Disruption of *dog-1* in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA

Iris Cheung[1,2], Michael Schertzer[1], Ann Rose[2] & Peter M. Lansdorp[1,3]

Published online: 8 July 2002, doi:10.1038/ng928

© 2002 Nature Publishing Group http://genetics.nature.com

Genetic integrity is crucial to normal cell function, and mutations in genes required for DNA replication and repair underlie various forms of genetic instability and disease, including cancer[1]. One structural feature of intact genomes is runs of homopolymeric dC/dG. Here we describe an unusual mutator phenotype in *Caenorhabditis elegans* characterized by deletions that start around the 3′ end of polyguanine tracts and terminate at variable positions 5′ from such tracts. We observed deletions throughout genomic DNA in about half of polyguanine tracts examined, especially those containing 22 or more consecutive guanine nucleotides. The mutator phenotype results from disruption of the predicted gene *F33H2.1*, which encodes a protein with characteristics of a DEAH helicase and which we have named *dog-1* (for deletions of guanine-rich DNA). Nematodes mutated in *dog-1* showed germline as well as somatic deletions in genes containing polyguanine tracts, such as *vab-1*. We propose that DOG-1 is required to resolve the secondary structures of guanine-rich DNA that occasionally form during lagging-strand DNA synthesis.

Guanine-rich single-stranded DNA *in vitro* forms stable secondary structures such as triplex structures and G quartets, in which four guanine bases adopt Hoogsteen base-pairing[2,3]. The G quartets are characterized by a variable stack of guanine quartet planes with strands in either a parallel or an antiparallel orientation[4]. Despite much speculation about their possible biological roles, G quartets have not been observed *in vivo*, and little is known about molecules that might interact with them *in vivo*. Here we show that the gene *dog-1* is required to maintain genetic stability of guanine-rich DNA *in vivo* in *C. elegans*. Mutation of *dog-1* or disruption of its expression by RNA-mediated interference (RNAi) results in a mutator phenotype characterized by deletions throughout genomic DNA that were initiated at tracts of consecutive dC/dG bases.

The gene *dog-1* encodes a predicted protein of 983 amino acids, DOG-1 (Fig. 1*a*), containing seven protein motifs characteristic of the DExH-box type of DNA/RNA helicases[5]. The mutation

