged bases and form DSBs. In the mitotic compartment of the *C. elegans* gonad DSBs can be visualized by staining for RAD51 foci. RAD51 foci in the mitotic compartment in the distal tip of the gonadal arms are rare events in healthy WT controls. However, after damage induction or as a result of spontaneous DNA damage in TLS mutants the number of foci increases while the brood size and embryonic lethality remain similar to WT control [46,47]. Both *rev-1* and *rev-1BRCT* mutants have significantly increased levels of spontaneous RAD51 foci in the mitotic compartment of the gonad, arguing that REV-1 suppresses the formation of DSBs ($p<0.01$; Fig 2A, B).

### **REV-1 protects against the formation of genomic deletions and rearrangements larger than 50 base pairs.**

While *rev-1* mutants displayed elevated levels of spontaneous DSBs no effect was observed on proliferation (Fig 1B, C & Fig. 2A), arguing for repair of RAD51-coated DSBs. Because repair of DSBs can be mutagenic, we asked whether the observed increase of spontaneous breaks affects mutation induction in *rev-1* mutants. To answer this we made use of an established mutagenesis assay; the *unc-93* reversion assay [48,49]. Animals carrying the toxic allele *unc-93(e1500)* have very poor capacity to move and also grow slowly. A mutation that kills the protein via mutation of an essential amino acid or via disruption of the ORF will lead to reversion to a WT-like phenotype




3

***Figure 2. Spontaneous DNA damage and mutagenesis in REV-1 deficient animals**. A) Quantification of spontaneous RAD51 foci in the mitotic compartment in the distal tip of the gonadal arms. B) Representative images of RAD51- and DAPI-stained distal tips of gonads. C) Mutations found in unc-93 assay in the different genetic backgrounds. D) Size of the deletions found the unc-93 assay. E) Properties of the 50-500 bp deletions found the unc-93 assay in both rev-1 mutants. F) The frequencies of the different categories of mutations accumulated over generations in WT and rev-1 mutants. G) SNP distribution in the mutation accumulation (MA) data. H) Size distribution of the deletions from MA data, excluding microsatellite indels. I) Characteristics of the 50-500 bp deletions in the MA data. See supplemental table 1 for sequence data.*

allowing for efficient selection of spontaneous mutants. Mutations in the suppressor genes sup-9, sup-10, sup-11 or sup-18 will also revert animals to a WT-like phenotype. We isolated 30 revertants for each genotype and subsequently sequenced the entire unc-93 locus. In WT, *rev-1BRCT* and *rev-1* backgrounds we found 8, 14, and 14 causative mutations in *unc-93(e1500)*, respectively. The other revertants likely carry mutations in one of the suppressor genes, but these genes were not sequenced. In the WT background, 7 of the 8 mutations are single nucleotide variations (SNVs) that disrupt gene function via amino acid substitution, introduction of an early stop or loss of a splice site. One deletion allele of 4 bp was found. Similar base substitutions and deletions of few bases were also found in the *rev-1BRCT* mutant, but 6 of the 14 mutations are larger deletions (>50 bp). In worms lacking any REV-1 activity all causative mutations were found to be >50 bp in size (Fig. 2C, D). From this data we conclude that the BRCT domain of REV-1 is involved in bypass of endogenous lesions and that REV-1 has an important role in spontaneous mutagenesis by suppressing the formation of small genomic deletions.

**REV-1 dependent TLS results in base substitutions but protects the genome from genomic deletions and rearrangements**

While the *unc-93* reversion assay is an established method to study mutagenesis it has a few drawbacks: it is limited to the genomic locus of *unc-93* and the only mutations that will be found are those that disrupt gene function. Deletions, for example, are more likely to disrupt gene function then SNVs, creating a bias towards detecting deletions over SNVs and making it difficult to measure mutations frequencies. In order to study spontaneous mutagenesis in an unbiased way and to allow us to determine mutation frequencies accurately we performed WGS of animals of populations that were grown for multiple generations. In WT animals we find on average ~0,23 base substitutions per animal generation, ~0,04 microsatellite indels and ~0,04 other small deletions. While the frequency of microsatellite indels remains more or less the same in the *rev-1* mutant the frequency of substitutions is halved to ~0,11 per generation and the frequency of deletion formation increases 3,5-fold to ~0,14 per generation. Additionally, few larger complex genomic rearrangements like tandem duplications and inversions occur in the *rev-1* mutant animals; these were not observed in WT animals (Fig. 2F). The distribution of substitutions in the *rev-1* mutant compared to the distribution in WT does not show marked difference (Fig. 2G); the variations are indicative of experimental variation in the small set (n=14) of substitutions that were found in the *rev-1* mutant. In addition to the striking increase in the frequency of non-microsatellite deletions there is also a marked difference in the size of the deletions between those observed in WT and *rev-1*: deletions of 50-500 bp are now the prominent category, while these are very rare in WT animals (Fig. 2H).

Assuming that lesions that absolutely require REV-1 for bypass will result in a deletion or complex rearrangement in the absence of REV-1, and estimating the number of cell divisions needed to go from one generation to the next based in the extensive knowledge on *C. elegans* cell lineage [50], we find that the occurrence of spontaneous lesions that depend on REV-1 action is surprisingly low: only 1 lesion in ~100 rounds of genome replication, which equates to 1 block per ~$10^{10}$ bases, requires REV-1 action.

**Loss of REV-1 does not lead to G-quadruplex instability in *C. elegans***

A potential role in replication of sequences that can form G-quadruplexes was described for REV1 in DT40 cells [25,26]. We tested whether *C. elegans* REV-1 is required for the maintenance of genome stability at G-quadruplexes as has been described previously for *C. elegans* FANCJ/DOG-1 [51,52]. We first searched for the presence of G-quadruplex motifs within the deleted sequence in *rev-1* mutant animals in the data from the *unc-93* and MA assays, but found none, arguing that spontaneous mutation induction is at sites of base damage and not on structured DNA. For further conformation, we tested if loss of functional REV-1 affected the nature and frequencies of G-quadruplex deletions that are formed in animals that lack the G-quadruplex resolving helicase DOG-1. In these animals G-quadruplex are more likely to constitute a block to replication [51,52]. However, we did not observe any significant difference on frequency and deletion size between *dog-1* and *dog-1; rev-1* mutant animals studying the endogenous G-quadruplex site qua1466, a very potent inducer of deletions in *dog-1* animals (Suppl. fig. 2).

3

**Absence of REV-1 leads to substrates for TMEJ/Polθ**

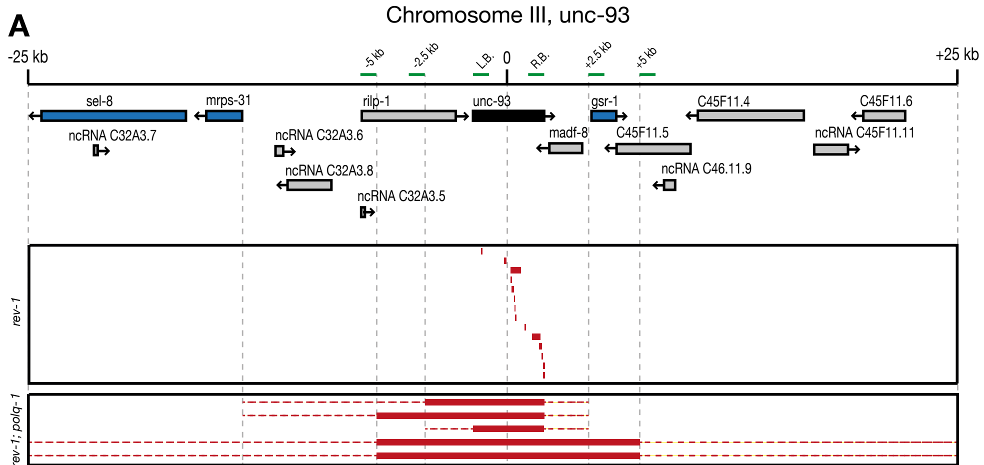
The size distribution of the spontaneous deletions that form over generations when lacking REV-1 activity is the same as for those found in TLS mutants lacking POLH-1 and POLK-1 [47]. This mutational footprint results from Polθ mediated end-joining (TMEJ) of replication associated DSBs and is further characterized by microhomology in the break sites and occasional inserts homologous to sequences flanking the deletion [47,52,53]. When we study the 50-500 bp deletions obtained in MA assay in the *rev-1* mutant we find similar characteristics: 36% shows the use of microhomology at the break sites and 23% contain templated inserts that originate from the flanks of the breaks (Fig 2I), comparable results are observed in the *unc-93* assay (Fig. 2I). In order to verify that TMEJ repairs spontaneous DSBs in *rev-1* mutants we employed the *unc-93* reversion assay in animals deficient for REV-1 and Polθ *(rev-1;polq-1)* and compared the repair footprints to the mutations found in the *rev-1* mutant. For the lager deletions the exact break points could not be determined by PCR and sequencing because all amplicons covering the unc-93 gene were deleted. We estimated the size of the large deletions by the loss of the amplicons in the flanks of unc-93 and found that all *rev-1;polq-1* revertants carry large deletions spanning several kb. Two of the five deletions extend into the neighboring

lethal genes indicting a deletion size >10 kb (Fig. 3A). We conclude that in a *rev-1* deficient worm small deletions form induced by stalled replication forks at endogenous lesions via TMEJ. To substantiate our conclusion that Polθ acts on substrates that result from the inability to bypass damaged DNA we tested Polθ dependency also in response to known bona fide replication blocks *i.e.* UV-C-induced photoproducts. We found that loss of just Polθ does not sensitizes animals to UV-C but combined loss of Polθ and REV-1 *(rev-1;polq 1)* results in a significantly higher sensitivity than the *rev-1* single mutant (Fig. 3B). This finding reinforces the notion that replication blocking lesions require TLS polymerases to prevent larger genomic insult, such as deletions resulting from DSB repair.

