DOI: 10.1534/genetics.104.038521

Mutation Rates, Spectra and Hotspots in Mismatch Repair-Deficient *Caenorhabditis elegans*

Dee R. Denver,*[,1] Seth Feinberg,* Suzanne Estes,[†] W. Kelley Thomas[‡] and Michael Lynch*

**Department of Biology, Indiana University, Bloomington, Indiana 47405, [†]Department of Zoology, Oregon State University, Corvallis, Oregon 97331 and [‡]Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824*

Manuscript received November 12, 2004
Accepted for publication January 20, 2005

ABSTRACT

Although it is clear that postreplicative DNA mismatch repair (MMR) plays a critical role in maintaining genomic stability in nearly all forms of life surveyed, much remains to be understood about the genome-wide impact of MMR on spontaneous mutation processes and the extent to which MMR-deficient mutation patterns vary among species. We analyzed spontaneous mutation processes across multiple genomic regions using two sets of mismatch repair-deficient (*msh-2* and *msh-6*) *Caenorhabditis elegans* mutation-accumulation (MA) lines and compared our observations to mutation spectra in a set of wild-type (WT), repair-proficient *C. elegans* MA lines. Across most sequences surveyed in the MMR-deficient MA lines, mutation rates were ~100-fold higher than rates in the WT MA lines, although homopolymeric nucleotide-run (HP) loci composed of A:T base pairs mutated at an ~500-fold greater rate. In contrast to yeast and humans where mutation spectra vary substantially with respect to different specific MMR-deficient genotypes, mutation rates and patterns were overall highly similar between the *msh-2* and *msh-6 C. elegans* MA lines. This, along with the apparent absence of a *Saccharomyces cerevisiae MSH3* ortholog in the *C. elegans* genome, suggests that *C. elegans* MMR surveillance is carried out by a single Msh-2/Msh-6 heterodimer.

GENOME stability is continually challenged by a diverse array of mutagenic forces that include errors during DNA replication, environmental factors such as UV radiation, and endogenous mutagens such as oxygen free radicals generated during oxidative metabolism (Lindahl 1993). Multiple DNA repair pathways have evolved to minimize the mutagenic consequences of DNA damage and erroneous DNA replication. Most of the major DNA repair pathways have been detected in all three domains of life, suggesting ancient origins (Eisen and Hanawalt 1999).

The mismatch repair (MMR) pathway corrects a wide range of base-base mismatches (some involving damaged bases) and small loop-outs in DNA molecules and has been extensively studied in multiple systems using a variety of genetic, biochemical, and biophysical approaches (reviewed in Harfe and Jinks-Robertson 2000). In eukaryotic MMR, heterodimeric complexes involving homologs of the *Escherichia coli* MutS protein [named MutS Homologs (Msh) 1-7] mediate error surveillance and recognition. In *Saccharomyces cerevisiae*, Msh2/Msh6 protein heterodimers recognize and repair nuclear base-base mismatches and small (1–2 bp) insertion-deletion (indel) loops, whereas Msh2/Msh3 heterodimers correct a range of small and larger loop-outs, but do not recognize most base-base mismatches (Alani 1996; Habraken *et al.* 1996). In humans, hMsh2/hMsh6 and hMsh2/hMsh3 heterodimers display partially overlapping damage recognition spectra similar to that observed in *S. cerevisiae* (Acharya *et al.* 1996). Msh1 proteins are involved in maintaining mitochondrial genome stability in *S. cerevisiae* (Chi and Kolodner 1994), but *msh1* orthologs have not been detected in any metazoan genomes surveyed thus far (Eisen 1998; unpublished genome database searches) and it is often assumed that MMR is absent from mitochondria in metazoans. Msh4 and Msh5 have roles in meiotic recombination in *S. cerevisiae*, *Caenorhabditis elegans*, and humans, with no apparent MMR-related functions (Zalevsky *et al.* 1999).

Despite the remarkable overall congruence in the MMR machinery and error recognition mechanisms between *S. cerevisiae* and humans, *msh3* orthologs are not detected in the genomes of *C. elegans* or *Drosophila melanogaster* (Eisen 1998; unpublished genome database searches). Furthermore, although the *Schizosaccharomyces pombe* genome encodes an *msh3* ortholog, *S. pombe* mutants deficient for this gene do not display defects in MMR but rather have reduced recombination frequencies (Tornier *et al.* 2001). The overall consequences of

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY863110–AY863143.

[1]*Corresponding author:* Department of Biology, Indiana University, 1001 East Third St., Bloomington, IN 47405.
E-mail: ddenver@bio.indiana.edu

lacking an *msh3* ortholog involved in MMR-mediated error recognition, however, are unclear. The Msh2/Msh6 dimer in species lacking an *msh3* ortholog may recognize and repair a spectrum of errors comparable to that repaired by both Msh2/Msh3 and Msh2/Msh6 dimers in species such as *S. cerevisiae* and humans. Alternatively, the Msh2/Msh6 dimer in *msh3* ortholog-deficient species may recognize a range of mismatches and small indel loops comparable to that of the Msh2/Msh6 dimer in *msh3* ortholog-proficient species. It also cannot be ruled out that Msh-2 and/or Msh-6 homodimers may be involved in *C. elegans* MMR-mediated error surveillance.