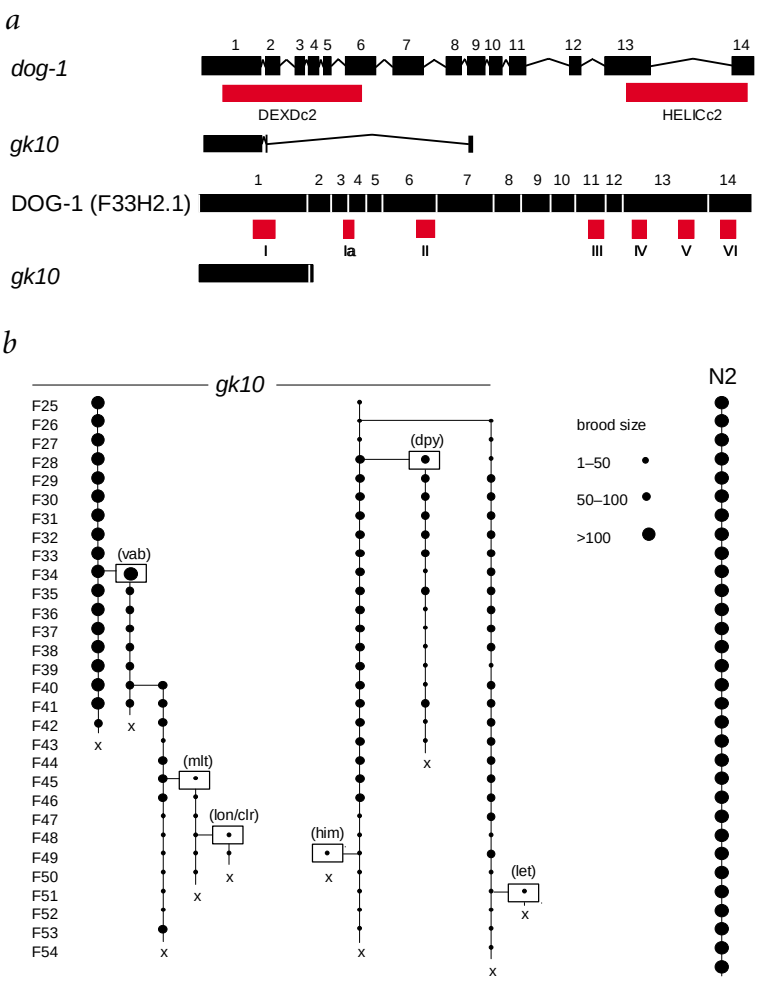

Fig. 1 Mutation of *dog-1* results in a mutator phenotype. ***a***, Intron-exon organization of *dog-1* (F33H2.1) and its *gk10*-null allele. The gene contains 14 exons encoding 7 conserved motifs of DNA/RNA helicases[5], encompassing a conserved DEXDc2 domain of the DEAH/DEAD-like helicase superfamily and a HELICc2 carboxy-terminal domain of the helicase superfamily. The deletion in *gk10* removes exons 2–8 and introduces a premature stop codon in exon 9. ***b***, Variable brood size and mutator phenotype in *gk10* nematodes. Generations F25–F54 of *gk10* nematodes were followed by transferring single animals. Brood size is indicated by size of circle. Horizontal lines indicate analysis of siblings in a particular generation; 'x' indicates extinction of the line. Heritable mutations are shown in boxes; the corresponding phenotypes were: mlt, molt defect; vab, variable abnormal; dpy, dumpy; lon/clr, long/clear; him, high incidence of males; let, lethal. Notably, the brood size is markedly variable and the frequency of heritable mutations is high relative to control nematodes (Bristol N2, right).

[1]*Terry Fox Laboratory, British Columbia Cancer Agency, 10th Avenue, Vancouver, British Columbia, V5Z 1L3, Canada.* [2]*Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.* [3]*Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada. Correspondence should be addressed to P.L. (e-mail: plansdor@bccancer.bc.ca).*

gk10 is a deletion of 2 kb, resulting in the absence of exons 2–8 of *dog-1* and in premature stop codon in exon 9 (Fig. 1*a*). Hence, *gk10* is a null mutation. Mutant offspring had brood sizes variably lower than the expected brood number of approximately 300 from a wildtype parent. We frequently recovered spontaneous mutants with characteristic phenotypes (Fig. 1*b*). Three such mutants, isolated independently, displayed a 'notched-head' phenotype[6] that was mapped by complementation to the *vab-1* locus[7] on chromosome III. All three mutations were deletions of exon 5 that were initiated in the same stretch of 20 continuous cytidines just outside exon 5 of *vab-1* (Fig. 2*a–c*). We observed no obvious location for the second breakpoint. We concluded that *gk10* nematodes have a defect in maintenance of the stability of a polycytidine tract upstream of exon 5 of *vab-1*. Using the polymerase chain reaction (PCR), we screened individual nematodes with no obvious Vab phenotype for deletions involving exon 5 of *vab-1*. In approximately 10% of the nematodes, we observed a band at a lower molecular weight in addition to the expected 'wildtype' band of 500 bp (data not shown). We determined the sequences of these bands and confirmed that they represented *de novo* deletions (Fig. 2*a–c*). Similar to the three germline deletions, all of these somatic-cell deletions had one breakpoint located just before or within the 5′ end of the polycytidine tract upstream of exon 5.