## Discussion

We here show that REV-1 safeguards survival in *C. elegans* upon exposure to UV light but is not needed under unchallenged conditions in the laboratory environment. During unperturbed growth, spontaneous DNA lesions arise that are bypassed by REV-1 dependent TLS, which can result in base substitutions. Without REV-1 these lesions cause persistent replication blocks that lead to the formation of small genomic deletions ranging from 50 to 500 bp in size and to the formation of larger complex genomic rearrangements. The characteristics of the deletions are consistent with TMEJ of replication associated DSBs. Accordingly, loss of TMEJ activity in a REV-1 deficient background further sensitizes animals to UV-C induced replication blocks and completely alters the mutational footprint of spontaneous replications block: the observed deletions become at least several kilobases in length, analogous to the mutational consequences of other persistent replication blocks in TMEJ deficient *C. elegans* [47,52].

Rev1 was named after its ‘reversion-less’ phenotype in yeast: in rev1 mutants UV-C induced mutagenesis is considerably lower than in WT [42,54]. Similar function has been assigned to human REV1 in response to UV induced DNA damage [43,55] and in mouse cells [19,56]. In line with the mutagenicity of REV1 dependent TLS we observe a reduction in base substitution frequency in the *rev-1* mutant. If the deoxycytidine transferase activity of REV-1 would be responsible for a substantial part of substitutions, *e.g.* the incorporation of dCs opposite abasic sites, REV-1 activity would be responsible for AT>GC, AT>CG and GC>CG substitutions. There are minor reductions in AT>GC and AT>CG substitutions, however we also see a loss of GC>TA substitutions in the *rev-1* mutant. Therefore, it is unlikely that the polymerase activity of REV-1 is responsible for a mayor part of base substitutions in WT animals. We propose that the lower levels in substitutions in the *rev-1* mutants are caused by a generally lower TLS efficiency of a mutagenic sub-pathway, due to the loss of the regulatory function of REV-1 in TLS. In previous studies it was reported that Polη and Polκ perform relatively error-free TLS



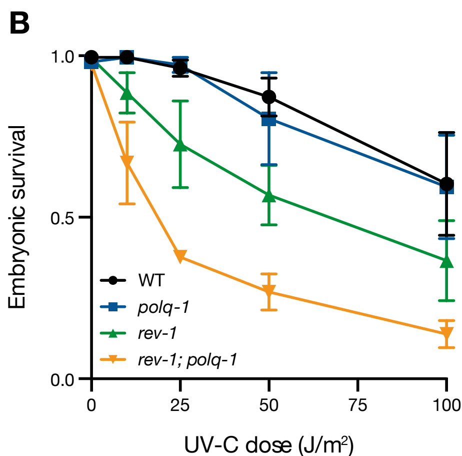




***Figure 3. In absence of REV-1 unresolved replication blocks lead to TMEJ substrates.*** *A) The genomic region surrounding the unc-93 locus on chromosome III; essential neighboring genes in blue, the unc-93 gene in black, and other genes in grey. PCR amplicons used to estimate the size of the large deletions are indicated as green bars. The genetic backgrounds are indicated. Deletions of which the junctions were defined by PCR and sequencing are indicated with red boxes. For sizable deletions the minimally deleted sequence established by PCR is indicated by a red box; the upper limit by red dotted lines. The bottom two revertants in the rev-1;polq-1 background were wild-type moving but sterile indicating that one or more essential genes were lost in addition to unc-93. B) Embryonic survival in response to UV-C exposure of the indicated genotypes.*

on lesions that are more abundant than those that require REV-1: loss of POLH-1 and POLK-1 induced ~2 deletions per generation, which is 14-fold more than observed in the *rev-1* mutant [44]. An explanation could be that REV-1 is required for a subset of lesions that needs the concerted efforts of multiple TLS polymerases: Polη or Polκ performs insertion of one or a few bases opposite the damage followed by REV-1 dependent extension of the aberrant primer terminus by Polζ or possibly Polκ. It is thought this complex sub-pathway of TLS induces more base substitutions [6].

It is interesting that REV1 deficient mouse embryonic fibroblasts, yeast and human fibroblast cell lines have no proliferative problems, yet Rev1-/- mice develop poorly arguing that the consequences of REV1 loss more prominently manifest on an

organismal level [19,42,43,56]. Unresolved replication blocks in Rev1 mutant mice may become a more serious problem due to accumulation of genomic rearrangements or cell death in tissues harboring cycling cells leading to loss of proliferative potential and ageing. The fact that we observe no proliferative defect in worms may be explained by the relatively high tolerance of worms to DNA damage and by relatively small size of the *C. elegans* genome resulting in less replication stress. Our observations suggest that the levels of spontaneous DSBs that form in REV-1 deficient worms are low and efficiently repaired. As a result, the level of replication-associated DSBs are insufficient to induce noticeable proliferative defects.

While it is widely established that REV1 acts in TLS, some studies have described roles in other pathways. Okada *et al.* and Sharma *et al.* describe involvement of REV1 in DNA repair via homologous recombination (HR) [27,28]. We cannot formally rule this out but have found no experimental support for a role for *C. elegans* REV-1 in HR; defective HR results in reduced brood size and high embryonic lethality [57]. Additionally, the spontaneous mutations we find in *rev-1* mutant genomes are consistent with the footprints found resulting from unresolved replication blocks in TLS mutants *polh-1* and *polh 1; polk-1*, that still have functional REV-1. Also, we have not found a role for *C. elegans* REV-1 in replicating through G-quadruplex structures as was observed in DT40 chicken B lymphocyte cells [25,26]. Both these additional functions could be reserved for vertebrate REV1 and dependent on C-terminal parts of the protein that are lacking in yeast and *C. elegans* REV-1. We do show a conservation of function of the BRCT domain from mice to worms. Jansen *et al.* describe that in mammalian cells the BRCT domain of Rev1 was important for bypass of UV-C induced lesions [18]. Here we confirm that loss of a functional BRCT domain in REV-1 sensitizes *C. elegans* for UV-C and we show a novel role of this domain in the bypass of endogenous lesions.

Based on our NGS data we conclude that during unchallenged *C. elegans* growth the number of spontaneous lesions that for their bypass depend on REV-1 action is remarkably low: in *rev-1* mutants we found a reduction of SNVs of ~0.1 per generation (each generation has at least 10-15 cell divisions), arguing that REV-1 action produces ~1 mutation per $10^{10}$ bases. Strikingly, we found a matching increase in the number of deletions in the absence of REV-1. This unexpected outcome argues that REV-1 action is mostly mutagenic during TLS of spontaneous base damage: if *e.g.* 99% of REV-1 dependent TLS would be error free, then we would have expected the number of deletions to go up manifold in a *rev-1* mutant background. The matching levels of SNV reduction and deletion increase is most simply explained by assuming that only 1 lesion per $10^{10}$ bases absolutely depend on REV-1 to be bypassed, while in the absence of REV-1 such lesions will constitute a persistent block to replication, leading to DSBs that require TMEJ. An important inference from our data is that although REV-1-dependent TLS introduces base substitutions, it protects the genome against the formation of

small genomic deletions and larger complex rearrangements and thereby helps to maintain genome stability. Because the latter outcomes are per case more detrimental to the integrity of our genetic information than SNVs, our study leads to the notion that the action of REV-1 during unperturbed growth is predominantly anti-mutagenic, challenging the current dogma that TLS action is harmful for accurate transmission of our genetic information. While it is so, its action prevents a far greater harm *i.e.* deletions and genomic rearrangements that have a much more prominent disturbing effect. Our findings thus support the notion that ensuring replication progression outweighs near-perfect replication of genetic material.

## Methods

3

### *C. elegans* genetics

All strains were cultured according to standard methods [58]. The N2 Bristol strain was used as WT control. The alleles *polq-1(tm2026), rev-1(gk455794), unc-93(e1500)* were obtained from the Caenorhabditis Genetics Center, Minnesota, USA. The *rev-1(BRCT)* allele was isolated in our own lab via a random mutagenesis approach described in [39]. The *rev-1(lf206)* and *rev-1(fl207)* KO alleles were obtained via CRISPR/Cas9 mediated genome editing as described in [32-37] and below in further detail.

### CRISPR/Cas9 mediated generation of rev-1 K.O. alleles

For CRISPR/Cas9 mediated targeting we used the following sequence (which includes the PAM site) in exon 2 of rev-1: AGTTTCATCCTCTTCGTCACTGG. Cloning in the appropriate vectors was done as described in [32-37]. Plasmids were injected using standard *C. elegans* microinjection procedures. Briefly, 1 day before injection, L4 animals were transferred to new plates and cultured at 15 degrees. Gonads of young adults were injected with a solution containing 20 ng/µl, pDD162 (Peft-3::Cas9, Addgene 47549; ref. [59]), 20 ng/µl pMB70 (u6::sgRNA with rev-1 target, 10 ng/µl pGH8, 2.5 ng/µl pCFJ90 and 5 ng/µl pCFJ104. Progeny (F1) animals that express mCherry were picked to new plates 3–4 days post injection and allowed to produce offspring. Of each F1 plate 10 F2 animals were pooled, lysed and genotyped. Genotyping was done by PCR amplification of a 480 bp product around the CRISPR/Cas9 target site. Subsequent restriction with MaeIII enzyme of the WT sequence would result in 2 fragments (91 bp + 389 bp). A mutation at the site of the predicted break site would likely disrupt the MaeIII recognition site resulting in an uncut PCR product. We isolated 4 alleles, 2 of which were small out of frame deletions. (Fig.1)

### Brood size and embryonic lethality assay

To determine the brood size, we singled L4 animals on OP50 plates. Every day, we

transferred the mother to a new plate and one day later quantified the number of embryos and larvae on the plate. We quantified the broods and embryonic lethality of at least 20 animals per genotype.