Mutation spectra in MMR-deficient backgrounds have been investigated in multiple eukaryotic species including *S. cerevisiae* (reviewed in Harfe and Jinks-Robertson 2000), *S. pombe* (Tornier *et al.* 2001), *C. elegans* (Degtyareva *et al.* 2002; Tijsterman *et al.* 2002), *D. melanogaster* (Harr *et al.* 2002), mice (Andrew *et al.* 2000), and humans (Malkhosyan *et al.* 1996; Ohzeki *et al.* 1997; Tauchi *et al.* 2000; Mark *et al.* 2002). Although these studies have provided important fundamental insights into MMR-deficient mutation processes, they have been limited to observations at one or a few reporter genetic loci and/or focus exclusively on mutations at known hotspot repetitive sequences, such as microsatellites. Furthermore, a general lack of direct and unbiased estimates of baseline (repair-proficient) spontaneous mutation spectra in almost all eukaryotic species has limited our ability to interpret mutation rates and patterns in DNA repair-deficient backgrounds. An accurate and comprehensive understanding of MMR's contributions to maintaining genome stability requires a broad-based analysis of MMR-deficient mutation spectra at multiple diverse genetic loci and in a system where baseline spontaneous mutation processes are well understood.

This study provides a direct and robust analysis of mutation rates and patterns in two MMR-deficient (*msh-2* and *msh-6*) strains of *C. elegans*. Mutation spectra were surveyed across multiple nuclear loci and one mitochondrial locus in the *msh-2* and *msh-6 C. elegans* mutation-accumulation (MA) lines to provide insights into the roles of MMR in maintaining eukaryotic genome stability. Mutational estimates from a set of long-term, wild-type (WT) *C. elegans* MA lines (Denver *et al.* 2004a,b) provide a unique mutational baseline for interpreting MMR-deficient mutation processes.

MATERIALS AND METHODS

Base strains and mutation-accumulation procedures: The *msh-2* strain of *C. elegans*, provided as a gift from Thomas D. Petes at the University of North Carolina-Chapel Hill, contains a Tc1 transposon insertion in the seventh exon of the *msh-2* gene and has been characterized as MMR-defective (Degtyareva *et al.* 2002). The *msh-6 C. elegans* strain, provided as a gift from Ronald H. Plasterk at the Hubrecht Laboratory, is missing the entire fifth and part of the sixth exon and is also MMR defective (Tijsterman *et al.* 2002). Before initiating MA experiments, each MMR knockout strain was backcrossed to N2 genomes six times so that the *msh-2* and *msh-6* experiments would be carried out on highly similar genetic backgrounds, other than the specific defective MMR genes of interest.

Fifty MA lines were initiated for each of the backcrossed *msh-2* and *msh-6* strains of *C. elegans*. Following standard MA procedures for *C. elegans* (Vassilieva *et al.* 2000), each MA line was propagated across an average of 18 generations in a benign environment (NGM agar plates seeded with the OP50 strain of *E. coli* as a food source, 20°) as single, randomly selected hermaphrodites picked at the L4 larval stage. This treatment resulted in an effective population size equal to one for each MA line throughout the experiment and ensured that all but the most deleterious mutations accumulated over time in an effectively neutral fashion. Sets of backups, maintained for each MA line (at 10°) for the preceding generation, were used in the event of sterile or dead worms. MA lines were declared extinct if all three consecutive attempts to transfer worms from the backup plate resulted in nonviable worms. Five *msh-2* MA lines and one *msh-6* line went extinct through the course of the MA experiment. For comparison, 26 of 100 original WT MA lines were extinct after an average of 214 generations (Denver *et al.* 2000).

Mutation detection and confirmation: Mutations were detected in the MMR-deficient *C. elegans* MA lines by polymerase chain reaction (PCR)-amplifying regions of the genome followed by direct DNA sequencing of the PCR products. The majority of loci sequenced were randomly distributed across *C. elegans* chromosomes by designing PCR primer pairs around chromosomal positions selected by a random number generator (Denver *et al.* 2004b); a subset of loci, however, were directly targeted to specific homopolymer (HP) runs to evaluate their mutational properties (Denver *et al.* 2004a). PCRs were performed using a large amount of genomic DNA (~25,000 diploid genomes per reaction) and 2 units Taq DNA polymerase (Eppendorf) to eliminate artifacts associated with initial amplification from small amounts of genomic DNA. PCR products were purified by solid phase reversible immobilization (Elkin *et al.* 2001), cycle sequenced, and analyzed on ABI3700 and ABI3730 DNA sequencers (Applied Biosystems) at the Indiana Molecular Biology Institute.