Subsequently, we examined other tracts of G/C-rich DNA in *gk10* nematodes for deletions. We detected deletions involving the sequence $C_4AC_{26}(CT)_2C_4$, located in an intergenic region in F55F3 on the X chromosome, in *gk10* nematodes (Fig. 2*d–f*). Such deletions were not detected on an extrachromosomal cosmid array in '*gk10* rescued', a transgenic strain carrying *gk10* but also containing a copy of wildtype *dog-1* (Fig. 3*a*). The size distribution of deletion products in *gk10* nematodes peaked at around 100–200 nucleotides, but we also detected longer deletion fragments (up to 1.7 kb; Fig. 3*b* and data not shown). These deletions were always initiated at the G/C tract and extended unidirectionally for various distances. Suppressing the expression of *dog-1* in wildtype nematodes by RNAi also resulted in deletions (Fig. 3*c*,*d*). These observations support the notion that the phenotype of *gk10* results from inactivation of *dog-1*. The effect of RNAi seemed to differ among individual nematodes (Fig. 3*d*), and was not evident in parents but was striking in the F1 progeny (Fig. 3*c*,*d*). Variable inhibition of *dog-1* expression by RNAi could explain these results.

To assess whether genomic instability in *gk10* nematodes was restricted to polyguanine tracts, we analyzed three A/T tracts, one AT/TA tract, two CT/GA tracts, two CA/GT tracts, four CTG/CAG repeats and three CCG/CGG repeats by PCR. We found no deletions in any of these sequences, suggesting that deletions are restricted to polyguanine tracts. To further study the instability of G/C tracts, we selected G/C tracts of different lengths at different locations in the *C. elegans* genome and screened by PCR for possible deletions. We observed deletions in approximately half of the tracts examined. We detected no deletions in any of the 5 G/C-tracts with 18 or fewer guanine nucleotides (Table 1), indicating that a threshold length is probably required for deletions to occur.

Analysis of individual *gk10* nematodes also supported a correlation between the length of the polyguanine tract and the