### UV-C survival assays

To measure germline sensitivity to UV, staged young adults (one day post L4) were transferred to empty NGM plates and exposed to different doses of UV-C. Per dose and genotype 3 plates with 3 adults were set up on fresh NGM plates with OP50 and allowed to lay eggs for 24 hours. Subsequently adults were discarded and the brood on the plate was allowed to hatch. 24 hours later the number of non-hatched eggs and surviving progeny was determined.

### RAD51 antibody staining, imaging and quantification

Animals were synchronized by picking L4 stage worms 22h before dissection. Worms were dissected in Egg buffer with 0.1% Tween and Levamisole to expose gonads. Most of the buffer was removed and the sample with cover glass was transferred to a Superfrost Plus slide and flash frozen on a metal block in dry ice. Upon complete crystallization of the sample the cover glass was quickly removed (freeze-cracking) followed by post-fixation in 4% PFA. After washing the samples were incubated with RAD51 antibody (rabbit polyclonal from SDIX/Novus Biologicals, cat# 29480002, used at 1:1000 in PBST+0,5%BSA) overnight at room temperature. Alexa anti-rabbit 488, 1:500, was used as secondary antibody and incubated for 2h at room temperature. DNA was visualized by DAPI staining and the slides were finished with Vectashield. Imaging and processing were done on a Leica DM6000 B microscope. The data obtained (Fig. 2A, B) are from at least 3 independent experiments. Each data point represents an average value of the mitotic zone of one gonad.

### Mutation accumulation lines and whole genome sequencing

Mutation accumulation (MA) lines were established by transferring single F1 animals that originated from a single parent, starting 6 clonal MA lines. Of each line three worms of the next generations were transferred to a new plate, marking every generation. MA lines were propagated for approximately 50 generations. At the end of MA timeline, of each line single animals were transferred to new plates and propagated to obtain full clonal plates for DNA isolation. Worms were washed off in M9 and to remove as much bacteria as possible from the intestines the worms wore incubated on a shaker for 2h at RT. Subsequently, genomic DNA was isolated using a Blood and Tissue Culture Kit (Qiagen). Whole genome sequencing and bio-informatics were performed as described in [53].

**G-quadruplex stability on qua1466**

Qua1466 is a genomic sequence [GGGAGGGCGGGCGGG] with genomic location IV:11326500..11326514 (build WBcel235), that can potentially form a G-quadruplex structure. To assay genomic instability and the formation of deletions at this site we perform a nested PCR reaction on lysed animals using the following primers: external forward CAAATAAGTATTGGGCCGAAACC; external reverse AAGGAACACCTTCAAGACTCC, internal forward CTGCGAACTTCTGACGAATTTG, internal reverse TTGACTCCTCCTCTTCTGGC. As template for the external PCR 1 µl of a 15 µl lysis with 5 worms was used. 0.5 µl of the external PCR was used as template for internal PCR. 10 µl of internal PCR product was run for 1 hour at 120V on a 1% agarose gel.

**unc-93(e1500) mutagenesis assay**

To pick up spontaneous mutations in the *rev-1(gk455794)* and *rev-1(gk455794);polq-1(tm2026)* backgrounds, we used a mutagenesis assay based on reversion of the so-called "rubber band" phenotype, caused by a dominant mutation in the muscle gene unc-93 [48,49]. Reversion of the *unc-93(e1500)* phenotype is caused by homozygous loss of *unc-93(e1500*) or one of the suppressor genes sup-9, sup-10, sup-11, and sup-18. For both genotypes 400 animals were singled to 9 cm plates. These plates were grown until starvation and of each plate an equal amount (chunks of 2 x 2 cm) were transferred to fresh 9 cm plates. Before these plates reached starvation, they were inspected for wild-type moving animals. From each starting culture, only one revertant animal was isolated to ensure independent events. Of each genetic background we randomly selected 30 revertants and sequenced the *unc-93* gene. When large deletions occurred that deleted the amplicons used for sequencing *unc-93*, we established the approximate size of the deletion with PCR amplicons of approximately 500 bp located at the borders of the gene and 2.5 kb and 5 kb up and downstream of the *unc-93* gene. The locations of these amplicons are indicated with green bars in figure 3A.

## Acknowledgements

Several strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs . This work is supported by the European Research Council (ERC) and the Netherlands Organization for Scientific Research (NWO).

# References


[1] McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res 2008;18:148–61. doi:10.1038/cr.2008.4.

[2] Fishel R. Mismatch repair. Journal of Biological Chemistry 2015;290:26395–403. doi:10.1074/jbc.R115.660142.

[3] Barnes DE, Lindahl T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet 2004;38:445–76. doi:10.1146/annurev.genet.38.072902.092448.

[4] Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell 2010;40:179–204. doi:10.1016/j.molcel.2010.09.019.

[5] Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nature Reviews Molecular Cell Biology 2014;15:465–81. doi:10.1038/nrm3822.

[6] Jansen JG, Fousteri MI, de Wind N. Send in the clamps: control of DNA translesion synthesis in eukaryotes. Mol Cell 2007;28:522–9. doi:10.1016/j.molcel.2007.11.005.

[7] Prakash S, Johnson RE, Prakash L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. Annu Rev Biochem 2005;74:317–53. doi:10.1146/annurev.biochem.74.082803.133250.

[8] Knobel PA, Marti TM. Translesion DNA synthesis in the context of cancer research. Cancer Cell Int 2011;11:39. doi:10.1186/1475-2867-11-39.

[9] Johnson RE, Washington MT, Haracska L, Prakash S, Prakash L. Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. Nature 2000;406:1015–9. doi:10.1038/35023030.

[10] Prakash S, Prakash L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. Genes & Development 2002;16:1872–83. doi:10.1101/gad.1009802.

[11] Gan GN, Wittschieben JP, Wittschieben BØ, Wood RD. DNA polymerase zeta (pol zeta) in higher eukaryotes. Cell Res 2008;18:174–83. doi:10.1038/cr.2007.117.

[12] Jansen JG, Tsaalbi-Shtylik A, de Wind N. Roles of mutagenic translesion synthesis in mammalian genome stability, health and disease. DNA Repair 2015;29:56–64. doi:10.1016/j.dnarep.2015.01.001.

[13] Nelson JR, Lawrence CW, Hinkle DC. Deoxycytidyl transferase activity of yeast REV1 protein. Nature 1996;382:729–31. doi:10.1038/382729a0.

[14] Haracska L, Unk I, Johnson RE, Johansson E, Burgers PM, Prakash S, et al. Roles of yeast DNA polymerases delta and zeta and of Rev1 in the bypass of abasic sites. Genes & Development 2001;15:945–54. doi:10.1101/gad.882301.

[15] Zhang Y, Wu X, Rechkoblit O, Geacintov NE, Taylor J-S, Wang Z. Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. Nucleic Acids Res 2002;30:1630–8. doi:10.1093/nar/30.7.1630.

[16] Zhou Y, Wang J, Zhang Y, Wang Z. The catalytic function of the Rev1 dCMP transferase is required in a lesion-specific manner for translesion synthesis and base damage-induced mutagenesis. Nucleic Acids Res 2010;38:5036–46. doi:10.1093/nar/gkq225.

[17] Wiltrout ME, Walker GC. The DNA polymerase activity of Saccharomyces cerevisiae Rev1 is biologically significant. Genetics 2011;187:21–35. doi:10.1534/genetics.110.124172.

[18] Jansen JG, Tsaalbi-Shtylik A, Langerak P, Calléja F, Meijers CM, Jacobs H, et al. The BRCT domain of mammalian Rev1 is involved in regulating DNA translesion synthesis. Nucleic Acids Res 2005;33:356–65. doi:10.1093/nar/gki189.

[19] Jansen JG, Tsaalbi-Shtylik A, Hendriks G, Gali H, Hendel A, Johansson F, et al. Separate domains of Rev1 mediate two modes of DNA damage bypass in mammalian cells. Mol Cell Biol 2009;29:3113–23. doi:10.1128/MCB.00071-09.

[20] Guo C, Sonoda E, Tang T-S, Parker JL, Bielen AB, Takeda S, et al. REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. Mol Cell 2006;23:265–71. doi:10.1016/j.molcel.2006.05.038.

[21] Guo C, Fischhaber PL, Luk-Paszyc MJ, Masuda Y, Zhou J, Kamiya K, et al. Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. Embo J 2003;22:6621–30. doi:10.1093/emboj/cdg626.

[22] Kikuchi S, Hara K, Shimizu T, Sato M, Hashimoto H. Structural Basis of Recruitment of DNA Polymerase by Interaction between REV1 and REV7 Proteins. Journal of Biological Chemistry 2012;287:33847–52. doi:10.1074/jbc.M112.396838.

[23] Wojtaszek J, Liu J, D'Souza S, Wang S, Xue Y, Walker GC, et al. Multifaceted recognition of vertebrate Rev1 by translesion polymerases $\zeta$ and $\kappa$. Journal of Biological Chemistry 2012;287:26400–8. doi:10.1074/jbc.M112.380998.

[24] Wojtaszek J, Lee CJ, D'Souza S, Minesinger B, Kim H, D'Andrea AD, et al. Structural Basis of Rev1-mediated Assembly of a Quaternary Vertebrate Translesion Polymerase Complex Consisting of Rev1, Heterodimeric Polymerase (Pol) , and Pol. Journal of Biological Chemistry 2012;287:33836–46. doi:10.1074/jbc.M112.394841.

[25] Sarkies P, Reams C, Simpson LJ, Sale JE. Epigenetic instability due to defective replication of structured DNA. Mol Cell 2010;40:703–13. doi:10.1016/j.molcel.2010.11.009.

[26] Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau M-N, et al. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. Embo J 2014;33:2507–20. doi:10.15252/embj.201488398.

[27] Okada T, Sonoda E, Yoshimura M, Kawano Y, Saya H, Kohzaki M, et al. Multiple Roles of Vertebrate REV Genes in DNA Repair and Recombination. Mol Cell Biol 2005;25:6103–11. doi:10.1128/MCB.25.14.6103-6111.2005.