DNA sequence text files from MMR-deficient MA lines, backcrossed MMR-deficient progenitor lines, and N2 (WT) were batch aligned using Clustal W (Higgins *et al.* 1994) to identify putative mutations in the MMR-deficient MA lines. Putative MA line-specific mutations identified in the alignments were then visually scrutinized on the electropherogram data to eliminate base-caller errors and other sequencing artifacts. All putative mutations showing unclear or ambiguous sequence data were resequenced. Putative mutations supported by clean, unambiguous electropherogram data were then evaluated on the opposite strand (sequencing reaction in opposite direction), using internal primers where necessary. Only those mutations supported by reliable electropherogram data on both strands of directly sequenced PCR products were considered for this study.

Calculation of mutation rates: Complex sequence mutation rates were calculated using the equation $\mu = m/(LnT)$, where μ is the mutation rate (per nucleotide site per generation), m is the number of observed mutations, L is the number of MA lines, n is the number of nucleotide sites, and T is the time in generations. The standard errors of mean (SEM) for complex sequence mutation rates were calculated using the equation $\text{SEM} = (\mu/(LnT))^{1/2}$. For HP loci, mutation rates were calculated using the equation $\mu = m/(LhT)$, where h is the number of HP loci, and SEM values were calculated using the equation $\text{SEM} = (\mu/(LhT))^{1/2}$. SEM values for mutation rates are shown in parentheses throughout the text and tables.

RESULTS AND DISCUSSION

We sequenced 20,469 bp of nuclear DNA, distributed across 24 PCR product loci, and 929 bp of mitochondrial DNA (one PCR locus) from each of 45 *msh-2* and 49 *msh-6 C. elegans* MA lines (see supplementary Table 1 at http://www.genetics.org/supplemental/). The majority of nuclear loci (21/24 total) was also assayed in the WT *C. elegans* MA lines (Denver *et al.* 2004b). Each MMR-deficient MA line was propagated across an average of 18 generations as single, randomly selected hermaphrodites to allow for the accumulation of all but the most deleterious mutations (Vassilieva *et al.* 2000). Both sets of MMR-deficient MA lines displayed severe fitness declines by the end of the mutation-accumulation phase (our unpublished data), consistent with a previous analysis of fitness declines in a separate set of *msh-2 C. elegans* MA lines (Estes *et al.* 2004). We investigated MMR-deficient *C. elegans* mutation processes first at HP loci ≥ 8 bp in length that are known indel hotspots, then across the remaining complex sequences. We compared the rates and spectra mutation in the MMR-deficient MA lines reported here to that observed in a long-term WT set of *C. elegans* MA lines (Denver *et al.* 2004a,b) to gain unique insights into relationships between MMR and spontaneous mutation processes in *C. elegans*.

Mutation rates and spectra at homopolymer loci: We first considered mutation rates and spectra at HP runs (defined here as any mononucleotide run ≥ 8 bp in length) as these simple sequences are known indel mutational hotspots in MMR-deficient backgrounds in yeast, humans, and *C. elegans* (Harfe and Jinks-Robertson 2000; Tijsterman *et al.* 2002). Fifteen nuclear HP loci (eight A:T HPs and seven G:C HPs) were assayed in the MMR-deficient MA lines in addition to a single mitochondrial $(A:T)_{11}$ run. We did not focus on other microsatellites (di- or trinucleotide, for instance) as previous studies have surveyed *C. elegans* MMR-deficient mutation processes at these types of loci (Degtyareva *et al.* 2002; Tijsterman *et al.* 2002), and HPs are a much more dominant component of the *C. elegans* genome than are other microsatellite types (Denver *et al.* 2004a).

We detected 149 mutations, dominated by single-nucleotide indels, at nuclear HP loci in the MMR-deficient *C. elegans* MA lines (Table 1). The majority of HP mutations were observed at G:C HP loci (115/149 total observed, as compared to 69.5/149 expected on the basis of an even distribution of mutations across the 15 assayed HP loci), and two-nucleotide indel mutations were exclusively observed at G:C HP runs. The mutation rates for G:C HPs were highly similar and not significantly different for the *msh-2* and *msh-6* MA lines (Table 2). Larger G:C HPs, 12–16 bp in length, mutated at approximately a twofold greater rate [$\mu = 1.3\ (\pm 0.1) \times 10^{-2}$ mutations per HP per generation, calculated for combined *msh-2* and *msh-6* mutations] than that of smaller, 8–11 bp, G:C HPs [$\mu = 4.7\ (\pm 1.0) \times 10^{-3}$ mutations per HP per generation, again calculated for both *msh-2* and *msh-6* data]. Mutation rates at A:T HPs were also highly similar between the *msh-2* and *msh-6* MA lines (Table 2). Larger A:T HPs, 12–16 bp in length, mutated at a higher rate [$\mu = 6.2\ (\pm 1.4) \times 10^{-3}$ mutations per HP per generation, calculated for combined *msh-2* and *msh-6* mutations] than that of smaller, 8–11 bp, HP loci [$\mu = 1.2\ (\pm 0.4) \times 10^{-3}$ mutations per HP per generation, also calculated for combined *msh-2* and *msh-6* mutations]. Overall, mutation rates at A:T HPs were lower (three- to fourfold) than rates observed at G:C HPs in the MMR-deficient MA lines, consistent with observations in *S. cerevisiae* (Tran *et al.* 1997; Gragg *et al.* 2002). All mutations observed at HP loci were indels, the majority of which were single-base pair indels (five 2-bp indels were observed). Deletions were more prevalent than insertions in the MMR-deficient MA lines at both A:T and G:C HPs (Table 1).