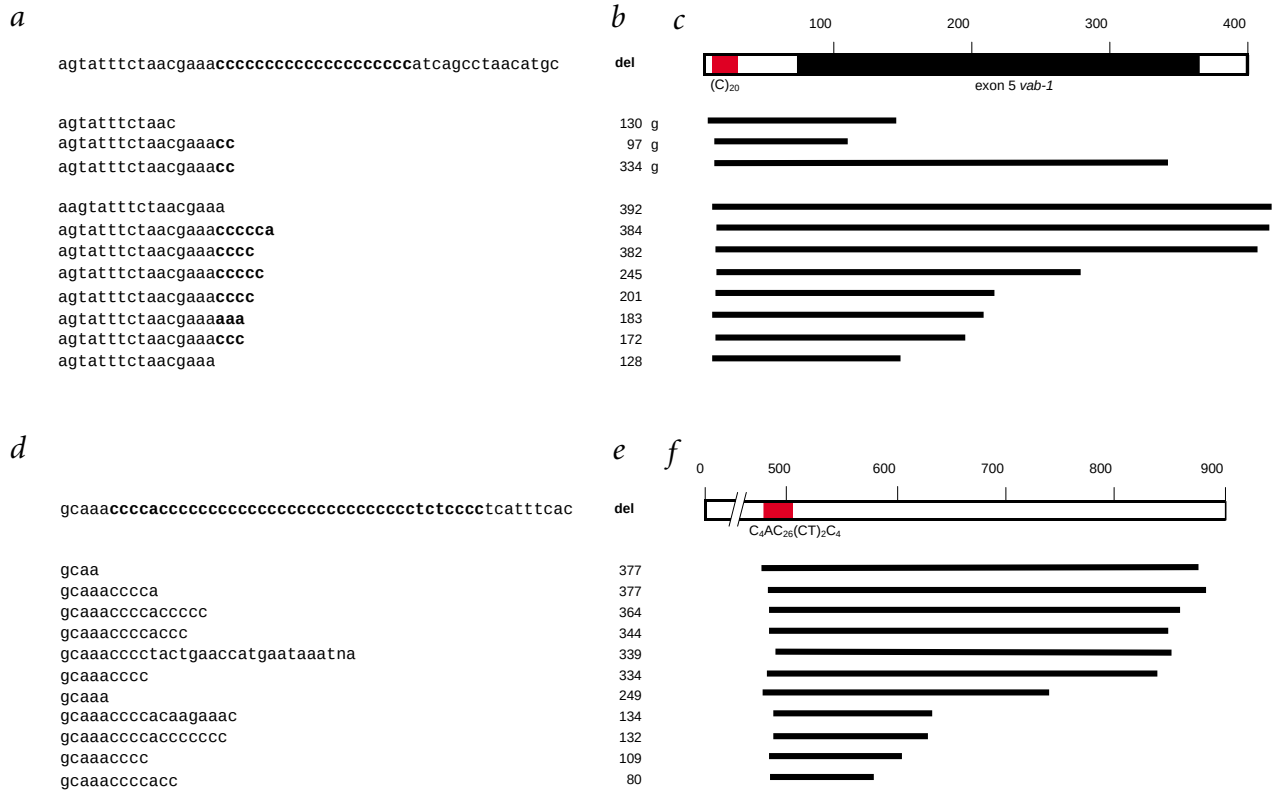

Fig. 2 Recurrent deletions in the *gk10* strain invariably start at tracts of C/G. ***a–c***, Germline and somatic deletions in *vab-1*. Top, nucleotides flanking a stretch of 20 Cs just outside exon 5 of *vab-1* (ref. 7) in which deletions were found to be initiated (red box in *c*). *a*, The deletion position relative to the tract of cytidines. *b*, Sizes of the deleted (del) fragments (g, germline deletions). *c*, Breakpoints relative to exon 5 of *vab-1*. ***d–f***, Deletions involving tracts of C/G in noncoding DNA. Nucleotides flanking the intergenic sequence $C_4AC_{26}(CT)_2C_4$ (red box in *f*) present on the cosmid F55F3, and breakpoints relative to this sequence are shown in *d* and *f*. The sizes of sequenced deletions are shown in *e*. Inserted sequences are underlined in *a* and *d*. For these experiments, we used genomic DNA from single *gk10* nematodes.