[28] Sharma S, Hicks JK, Chute CL, Brennan JR, Ahn J-Y, Glover TW, et al. REV1 and polymerase $\zeta$ facilitate homologous recombination repair. Nucleic Acids Res 2012;40:682–91. doi:10.1093/nar/gkr769.

[29] Sale JE. Translesion DNA synthesis and mutagenesis in eukaryotes. Cold Spring Harb Perspect Biol 2013;5:a012708. doi:10.1101/cshperspect.a012708.

[30] Ghosal G, Chen J. DNA damage tolerance: a double-edged sword guarding the genome. Transl Cancer Res 2013;2:107–29. doi:10.3978/j.issn.2218-676X.2013.04.01.

[31] Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, et al. The million mutation project: a new approach to genetics in Caenorhabditis elegans. Genome Research 2013;23:1749–62. doi:10.1101/gr.157651.113.

[32] Chiu H, Schwartz HT, Antoshechkin I, Sternberg PW. Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. Genetics 2013;195:1167–71. doi:10.1534/genetics.113.155879.

[33] Cho SW, Lee J, Carroll D, Kim J-S, Lee J. Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 2013;195:1177–80. doi:10.1534/genetics.113.155853.
[34] Katic I, Großhans H. Targeted heritable mutation and gene conversion by Cas9-CRISPR in Caenorhabditis elegans. Genetics 2013;195:1173–6. doi:10.1534/genetics.113.155754.
[35] Lo T-W, Pickle CS, Lin S, Ralston EJ, Gurling M, Schartner CM, et al. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 2013;195:331–48. doi:10.1534/genetics.113.155382.
[36] Tzur YB, Friedland AE, Nadarajan S, Church GM, Calarco JA, Colaiácovo MP. Heritable custom genomic modifications in Caenorhabditis elegans via a CRISPR-Cas9 system. Genetics 2013;195:1181–5. doi:10.1534/genetics.113.156075.
[37] Waaijers S, Portegijs V, Kerver J, Lemmens BBLG, Tijsterman M, van den Heuvel S, et al. CRISPR/Cas9-targeted mutagenesis in Caenorhabditis elegans. Genetics 2013;195:1187–91. doi:10.1534/genetics.113.156299.
[38] Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nature Methods 2013;10:741–3. doi:doi:10.1038/nmeth.2532.
[39] Cuppen E, Gort E, Hazendonk E, Mudde J, van de Belt J, Nijman IJ, et al. Efficient target-selected mutagenesis in Caenorhabditis elegans: toward a knockout for every gene. Genome Research 2007;17:649–58. doi:10.1101/gr.6080607.
[40] Lawrence CW. Cellular functions of DNA polymerase zeta and Rev1 protein. Adv Protein Chem 2004;69:167–203. doi:10.1016/S0065-3233(04)69006-1.
[41] Pryor JM, Gakhar L, Washington MT. Structure and Functional Analysis of the BRCT Domain of Translesion Synthesis DNA Polymerase Rev1. Biochemistry 2013;52:254–63. doi:10.1021/bi301572z.
[42] Larimer FW, Perry JR, Hardigree AA. The REV1 gene of Saccharomyces cerevisiae: isolation, sequence, and functional analysis. J Bacteriol 1989;171:230–7.
[43] Gibbs PE, Wang XD, Li Z, McManus TP, McGregor WG, Lawrence CW, et al. The function of the human homolog of Saccharomyces cerevisiae REV1 is required for mutagenesis induced by UV light. Proc Natl Acad Sci U S A 2000;97:4186–91. doi:10.1073/pnas.97.8.4186.
[44] Roerink SF, Koole W, Stapel LC, Romeijn RJ, Tijsterman M. 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 2012;8:e1002800. doi:10.1371/journal.pgen.1002800.
[45] van Bostelen I, Tijsterman M. Combined loss of three DNA damage response pathways renders *C. elegans* intolerant to light. DNA Repair 2017. doi:10.1371/journal.pbio.1001378.
[46] Alpi A, Pasierbek P, Gartner A, Loidl J. Genetic and cytological characterization of the recombination protein RAD-51 in Caenorhabditis elegans. Chromosoma 2003;112:6–16. doi:10.1007/s00412-003-0237-5.
[47] Roerink SF, van Schendel R, Tijsterman M. Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. Genome Research 2014;24:954–62. doi:10.1101/gr.170431.113.
[48] De Stasio E, Lephoto C, Azuma L, Holst C, Stanislaus D, Uttam J. Characterization of revertants of unc-93(e1500) in Caenorhabditis elegans induced by N-ethyl-N-nitrosourea. Genetics 1997;147:597–608.
[49] Greenwald IS, Horvitz HR. unc-93(e1500): A behavioral mutant of Caenorhabditis elegans that defines a gene with a wild-type null phenotype. Genetics 1980;96:147–64.
[50] Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 1983;100:64–119. doi:10.1016/0012-1606(83)90201-4.
[51] Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective *C. elegans*. Curr Biol 2008;18:900–5. doi:10.1016/j.cub.2008.05.013.
[52] Koole W, van Schendel R, Karambelas AE, van Heteren JT, Okihara KL, Tijsterman M. A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. Nat Comms 2014;5:3216. doi:10.1038/ncomms4216.
[53] van Schendel R, Roerink SF, Portegijs V, van den Heuvel S, Tijsterman M. Polymerase Θ is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nat Comms 2015;6:7394. doi:10.1038/ncomms8394.
[54] Lemontt JF. Mutants of yeast defective in mutation induced by ultraviolet light. Genetics 1971;68:21–33.
[55] Clark DR, Zacharias W, Panaitescu L, McGregor WG. Ribozyme-mediated REV1 inhibition reduces the frequency of UV-induced mutations in the human HPRT gene. Nucleic Acids Res 2003;31:4981–8. doi:10.1093/nar/gkg725.
[56] Jansen JG, Langerak P, Tsaalbi-Shtylik A, van den Berk P, Jacobs H, de Wind N. Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med 2006;203:319–23. doi:10.1084/jem.20052227.
[57] Lemmens BBLG, Tijsterman M. DNA double-strand break repair in Caenorhabditis elegans. Chromosoma 2011;120:1–21. doi:10.1007/s00412-010-0296-3.
[58] Brenner S. The genetics of Caenorhabditis elegans. Genetics 1974;77:71–94.
[59] Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nature Methods 2013;10:1028–34. doi:10.1038/nmeth.2641.

# Supplemental information

## Supplemental figure 1.

*Alignment of REV1 protein sequences and phylogenetic tree.*

Sc = Saccharomyces cerevisiae
Ce = Caenorhabditis elegans
Dm = Drosophila melanogaster
Dr = Danio rerio
Xl = Xenopus leavis
Gg = Gallus gallus
Mm = Mus musculus
Hs = Homo sapiens

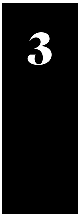
### A) CLUSTAL multiple sequence alignment of REV1 protein sequences

Known domains in human REV1 are indicated:

**BRCT** With mutated residues in this and other studies in BRCT mutants indicated as: G
**Pol** Y-family polymerase domain
**UMB** Ubiquitin binding motifs (UBM)
**TLSinteract** C-terminal domain interacting with other TLS polymerases