No mutations were observed at the mitochondrial $(A:T)_{11}$ HP run. This is notable as 4.5 (±1.4) mutations

TABLE 1

Mutations detected at HP loci in the MMR-deficient MA lines

		msh-2 MA lines			*msh-6* MA lines		
Locus	HP length	+1	−1	Other	+1	−1	Other
		(A:T) HPs					
Y56A3A	8	0	1	0	0	1	0
C48B6	9	0	0	0	0	1	0
M106	9	1	0	0	0	0	0
B0284	10	0	1	0	0	1	0
K07G5	10	0	0	0	0	1	0
R10E4	11	0	2	0	0	4	0
C48B6(a)	12	0	3	0	0	3	0
F02E9	15/13[a]	0	12	0	1	2	0
	Totals:	1	19	0	1	13	0
		(G:C) HPs					
Y73B6BL	9/10[b]	1	0	0	3	2	0
F35D2	10	0	2	0	2	2	0
C56E6	11	1	4	0	1	5	(−2)
C01B12	13	3	4	(−2)	5	8	0
Y113G7A	14	2	14	0	6	9	(+2, −2)
Y87G2A	14	2	8	0	2	7	0
Y48G9A	16	3	6	0	4	5	0
	Totals:	12	38	1	23	38	3

Locus refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found (PCR loci surveyed for mutation are named according to *C. elegans* cosmids, fosmids, and yeast artificial chromosomes; see supplementary Table 1 at http://www.genetics.org/supplemental/ for details). +1 refers to the number of single-base pairs insertions observed; −1 refers to the number of single-base pairs deletions observed. Other specific types of mutations are indicated in parentheses in the "Other" columns.

[a] The *msh-2* progenitor locus was 15 bp long whereas the *msh-6* progenitor locus was 13 bp long.

[b] The *msh-2* progenitor locus was 9 bp long and the *msh-6* progenitor locus was 10 bp long.

TABLE 2

Mutation rates observed in the MMR-deficient MA lines

	N2 (WT)	*msh-2*	*msh-6*
		HP loci	
A:T	4.5 ($\pm$3.2) $\times$ 10^{-6}	3.1 ($\pm$0.7) $\times$ 10^{-3}	2.2 ($\pm$0.6) $\times$ 10^{-3}
G:C	9.0 ($\pm$1.7) $\times$ 10^{-5}	8.1 ($\pm$1.1) $\times$ 10^{-3}	1.0 ($\pm$0.1) $\times$ 10^{-2}
		CS loci	
BS	9.0 ($\pm$3.0) $\times$ 10^{-9}	6.4 ($\pm$2.0) $\times$ 10^{-7}	5.0 ($\pm$1.7) $\times$ 10^{-7}
Indel	1.2 ($\pm$0.4) $\times$ 10^{-8}	3.6 ($\pm$1.5) $\times$ 10^{-7}	5.5 ($\pm$1.8) $\times$ 10^{-7}
Total	2.1 ($\pm$0.7) $\times$ 10^{-8}	1.0 ($\pm$2.5) $\times$ 10^{-6}	1.0 ($\pm$2.4) $\times$ 10^{-6}

Mutation rates shown for HP loci are reported as the number of mutations per HP per generation. Mutation rates for complex sequence (CS) loci are reported as the number of mutations per nucleotide per generation. SEM is shown in parentheses. BS indicates base substitution mutations. Data for N2 are from Denver *et al.* (2004a,b).

were expected among the 94 MMR-deficient MA lines, under the assumption that nuclear and mitochondrial A:T HPs mutate at comparable rates in MMR-deficient backgrounds. Alternatively, under a null expectation that mitochondrial HP loci mutate at the same rate in WT (Denver *et al.* 2000) and MMR-deficient backgrounds, 0.5 ($\pm$0.2) mutations would be expected at the $(A:T)_{11}$ HP. The number of mutations observed at this mitochondrial HP locus in the MMR-deficient MA lines (zero) was much closer to the latter expectation, suggesting that *msh-2* and *msh-6* do not function in mitochondrial MMR in *C. elegans*.