© 2002 Nature Publishing Group http://genetics.nature.com

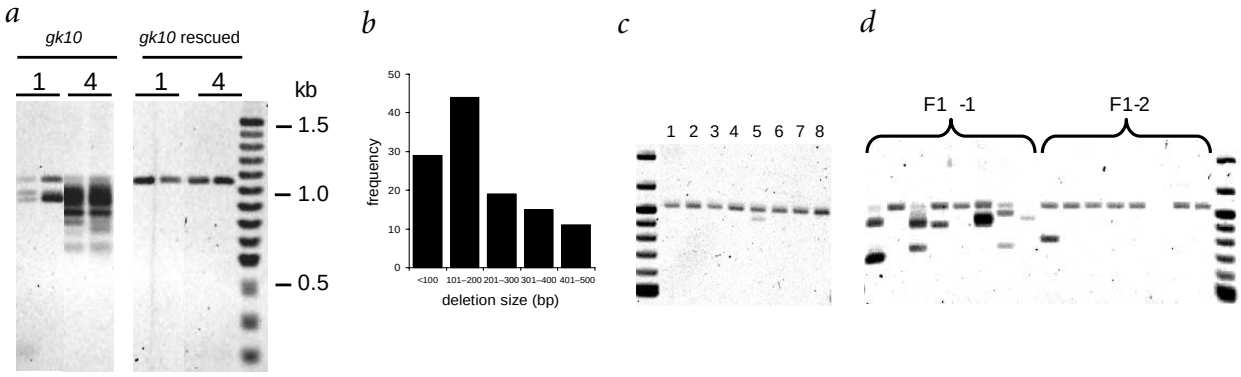

Fig. 3 Disruption of DOG-1 expression leads to frequent deletions involving the polyguanine tract in F55F3. ***a***, Expression of wildtype *dog-1* rescues the mutator phenotype of *gk10* nematodes. We lysed the indicated number of nematodes (*gk10* or transgenic *gk10* carrying full-length *dog-1*, indicated as *gk10* rescued) in a single lysis reaction. We divided each lysate in two and carried out two independent PCRs to amplify the F55F3 fragment. ***b***, Size distribution of deletions in F55F3 genomic DNA in *gk10*. We observed deletions of 100–200 nt most frequently. ***c,d***, Inhibition of *dog-1* expression using RNAi. We observed no deletions in the DNA from parents fed with *E. coli* expressing *dog-1* dsRNA (*c*) and a variable number of deletions in their F1 progeny (*d*). The single apparent deletion in (*c*) was DNA of *E. coli* origin. Failure to extract DNA from single nematodes is a plausible explanation for the empty lane in F1-2. F1-1 are the progeny of parent 1 and F1-2 are the progeny of parent 2 (lanes 1 and 2, respectively, in *c*).

© 2002 Nature Publishing Group http://genetics.nature.com

frequency of deletions. On average, we amplified fewer than one deletion fragment by PCR in tracts of fewer than 22 guanine nucleotides, whereas we typically observed several deletion fragments in vulnerable polyguanine tracts of 22 or more guanine nucleotides. An exception was G/C_{25} in the locus R144, where we detected single deletion bands in the DNA of roughly half the mutant nematodes. For each G/C tract in which deletions were observed, we randomly sampled and sequenced deletion bands. In all but two cases, all deletions we sequenced (>40) were initiated around the 3′ end of the polyguanine tracts and extended upstream for various lengths. The two unusual deletions were initiated 100 bp downstream of the polyguanine tract but, like the other deletions, extended several hundred base pairs upstream.

Telomeres in *C. elegans* consist of up to several kilobases of TTAGGC repeats[8]. Tracts of this sequence are also present elsewhere in the genome. Although telomere length in *gk10* nematodes was compatible with telomere instability (see Web Fig. A online), several internal tracts of 10–14 TTAGGC repeats did not show deletions (data not shown). These results indicate that internal TTAGGC repeats are not as vulnerable to deletions as are consecutive tracts of guanine nucleotides, perhaps because TTAGGC tracts form secondary structures less readily *in vivo*.

Because genetic deletions in *gk10* nematodes seem to be initiated at the 3′ end of polyguanine tracts and to involve about half of the polyguanine tracts examined, we propose that these deletions result from failure to resolve secondary structures of guanine-rich DNA that may arise sporadically during lagging-strand DNA synthesis (Fig. 4). A mechanism dependent on length and orientation and suggestive of *in vivo* lagging-strand secondary structures has been proposed to explain the instability of trinucleotide repeats in *Escherichia coli*[9–11] and *Saccharomyces cerevisiae*[12,13]. On the basis of the specificity of the deletion phenotype involving polyguanine tracts in *gk10*, and of the highly conserved helicase motifs in DOG-1 (Fig. 1*a*), it seems plausible that DOG-1 is required to unwind secondary structures of guanine-rich DNA. Alternatively, DOG-1 might be part of a response to DNA damage triggered by DNA lesions involving guanine-rich DNA, perhaps in a manner similar to MRT-2 (refs 14,15). Destabilization of repetitive DNA can also result from mutations in DNA mismatch-repair genes[16]. The mismatch-repair gene *msh-2* of *C. elegans* is required for microsatellite stability and maintenance of genome integrity[17]. Nematodes bearing mutations in either *msh-2* or *dog-1* have reduced brood sizes and a mutator phenotype, suggesting that these abnormalities could be common features of deficiencies in DNA repair in this organism.