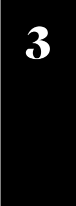
```
Sc      --------------------------------------MGEHGGLVDLLDSD--------    14
Ce      MHSPDPFTDHPDSDDSFEKFVPTPNVNRRRSDAIIVSDEGSFDGITGYSQVTVIENCPNT    60
Dm      ------------------------------------------------------------    0
Dr      ------------------------------------------------------------    0
Xl      ------------------------------------------------------------    0
Gg      ------------------------------------------------------------    0
Mm      ------------------------------------------------------------    0
Hs      ------------------------------------------------------------    0

Sc      --LEYSINRETPDKN--------------------NCLSQ----QSVNDS---HLTA-KT    44
Ce      TVLEDENKRANPDDTITRMLESSDEEDETENNSDNDALADLLKSFRTEDQKRKEVDIEQA    120
Dm      ------------------------------------------------------------    0
Dr      ------------------------------------------------------------    0
Xl      ------------------------------------------------------------    0
Gg      ------------------------------------------------------------    0
Mm      ------------------------------------------------------------    0
Hs      ------------------------------------------------------------    0

Sc      GGLNARSFLSTL----SDDSLIEYVNQLSQ--------------TNKNNSNPTAGTLRFT    86
Ce      ANISIDIFGDLIEEVDSEHEAAEYPDREGGNSMDKQGTMLPVEVPIQNNIKLIPRKRRLE    180
Dm      ------------------------------------------------------------    0
Dr      ------------------------------------------------------------    0
Xl      ------------------------------------------------------------    0
Gg      ------------------------------------------------------------    0
Mm      ------------------------------------------------------------    0
Hs      ------------------------------------------------------------    0

Sc      TKNISCDELHADLGGGEDSPIARSVIEIQESDSNGDDVKK---------NTVYTREAYFH    137
Ce      PKSYDCEPSTSAAHG----P-----VGVVSSDSGEYWNQQKTITIHGQNFEINNFNDYMR    231
Dm      --------------------------------------MT-----RDEDNGFSEWGGYFE    17
Dr      -------------------------------MSRDGWRKK-----ASENDGWGGQGGYMA    24
Xl      -------------------------------MRQGGWRNR-----ASEKDGWTAWGGYIA    24
Gg      -------------------------------MRRGGWRKR-----AGEGDGWGGWGGYMS    24
Mm      -------------------------------MRRGGWRKR-----T-ENDGWEKWGGYMA    23
Hs      -------------------------------MRRGGWRKR-----A-ENDGWETWGGYMA    23
                                                                 *:
```

```
                            BRCT>                       *
Sc      EKAHGQTLQDQILKDQYK-DQISSQSSKIFKNCVIYINGYTKPGRLQLHEMIVLHGGKFL      196
Ce      MK--ITKLNHQVNHGYAK---PLENVSRIMEGFSVFVNGYTDPPALVIRDLMISHGGEYH      286
Dm      AK--KSKLEEQFAAASDP-----FRKSDLFQGISIFVNGRTDPSADELKRIMMVHGGTFH      70
Dr      AK--VSKLEEQFQKDAPREQKKDGKSSCIFSGVAIYVNGYTDPSAAELRRLMMLHGGQFH      82
Xl      AK--VQKLDEQFRTDAPLQQQKDGTSSKIFNGVSIYVNGYTDPTADQLRHLMMLHGGQYH      82
Gg      AK--VKKLEDQFRSDSAIQHQRDGNSSSIFSGVAIYVNGFTDPSADELRRLMMLHGGQYH      82
Mm      AK--VQKLEEQFRTDAAN--QKDGTASAIFSGVAIYVNGYTDPSAEELRNLMMLHGGQYH      79
Hs      AK--VQKLEEQFRSDAAM--QKDGTSSTIFSGVAIYVNGYTDPSAEELRKLMMLHGGQYH      79
         *    .*:.*.              * ::.   :::** *.*    :: :::  *** :

Sc      HYLSSKKTVTHIVASNLPLKKRIEFAN-YKVVSPDWIVDSVKEARLLPWQNYSLTSKLDE      255
Ce      CYYQHGI-TSYTIASSIATAKINRIRENEIFIKADWITESIAAGKPLDYRDFLIYEKGSV      345
Dm      HY-ERSH-TTYIIASVLPDVKVRNMNL-SKFISAKWVVDCLEKKKIVDYKPYLLYTNQKT      127
Dr      LYYSRSK-TTHIIATNLPNFKIQELKG-EKVVRPEWIIDSIKAGRQLSYIQYQLYAKQ--      138
Xl      VYYSRSK-TTHIIATNLPNNKVNELKD-EKVVRPEWITESIKAGRLLSYVPFQLYTKKSS      140
Gg      VYYSRSK-TTHIIATNLPNAKIKELKG-EKVVRPEWIVESIKAGRLLSHIPYQLYTKQSS      140
Mm      VYYSRSK-TTHIIATNLPNAKIKELKG-EKVIRPEWIVESIKAGRLLSSAPYQLYTKPSA      137
Hs      VYYSRSK-TTHIIATNLPNAKIKELKG-EKVIRPEWIVESIKAGRLLSYIPYQLYTKQSS      137
         * .    .:: :*: :   *  .:     .:  .*: :.:   : :    : :  :

Sc      QQKKLDN---CKTVNSIPLPSE--------------------------------------      274
Ce      EKGQMQ------------------------------------------------------      351
Dm      SQPMLIFGKP---KDNGANE--SKSDVEPPK-----------------------DKAEVE      159
Dr      --KGLNFTRVSGIEGQDPSSN-HGPDIKPQPSHTLSKDVLK--STQANHMANYTEKTDHQ      193
Xl      VQKGLNFTSICKRDDSLPGPSNISRDFSQRENSTIKDFSARPISVLLNGVERL-EEDEHK      199
Gg      VQKGLSFNSICKPEDAMPGPSNIAKDLNR-VNH-IKQCEMES-EITPNGISSWNEEEEED      197
Mm      AQKSLNFNPVCKPEDPGPGPSNRAKQLNNRVNHIIKKIETES-EVKANGLSSWNEDGVN-      195
Hs      VQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNHIVKKIETEN-EVKVNGMNSWNEEDEN-      195
            :

Sc      ------------------------------------------------------------      274
Ce      ------------------------------------------------------------      351
Dm      VD--------------------------------------STKDETQM-------ELGG      173
Dr      LDIRLRNGLHSIISEVDAEKSRVNGIHDDDDDDDIACSILPRGSQDTLLTNGHVHPV-NG      252
Xl      NG-----DFEMLDLEQIFSDHKPNGL--------------QKHIGNTDVCSKHIPSS-NG      239
Gg      SD-----GLGFTKLDQILPERKQNGI--------------QSHKDSTAIFNGHTHNTCTS      238
Mm      -D-----DFSFEDLEHTFPGRKQNGV--------------MHPRDTAVIFNGHTHSS-NG      234
Hs      -N-----DFSFVDLEQTSPGRKQNGI--------------PHPRGSTAIFNGHTPSS-NG      234

Sc      ------------------------------------------------------------      274
Ce      ----------------------------------------------QFLT-----------      355
Dm      ILKNLQQA----------------------------------------------------      181
Dr      ALKPQDFADHHPHVVDKASNQSHLPQDKYREGEEPCCSYTESKANSHLRSLISAASPAKQ      312
Xl      ALKSLDSTVHFNSSFESGL--DLLQ--QQEKPDQCCANITDCTV-EYLQ-----------      283
Gg      ALKTQDCLVPSSNSVASRFSPGPVQ--EEGKPEKGIVDFRDCTM-QQLQ-----------      284
Mm      ALKTQDCLVPVGNSVASRLSLDSTQ--EEKRAEKSNADFRDCTV-QHLQ-----------      280
Hs      ALKTQDCLVPMVNSVASRLSPAFSQ--EEDKAEKSSTDFRDCTL-QQLQ-----------      280

Sc      --------------------------TSLHK--GSKCVGSALLPVEQQ-------SPVNL      299
Ce      ------------ISTRNT--------------------GGN------------------N      365
Dm      ---------VATSPEKEAS----AS--------ESKIT-----------------NLST-      202
Dr      PGPHNMQTDLHQKPTSNSGTLGAPQKASLHDQTTVRLNGSYHVTS----------NSSTL      362
Xl      --PNSRQTDVSCNPHRTVPL--P-SSSCLHT--NSKINGANHSLL-GPSSTNALS----P      331
Gg      --QSNKNTDFSWNPHRTMSN--SSSSSSLHS--NTKINGAHHSTVQGPSSTKSTS-VPTP      337
Mm      --HSTRSADALRSPHRTNS----LSP-SLHS--NTKINGAHHSTVQGPSSTKSTS-VLTL      330
Hs      --QSTRNTDALRNPHRTNS----FSLSPLHS--NTKINGAHHSTVQGPSSTKSTSSVSTF      332
```