The remarkable similarity between the *msh-2* and *msh-6* MA lines in terms of overall nuclear HP mutation rate and pattern suggests that a single Msh-2/Msh-6 heterodimeric complex mediates MMR surveillance for postreplicative HP loop-outs in *C. elegans*. This is supported by the apparent absence of an *msh3* ortholog in the *C. elegans* genome (Eisen 1998) and indications from yeast two-hybrid experiments that Msh-2 and Msh-6 may interact exclusively with one another (Boulton *et al.* 2002). The mutational similarities between *msh-2* and *msh-6 C. elegans* reported here contrast with observations in *S. cerevisiae* where HP mutation patterns observed in *msh2*, *msh3*, and *msh6* strains differ markedly with respect to one another (Gragg *et al.* 2002). Compared to G:C HP mutation rates in the WT set of *C. elegans* MA lines (Denver *et al.* 2004a), a ~100-fold elevated rate was observed for G:C HPs in the MMR-deficient MA lines. The mutation rate disparity between the MMR-deficient and WT MA lines was more pronounced for A:T HPs where the rate was elevated ~500-fold in the MMR-deficient MA lines. This observation suggests that the *C. elegans* MMR surveillance machinery may have evolved a greater ability to recognize and repair loop-outs specific to A:T HPs as compared to G:C HPs. The selective pressure to maintain stability specifically at A:T HP runs may be related to the extreme dominance of A:T HPs in the *C. elegans* genome (146,224 A:T HPs $\geq$ 8 bp were detected in the genome) as compared to G:C HPs (only 2,401 G:C HPs $\geq$ 8 bp were detected) (Denver *et al.* 2004a).

Mutation rates and spectra in complex sequence: Although mutations at known hotspot repetitive loci (such as HPs and microsatellites) in MMR-deficient backgrounds provide important insights into the MMR process, the vast majority of the *C. elegans* genome is composed of more complex sequences where the impacts of MMR on mutation processes are less clear. Across >20 kb of nuclear DNA surveyed from each MMR-deficient MA line, we detected 17 nuclear complex sequence mutations (not at HP loci, defined previously) in the *msh-2* MA lines and 19 nuclear complex sequence mutations in the *msh-6* MA lines (Table 3), resulting in strikingly similar total mutation rates for the *msh-2* and *msh-6* MA lines (Table 2) and further suggesting that MMR surveillance in *C. elegans* is carried out by a single Msh-2/Msh-6 heterodimer. No complex sequence mutations were observed at the mitochondrial locus. Highly similar mutation rates specific for base substitutions and indels were observed for the *msh-2* and *msh-6* MA lines (Table 2).

The distributions of mutations across *msh-2* and *msh-6* MA lines were very close to Poisson expectations (for *msh-2*, 30.8, 11.7, 2.2, and 0.3 instances of MA lines with 0, 1, 2, and 3 mutations were expected, respectively, and 31, 12, 1, and 1 lines were observed with 0, 1, 2, and 3 mutations, respectively; for *msh-6*, 33.2, 12.9, 2.5, and 0.3 instances of MA lines with 0, 1, 2, and 3 mutations were expected, respectively, and 33, 14, 1, and 1 MA lines were observed with 0, 1, 2, and 3 mutations, respectively). For the *msh-2* MA lines, the distribution of mutations across the 24 assayed nuclear PCR product loci were also very close to Poisson expectations (11.8, 8.4, 3.0, and 0.7 loci with 0, 1, 2, and 3 mutations were expected, respectively, and 11, 8, 4, and 1 loci with 0, 1, 2, and 3 mutations were observed, respectively). For the *msh-6* MA lines, however, seven mutations were ob-

TABLE 3

Complex sequence mutations observed in the MMR-deficient MA lines

Chr.	Locus	Locus position	Mut.	Context	Line	Cod.
			msh-2			
I	C48B6(b)	41,650	T → C	ATTAT → ATCAT	2-47	IN
I	Y87G2A	8,344	T → C	GATAA → GACAA	2-43	IG
I	Y87G2A	8,347	G → A	AAGCT → AAACT	2-6	IG
II	C17G10	23,624	T → C	GGTCT → GGCCT	2-13	IG
II	C17G10	23,950	+C	TGCAG → TGCCAG	2-16	IG
II	M106	19,689	G → A	GGGCA → GGACA	2-21	IG
III	Y56A3A	139,797	−T	ATTTTTC → ATTTTC	2-38	EX
III	Y56A3A	139,905	+T	TCTCG → TCTTCG	2-38	EX
IV	Y73B6BL	40,415	−C	AACTC → AATC	2-46	EX
IV	C06A6	18,654	+C	GGGAA → GGGCAA	2-23	IG
IV	C06A6	18,763	A → C	TCAAA → TCCAA	2-5	IG
V	BO240	2,378	A → G	AAATT → AAGTT	2-23	IN
V	B0240	2,684	G → A	TAGTT → TAATT	2-26	IN
V	Y113G7A	5,294	A → C	TGACG → TGCCG	2-35	IG
V	Y113G7A	5,345	−T	TATCG → TACG	2-13	IG
V	Y113G7A	5,390	A → G	CGATT → CGGTT	2-23	IG
X	F19C6	3,499	A → C	AAACT → AACCT	2-11	IN
			msh-6			
I	C48B6(a)	22,768	−8 bp	GCAAATCCGTATT → GCTTT	6-17	IN
I	K07G5	9,681	G → A	CAGAT → CAAAT	6-43	IN
I	ZK337	26,111	+AA	GGAAAAGC → GGAAAAAAGC	6-17	EX
I	ZK337	26,111	+A	GGAAAAGC → GGAAAAAGC	6-34	EX
I	ZK337	26,111	+A	GGAAAAGC → GGAAAAAGC	6-25	EX
I	ZK337	26,159	+G	GTGGCA → GTGGGCA	6-9	EX
I	ZK337	26,159	+G	GTGGCA → GTGGGCA	6-17	EX
I	ZK337	26,203	+T	CATTCG → CATTTCG	6-44	EX
I	ZK337	26,239	+A	AGAAAAGT → AGAAAAAGT	6-43	EX
II	C01B12	1,059	T → C	ATTTC → ATCTC	6-7	IG
III	Y56A3A	139,883	A → G	TGACC → TGGCC	6-37	EX (S)
IV	Y73B6BL	40,011	A → G	TGAAG → TGGAG	6-30	IG
IV	Y73B6BL	40,700	A → G	TCAAT → TCGAT	6-49	IN
V	W05B10	21,037	T → C	GGTGG → GGCGG	6-35	IG
V	W05B10	21,436	C → T	TACGA → TATGA	6-4	IG
V	W05B10	21,447	−A	CAAAAAT → CAAAAT	6-41	IG
V	Y113G7A	5,313	+C	TGTGT → TGTCGT	6-2	IG
X	F59F5	22,024	G → C	CAGTG → CACTG	6-45	EX (R: T→S)
X	F19C6	3,938	T → G	AATGT → AAGGT	6-48	IN