Around half of large polyguanine tracts in genomic DNA in *gk10* nematodes seemed to be vulnerable to deletion events (Table 1). This observation is compatible with our model (Fig. 4), which predicts that deletion events depend on the direction of DNA replication. The frequency of actual deletions in vulnerable polyguanine tracts varied, however, as we observed either a single deletion event in up to half of the offspring or, more typically, multiple deletions in all offspring (Table 1). These observations may point to changes in the direction of DNA replication during development, as have been described in *Schizosaccharomyces pombe*[18]. Such switches, perhaps together with apoptosis of aberrant cells in

Table 1 • Frequency and size dependence of deletions in genomic DNA containing G/C tracts in *dog-1* mutants

Cosmid	Chr	Tract length	Presence of deletions	Estimated deletion alleles per adult nematode
Y77E11A	IV	G_{14}	no	0
Y75B7AL	V	G_{15}	no	0
Y39A3CR.6	III	G_5TG_{12}	no	0
Y41E3	IV	C_{17}	no	0
ZC123.3-a	I	G_{18}	no	0
ZC123.3-b	I	C_{19}	yes	0.2
M03A1.1	II	C_{20}	yes	0.1
M01E10.2	III	C_{20}	no	0
C04C11	X	C_{20}	no	0
Y15E3A	X	C_{20}	no	0
F42C5	IV	C_{21}	yes	0.4
R03G5	X	C_{22}	no	0
F46H6/C07A12	X	C_{22}	yes	multiple
Y41D4A	IV	G_{24}	yes	multiple
C18F3	IV	G_2AG_{25}	yes	multiple
R144	III	C_{25}	yes	0.5
R11B5	X	C_{25}	yes	multiple
F55F3	X	C_{26}	yes	multiple
F49E10	X	G_{26}	no	0
F38A6	V	C_{29}	yes	multiple
B0524.1	III	C_{32}	no	0

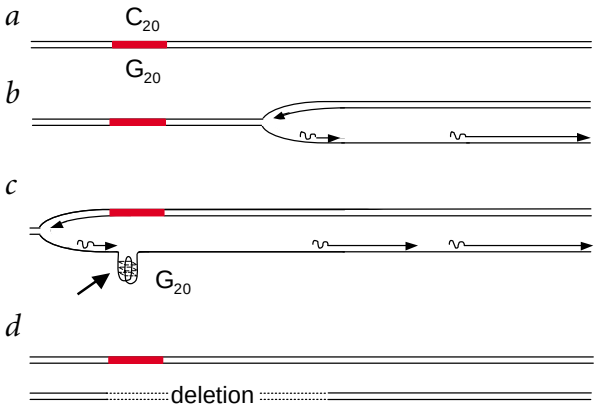

Fig. 4 Model of recurrent deletions in *gk10* nematodes. The invariable occurrence of breakpoints at the 5′ end of poly(dC) or the 3′ end of poly(dG) (Fig. 2) is compatible with deletions resulting from the failure of lagging-strand DNA synthesis. According to this model, the replication fork moves in from the right toward the poly(dC/G) sequences (***a***–***c***). Single-stranded poly(dG) sequences used as a template for lagging-strand DNA synthesis are postulated to give rise occasionally to stable secondary structures such as G quartets[2,3] (*c*, arrow). We propose that resolution of such secondary structures requires specialized DNA replication or repair machinery involving DOG-1. Failure to resolve secondary structures of guanine-rich DNA results in deletions by a mechanism that is unknown but is likely to involve excision and recombination, trans-lesion synthesis, or both.

the germline, could allow the polyguanine tracts in F55F3 to be maintained over several generations in *gk10* nematodes despite a high frequency of deletions in single nematodes (Fig. 3*b*). Alternatively, the apparent germline stability of this polyguanine tract could be related to changes in factors that protect the germline[19], such as the expression of other genes involved in the processing of secondary structures of guanine-rich DNA. Previous studies have shown that *dog-1* is highly expressed in the germline[20], perhaps reflecting a role of DOG-1 and polyguanine tracts in germline-specific processes such as recombination, as was originally proposed[2]. Such a role is supported by the similar chromosomal distribution of polyguanine tracts and regions of high recombination frequencies[21] in *C. elegans*. Polyguanine tracts have also been implicated in recombination during the development of human B cells[22].