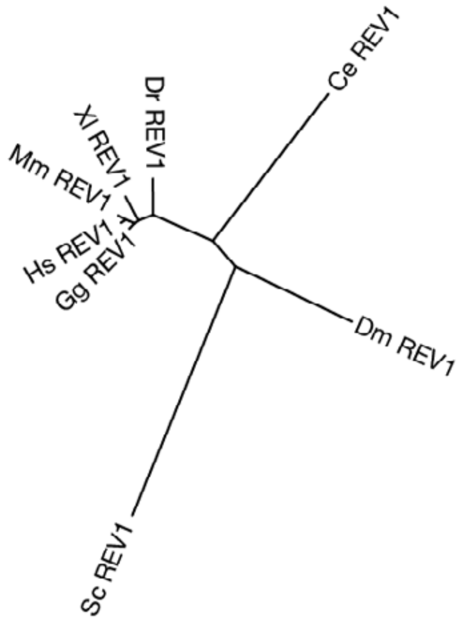
```
Sc      NNLEAKRIVACDDPDFLTSYFAHSRLHHLSAWKANLKDKFLNENIHK-------------   346
Ce      ETTSSNQFSDARNPNFIRDYYARSRLHLISTLAQDMKDFVANLKLEGKLT-EKCFEEKEL   424
Dm      TSNNSTTARTAADPNFLSEFYKNSRLHHIATLGAGFKQYVCRLRQKH---GTQGFPKRET   259
Dr      ANSSNQSGKSSAEAGIISEFFSHSRLHHISTWRNEFSEYVNSLQSRRRAAGGAVFSGREK   422
Xl      AKLHSPQLPKTADPNFISDFYSHSRLHHISTWKCEFTEFVRNLQTQS----NRGFPGRER   387
Gg      SKAASLSVSKPSDCSFISDFYSRSRLHHISTWKCELTEFVNSLQRKN----SGVFPGREK   393
Mm      SKVAPSVPSKPSDCNFISDFYSRSRLHHISTWKCELTEFVNTLQRQS----SGIFPGREK   386
Hs      SKAAPSVPSKPSDCNFISNFYSHSRLHHISMWKCELTEFVNTLQRQS----NGIFPGREK   388
         .          :  :: .:: .**** ::     :.: .   . .

                                         Pol>________________________
Sc      YTK-------------------ITDKDTYIIFHIDFDCFFATVAYLCRSSSFSACDFKRD   387
Ce      ID-MK--------------SLSNEISRESTVFHVDLDCFFVSVAVRNRID------LKHK   463
Dm      LKSLANSH---------------HNCLERYVMHIDMDCFFVSVGLRTRPE------LRGL   298
Dr      LKKLKANCNSVSHFDPGSLMA-APQVRQSCVLHVDMDCFFVSVGIRHRPD------LIGK   475
Xl      LKKLKPGN----------LPPSAFPKCQNCIIHVDMDCFFVSVAIRNHAD------LKGK   431
Gg      LKKWKAGRSAL-KTDTGNVSVASSAKPQSCIMHVDMDCFFVSVAIRNRPD------LKGK   446
Mm      LKKVKTGRSSLVVTDTGTMSVLSSPRHQSCVMHVDMDCFFVSVGIRNRPD------LKGK   440
Hs      LKKMKTGRSALVVTDTGDMSVLNSPRHQSCIMHVDMDCFFVSVGIRNRPD------LKGK   442
                                      ::*:*:****.:*.   :  .      :

        ____________________________________________________________
Sc      PIVVCH------------------------------------------------------   393
Ce      EVAITHSKGTIS------------------------------------------------   475
Dm      PIAVTHSKGGNAATDVPVHPQADRKAELELFAQRFEHHFHD---GD---------KAEKV   346
Dr      PVAVTSNRGPGRV---AQRPGANPQLEFQYYQNKQKHYRKEKTDGDL-EMTPSPQDGEVP   531
Xl      PVAVTSNRGAGTT---ITREGSNPQLEFQYYQNRILKGKAGQPMS--------ATSDSAQ   480
Gg      PVAVTSNRGAGKA---PLRPGANPQLEWQYYQNKLLNGKAEIRIPDKLDSWVWEHSDSAH   503
Mm      PVAVTSNRGTGTA---PLRPGANPQLEWQYYQNRALRGKAA----DIPDSSVWENQDSTQ   493
Hs      PVAVTSNRGTGRA---PLRPGANPQLEWQYYQNKILKGKAA----DIPDSSLWENPDSAQ   495
         :.:

        ____________________________________________________________
Sc      ---------GTKNSDIASCNYVARSYGIKNGMWVSQAEKMLPNGIKLISLPYTFEQFQLK   444
Ce      ----------NSMSEVASCSYAARDCGVKNGMLVRDALQKCPQ---LTLLPYQFEDYVQV   522
Dm      RSGFDK---KMSLSEIASCSYEAREKGIRNGMFVGQALKLCPE---LKTIPYDFEGYKEV   400
Dr      SNGVHEDLAALSMAEIASCSYEARQAGVRNGMFFGRAKQLCPD---LQSVPYDFHAYKEV   588
Xl      QNGLDQDISHLSMAEIASCSYEARQAGVKNGMFFGRAKQLCPE---LQAVPYDFDAYKEV   537
Gg      MNGVDCDLTVLSMAEIASCSYEARQAGIKNGMFFGQAKKLCPN---LQAVSYDFNAYKEV   560
Mm      TNGIDS---VLSKAEIASCSYEARQVGIKNGMFFGYAKQLCPN---LQAVPYDFHACREV   547
Hs      ANGIDS---VLSRAEIASCSYEARQLGIKNGMFFGHAKQLCPN---LQAVPYDFHAYKEV   549
                   . :::***.* **. *::*** .  * :  *:   *  : * *.

        ____________________________________________________________
Sc      SEAFYSTLKRLNIFNLILPISIDEAVCVRIIPDNIHNTNT-LNARLCEEIRQEIFQGTNG   503
Ce      SRKIYEILASYTL--EVRAVSCDEMYIN--MSSFCEKYEINDPTILAEHIRKVIR-EKTQ   577
Dm      AFTLYDTVAQYTL--NIEAVSCDEMFVE--LTDLAHELNV-DVMAFVSHLRQEVY-SKTG   454
Dr      ALAMYEILASYTH--NIEALSCDEALVD--ATALLVELGV-SPDELARSIREDIK-EKTG   642
Xl      AMNMYKILASYTH--DIEAVSCDEALAD--ITGILTETRL-TPDEASNAIRTEIK-EKTG   591
Gg      AQTVYEILASYTH--NIEAVSCDEALVD--ITEILTETRL-TPDELANAIRDEIK-AQTK   614
Mm      AQAMYETLASYTH--SIEAVSCDEALID--VTDILAETKL-SPEEFAAALRIEIK-DKTK   601
Hs      AQTLYETLASYTH--NIEAVSCDEALVD--ITEILAETKL-TPDEFANAVRMEIK-DQTK   603
        :  .*. :   .    :  :* **            :            :*  :    .

        _________________________________________________
Sc      CTVSIGCSDSLVLARLALKMAKPNGYNITFKSNLSEEFWSSFKLDDLPGVGHSTLSRLES   563
Ce      CPASVGIGSTSLLARLATRHAKPDGVFWVNAH-KKNEFISEEKVKDLPGFGYEMMNRLTS   636
Dm      CPCSAGVAGNKLLARMATKEAKPNGQFLLDSSNDILAYMAPMSLDLLPGVGSSISHKLKQ   514
Dr      CCASVGMSSNILLARMATRKAKPKGQYLLRSE-EVDDFIRDQPVSSLPGVGRSMSSKLTS   701
Xl      CAASIGIGSNILLARMATRRAKPDGQYHLKPE-EVDDFIRGQLVNNLPGVGRTMDCKLSS   650
Gg      CTASVGMGSNILLARMATRKAKPDGQYHLKPE-EVDDFIRGQLVTNLPGVGRSMESKLAS   673
Mm      CAASVGIGSNILLARMATKKAKPDGQYHLQPD-EVDDFIRGQLVTNLPGVGRSMESKLAS   660
Hs      CAASVGIGSNILLARMATRKAKPDGQYHLKPE-EVDDFIRGQLVTNLPGVGHSMESKLAS   662
        *  * * . . :***:* : ***.*            :     :  ***.*       :* .
```

```
Sc      TFDSPHSLNDLRKRYTLDALKASVGSKLGMKIHLALQGQDDEESLKILYDPKEVLQRKSL     623
Ce      FFGDITKCRELQL-KTERELVPVFGPKLASKILRQCRGIEEDPDD-FW----ATHVRKSV     690
Dm      A--GLNNCGDVQN-TTLEKMEKVLGKKLGQNLFQNCRGIDDRPLA-------YEQIRKTV     564
Dr      L--GVSTCGDLQQ-LSLSQLQREFGPRTGQTLFRFCRGLDDRPVR-------SEKERKSV     751
Xl      L--GVKTCGELQN-ITMAKLQKEFGPKTGQMLYRFCRGLDDRPIR-------KEKERKSV     700
Gg      L--GIRTCGDLQC-ASMSKLQKEFGPKTGQMLYRFCRGLDDRPVR-------TEKERKSV     723
Mm      L--GIKTCGDLQC-LTMAKLQKEFGPKTGQMLYRFCRGLDDRPVR-------TEKERKSV     710
Hs      L--GIKTCGDLQY-MTMAKLQKEFGPKTGQMLYRFCRGLDDRPVR-------TEKERKSV     712
              .  :::   :   :   .* : .  :    :* ::               **::

Sc      SIDINWGIRFKNITQVDLFIERGCQYLLEKLNEINKTTSQITLKLMRRCKDAPIEPPKYM     683
Ce      SCDINYGIRFTKRGEVIQLMTAIGAELERKLIDSKLTAGSITLKLMVRSANAPIQTSKFM     750
Dm      SAEMNFGIRFTNSVECEQFLCQLSEEVTKRLVEIRRKARSINLKIMVRAAEAPVETSKYM     624
Dr      SAEMNYNIRFTQVEEAESFLNNLSMEVQKRLEGAGLRGRRVTLKVMMRKPGAPVEPAKYG     811
Xl      SAEINYGIRFTQTSEAEVFLMNLSEEIQRRLEVVGMKGKKLTLKIMVRKAGAPIESAKYG     760
Gg      SAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEAAGMKGKRLTLKIMVRKAGAPVEPAKYG     783
Mm      SAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEAAGMKGKRLTLKIMVRKPGAPIETAKFG     770
Hs      SAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEATGMKGKRLTLKIMVRKPGAPVETAKFG     772
        * ::*: ***.:  :   ::      : .:*         :.**:* *   **::  *:

Sc      GMGRCDSFSRSSRLGIPTNEFGIIATEMKSLYRTLGCPPMELRGLALQFNKLVDVGPDNN     743
Ce      GHGICDTFTKTCNLNVPTTRGESLTSEAMKLYAKVSPKVEDLRGVGVTCGKLKSKLKKDA     810
Dm      GHGVCDIINKSSLIKYATDDVNVITTVVLDLMKDADIPPDELRGLGIHLTRLEDANEVRK     684
Dr      GHGICDNFARSVLLAQPTDSGRVIASEAIKLFHAMKLNVKDMRGVGLQVQQLDGSHAD--     869
Xl      GHGICDNIARTVTLHQATASAKVIGKEAVDMFHTMKLNISDMRGVGLQVHQLIPVGGLSI     820
Gg      GHGICDNIARTVTLDHATDSAKVIGKETLNMFHTMKLNISDMRGVGIQVQQLVPISKTT-     842
Mm      GHGICDNIARTVTLDQATDSAKIIGKATLNMFHTMKLNISDMRGVGIQVNQLVPANSNLS     830
Hs      GHGICDNIARTVTLDQATDNAKIIGKAMLNMFHTMKLNISDMRGVGIHVNQLVPTNLNPS     832
        * * ** : ::  :   *     : .   .:         ::**:.:   :*

Sc      QLKLRLPFKTIVTNRAFEALPEDVKN-----------------------DINNEFEKRN     779
Ce      --------------------ATAVQEMFGKTSRVGNMARTDEQLNIIPRNEDELDK----     846
Dm      E--------------------NNIKEMFGKMSEMR---KDKPIPQG-AVGDKSIGDDKVN     720
Dr      ------------PSGQGPSRGRSIRDLLLAKQSAHSPSKESPSQDGLISASSSRAFSSNQ     917
Xl      --FGSTL----VKSGHLPGGSRSLMDMFQGQKMKS-FSEDNDGKRDTITAVDEIGFASGT     873
Gg      --SAQSA----VQSGRLPGGSHSVIDLLHVQKAKK-CSEEEHKEVFVAAMDLEISSDSRT     895
Mm      TCSSRPS----AQSSLFSGRPHSVRDLFQLQKAKK-PTEEEHKEVFLAAVDLEVSSTSRA     885
Hs      TCPSRPS----VQSSHFPSGSYSVRDVFQVQKAKK-STEEEHKEVFRAAVDLEISSASRT     887
                               : :                          .

Sc      Y-------KRKESGLTSNS---------------LSSKKKGFAIS--------------R     803
Ce      ---EPVIIPAVTEDLEVQS-----IPRQI-----NFRVANDIDIPEIVKSTLLNGRSDNN     893
Dm      KPLVFEN-KPKPRE-PRNVLSMLTAAAVSRKSVTEDRSQRGTSKPIT-------------     765
Dr      KRVTPALSPPPSTSSSSSPPPSFTIIDPMP--GTSKGDYHHPHTPNHVRA-------CLN     968
Xl      ---CYSI-PPYNKS---SQ-----APSTSK--AASTGQKNGFSSPNNVKS-------GLR     912
Gg      ---CTVL-PSRGTHLTAGL-----NSNVSK--TDSTVKLNGLHSPISVKS-------RLN     937
Mm      ---CGLL-SPLSAHLAASV-----SPDTNS--GECSRKWNGLHSPVSGQS-------RLN     927
Hs      ---CTFL-PPFPAHLPT-------SPDTNK--AESSGKWNGLHTPVSVQS-------RLN     927
                                               .

             UBM1>
Sc      L-EVNDLPSTMEEQFMNELPTQIRAEVRHDLRIQKKIQQTKLGNLQEKIKRREE-SLQNE     861
Ce      LELEDLEDDGLVENRLA----------ELECIIQ-NEPTKESVT--NLQEM-LSTLLNYG     939
Dm      --RPLSLVPKLDEDVLAQLPEDIRLEVIANR-----------------------------     794
Dr      ISIEVPSPSQVDPSVLKALPIDIRRQVEETWRYREEQPSTSSHY--STPPR--PPSTP--     1022
Xl      FAIEVPSPSQIDPSVLEALPSDIREQVEQTFGFQQKTLLTESKN--DPAQKGSSELKQ--     968
Gg      LSIEVPSASQLDKSVLEALPPDLREQVEQIYTIQQGETYGDSKR--EPINGCNTALLS--     993
Mm      LSIEVPSPSQIDQSVLEALPLDLREQIEQVCAAQQGEPRG-KKK--EPVNGCSSGVLP--     982
Hs      LSIEVPSPSQLDQSVLEALPPDLREQVEQVCAVQQAESHGDKKK--EPVNGCNTGILP--     983
                  :  . :
```



```
                                                  UBM2>___________________________
Sc        -----------------------------------------------------------   861
Ce        KLSAFETVFRKFEEISMDSSRCDTAWFSVFHLMVPFIEEESKDLLEFPIATTRNLGAMIA   999
Dm        -------------------------------------------------EEHLCIAEY   803
Dr        --PPGPSLVLQLPNQ--PGQPCTT---GIILELPDFSQVDPDVFAALPRELQEELCSA-Y   1074
Xl        --QPVATVLLQIPNL---SDQ-GEEQGINVIALPAFSQVDPEVFAALPADLQEELRAA-Y   1021
Gg        --QPVGTVLLQVPEL---QEP-NANMGINVIALPAFSQVDPEVFAALPAELQAELKDA-Y   1046
Mm        --HPVGTVLLQIPEP---QEPCNSDSKISVIALPAFSQVDPDVFAALPAELQKELKAA-Y   1036
Hs        --QPVGTVLLQIPEP---QES-NSDAGINLIALPAFSQVDPEVFAALPAELQRELKAA-Y   1036

          ____
Sc        ---------------KNHFMGQNSI-----------------------------------   871
Ce        S-----NTDNTSISSPEPFVPKEVFVI------DSTKHT---------------------   1027
Dm        DGY-RSPQ----------------YPTLRSPPLLN---PYVTTNV-------SPLKATDL   836
Dr        RNKGNAQA-----QASTVVEQKNSFPQLKQPAVGKLKRRYKKRNTSPAKNGSSPLKKMFP   1129
Xl        GQRNKQTNNI---NINPTFVSKNPLLQLRKPLEKSKKRSRKKNKGSPTKNIHSPLKPKLF   1078
Gg        DQRQKQPE----QQPANAFVSKNPCLQLKHATTKNKKKIRKKNPVSPVKKIQSPLKNKLL   1102
Mm        DQRQRQGEDTTHQQPTSTSVPKNPLLQLKPPAMKDKR-NKRKNLIGSPRK--SPLKNKLL   1093
Hs        DQRQRQGENSTHQQSASASVPKNPLLHLKAAVKEKKR-NKKKKTIGSPKRIQSPLNNKLL   1095

                                                            TLSinteract>_____
Sc        ------------------------------------------------------------   871
Ce        ------------------------------------------------------------   1027
Dm        KPSTSRA---AVARLQKRKERKEQEHYIRSDQI-----VADY------IDDLPDFVNPHI   882
Dr        GNSPAKTS------PSK-----TLPLPLREQDLKCPSSL---------STDVPET---LP   1166
Xl        S-SPLKNGF-SAGSPQK-----LKDGPLKQEAKPSGQSQVELGPSTSNDA---KSLAQSH   1128
Gg        G-SPAKNMPAASGSPQK-----LIDGFLKQEGAAAQLE---AVPSTSDASDPSALQTEQC   1153
Mm        S-SPAKTLPGAYGSPQK-----LMDGFLQHEGMASERPLEEVSASTPGAQDLSSLLPGQS   1147
Hs        N-SPAKTLPGACGSPQK-----LIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPA   1149

Sc        ----FQPIKFQNLTRFKKICQLVKQWVAETLGDGGPHEKDVKLFVKYLIKLCDSNRVHLV   927
Ce        ------------------------------------------------------------   1027
Dm        LKLISHPVEMPELLMGDNYKDLLNDWVSRE---EVPKPNDVDLILKQVSRMIKNDQLDHV   939
Dr        KPIPRPAPTLAGAHEFSEIRTLLREWVTTI---SEPMEEDILQVVKYCTELVEDKDLEKL   1223
Xl        SSFKPKPPNLAGAIEFSDVKTLLREWITTI---SDPMEEDILQVVRYCTDLIDEKDLEKL   1185
Gg        GSFRPQAPNLAGAVEFNDVKTLLKEWITTI---SDPMEEDILQVVKYCTDLIEEKDLEKL   1210
Mm        SCFRPAAPNLAGAVEFSDVKTLLKEWITTI---SDPMEEDILQVVRYCTDLIEEKDLEKL   1204
Hs        GCVRPPAPNLAGAVEFNDVKTLLREWITTI---SDPMEEDILQVVKYCTDLIEEKDLEKL   1206

Sc        LHLSNLISRELNLCAFLNQDHSGFQTWERIL---LNDIIPLLNRNKHTYQTVRKLDMDFE   984
Ce        ------------------------------------------------------------   1027
Dm        CDVMKYW------CRIINMKRSSSCCWHVAYKHIEESIQNQM---LTIEGYSLLFIEYIR   990
Dr        YLVIKYM------KRL--MQQSSESVWSMAFDFVLDNVQVVV---QQTYGSTLKIT----   1268
Xl        DLVIKYM------KRL--MQQSVESVWNMAFDFILDNIQVVL---QQTYGSTLKVV----   1230
Gg        DLVVKYM------KRL--MQSSVESVWNMAFDFILDNVQVVL---QQTYGSTLKVI----   1255
Mm        DLVIKYM------KRL--MQQSVESVWNMAFDFILDNVQVVL---QQTYGSTLKVT----   1249
Hs        DLVIKYM------KRL--MQQSVESVWNMAFDFILDNVQVVL---QQTYGSTLKVT----   1251

Sc        V----   985
Ce        -----   1027
Dm        CIKCS   995
Dr        -----   1268
Xl        -----   1230
Gg        -----   1255
Mm        -----   1249
Hs        -----   1251
```