Locus refers to the sequenced *C. elegans* cosmid or yeast artificial chromosome in which the mutation was found (PCR loci surveyed for mutation are named according to *C. elegans* cosmids, fosmids, and yeast artificial chromosomes; see supplementary Table 1 at http://www.genetics.org/supplemental/ for details). Context provides information regarding the bases surrounding the observed mutation with respect to the (+) strand of *C. elegans* chromosomes. Line indicates the specific MA line in which the mutation was detected. Chr., the chromosome in which the mutation was found. Mut., the observed mutation. Cod., the coding context of the sequence in which the mutation was found: EX, exon, IG, intergenic, IN, intron. For exon base substitution mutations: S, a silent base substitution, R, a replacement base substitution (amino acid change is indicated after R).

served at a single locus (ZK337; see Table 3), resulting in a significant deviation from Poisson expectations ($P < 0.005$) for the distribution of mutations across PCR loci (10.9, 8.6, 3.4, 0.9, and 4.2×10^{-4} loci with 0, 1, 2, 3, and 7 mutations were expected, respectively, and 14, 7, 1, 1, and 1 loci with 0, 1, 2, 3, and 7 mutations were observed, respectively).

The complex sequence mutation rates observed in MMR-deficient backgrounds were ~100-fold higher than complex sequence mutation rates in the WT MA lines (Denver *et al.* 2004b). This observation is in contrast to suggestions that the relatively high mutation rates observed in the WT *C. elegans* MA lines (as compared to previous lower mutation rate estimates for *C.*

elegans based on indirect phenotypic assays) may be due to mutation-induced stress responses where mismatch repair is inactivated (Rosenberg and Hastings 2004). If the high mutation rate observed in the WT MA lines was due to stress-induced MMR-deficiency, we would expect roughly similar mutation rates between the WT and MMR-deficient *C. elegans* MA lines rather than the observed ~100-fold disparity.

Similar mutation patterns were also observed between the *msh-2* and *msh-6 C. elegans* MA lines. Among base substitutions, eight transitions and three transversions were observed in the *msh-2* MA lines; seven transitions and two transversions were observed in the *msh-6* MA lines (Table 3). These transition biases were also similar to that observed in the WT *C. elegans* MA lines (Denver *et al.* 2004b). Six complex sequence indel mutations were observed in the *msh-2* MA lines: three single-bp insertions and three single-bp deletions. Ten complex sequence indels were observed for the *msh-6* MA lines: seven single-base pair insertions, one 2-bp insertion, one single-base pair deletion, and one 8-bp deletion. Seven of the *msh-6* indels, all insertions, were observed at a single locus (ZK337). Six of the 16 total complex sequence indels in the MMR-deficient MA lines (both *msh-2* and *msh-6*) were in short mononucleotide run stretches below our cutoff threshold for HP loci (≥8 bp). Whereas complex sequence base substitutions were more frequent than indels in MMR-deficient backgrounds, complex sequence indels occurred at a higher rate than did base substitutions in the WT MA lines (Denver *et al.* 2004b), suggesting that the *C. elegans* MMR machinery may be more efficient at eliminating base substitutions than the indels in complex sequence.

The overall similarity in complex sequence mutation spectra observed between the *msh-2* and *msh-6 C. elegans* MA lines contrasts with mutation patterns observed in MMR-deficient *S. cerevisiae* strains at the Can^r locus (Marsischky *et al.* 1996) that vary with respect to differing specific genotypes (Table 4). Similarly, mutation spectra differed in MMR-deficient human colon carcinoma cell lines at the hprt locus depending on the specific MMR-deficient background (Malkhosyan *et al.* 1996; Ohzeki *et al.* 1997; Tauchi *et al.* 2000). The distinctive similarity between *msh-2* and *msh-6 C. elegans* mutation spectra, as compared to yeast and humans, may be due to the apparent absence of an *msh3* homolog and the corresponding presence of a single major MMR error surveillance complex (the Msh-2/Msh-6 heterodimer) in *C. elegans*.