Although many genes are involved in the maintenance of genetic stability, none have previously been identified that are functional homologs of *dog-1*. The genes involved in Bloom and Werner syndromes encode RecQ helicases capable of unwinding aberrant DNA structures, including G quartets[23–25]. The DOG-1 protein is not, however, a member of the RecQ family. Indeed, the human gene with the highest sequence similarity to *dog-1* is *BRIP1*, encoding a protein that binds to BRCA1, participates in the DNA repair function of the BRCA1 complex and is abnormal in some individuals with breast cancer[26]. Thus, our work provides the first description of an activity required specifically for the maintenance of guanine-rich DNA *in vivo*. Homologs of DOG-1 in other species may have a similar role in maintaining genome stability through interactions with guanine-rich tracts of DNA.

Methods

Nematode strains. We maintained strains of *C. elegans* as described[6]. We obtained *C. elegans* strain VC13, bearing the *gk10* allele of *dog-1*, from the Reverse Genetics Core Facility at the University of British Columbia (Vancouver, Canada). We constructed the transgenic strain KR3847 using *gk10* males (KR3792) and a *gk10* strain marked with *dpy-5* (KR3838). Outcrossed males were crossed to a transgenic strain carrying the cosmid ZK340 as an extrachromosomal array (BC5865), kindly provided by D. Baillie, Simon Fraser University, Vancouver, Canada. The ZK340 cosmid contains the wildtype *dog-1* along with the dominant gene *rol-6*, which confers a Roller phenotype. We subsequently examined resulting Roller progeny for the presence of full-length *dog-1* using PCR.

Analysis of deletion fragments. We obtained genomic DNA by lysis of nematodes with proteinase K. Polyguanine tracts in the genome were identified using euGenes (http://iubio.bio.indiana.edu:8089/). We carried out PCR using standard conditions. Primer sequences of *vab-1* exon 5, F55F3 (Fig. 2) and the G/C tracts (Table 1) are available on request. After separating the PCR products on agarose gels, we stained them with SYBR Green and then visualized them using a PhosphorImager. We randomly sampled and sequenced bands that were smaller in size than the wildtype band. Sequencing was done using Big Dye Terminator cycle sequencing V2.0 on a 377XL DNA Sequencer (Applied Biosystems) at the Nucleic Acids Protein Services facility at the University of British Columbia.

RNAi. We carried out RNAi assays as described[27]. We transferred single L4 larvae to NGM plates (with 50 μg ml^{-1} ampicillin, 25 μg ml^{-1} carbenicillin and 1 mM isopropyl-β-D-thiogalactoside) seeded with *E. coli* expressing *F33H2.1* obtained from the *C. elegans* Chromosome 1 RNAi library[28] (provided by J. Ahringer, United Kingdom Human Genome Mapping Project Resource Center). After 2.5 d, we analyzed the parents for deletions and transferred 8 L4 progeny from each parent to a fresh RNAi plate. We analyzed the DNA of the progeny for deletions 2 d later.

Note: Supplementary information is available on the Nature Genetics website.

Acknowledgments

This paper is dedicated to the memory of M. Smith, who died in October 2000. We thank D. Baillie for discussion and help with the analysis, R. Kay for critical reading of the manuscript and V. Vijayaratnam for outcrossing the gk10 strain. This work was supported by a grant from the National Cancer Institute of Canada with funds from the Terry Fox Run and by grants from the Canadian Institute of Health Research (A.R. and P.L.). I.C. was supported by a studentship from the University of British Columbia.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 16 April; accepted 24 May 2002.