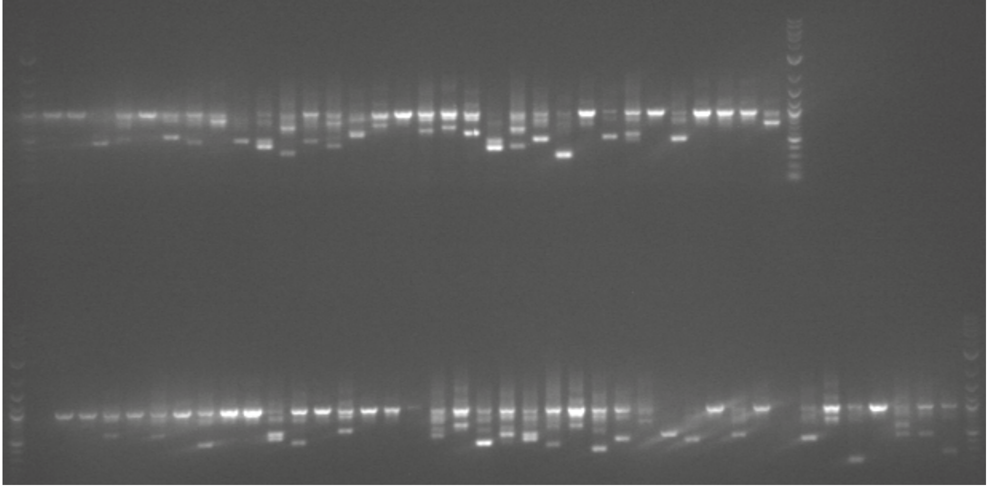
**B) Phylogenetic tree of REV1 genes based on the aligned sequences above**



## Supplemental figure 2.

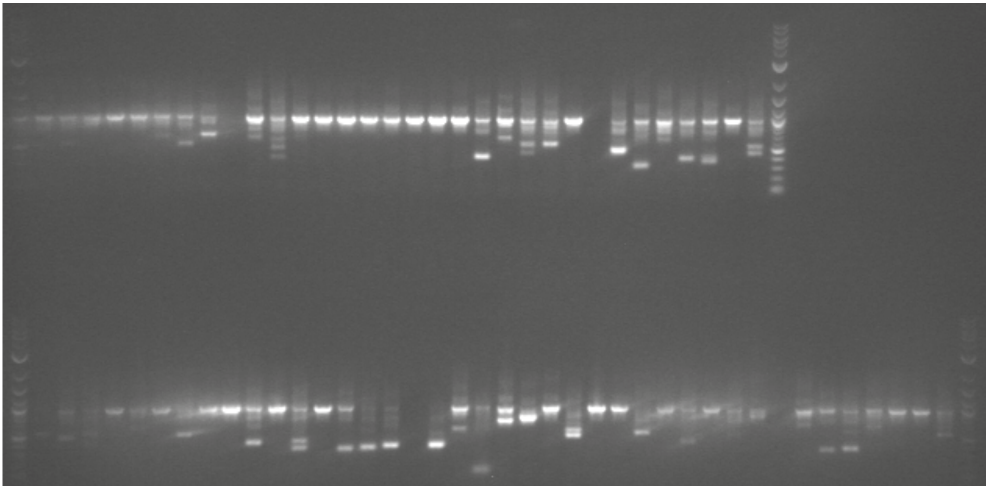
*PCR assay on genomic G4 site qua1466. Lyses were made by picking 5 adult worms in 15 ul lyses buffer. Of the lyses 1 ul was used as a template for the PCR. In the dog-1 background we did not observe additiona genomic instability at qua1466 after loss of REV-1.*