An exception to the trend of mutational similarity between *msh-2* and *msh-6 C. elegans* was the ZK337 locus where seven insertion mutations were observed exclusively in the *msh-6* MA lines (Table 3); no mutations were found at this locus in the *msh-2* MA lines. This observation deviated significantly ($P < 0.05$) from expectations based on an even distribution of ZK337 mutations across *msh-2* and *msh-6* MA lines. This *msh-6*-specific mutational hotspot locus was also distinctive from other assayed nuclear loci as all seven observed mutations (occurring across four distinct specific locus positions; Table 2) were insertions, whereas an approximately equal number of insertions and deletions were detected at the other 23 loci surveyed in the MMR-deficient MA lines. This finding suggests that Msh-6 may be involved in mutation deterrence in an Msh-2-independent fashion that is specific to this region (and perhaps other regions) of the *C. elegans* genome. Further studies are required to understand the significance of the unusual mutation patterns at this single locus.

TABLE 4

MMR-deficient mutation spectra in different species

Genotype	BS/ID	Ts/Tv	Ins/Del
Ce WT (76 loci)	13/17	8/5	13/4
Ce *msh-2* (24 loci)	11/6	8/3	3/3
Ce *msh-6* (23 loci)	9/3	7/2	1/2
Ce *msh-6* (ZK337)	0/7	0/0	7/0
Sc *msh2* (Can^r)	3/17	0/3	1/16
Sc *msh6* (Can^r)	18/3	12/6	0/3
Sc *msh3 msh6* (Can^r)	7/15	3/4	1/14
Hs *msh2* (hprt)	16/4	16/0	0/4
Hs *msh3* (hprt)	38/32	37/1	8/24
Hs *msh6* (hprt)	59/3	20/39	2/1

Genotype refers to the species (Ce, *C. elegans*; Sc, *S. serevisiae*; Hs, *Homo sapiens*; Mm, *Mus musculus*) and specific MMR-deficient genetic background for which mutation data were reported. The specific loci assayed are listed in parentheses. BS/ID indicates the observed ratio of base substitution to indel mutations, Ts/Tv refers to the ratio of transitions to transversions, and Ins/Del indicates the ratio of insertions to deletions (excludes HP loci ≥ 8 bp). The WT *C. elegans* data are from Denver *et al.* (2004b); the *S. cerevisiae* data are from Marsischky *et al.* (1996); and *H. sapiens* data are from Malkhosyan *et al.* (1996), Ohzeki *et al.* (1997), Tauchi *et al.* (2000), and Mark *et al.* (2002). The data for *msh-6 C. elegans* are divided into two sets to distinguish between patterns observed at the majority of loci *vs.* the ZK337 hotspot locus.

Conclusion: In this study we provide important insights into the relationships between MMR and spontaneous mutation processes in *C. elegans*. We find that, for the most part, mutation rates and spectra in two distinct MMR-deficient *C. elegans* backgrounds (*msh-2* and *msh-6*) are highly similar to one another (the ZK337 locus being an exception), but differ when compared to rates and patterns of spontaneous mutation in a WT strain of *C. elegans*. The nuclear MMR-deficient mutation spectra reported here for *C. elegans* also differ from that observed in other eukaryotes that encode an *msh3* ortholog (*S. cerevisiae* and humans). A broad-based understanding of the evolution of MMR pathways and their contributions to maintaining genome stability across eukaryotic phylogeny will require comparable surveys of MMR-deficient mutation spectra in species such as *D. melanogaster*, which also appears to lack an *msh3* or-

tholog, and *Arabidopsis thaliana*, which encodes AtMsh2, AtMsh3, AtMsh6, and AtMsh7, carrying out MMR surveillance with three different error surveillance complexes (Culligan and Hays 2000).

We thank T. D. Petes at the University of North Carolina, Chapel Hill and R. H. Plasterk at the Hubrecht Laboratory for providing the *msh-2* and *msh-6 C. elegans* strains, respectively. We thank Lawrence Washington at the Indiana Molecular Biology Institute for DNA sequencing. We thank two anonymous reviewers for helpful comments. Funding for this work was provided by National Institutes of Health grant R01 GM36827 to M.L. and W.K.T., and fellowship F32 GM66652 to D.R.D.

LITERATURE CITED

Acharya, S., T. Wilson, S. Gradia, M. F. Kane, S. Guerrette *et al.*, 1996 hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. Proc. Natl. Acad. Sci. USA **93:** 13629–13634.

Alani, E., 1996 The *Saccharomyces cerevisiae* Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. Mol. Cell. Biol. **16:** 5604–5615.

Andrew, S. E., X. S. Xu, A. Baross-Francis, L. Narayanan, K. Milhausen *et al.*, 2000 Mutagenesis in PMS2- and MSH2-deficient mice indicates differential protection from transversions and frameshifts. Carcinogenesis **21:** 1291–1295.