1. Hoeijmakers, J.H.J. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
2. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
3. Williamson, J.R., Raghuraman, M.K. & Cech, T.R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871–880 (1989).
4. Simonsson, T. G-quadruplex DNA structures—variations on a theme. *Biol. Chem.* **382**, 621–628 (2001).
5. Bork, P. & Koonin, E.V. An expanding family of helicases within the 'DEAD/H' superfamily. *Nucleic Acids Res.* **21**, 751–752 (1993).
6. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
7. George, S.E., Simokat, K., Hardin, J. & Chisholm, A.D. The VAB-1 Eph receptor tyrosine kinase functions in neural and epithelial morphogenesis in *C. elegans*. *Cell* **92**, 633–643 (1998).
8. Wicky, C. *et al.* Telomeric repeats $(TTAGGC)_n$ are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 8983–8988 (1996).
9. Trinh, T.Q. & Sinden, R.R. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**, 544–547 (1991).
10. Kang, S., Jaworski, A., Ohshima, K. & Wells, R.D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* **10**, 213–218 (1995).
11. Hirst, M.C. & White, P.J. Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of *in vivo* lagging strand secondary structure. *Nucleic Acids Res.* **26**, 2353–2358 (1998).
12. Rolfsmeier, M.L. *et al. Cis*-elements governing trinucleotide repeat instability in *Saccharomyces cerevisiae*. *Genetics* **157**, 1569–1579 (2001).
13. Balakumaran, B.S., Freudenreich, C.H. & Zakian, V.A. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum. Mol. Genet.* **9**, 93–100 (2000).
14. Ahmed, S. & Hodgkin, J. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**, 159–164 (2000).
15. Gartner, A., Milstein, S., Ahmed, S., Hodgkin, J. & Hengartner, M.O. A conserved checkpoint pathway mediates DNA damage–induced apoptosis and cell cycle arrest in *C. elegans*. *Mol. Cell* **5**, 435–443 (2000).

© 2002 Nature Publishing Group http://genetics.nature.com

16. Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
17. Degtyareva, N.P. *et al. Caenorhabditis elegans* DNA mismatch repair gene *msh-2* is required for microsatellite stability and maintenance of genome integrity. *Proc. Natl Acad. Sci. USA* **99**, 2158–2163 (2002).
18. Dalgaard, J.Z. & Klar, A.J. A DNA replication-arrest site RTS1 regulates imprinting by determining the direction of replication at *mat1* in *S. pombe*. *Genes Dev.* **15**, 2060–2068 (2001).
19. Seydoux, G. & Schedl, T. The germline in *C. elegans*: origins, proliferation, and silencing. *Int. Rev. Cytol.* **203**, 139–185 (2001).
20. Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
21. Zetka, M.C. & Rose, A.M. Mutant *rec-1* eliminates the meiotic pattern of crossing over in *Caenorhabditis elegans*. *Genetics* **141**, 1339–1349 (1995).
22. Dempsey, L.A., Sun, H., Hanakahi, L.A. & Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D. A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.* **274**, 1066–1071 (1999).
23. Sun, H., Karow, J.K., Hickson, I.D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273**, 27587–27592 (1998).
24. Mohaghegh, P., Karow, J.K., Brosh, J.R. Jr, Bohr, V.A. & Hickson, I.D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **29**, 2843–2849 (2001).
25. Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M. & Fry, M. Interactions between the Werner syndrome helicase and DNA polymerase δ specifically facilitate copying of tetraplex and hairpin structures of the $d(CGG)_n$ trinucleotide repeat sequence. *J. Biol. Chem.* **276**, 16439–16446 (2001).
26. Cantor, S.B. *et al.* BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**, 149–160 (2001).
27. Kamath, R.S., Martinez-Campos, M., Zipperlen, P., Fraser, A.G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**, research0002.1–research0002.10 (2000).
28. Fraser, A.G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).

© 2002 Nature Publishing Group http://genetics.nature.com