*dog-1(gk10)*



69 succesful PCRs with 50 samples positive for one or more deletions

*dog-1(gk10); rev-1(splice)*



67 succesful PCRs with 36 samples positive for one or more deletions

## Supplemental table 1

| Genotype | Chromosome | Size of deletion | Start | End | 5' flank | Deleted sequence | 3' flank | Inserted sequence |
|---|---|---|---|---|---|---|---|---|
| N2 | CHROMOSOME_I | 0 | 7667699 | 7667699 | CACGACAAAAGGTCA | | CATACACAAGAAAAA | caaaaggt |
| N2 | CHROMOSOME_II | 2 | 11519275 | 11519277 | ATATAAAATTTTTGA | tg | AAATTTTTTTGTTTT | |
| N2 | CHROMOSOME_V | 2 | 12400560 | 12400562 | AATTATCTCAAAATT | tc | AAAACGTATAATGAA | caaagcg |
| N2 | CHROMOSOME_II | 2 | 13447128 | 13447130 | AATCCTACTTTTTCG | ga | GAAATATCCTTTTTA | |
| N2 | CHROMOSOME_V | 6 | 18324061 | 18324067 | TTGAAATAGATTATT | tcaggc | TCAACGTCGTTGAAA | |
| N2 | CHROMOSOME_II | 7 | 10938399 | 10938406 | GGGCAAAGGGTAATT | atctgga | ATAGTTACCTCATTC | |
| N2 | CHROMOSOME_X | 11 | 16256717 | 16256728 | TTATGCCAATTATTG | tacatatatct | TACATGGATCAATTT | |
| N2 | CHROMOSOME_I | 18 | 4336436 | 4336454 | GTTTTTTGAAGTTTC | atatttcaaatttcagaa | ATATCGTAATCGTGG | |
| N2 | CHROMOSOME_III | 128 | 1407853 | 1407981 | AATTTTCGATCACAG | tagatggttgagcaatttttttctcagtgactgcatgcctgaaacagttagcaaaacgtgtcaggtaaatgctctgataactgtcgaaaatatttatcagtagagagctcagttatgagcccataatg | AGTTTGTTAGTTGAA | agtcac |
| rev-1 | CHROMOSOME_X | 1 | 4211498 | 4211499 | AAGAAAAGAAAAGTG | a | TTATTTCAAAAAAAA | |
| rev-1 | CHROMOSOME_X | 3 | 6824150 | 6824153 | AGAATATCGAGGAGT | taa | TAAGTAGTGGTGTAA | |
| rev-1 | CHROMOSOME_IV | 7 | 14354880 | 14354887 | ATTTTCAGAATCAGT | ttaaaac | TATGAAACAAATGTA | |
| rev-1 | CHROMOSOME_I | 18 | 3281427 | 3281445 | AGGCTTAGGTTGGGT | ttaggcttaggctgaggc | TTAGGCTTAGTTACA | |
| rev-1 | CHROMOSOME_V | 52 | 3945436 | 3945488 | AAAACTCCCACCGGG | tcccatagctttcccataatcggaaatattcctcaagtggcttattatgtc | TGGAAGACGGAGGGA | |
| rev-1 | CHROMOSOME_V | 58 | 1066424 | 1066482 | CGAGAAGCTTTCGTT | tttcctcgaatgtcatttcttttgtcggaactttgacctccttctcttttctgaaaga | AAGTTTTTAACATTT | |
| rev-1 | CHROMOSOME_III | 80 | 1418981 | 1419061 | TTCGCACAAAAATCA | ctggttttcactatttttacgctatttttttggttttttcttaattttcaggcacaaatatcgaatttaagaaggattag | CTCATAAAAATCAAG | |

| Geno-type | Chromo-some | Size of deletion | Start | End | 5' flank | Deleted sequence | 3' flank | Inserted sequence |
|---|---|---|---|---|---|---|---|---|
| rev-1 | CHROMO-SOME_IV | 91 | 11887607 | 11887698 | TTCAACTTTTACAAG | atctgggatgttcaaaa-gaccgttatgctattttc-caaattatatttttacaa-gaaagcatcaaatat-tataaaattatgtttgt-ttctac | TGTATTATCT-TTGTT | |
| rev-1 | CHROMO-SOME_V | 102 | 14695448 | 14695550 | ACAGCGAATGTCCCA | tccgaaacaatcagct-tatgtagatgcgcacg-cgtgcactcgaatagt-caacattaacaggggt-taggatgaaaagaaaaa-gaagaaagagaagag-caactg | GTAGGAGAT-GAGAAG | |
| rev-1 | CHROMO-SOME_X | 119 | 9279675 | 9279794 | ATCATTATTCAAAAA | ctatgactaacatcccac-tatctatttttcagattgt-ggtcacaatctatg-taggagctgcatcaa-caaactcactggcaacg-gaatcgtcaagtgtccat-ttgaccgtttggatacgc | GCGTCAGAGT-GACTG | |
| rev-1 | CHROMO-SOME_V | 133 | 625968 | 626101 | GTAATTTGGATCCGT | acacaaat-ttcgtcgcattgca-caatcgctaactcct-gtcgcattgtttattttgt-catggcatggatcagag-gtagtggatgttgatg-gggaaagtgtggtgc-cattgcccgatatgactg-gaagcaaaaat | TTTTTAGTGTG-GTAG | g |
| rev-1 | CHROMO-SOME_III | 155 | 10687677 | 10687832 | GGGTGTATTGCAGTG | ctgtgcggctctcggc-gagagccgaaaaaaat-tttgactcggctctcgg-cgcgagccga-gagtcgactttttcaa-gagccgcacagcactg-gtgtattgccttaaagaa-ggagtatcgtcaatgg-ggaaattgttttaaaatg-tagtatttgta | CTCAAACTTC-CAATT | |
| rev-1 | CHROMO-SOME_V | 174 | 16485754 | 16485928 | AAGGTACTGTAGTGG | gtccgcaagggatact-gtaaaattactgtaa-ggttcctgtagctcg-ggaaaatttgaat-tttcagcttttgaagag-gtttttttgttattttgttgt-cagtttggaatcttagcct-gactcgagtgcacttttc-caaaaaaaaaactcgc-tgtcaataacaagtagt-tttttaa | TCAACAATGT-GAATT | tcaacaatg |
| rev-1 | CHROMO-SOME_I | 194 | 10249974 | 10250168 | AACTGGCAGTTCTTT | ttaagcttgtctccac-cattatcaaggtttttc-tataagttgagcggt-gactcagatgaaatt-agatattcatagtcgc-tattttctcacgtctaact-tatactgataagcccat-ttgcaaataagatacg-cagatccctaaaaatca-catgctcacacctctt-gaatgtgtccccatcat-gagcccctgtcag | TTG-CACCTGCCTCTT | ctttt |

| Genotype | Chromosome | Size of deletion | Start | End | 5' flank | Deleted sequence | 3' flank | Inserted sequence |
|---|---|---|---|---|---|---|---|---|
| rev-1 | CHROMOSOME_I | 224 | 14832621 | 14832845 | TCCACCTTATCCTTT | ttcaattcaactagcaattagtaatatcttgaacaacagcagctggcggcttgctgattcctctttttccttgatttctctttaaaaaacgacacctagcaattaacagagaggagaacaggcaatatccaatctggcttcttaagaaataagaaagcatcagttccaattatgcgctgtgcttagtgaactggaaaaaacagcggcatgcgaaacctacccag | GCGAGCCCGAACAAG | |
| rev-1 | CHROMOSOME_V | 266 | 12384944 | 12385210 | TTTTTAGAACGAACT | cttgactatggttactctgggaattgggcatgatgtaggtgcagaggttaaactgaaagaggtaattgatgaatgccacgcataccataaaaacagtttctaactttcagctttatccaaaaattgattttttcggagcagatccagcagcagtgatcaataaagatctttatgagaatacattaggaggaaaatatttccaatatgctgtcagcggagaaagtggaatgcaacgctcaagtgtatatgaaagtaagccaaatggt | TATCCATAAAATTAA | tccatcaaa |
| rev-1 | CHROMOSOME_V | 306 | 15582024 | 15582330 | CTATACGCCACCCTG | agatttttgttgaaaatcagtgatattgccaaattatcaacctttttcaggaccaactttttttaattgttctccaattttcgagtacccgtgtatatttcaactattctcataagattttcacgtcacagaatttttataattttttttggtcctgtaagccaactttgggagccgaccactgatggcgcccttatctcacgttccagccctccaaaaattctgaatttttttcagtgctaaaggagaatgcttacatttcatactataactgccaaagttgtaggttctgtttaaataattctgt | AATTTGTTTTGTGTT | |
| rev-1 | CHROMOSOME_V | 312 | 1010572 | 1010884 | TAAAACACTTTTTTC | aatttcaatttccaatatttcagacttgccccaccaaagtacactgtgttcaaaagattaacatatctaatttgcgaagaagcgtacgaacccgagcttccggatgtgtcactgtgtcataaactgagaattggtacagacgcgagagttggagtttcgattgctattcgcacaaatctgagagccggaatcactgcatgctttgatcgtgagtttttttgcagagatttcttgacaattttcagccgacgacatccccgaattaaacccctatttcatttgaaagaggatcttcgaattaggttcctacaaa | AATTTTAGCCACAAA | caattaaaacacttttttt |