Boulton, S. J., A. Gartner, J. Reboul, P. Vaglio, N. Dyson *et al.*, 2002 Combined functional genomics maps of the *C. elegans* DNA damage response. Science **295:** 127–131.

Chi, N. W., and R. D. Kolodner, 1994 Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. J. Biol. Chem. **269:** 29984–29992.

Culligan, K. M., and J. B. Hays, 2000 *Arabidopsis* MutS homologs—AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7—form three distinct protein heterodimers with different specificities for mismatched DNA. Plant Cell **12:** 991–1002.

Degtyareva, N. P., P. Greenwell, E. R. Hofmann, M. O. Hengartner, L. Zhang *et al.*, 2002 *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.

Denver, D. R., K. Morris, M. Lynch, L. L. Vassilieva and W. K. Thomas, 2000 High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science **289:** 2342–2344.

Denver, D. R., K. Morris, A. Kewalramani, K. E. Harris, A. Chow *et al.*, 2004a Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. J. Mol. Evol. **58:** 584–595.

Denver, D. R., K. Morris, M. Lynch and W. K. Thomas, 2004b High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature **430:** 679–682.

Eisen, J. A., 1998 A phylogenomic study of the MutS family of proteins. Nucleic Acids Res. **26:** 4291–4300.

Eisen, J. A., and P. C. Hanawalt, 1999 A phylogenomic study of DNA repair genes, proteins, and processes. Mutat. Res. **435:** 171–213.

Elkin, C. J., P. M. Richardson, H. M. Fourcade, N. M. Hammon, P. F. Pollard *et al.*, 2001 High-throughput plasmid purification for capillary sequencing. Genome Res. **11:** 1269–1274.

Estes, S., P. C. Phillips, D. R. Denver, W. K. Thomas and M. Lynch, 2004 Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics **166:** 1269–1279.

Gragg, H., B. D. Harfe and S. Jinks-Robertson, 2002 Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **22:** 8756–8762.

Habraken, Y., P. Sung, L. Prakash and S. Prakash, 1996 Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Curr. Biol. **6:** 1185–1187.

Harfe, B. D., and S. Jinks-Robertson, 2000 DNA mismatch repair and genetic instability. Annu. Rev. Genet. **34:** 359–399.

Harr, B., J. Todorova and C. Schlotterer, 2002 Mismatch repair-driven mutational bias in *D. melanogaster*. Mol. Cell **10:** 199–205.

Higgins, D. G., J. D. Thompson and T. J. Gibson, 1994 Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266:** 383–402.

Lindahl, T., 1993 Instability and decay of the primary structure of DNA. Nature **262:** 709–715.

Malkhosyan, S., A. McCarty, H. Sawai and M. Perucho, 1996 Differences in the spectrum of spontaneous mutations in the hprt gene between tumor cells of the microsatellite mutator phenotype. Mutat. Res. **316:** 249–259.

Mark, S. C., L. E. Sandercock, H. A. Luchman, A. Baross, W. Edelmann *et al.*, 2002 Elevated mutation frequencies and predominance of G:C to A:T transition mutations in *Msh6*$^{-/-}$ small intestinal epithelium. Oncogene **21:** 7126–7130.

Marsischky, G. T., N. Filosi, M. F. Kane and R. Kolodner, 1996 Redundancy of *Saccharomyces cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev. **10:** 407–420.

Ohzeki, S., A. Tachibana, K. Tatsumi and T. Kato, 1997 Spectra of spontaneous mutations at the hprt locus in colorectal carcinoma cell lines defective in mismatch repair. Carcinogenesis **18:** 1127–1133.

Rosenberg, S. M., and P. J. Hastings, 2004 Worming into genetic instability. Nature **430:** 625–626.

Tauchi, H., K. Komatsu, K. Ishizaki, F. Yatagai and T. Kato, 2000 Mutation spectrum of MSH3-deficient HHUA/chr.2 cells reflects *in vivo* activity of the MSH3 gene product in mismatch repair. Mutat. Res. **447:** 155–164.

Tijsterman, M., J. Pothof and R. H. Plasterk, 2002 Frequent germline mutations and somatic repeat instability in DNA mismatch repair-deficient *Caenorhabditis elegans*. Genetics **151:** 651–660.

Tornier, C., S. Bessone, I. Varlet, C. Rudolph, M. Darmon *et al.*, 2001 Requirement for Msh6, but not for Swi4 (Msh3), in Msh2-dependent repair of base-base mismatches and mononucleotide loops in *Schizosaccharomyces pombe*. Genetics **158:** 65–75.

Tran, H. T., D. Keen, M. Kricker, M. A. Resnick and D. A. Gordenin, 1997 Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol. Cell. Biol. **17:** 2859–2865.

Vassilieva, L. L., A. M. Hook and M. Lynch, 2000 The fitness effects of spontaneous mutation in *Caenorhabditis elegans*. Evolution **54:** 1234–1246.

Zalevsky, J., A. J. MacQueen, J. B. Duffy, K. J. Kemphues and A. J. Villenueve, 1999 Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. Genetics **153:** 1271–1283.

Communicating editor: K. Kemphues